UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

**UNIVERSITE PARIS I PANTHÉON SORBONNE**
**ED de 465**
Laboratoire de rattachement : Économie Panthéon Sorbonne

THÈSE
Pour l'obtention du titre de Docteur en Économie
Présenté(e) et soutenu(e) publiquement
le 22 mai 2024 par

**Duncan WEBB**

# Essays in Development Economics

**Sous la direction de :**
Karen MACOURS et Suanna OH, Paris School of Economics

**Membres du Jury :**

**Rapporteurs :**
Jonathan de Quidt, Queen Mary University
Manisha Shah, University of California Berkeley

**Jury :**
Abhijit Banerjee, Massachusetts Institute of Technology
Frank Schilbach, Massachusetts Institute of Technology

**UNIVERSITE PARIS I PANTHÉON SORBONNE**
**Doctoral School 465**
Laboratory: Economics, Panthéon Sorbonne

Thesis
For the title of Doctor in Economics
Presented and defended publicly
22 May 2024 by

**Duncan WEBB**

# Essays in Development Economics

**Under the direction of:**
Karen MACOURS and Suanna OH, Paris School of Economics

**Members of the jury :**

**Referees :**
Jonathan de Quidt, Queen Mary University
Manisha Shah, University of California Berkeley

**Jury :**
Abhijit Banerjee, Massachusetts Institute of Technology
Frank Schilbach, Massachusetts Institute of Technology

# Acknowledgements

I'd like to say an enormous thank you to everyone who helped me throughout this PhD. I feel incredibly lucky to have been supported by so many amazing people along the way.

First, to Karen – you have been such an unwaveringly dedicated supervisor and mentor ever since I kicked off work on my Master thesis. I never cease to be amazed at how much time you dedicate to me and your other students, how unerringly sharp your advice is, and how committed you are to guiding my research to be the best it possibly can be. I've learned so much from you about doing research that aims to have a big policy impact, and I can't wait to continue working together in Madagascar.

Second, to Suanna – I feel very lucky that you joined PSE just at the right moment! You have been such a kind and astute mentor, always looking out for me, giving me incredibly good advice on my overly complicated experimental designs, and looping me into a very exciting project with you and Evan. Thank you for being endlessly encouraging and supportive.

To Frank – thank you so much for welcoming me with such enthusiasm to MIT, and for being unbelievably generous with your time even during such a busy period for you. My discussions with you have always been inspiring, and stimulate many new ideas, and you have also been such a kind and encouraging mentor throughout my job market applications.

To Abhijit – an enormous thank you for making my time at MIT so exciting, for being endlessly creative and encouraging whenever we meet, and for being kind and willing to dedicate so much time to helping me and cheerleading for me during my job applications.

To Esther – I am so grateful that you welcomed me to MIT. I had such an amazing time in Cambridge, met so many inspiring researchers and lifelong friends, and transformed my half-baked ideas into exciting new projects. Your guidance has always been an inimitable combination of precision and ambition. I hope the "air-bridge" from PSE to Cambridge (as you once called it) carries on for much longer!

To Manisha and Jon, thank you so much for agreeing to be my referees for this defence. My projects have already benefited so much from your careful comments and excellent advice during the pre-defence.

To Supreet, thank you for welcoming me to Berkeley for my fantastic visit. My time there was inspiring, and our discussions were always creative and able to generate step-changes in how ambitious my research was.

To all the other professors whose comments and ideas were invaluable in improving my work, thank you – in particular, to David Atkin, Luc Behaghel, Denis Cogneau, Pascaline Dupas, Evan

Friedman, Julien Grenet, Eliana La Ferrara, Sylvie Lambert, David Margolis, Gautam Rao, Akiko Suwa-Eisenmann, Oliver Vanden Eynde, Liam Wren-Lewis, and Ekaterina Zhuravskaya.

To my other amazing coauthors – Rachid Laajaj, Julieta Rueda Vera, and Evan Friedman. Thank you for being inspiring and supportive, and making research such a fun, collaborative project. I've learned a lot from all three of you.

To Maya, Oscar, and Alex – thank you for being such stars on the project in Madagascar. It has been a joy to work with people who are so good at getting things done, and so driven to make our project work.

Thank you to everyone on the teams of projects in this thesis, without whom all this work would have been impossible. To the team at CARE – particularly Emilie, Marina, Hasina, Emma, Ifaliana, and Andrianarilaza – thank you for working together in such a collaborative spirit and delivering such an amazingly effective program. To Eliane and the team at Consultant Associates, you have been a dream to work with and have shown such dedication to our project in Madagascar. Akhilan, Arun, Manvi, Bhala, Balakumar, and Murali, and all the team at J-PAL South Asia, thank you for making our fantastically complicated project work so well.

Thank you to all the organizations and people who funded the work in this thesis, including EUR-PgSE, the Weiss Fund, the Forethought Foundation, J-PAL's Crime and Violence Initiative, the Fonds d'Innovation pour le Développment, CARE, the Development Research Group at PSE, the ED 465, and CEPREMAP.

The real joy of the PhD has been all the amazing friends I have made along the way. Anna, Seámus, Tom, Marion, Ivan – what an absolute delight it's been to be friends with you ever since those heady days of doing measurement homeworks until 3am at PSE. I'm glad those days are gone, and we instead get to dance the night away in Barcelona, Paris, London, Lisbon, and beyond.

To all the other wonderful people from PSE: Amory, Zhexun, Ismael, Andrea, Charlotte, Juliette, Sarah, Melanie, Hector, Marcelo, Louise, Eric, Sebastian, Julia, Paola, Kim Lan, Vrinda, Balasai, Billy, Maria, Oscar Diaz, Mark, Alvaro, Oscar Barrera, Christian, Juni, Dave, Laura, Justine, Yagan, Manpreet, Morten, Thomas, and Kelsey. Thanks so much for making my time throughout the PhD light-hearted, fun, and sunny.

To all the people I met while at Berkeley and MIT, thank you for making my time in the US both joyful and inspiring: Eva, Wouter, Sebastian, Deivy, Advik, Mohit, Max, Dean, Matthew, Garima, Charlie, Elisa, Lisa, Lucy, Hannah, Paul, and many others.

To Lesley, thank you so much for being so endlessly supportive, and for giving me amazing travel tips in Madagascar, Paris, and Boston.

An enormous thank you to all my friends that have supported me and given me so much joy for the last 5 years. There are too many name, but I couldn't have done it without Sunaina, Miles, Alex, Simps, Andy, Oda, Simon, Julia, Liv, Kailash, Sacha, Priyanka, AJ, Miranda, Fanny, Jean and Ada.

Finally, to my family – Mum, Dad, Nai, Emily, Pip, Max, and Sol. You have been absolute rocks, keeping me going during lockdown when I was on an endless cycle of Indonesian wet- and dry-seasons with lockdown loops and crazy golf, giving me worldly advice about my life plans, sharing in my excitement about bouncing across the world, helping me get my passport back in time to go to Madagascar, nodding sagely when I try to explain my unexplainable theoretical model on a cycling trip, and giving me love and support every step of the way. Thank you.

# Summary

This dissertation studies how social interactions have impacts on economic outcomes for marginalized groups in lower- and middle-income countries, and how inequalities in human capital can propagate and be addressed.

The first chapter, *Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India*, studies how discrimination is affected by communication about a marginalized minority. 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. Non-transgender 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 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.

The second chapter, *Menstrual Stigma, Hygiene, and Human Capital: Experimental Evidence from Madagascar*, joint with Karen Macours and Julieta Vera Rueda, studies how menstrual stigma can be addressed using *positive deviants* (those who are willing to speak out against such stigma), and how hygiene-focused programs can generate human capital improvements in low income schools. The chapter presents results from an RCT in 140 schools in Madagascar that targets both hygiene practices and menstrual stigma. First, we show that a bundle of interventions (sanitation infrastructure, menstrual products, and teacher sensitization) leads to substantial (0.15 SD) improvements in learning tests and school marks, without affecting attendance or health. Girls' heart rate during the endline survey is 0.12 SD lower in treatment schools, suggesting that learning benefits may come from reduced stress and an improved psychosocial environment at school. Second, we evaluate the additional effect of nominating and coaching "young girl leaders" - school girls willing to speak out against menstrual stigma - to spread positive messages about hygiene and menstruation. The combined program generates substantial improvements in hygiene knowledge and behavior (0.32-0.56 SD) and in menstrual stigma (0.5 SD), and the Young Girl Leader component significantly increases the impact on all of these dimensions.

The third chapter, *Critical Periods in Cognitive and Socioemotional Development: Evidence from Weather Shocks in Indonesia*, studies how the timing of shocks in early childhood can have long-lasting consequences on the development of an individual's human capital. Early life circumstances are important determinants of long-run human capital and wellbeing outcomes. The first 1000 days of life are often cited as a 'critical period' for child development, but this notion has rarely been directly tested. In a setting where children are potentially subject to shocks in every year of their childhood, I estimate the impact of early life weather shocks on adult cognitive and socioemotional outcomes for individuals born in rural Indonesia between 1988 and 2000. There is a strong critical period for these shocks at age 2 for cognitive development, but no evidence for a similar critical period for socioemotional development. The effects appear to be driven by changes in agricultural income and nutritional investment. The impacts are initially latent, only appearing after age 15. I show suggestive evidence for dynamic complementarity in early life investments.

# Résumé

Cette thèse étudie l'impact des interactions sociales sur les résultats économiques des groupes marginalisés dans les pays à revenus faibles et moyens, ainsi que la manière dont les inégalités en matière de capital humain peuvent se propager et être résolues.

Le premier chapitre, *Silence to Solidarity : Using Group Dynamics to Reduce Anti-Transgender Discrimination in India*, étudie la manière dont la discrimination est affectée par la communication sur une minorité marginalisée. La discrimination au niveau individuel est souvent attribuée à des préjugés profondément ancrés qui sont de changer. Mais au niveau sociétal, nous observons parfois des réductions rapides des préférences discriminatoires, ce qui suggère que les interactions sociales et la communication qu'elles impliquent peuvent être à l'origine de ces changements. J'étudie si la discrimination peut être réduite par deux types de communication sur une minorité : (i) *la communication horizontale* entre les membres du groupe majoritaire, ou (ii) *la communication top-down* de la part d'agents d'autorité (par exemple, le système juridique). Je mène une expérience sur le terrain dans une zone urbaine de l'Inde ($N$=3 397) qui mesure la discrimination contre une communauté marginalisée de personnes transgenres. Les participants non transgenres sont très discriminatoires : dans une condition de contrôle, ils sacrifient 1,9 fois leurs dépenses alimentaires quotidiennes pour éviter d'embaucher un travailleur transgenre pour livrer les courses à leur domicile. Mais la communication horizontale entre les participants réduit fortement la discrimination : les participants qui ont participé à une discussion de groupe avec deux de leurs voisins ne font plus de discrimination en moyenne, même lorsqu'ils font des choix privés après la discussion. Cet effet est 1,7 fois plus important que l'effet de la communication top-down qui informe les participants des droits légaux des personnes transgenres. Les effets de la discussion ne sont pas dus à un signal de vertu ou à la correction d'une norme mal perçue. Au contraire, les participants semblent s'inciter mutuellement à être plus favorables aux transgenres, en partie parce que les participants favorables aux transgenres sont ceux qui s'expriment le plus dans les discussions.

Le deuxième chapitre, *Menstrual Stigma, Hygiene, and Human Capital : Experimental Evidence from Madagascar*, en collaboration avec Karen Macours et Julieta Vera Rueda, étudie la manière dont la stigmatisation menstruelle peut être abordée en utilisant *les déviants positifs* (ceux qui sont prêts à s'exprimer contre une telle stigmatisation), et comment les programmes axés sur l'hygiène peuvent générer des améliorations du capital humain dans les écoles pauvres. Ce chapitre présente les résultats d'un essai contrôlé randomisé mené dans 140 écoles de Madagascar et ciblant à la fois les pratiques d'hygiène et la stigmatisation menstruelle. Tout d'abord, nous montrons qu'un ensemble d'interventions (infrastructures sanitaires, produits menstruels et sensibilisation des enseignants) conduit à des améliorations substantielles (0,15 E.T.) des tests d'apprentissage et des notes scolaires, sans affecter l'assiduité ou la santé. Le rythme cardiaque des filles lors de l'enquête finale est inférieur de 0,12 écart-type dans les écoles traitées, ce qui suggère que les bénéfices de l'apprentissage peuvent provenir d'une réduction du stress et d'une amélioration de l'environnement

psychosocial à l'école. Deuxièmement, nous évaluons l'effet supplémentaire de la nomination et de l'encadrement de "Jeunes Filles Leaders" - des écolières prêtes à s'exprimer contre la stigmatisation menstruelle - pour diffuser des messages positifs sur l'hygiène et la menstruation. Le programme combiné génère des améliorations substantielles dans les connaissances et les comportements en matière d'hygiène (0,32-0,56 SD) et dans la stigmatisation menstruelle (0,5 SD), et la composante des jeunes filles leaders augmente de manière significative l'impact sur toutes ces dimensions.

Le troisième chapitre, *Critical Periods in Cognitive and Socioemotional Development: Evidence from Weather Shocks in Indonesia*, étudie comment le timing des chocs dans la petite enfance peut avoir des conséquences durables sur le développement du capital humain d'un individu. Les circonstances de la petite enfance sont des déterminants importants du capital humain et du bien-être à long terme. Les 1000 premiers jours de la vie sont souvent considérés comme une "période critique" pour le développement de l'enfant, mais cette notion a rarement été testée directement. Dans un contexte où les enfants sont potentiellement soumis à des chocs chaque année de leur enfance, j'estime l'impact des chocs climatiques au début de la vie sur les résultats cognitifs et socio-émotionnels à l'âge adulte pour les individus nés dans l'Indonésie rurale entre 1988 et 2000. Il existe une période critique importante pour ces chocs à l'âge de 2 ans pour le développement cognitif, mais aucune preuve d'une période critique similaire pour le développement socio-émotionnel. Les effets semblent être dus à des changements dans le revenu agricole et l'investissement nutritionnel. Les impacts sont initialement latents et n'apparaissent qu'après l'âge de 15 ans. Je montre des indications de la complémentarité dynamique des investissements au début de la vie.

# Contents

**General Introduction** **11**

**1 Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India** **14**
1.1 Introduction . . . 14
1.2 Context: Transgender community in India . . . 19
1.3 Experimental design . . . 20
1.4 Results . . . 30
1.5 Intermediate outcomes: effects on attitudes, beliefs, norms, and behavior during discussion . . . 37
1.6 Mechanisms: What is behind the effect of the discussion? . . . 40
1.7 Alternative mechanisms . . . 47
1.8 Conclusion . . . 50

**2 Menstrual Stigma, Hygiene, and Human Capital: Experimental Evidence from Madagascar** **74**
2.1 Introduction . . . 74
2.2 Study design . . . 78
2.3 Results . . . 87
2.4 Mechanisms . . . 99
2.5 Discussion . . . 103

**3 Critical Periods in Cognitive and Socioemotional Development: Evidence from Weather Shocks in Indonesia** **127**
3.1 Introduction . . . 127
3.2 Data and measurement system . . . 130
3.3 Empirical strategy . . . 134
3.4 Main results . . . 137
3.5 Mechanisms . . . 146
3.6 Discussion & conclusion . . . 154

**A Appendix for Chapter 1** **177**
A.1 Additional tables and figures . . . 177
A.2 Estimation of WTP to avoid transgender workers . . . 223
A.3 Distribution of discrimination: positive and negative discrimination, and polarisation 228
A.4 Ethical considerations . . . 229
A.5 Background: Legal context . . . 231
A.6 Proofs . . . 233
A.7 Pre-analysis plan . . . 241

A.8 Discussion design details . . . 243
A.9 Video scripts . . . 245
A.10 Data and measurement . . . 248

**B Appendix for Chapter 2** **255**
B.1 Supplementary tables and figures . . . 255
B.2 Data . . . 277
B.3 Spillovers . . . 289

**C Appendix for Chapter 3** **291**
C.1 Data Appendix . . . 291
C.2 Construction of indexes . . . 298
C.3 Descriptive statistics . . . 309
C.4 Construction of predicted harvest measure . . . 313
C.5 Empirical specification for timing . . . 318
C.6 Conceptual framework . . . 321
C.7 Cognitive placebo tests . . . 325
C.8 Power simulations . . . 328
C.9 Heterogeneous effects . . . 329
C.10 Multiple testing adjustment . . . 331
C.11 Sample selection . . . 333
C.12 Additional tables: cognitive outcomes . . . 337
C.13 Additional tables: alternative non-linear models . . . 351
C.14 Additional tables: socioemotional outcomes . . . 357
C.15 Additional tables: household expenditure . . . 361
C.16 Additional tables: alternative mechanisms . . . 365
C.17 Height-for-age . . . 367

# General Introduction

An expanding body of evidence shows that economic behavior can be shaped by social forces. People are influenced by prevailing social norms and their social image concerns (Bursztyn & Jensen, 2017), the spread of specific narratives about other groups or behavior (Shiller, 2017), and the structure of their social networks (Breza, 2016). While we now have good reason to believe that these social forces *do* shape behavior, there are two key gaps in the existing literature. First, how can these social forces change over time, and can we design policies that guide those social forces towards prosocial behavior? Second, how does social forces play out in lower- and middle-income countries, where they might be relatively more influential, given the weaker role of formal institutions. To begin to address these gaps, two of the chapters in this thesis focus on the role of social interactions and interpersonal communication in affecting economic outcomes for marginalized groups in lower- and middle-income countries.

Marginalization and discrimination have typically been framed by economists as an individual decision based on unmalleable prejudice. However, at the *societal* or *group* level, we sometimes observe relatively rapid shifts in social attitudes towards outgroups. Consider, for example, the speed with which attitudes towards same-sex relations improved in the US, with acceptance of homosexuality jumping up 65% in the course of one year in 1993 (Fernández et al., 2019). Or opposition to interethnic marriage, which halved in only 10 years in the US (GSS, 2021). Motivated by these episodes, I seek to understand what kind of interpersonal interactions might generate such shifts.

In the first chapter, I examine two types of communication about a minority that could generate changes in discrimination. The first, which I term *horizontal communication*, is when peers from a majority group talk about a minority. There are a number of reasons why this kind of communication could generate changes in discrimination, rather than just reinforcing whatever discrimination already exists. For example, even if people are privately discriminatory, they may not want to reveal to their peers that they are discriminatory, thereby biasing their communication in favour of the minority group (and possibly persuading others to discriminate less in the process). Alternatively, peer-to-peer communication may correct a misperceived social norm (in the style of Bursztyn, González, & Yanagizawa-Drott, 2020), making discrimination more or less acceptable once the discussion has finished.

In my context, I find that horizontal communication (taking the form of a 3-person discussion among neighbours, centred around a collective hiring choice) leads to large *reductions* in discrimination against transgender workers in south India. In other words, privately discriminatory people persuade each other to discriminate less. This contrasts to a long-standing literature in social psychology which suggests that intragroup communication will tend to make existing prejudiced attitudes more extreme (see e.g., Moscovici & Zavalloni, 1969; Myers & Bishop, 1970; Smith & Postmes, 2011). Given this evidence, and given that horizontal communication could *a priori* both exacer-

bate or ameliorate discrimination, I seek to document evidence on the mechanisms at play to help understand how my results generalize to other settings. To this end, I show that the effects do not seem to be driven by virtue signalling or misperceived social norms. Instead, participants persuade each other, i.e. change each others' private attitudes, during the discussion.

I show suggestive evidence that this effect is driven by *positive deviants* - pro-trans participants in the discussions who are particularly vocal in the discussions. These participants speak up and persuade others to discriminate less, while anti-trans participants stay silent. I derive a model to understand when this occurs. Under the right conditions, pro-trans participants are more vocal because they start further from the existing group norm, and therefore have a greater incentive to pay a cost to change others' attitudes to match their own. Importantly, this occurs in a "sweet spot", when overall levels of discrimination are not too strong (otherwise no-one is privately pro-trans), and not too weak (otherwise pro-trans people don't need to be vocal, or anti-trans people might speak up in a form of backlash).

The model provides a potential framework for thinking about when horizontal communication can be helpful, but also makes clear that horizontal communication can also be harmful when groups are outside the sweet spot. This implies that it is important to do further research on how to identify this sweet spot, i.e., identify cases where the pro-minority individuals are likely to be more vocal. Once these cases are identified, my work implies that we should then design decision environments (e.g. hiring practices) to involve groups rather than individuals, and develop discussion-based sensitization programs, similar to those seen in Dhar et al. (2022) and Broockman & Kalla (2016).

An alternative, more directed approach suggested by the results is to actively seek out *positive deviants* – individuals willing to act in defiance of a harmful social norm – and give them the tools needed to encourage prosocial behavior among their peers. This is the approach we evaluate in the second chapter. We partner with a large NGO in Madagascar to evaluate their "Young Girl Leader" program, which identifies, nominates, and trains girls in rural schools to act as ambassadors for hygiene, menstrual hygiene, and to proactively reduce the stigma that surrounds menstruation in this context. These girls are selected based on whether they are positive deviants, i.e., are already willing to speak openly and positively about menstruation in defiance of menstrual stigma. We find that they are highly effective at spreading knowledge and behavior about hygiene and menstrual hygiene, and at reducing a variety of measures of the stigma that surrounds menstruation. Our results imply that positive deviants can be an effective means for generating widespread improvements in prosocial behaviour in a policy-relevant setting.

The second theme of my work is human capital accumulation in lower- and middle-income countries (LMICs), and how socioeconomic circumstance can affect that human capital accumulation. We examine this in the second chapter by also evaluating the effect of a bundle of hygiene and menstrual hygiene interventions on human capital outcomes for girls. We find, strikingly, that these hygiene interventions generate an approximately 0.15 SD improvement in learning test scores for girls in treatment schools, an effect size that is comparable to successful programs in LMICs that specifically focus on education and learning (Muralidharan & Sundararaman, 2015). There are no measurable effects on school attendance or self-reported health, indicating that these are not the main channels of the effects. Instead, we find that the learning improvements are likely to be driven by psychosocial channels: girls in treatment schools have lower heart rates during the endline survey, implying lower levels of stress; they are less exposed to severe bullying; and they have more network connections with other students in schools. The program, which included both "hard" interventions (e.g., building sanitary infrastructure and distributing menstrual pads) and "soft" interventions (teacher

sensitization on hygiene and menstruation) appears to have generated important improvements in the overall school environment. In this very low-income and constrained setting, a school-level intervention can plausibly crowd in people's excitement and motivation, leading to better concentration while at school, and thus improved learning. We believe these psychosocial channels are typically neglected in the literature on hygiene interventions, but may be an important source of benefits to such programs.

In the third chapter, I build on this focus on human capital by examining how early life shocks affect long-run human capital development. While the first 1000 days of life are often cited as a 'critical period' for child development, this notion has rarely been directly tested. I therefore use a setting where children are potentially subject to shocks in every year of their childhood, and estimate the impact of early life weather shocks on human capital outcomes for individuals born in rural Indonesia. I examine cognitive and socioemotional skills as my primary measures of human capital. I find that there is a strong critical period for these shocks at age 2 for cognitive development, but no evidence for a similar critical period for socioemotional development, and that the effects of the shock appear to be driven by changes in agricultural income and nutritional investment. My results suggest that the timing of shocks and interventions in childhood may be very important for determining their effects on human capital accumulation, and that the dynamics are strongly impacted by how parents respond to such shocks. Future research should therefore document carefully how parents respond to early childhood interventions, and how impacts may differ across different age groups.

# Chapter 1

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

## 1.1 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

[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).

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 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 (Department, 2022), 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

[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.

[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).

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 ($p<0.001$), implying that anti-transgender discrimination is reduced to 0 on average ($p$ of difference between transgender and non-transgender: 0.30). The effects are also partially persistent: 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 the 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 $\in [0.01, 0.12]$, depending on the specification), 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. ($p<0.001$, as measured by the predicted probability of selecting a transgender worker). 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

[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.

[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.

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 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 could 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 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

[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. This differs from a virtue-signalling motive, which I conceptualize as wanting to signal that one is not discriminatory (regardless of the preferences of others in one's group).

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, Jindani, et al., 2023; 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 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,

[8]Other literature also examines ways of exogenously changing social norms in order to reduce discrimination (Dhar et al., 2022; Jayachandran, 2021; Gómez et al., 2018; Beaman et al., 2009; La Ferrara et al., 2012; Banerjee, La Ferrara, & Orozco-Olvera, 2019; Andrew et al., 2022).

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, Lombardi, & Ramos (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.

## 1.2 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 particularly 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; Department, 2022; Mal, 2015; Ganju & Saggurti, 2017; Shaikh et al., 2016; Baba & Sogani, 2018; Chakrapani et al., 2004; V. Chakrapani et al., 2011; Sangama, 2015).

[9] Gulesci, Lombardi, & Ramos (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".

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 A.5). 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 (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]

## 1.3 Experimental design

### 1.3.1 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

[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.

[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.

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 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.

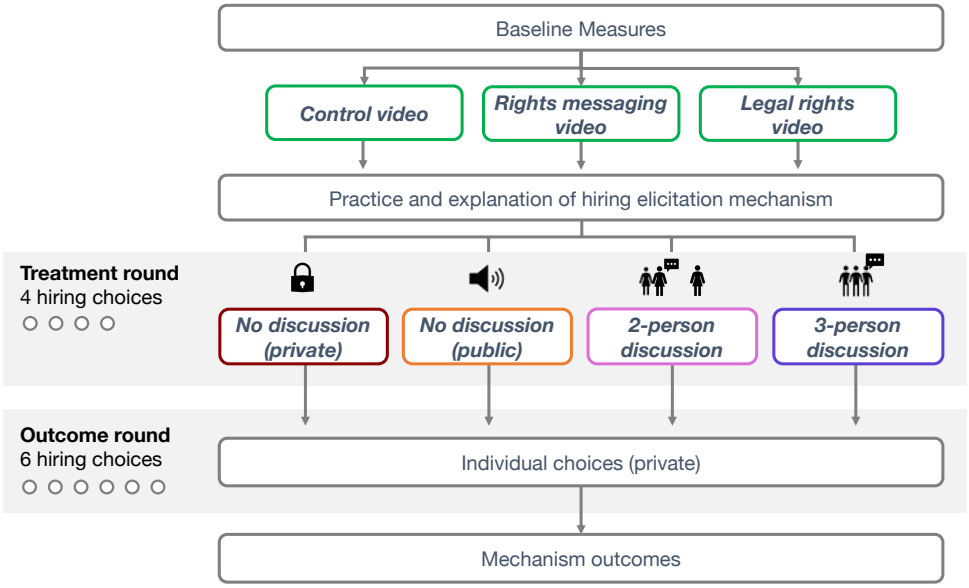
Figure 1.1 shows a summary of the experimental design for the main session (Figure A.1.1 gives further detail). I describe the design in more detail below.

### 1.3.2 Sample and recruitment

Participants were recruited from urban areas in Chennai between March and July 2023 (see Figure A.1.2 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, and the median per capita food expenditure in the sample (Rs. 2000 per month; 87.40 USD PPP) is approximately the same as a representative sample of urban Tamil Nadu residents from 2012 (National Sample Survey Office, India, 2012).

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

**Figure 1.1:** *Summary of experimental design*




*Notes*: More detail is given in Appendix Figure A.1.1.

group were recruited as a group of 3. All 3 members of a group were interviewed simultaneously.[13]

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%).

### 1.3.3 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 1.7).

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 1.2 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

[13] Participants willing to be recruited as a group may be more sociable or socially sensitive than the average urban resident of Chennai. However, this does not appear to moderate the treatment effects of the discussion: there is no heterogeneity by intragroup relations or an index of individual sociability (Table A.1.38).

preferred items.[14]

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 A.1.1).

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*" option who was either cisgender male, cisgender female, or transgender.[15] 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.[16]

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).[17] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

**Figure 1.2:** *Example of one of the binary choices participants face*

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 section A.4 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.[18]

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.

### 1.3.4 Treatments

*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

[18]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 section A.4 for discussion of the ethical considerations).

delivery options, regardless of treatment status.

In the treatment round, groups were randomized into one of the four conditions described 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.[19]

*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 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

[19] *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.

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 A.8 for the discussion script and further details).

*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.*"

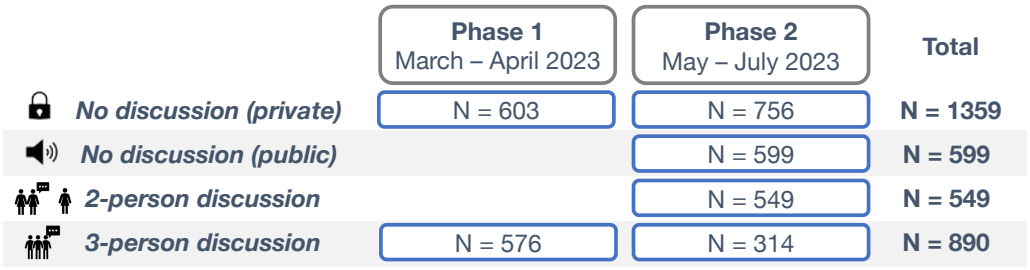
3. **Control video** ($N$=1135). Participants were not given information about the rights of transgender people. Instead, the video included placebo information about voting rights: *"As another example, some people have the right to vote. If you have the right to vote, you can elect your representatives. That means you can choose who should be in power and who should make decisions on your behalf.*"

Appendix A.9 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.

### 1.3.5 Data collection phases and samples

Data collection was divided into two phases (see Figure 2.1). 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

**Figure 1.3:** *Sample sizes and timeline*




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.[20]

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.

### 1.3.6 Pre-analysis plan

I preregistered the design of both phases, and document any deviations from the pre-analysis plan in Appendix A.7. 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 A.1.10 and Figure 1.4). 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.

### 1.3.7 Balance checks

For all samples and relevant treatment comparisons, the treatment groups were well-balanced across key characteristics (Tables A.1.3, A.1.4, and A.1.5). 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 A.1.3), and *rights messaging video* participants came from slightly larger households, with a slightly lower per capita food expenditure (Table A.1.5). I use LASSO to

[20]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 A.1.2).

select all controls that predict both treatment status and outcomes (as per Belloni et al. (2014), see Appendix A.10.9), so these imbalances are unlikely to affect my results.

### 1.3.8 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 1.7).

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:

$$\begin{aligned} ChooseAlternative_{ijk} = &\sum_{\tau \in \mathcal{T}} \beta_\tau \left( Treat_{\tau ij} \times Trans_{ijk} \right) + \gamma Trans_{ijk} + \sum_{\tau \in \mathcal{T}} \delta_\tau Treat_{\tau ij} \\ &+ \mathbf{X}'_{ijk}\Gamma_0 + (\mathbf{X}'_{ijk}\Gamma_1 \times Trans_{ijk}) + \varepsilon_{ijk} \end{aligned} \tag{1.1}$$

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 $\tau$, 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 A.1.10 and Figure 1.4.
- $\mathbf{X}_{ijk}$ 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 A.10.9 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 $\beta_\tau$, which describes the reduction in discrimination caused by the treatment. When interacted controls are not included, $\gamma$ 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.

### 1.3.9 Mechanism outcomes

In addition to the main outcome variable, I elicited other measures designed to understand the mechanisms behind the results (see Appendix A.10 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.[21]

**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.[22]

**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

[21] 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.

[22] We also asked for consent to record the audio of the discussion. Consent was refused in 16% of discussions. Table A.1.6 shows that the treatment effects of the 3-person discussion are not significantly different for groups that refused and consented to the audio recording.

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 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).

## 1.4 Results

### 1.4.1 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 1.1 and Figure 1.4). 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 1.1, 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 A.1.7), and to dropping the 6% of cases where the outcome round was overheard by neighbors (Table A.1.8). 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 A.1.9). 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.

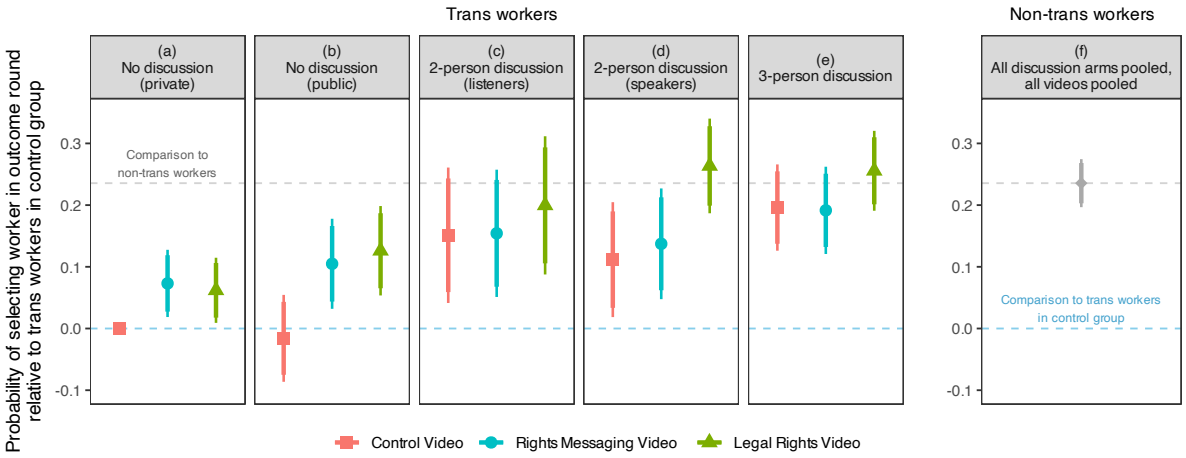
Relatedly, there is no direct evidence of interactions between the *legal rights* video and discussions (Figure 1.4 and Table A.1.10), but weak evidence of a negative interaction effect between the *rights messaging* video and the discussions, suggesting that these two interventions are substitutes. Throughout the paper, I present results that control for uninteracted treatments. The coefficients on the discussion should therefore be interpreted as conditional on the distribution of the other video

**Table 1.1:** *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 1.1. 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 A.10.9). 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 1.4:** *Summary of results: effect of all treatment arms on probability of selecting worker in the outcome round*




*Notes*: Shows the probability of selecting a worker in the outcome round, relative to the probability of selecting a transgender worker in the *No discussion (private)*, *Control video* arm. Panels (a) to (e) show the probability of selecting a transgender worker in each treatment arm. Panel (f) shows the probability of selecting a non-transgender *alternative* worker, pooling all treatment arms. 95% confidence intervals are base on standard errors clustered at the group-of-3 level. Controls include stratum fixed effects; whether 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 number of items offered; relative reliability score; whether the relative reliability score was shown; and the controls selected by double LASSO (see Section A.10.9). Unit of observation is the participant × choice level. As a placebo test, the effect of each treatment arm on the probability of selecting *non-transgender* alternative workers is seen in Appendix Figure A.1.3.

treatment.[23]

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 choosing a transgender worker (Figure A.1.4). 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 A.2, 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 A.1.11, Figure A.1.5). In the *No discussion (private)* arm, even when shown a transgender worker that is dominating on items or reliability score, or both, participants still select the non-transgender worker 47% of the time. By contrast, in the *3-person discussion* arm, this figure has reduced to 29% (difference: 17.9 p.p., $p$<0.001).

The pattern of choices indicates that participants traded off a preference for avoiding transgender

[23]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). 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.

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 1.1, column 2). And people were less sensitive to items when shown a transgender person (Table A.1.12, column 1), a result that holds for both treatment conditions (Table A.1.12, column 2).[24]

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 1.4, 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 A.1.13, column 1). Anti-transgender discrimination in the control group therefore appears to be partially driven by a form of inaccurate statistical discrimination in this context (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 1.4, 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 A.1.13, column 2). While the point estimate of the interaction of (*Worker is trans* × *Reliability score is shown* × *3-person discussion*) is negative and large enough to negate the effect of (*Worker is trans* × *Reliability score*), I cannot reject that it is different from 0 ($p$=0.24).

A heterogeneity analysis (Table A.1.14) 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).[25]

Finally, the discussion consistently reduces discrimination across all 13 transgender worker photos used (Figure A.1.7). 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 the opposite direction for a subset of photos – something that is not observed empirically.

[24]The fact that the sensitivity to items did not vary across treatment conditions (Table A.1.12, 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.

[25]There is heterogeneity in levels separating the analysis of non-transgender workers into males and females (Figure A.1.6). 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).

### 1.4.2 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 1.2 and Figure 1.4). 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.[26] 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 $\in [0.002, 0.04]$).

**Table 1.2:** *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 | 20382 | 20382 | 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 1.1.

I do not find interaction effects between the *legal rights* video and group discussions (Table A.1.10 and Figure 1.4); 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

[26] 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 A.1.15).

content.

### 1.4.3 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,[27] and there was no evidence of differential attrition (Table A.1.16). 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 1.3, panel A). Participants were approximately 4 p.p. more likely to select transgender workers in the hypothetical follow-up choices ($p$=0.03). Approximately 25% of the short-run effect thus remains after around 1 month. This is comparable to persuasion decay rates seen in the political science literature, such as Hill et al. (2013), who estimate approximately 10-15% of the persuasive effects of US TV political advertisements on voting remain after a similar time lag.

[27] 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 1.3:** *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 1.1 and 1.2. 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 1.1 and 1.2.

## 1.5 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 A.10 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 1.4), 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 A.1.19, Table A.1.20).

### 1.5.1 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 1.4, 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 1.4, 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 A.1.18).

**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 A.10.5 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 did not have a significant effect on the proportion of people who agreed with the sensitive anti-transgender statement in the list experiment (Table 1.4, 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 1.4, panel A, column 2). The probability

**Table 1.4:** *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 A.10.9). 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.

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 1.4, 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).

### 1.5.2 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 A.1.19), 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 A.1.20, 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 A.1.20, columns 1-2).

### 1.5.3 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 A.1.21 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 work-

[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.

ers 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 A.1.9).

**Reasons cited in discussions.** Figure A.1.10 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 A.1.11). 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 A.1.12).

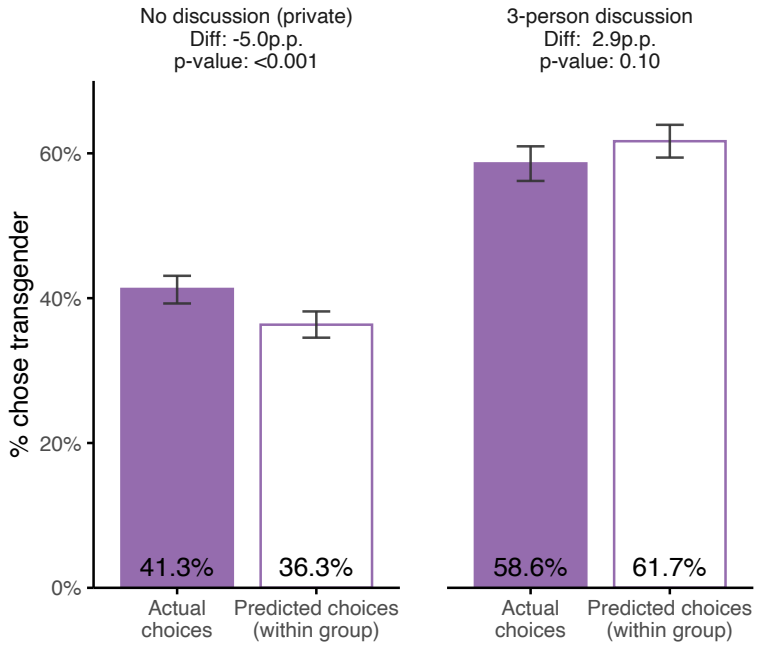
## 1.6 Mechanisms: What is behind the effect of the discussion?

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 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.

**Figure 1.5:** *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).

### 1.6.1 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 1.5 examines this hypothesis by displaying participants' *within-group* predictions about others' private choices (described above in Section 1.5), 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

[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 1.5 shows that this does not fit the data.

would be the case if the control group's misperceptions were simply 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.

### 1.6.2 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 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 alone 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 A.1.22, $p$=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 A.1.23). There were also small or null effects on the *outcome* round (Table 1.5).[32]

### 1.6.3 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.

*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 1.5). This effect is not significantly different from the effect of speaking

[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 A.1.8, 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 1.5).

[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 the *No discussion (public)* arm has no effect even when restricting to choices where there is little plausible deniability for participants, namely, when transgender workers offer more items ($\beta$=0.01, $p$=0.82). I also cannot rule out that virtue signalling *in combination* with other mechanisms contributes to the discussion's effects, e.g., if pro-trans participants are initially more vocal for other reasons and subsequently induce other participants to virtue-signal.

[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 ($p \in [0.037, 0.096]$).

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]

**Table 1.5:** *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 1.1, and is otherwise the same as Tables 1.1 and 1.2.

*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

[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 1.7 and Table A.1.25). I can also rule out that the listener effect is due to a correction of misperceptions, as per the arguments in Section 1.6.1.

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 A.1.24). 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 1.5.3 (for example, statements about transgender workers were 5.7x more likely to say something positive than to say something negative).

### 1.6.4 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 A.1.23). 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]

*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

[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 A.1.13), 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.

$R > 0$. 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 $\alpha$, where $\alpha$ 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 $\alpha$.[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 an action to maximize expected utility, taking as given the strategies of others $\sigma_{-i}$:

$$\begin{aligned}\max_{Y_i, S_i} \mathbb{E}_i[U_i(Y_i, S_i | P_i, \sigma_{-i}))] = V(Y_i) + P_i \cdot \mathbb{1}\{Y_i = 1\} + \gamma_0 \, \mathbb{E}_{-i}[\tilde{P}_i | Y_i, S_i, \sigma_{-i}] \\ - \gamma_1 \Big( \mathbb{E}_{-i}[\tilde{P}_i | Y_i, S_i, \sigma_{-i}] - \mathbb{E}_i[\tilde{P}_{-i} | S_i, \sigma_{-i}] \Big)^2 \\ - c \cdot \mathbb{1}\{S_i \in \{-1, 1\}\}\end{aligned} \tag{1.2}$$

$V(Y_i) \in \mathbb{R}$ 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 $\alpha$). 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 $i$.

*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).

[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.

[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 A.6.2), explaining the null result on the *No discussion (public)* treatment in the data.

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) 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 A.6. 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 A.1.14 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 $\mu_P$ is low. When $\mu_P$ 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.

*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:

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

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

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 $\mu_P$ 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 have the same discriminatory preferences.[40] The value of $\mu_P$ 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 $\mu_P$ that maximizes the number of people sending pro-trans messages *without* inducing anti-trans participants to also start persuading (see Figure A.1.15).

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. The model equilibrium in the sweet spot also finds suggestive support in the data: only 9% of discussions include negative mentions of transgender workers (Figure A.1.9), while the 49% of discussions in which there are positive mentions of transgender workers matches well with the 53% of control groups with at least one privately pro-trans member who always selects transgender workers.

## 1.7 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 A.10.7) 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.

[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 $\tilde{Y}_i$.

[40] I also show that a "NoSend-NoSend" equilibrium exists when the cost of communication $c$ is sufficiently high (Appendix A.6). 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 $c$.

The 3-person discussion still reduced discrimination for the private grocery pick-up choices (Table A.1.25).[41] The discussion treatment effect on this outcome is large, although slightly 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 A.1.26). 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 A.1.16, 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 A.1.11, 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 A.1.27, 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 A.1.28). 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 A.1.29, column 1), and the effect on discrimination is not significantly stronger for participants who remembered the word transgender (Table A.1.30,

[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.

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 A.1.29, columns 2 and 3; Figure A.1.17), and there is no detectable difference in the treatment effects for people that do and do not correctly guess the study purpose (Table A.1.30, columns 1 and 2). While the rights videos did increase the likelihood of a participant correctly guessing the purpose of the experiment (Table A.1.31), those who correctly guessed did not drive the reductions in discrimination seen in the discussion groups (Table A.1.32, columns 1 and 2). These tests do not fully rule out *subconscious* demand effects, but the *Legal rights video* likely represents the upper bound on such demand effects, and has a substantially smaller treatment effect than the *3-person discussion*, suggesting that the latter's effects are not driven by experimenter demand.

**Social desirability bias**. To measure a participant's propensity to give socially desirable answers, at baseline I elicited a shortened version of the 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 A.10.3). 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 A.1.30, column 3; Table A.1.32, 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 A.1.33, 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 A.1.34). 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 be-

tween 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.

## 1.8 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 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

[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.

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.

Aboud, F. E., & Singla, D. R. (2012, August). Challenges to changing health behaviours in developing countries: A critical overview. *Social Science & Medicine (1982)*, *75*(4), 589–594. doi: 10.1016/j.socscimed.2012.04.009

Adhvaryu, A., Fenske, J., & Nyshadham, A. (2019, August). Early Life Circumstance and Adult Mental Health. *Journal of Political Economy*, *127*(4), 1516–1549. doi: 10.1086/701606

Adhvaryu, A., Nyshadham, A., Molina, T., & Tamayo, J. (2018). Helping Children Catch Up: Early Life Shocks and the PROGRESA Experiment. *NBER Working Paper Series*, 1–75.

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

Aiyar, A., & Cummins, J. R. (2017). Age-Profile Estimates of the Relationship Between Economic Growth and Child Health. , 46.

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

Akyeampong, K., Andrabi, T., Banerjee, A., Banerji, R., Dynarski, S., Glennerster, R., … Yoshikawa, H. (2023). *Cost-Effective Approaches to Improve Global Learning - What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries?* London, Washington D.C., New York: FCDO, the World Bank, UNICEF, and USAID.

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021, November). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics*, *136*(4), 2147–2194. doi: 10.1093/qje/qjab009

Alan, S., & Kubilay, E. (2024). Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting. *European University Institute, mimeo*.

Ali, T. S., & Rizvi, S. N. (2010, August). Menstrual knowledge and practices of female adolescents in urban Karachi, Pakistan. *Journal of Adolescence*, *33*(4), 531–541. doi: 10.1016/j.adolescence.2009.05.013

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

Almond, D., Currie, J., & Duque, V. (2018, December). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. doi: 10.1257/jel.20171164

Almond, D., & Mazumder, B. (2013). Fetal Origins and Parental Responses. *Annual Review of Economics*, *5*(1), 37–56. doi: 10.1146/annurev-economics-082912-110145

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

Ashraf, N., Bau, N., Low, C., & McGinn, K. (2020). Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment*. *The Quarterly Journal of Economics*. doi: 10.1093/qje/qjz039

Attanasio, O. (2015, December). The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies. *Journal of the European Economic Association*, *13*(6), 949–997. doi: 10.1111/jeea.12159

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., & Rubio-Codina, M. (2020, January). Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia. *American Economic Review*, *110*(1), 48–85. doi: 10.1257/aer.20150183

Attanasio, O., Meghir, C., & Nix, E. (2020, November). Human Capital Development and Parental Investment in India. *The Review of Economic Studies*, *87*(6), 2511–2541. doi: 10.1093/restud/rdaa026

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

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2020, January). Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. *American Economic Journal: Applied Economics*, *12*(1), 210–259. doi: 10.1257/app.20170416

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013, July). The Diffusion of Microfinance. *Science*, *341*(6144), 1236498–1236498. doi: 10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019, November). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *The Review of Economic Studies*, *86*(6), 2453–2490. doi: 10.1093/restud/rdz008

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

Barham, T., Macours, K., & Maluccio, J. A. (2013). Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *American Economic Review*, *103*(3), 467–471. doi: 10.1257/aer.103.3.467

Barham, T., Macours, K., & Maluccio, J. A. (2018). Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes. (August).

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., & Shapiro, J. S. (2016, February). Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy*, *124*(1), 105–159. doi: 10.1086/684582

Barreca, A. I. (2012, January). Climate change, humidity, and mortality in the United States. *Journal of Environmental Economics and Management*, *63*(1), 19–34. doi: 10.1016/j.jeem.2011.07.004

Bazzi, S., & Gudgeon, M. (2016). Local Government Proliferation , Diversity, and Conflict.

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021, June). Can Network Theory-Based Targeting Increase Technology Adoption? *American Economic Review*, *111*(6), 1918–1943. doi: 10.1257/aer.20200295

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.

Beguería, S., Vicente-Serrano, S. M., Reig, F., & Latorre, B. (2014, August). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, *34*(10), 3001–3023. doi: 10.1002/joc.3887

Beguería, S., & Vicente Serrano, S. M. (2017). *SPEIbase v.2.5 Global 04-month 1901-2015 SPEI, 2017*. http://digital.csic.es/handle/10261/153475?locale=en. doi: 10.20350/digitalCSIC/8508

Belachew, T., Hadley, C., Lindstrom, D., Getachew, Y., Duchateau, L., & Kolsteren, P. (2011, September). Food insecurity and age at menarche among adolescent girls in Jimma Zone Southwest Ethiopia: A longitudinal study. *Reproductive biology and endocrinology: RB&E*, *9*, 125. doi: 10.1186/1477-7827-9-125

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

Benjamin, D. J., Debnam, J., Fleurbaey, M., Heffetz, O., & Kimball, M. (2020). What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents. , 41.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Benshaul-Tolonen, Anja, Aja, Zulaika, G., Sommer, M., & Phillips-Howard, P. A. (2020). Measuring menstruation-related absenteeism among adolescents in low-income countries. In *The Palgrave Handbook of Critical Menstruation Studies* (pp. 705–723). Palgrave Macmillan.

Benshaul-Tolonen, A., Zulaika, G., Nyothach, E., Odour, C., Mason, L., Obor, D., & Laserson, K. F. (2021). *Sanitary products, absenteeism and psychosocial well-being: Evidence from a three-arm cluster randomized controlled feasibility study in Western Kenya* (Working Paper No. WP No.93). Columbia University, SIPA, CDEP-CGEG.

Bergstrom, K., & Ozler, B. (2021). *Improving the Well-Being of Adolescent Girls in Developing Countries*. The World Bank. doi: 10.1596/1813-9450-9827

Bernard, T., Dercon, S., Orkin, K., & Taffesse, A. S. (2019, May). Parental Aspirations for Children's Education: Is There a "Girl Effect"? Experimental Evidence from Rural Ethiopia. *AEA Papers and Proceedings*, *109*, 127–132. doi: 10.1257/pandp.20191072

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

Bharadwaj, P., Løken, K. V., & Neilson, C. (2013, August). Early Life Health Interventions and Academic Achievement. *American Economic Review*, *103*(5), 1862–1891. doi: 10.1257/aer.103.5.1862

Bharati, T., & Chin, S. (2016). Does Education Affect Time Preference? *SSRN Electronic Journal*. doi: 10.2139/ssrn.2880879

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. doi: 10.1093/acprof:oso/9780190622046.001.0001

Bicchieri, C., & Funcke, A. (2018, March). Norm Change: Trendsetters and Social Structure. *Social Research: An International Quarterly*, *85*(1), 1–21. doi: 10.1353/sor.2018.0002

Binswanger, H. P. (1980, August). Attitudes Toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics*, *62*(3), 395–407. doi: 10.2307/1240194

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

Breza, E. (2016, April). Field Experiments, Social Networks, and Development. In Y. Bramoullé, A. Galeotti, & B. W. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks* (pp. 411–439). Oxford University Press. doi: 10.1093/oxfordhb/9780199948277.013.25

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

Brown, S., & Taylor, K. (2008). Bullying, education and earnings: Evidence from the National Child Development Study. *Economics of Education Review*, *27*(4), 387–401.

Buehren, N., Goldstein, M. P., Gulesci, S., Sulaiman, M., & Yam, V. (2017, February). Evaluation of an adolescent development program for girls in Tanzania. *Policy Research Working Paper Series*(7961).

Burgess, R., Deschênes, O., Donaldson, D., & Greenstone, M. (2017). Weather, climate change and death in India. *Unpublished Manuscript*, 71.

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

Buss, A. T., Fox, N., Boas, D. A., & Spencer, J. P. (2014, January). Probing the early development of visual working memory capacity with functional near-infrared spectroscopy. *NeuroImage*, *85*, 314–325. doi: 10.1016/j.neuroimage.2013.05.034

Buss, A. T., & Spencer, J. P. (2014, June). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii, 1–103. doi: 10.1002/mono.12096

Buss, A. T., & Spencer, J. P. (2018, July). Changes in frontal and posterior cortical activity underlie the early emergence of executive function. *Developmental Science*, *21*(4), e12602. doi: 10.1111/desc.12602

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

Carneiro, P., Kraftman, L., Mason, G., Moore, L., Rasul, I., & Scott, M. (2021, August). The Impacts of a Multifaceted Prenatal Intervention on Human Capital Accumulation in Early Life. *American Economic Review*, *111*(8), 2506–2549. doi: 10.1257/aer.20191726

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

Cattell, R. B. (1966, April). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. doi: 10.1207/s15327906mbr0102_10

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

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2018, June). *Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India* (Working Paper No. 24678). National Bureau of Economic Research. doi: 10.3386/w24678

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

Chirgwin, H., Cairncross, S., Zehra, D., & Sharma Waddington, H. (2021, December). Interventions promoting uptake of water, sanitation and hygiene (WASH) technologies in low- and middle-income countries: An evidence and gap map of effectiveness studies. *Campbell Systematic Reviews*, *17*(4), e1194. doi: 10.1002/cl2.1194

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

Chuang, Y., & Schechter, L. (2015, November). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, *117*, 151–170. doi: 10.1016/j.jdeveco.2015.07.008

Clark, L. A., & Watson, D. (1995, September). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. doi: 10.1037/1040-3590.7.3.309

Cordero, M. E., D'Acuña, E., Benveniste, S., Prado, R., Nuñez, J. A., & Colombo, M. (1993). Dendritic development in neocortex of infants with early postnatal life undernutrition. *Pediatric Neurology*, *9*(6), 457–464. doi: 10.1016/0887-8994(93)90025-8

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*.

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, *97*(2), 25.

Cunha, F., & Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, *43*(4), 738–782. doi: 10.3368/jhr.43.4.738

Cunha, F., & Heckman, J. J. (2009, April). The Economics and Psychology of Inequality and Human DEvelopment. *Journal of the European Economic Association*, *7*(2-3), 320–364. doi: 10.1162/JEEA.2009.7.2-3.320

Cunha, F., Heckman, J. J., & Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931. doi: 10.3982/ECTA6551

Currie, J., & Almond, D. (2011). Human capital development before age five. In *Handbook of Labor Economics* (Vol. 4, pp. 1315–1486). Elsevier B.V. doi: 10.1016/S0169-7218(11)02413-0

Czura, K., Menzel, A., & Miotto, M. (2019). Menstrual Health, Worker Productivity and Well-being among Female Bangladeshi Garment Workers. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3519895

Dal Bó, E., Finan, F., Li, N. Y., & Schechter, L. (2018). Government Decentralization Under Changing State Capacity: Experimental Evidence From Paraguay. , 66.

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

Das, J., Do, Q. T., Friedman, J., & McKenzie, D. (2009). Mental health patterns and consequences: Results from survey data in five developing countries. *World Bank Economic Review*, *23*(1), 31–55. doi: 10.1093/wber/lhn010

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 Chaisemartin, C., & D'Haultfœuille, X. (2018, April). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, *85*(2), 999–1028. doi: 10.1093/restud/rdx049

de Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way Fixed Effects Regressions with Several Treatments. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3751060

de Quidt, J., & Haushofer, J. (2019). Depression through the Lens of Economics. In *The Economics of Poverty Traps* (pp. 127–152). doi: 10.7208/chicago/9780226574448.003.0003

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

Dean, J. T., & Jayachandran, S. (2020). Attending kindergarten improves cognitive development in India, but all kindergartens are not equal. , 64.

De Datta, S. K. (1981). *Principles and practices of rice production*. New York: Wiley.

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

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–798. doi: 10.1257/jel.52.3.740

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

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, *1*(3), 111–34. doi: 10.1257/app.1.3.111

Deoni, S., Dean, D., Joelson, S., O'Regan, J., & Schneider, N. (2018, September). Early nutrition influences developmental myelination and cognition in infants and young children. *NeuroImage*, *178*, 649–659. doi: 10.1016/j.neuroimage.2017.12.056

Department, U. S. S. (2022). *2022 Country Reports on Human Rights Practices*.

Deschênes, O., & Greenstone, M. (2011, October). Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US. *American Economic Journal: Applied Economics*, *3*(4), 152–185. doi: 10.1257/app.3.4.152

Deschênes, O., Greenstone, M., & Guryan, J. (2009, April). Climate Change and Birth Weight. *American Economic Review*, *99*(2), 211–217. doi: 10.1257/aer.99.2.211

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, March). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, *112*(3), 899–927. doi: 10.1257/aer.20201112

Dinarte, L., & Egana del Sol, P. (2018). Preventing Violence in the Most Violent Contexts. Behavioral and Neurophysiological Evidence. (May).

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

Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. In *Handbook of economic field experiments* (Vol. 2, pp. 3–93). Elsevier.

Edmonds, E., Feigenberg, B., & Leight, J. (2020, July). *Advancing the Agency of Adolescent Girls* (Tech. Rep. No. w27513). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w27513

El-Gilany, A.-H., Badawi, K., & El-Fedawy, S. (2005, January). Menstrual Hygiene among Adolescent Schoolgirls in Mansoura, Egypt. *Reproductive Health Matters*, *13*(26), 147–152. doi: 10.1016/S0968-8080(05)26191-8

Eriksen, T. L. M., Nielsen, H. S., & Simonsen, M. (2014). Bullying in Elementary School. *Journal of Human Resources*, *49*(4), 839–871.

Evans, D. K., & Yuan, F. (2022, February). What We Learn about Girls' Education from Interventions That Do Not Focus on Girls. *The World Bank Economic Review*, *36*(1), 244–267. doi: 10.1093/wber/lhab007

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

Fernández, R., Parsa, S., & Viarengo, M. (2019). Coming out in America: AIDS, Politics, and Cultural Change. , 65.

Ferrer, E. (2009, May). Fluid reasoning and the developing brain. *Frontiers in Neuroscience*, *3*(1). doi: 10.3389/neuro.01.003.2009

Fishman, R. (2016, February). More uneven distributions overturn benefits of higher precipitation for crop yields. *Environmental Research Letters*, *11*(2), 024004. doi: 10.1088/1748-9326/11/2/024004

Fishman, R. (2018, March). Groundwater depletion limits the scope for adaptation to increased rainfall variability in India. *Climatic Change*, *147*(1-2), 195–209. doi: 10.1007/s10584-018-2146-x

Fiske, A., & Holmboe, K. (2019, June). Neural substrates of early executive function development. *Developmental Review*, *52*, 42–62. doi: 10.1016/j.dr.2019.100866

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

Food and Agriculture Organisation. (2005). *Fertilizer use by crop in Indonesia.*

Frankenberg, E., & Karoly, L. (1995). *The 1993 indonesian family life survey: Overview and field report, publication no* (Tech. Rep.). DRU-1195/1-NICHD/AID, RAND, Santa Monica, CA.

Frederick, W. H., Worden, R. L., & of Congress. Federal Research Division, L. (2011). *Indonesia: A Country Study*. Federal Research Division, Library of Congress.

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

Garces, E., Thomas, D., & Currie, J. (2002). Longer-Term Effects of Head Start. *The American Economic Review*, *92*(4), 32.

Gertler, P., Heckman, J. J., Pinto, R., Chang, S. M., Grantham-McGregor, S., Vermeersch, C., ... Wright, A. (2021). Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31. , 50.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018, March). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, *19*(3), 123–137. doi: 10.1038/nrn.2018.1

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

Goodman-Bacon, A. (2018, September). *Difference-in-Differences with Variation in Treatment Timing* (Tech. Rep. No. w25018). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w25018

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

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007,

January). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60–70. doi: 10.1016/S0140-6736(07)60032-4

Grantham-McGregor, S., Powell, C., Walker, S., & Himes, J. (1991, July). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican Study. *The Lancet*, *338*(8758), 1–5. doi: 10.1016/0140-6736(91)90001-6

Groh, M., McKenzie, D., & Vishwanath, T. (2015). Reducing Information Asymmetries in the Youth Labor Market of Jordan with Psychometrics and Skill Based Tests. *The World Bank Economic Review*, *29*(suppl 1), S106-S117. doi: 10.1093/wber/lhv005

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., & Young, H. (2023). A Stepping Stone Approach to Norm Transitions. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4425503

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

Gunnsteinsson, S., Adhvaryu, A., Christian, P., Labrique, A., Sugimoto, J., Shamim, A. A., & West, K. P. (2014). Resilience to early life shocks: Evidence from the Interaction of a Randomized Controlled Trial and a Natural Experiment. *Working Paper*, 1–48.

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

Hamory, J., Baird, S., Das, S., Jones, N., Woldehanna, T., & Yadete, W. (2023). Do layered adolescent-centric interventions improve girls' capabilities? Evidence from a mixed-methods cluster randomised controlled trial in Ethiopia.

Hanson, J. L., Hair, N., Shen, D. G., Shi, F., Gilmore, J. H., Wolfe, B. L., & Pollak, S. D. (2013, December). Family Poverty Affects the Rate of Human Infant Brain Growth. *PLoS ONE*, *8*(12), e80954. doi: 10.1371/journal.pone.0080954

Harari, M., & Ferrara, E. L. (2018, October). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics*, *100*(4), 594–608. doi: 10.1162/rest_a_00730

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

Haushofer, J., & Fehr, E. (2014, May). On the psychology of poverty. *Science*, *344*(6186), 862–867. doi: 10.1126/science.1232491

Heckman, J. J. (2006, June). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(5782), 1900–1902. doi: 10.1126/science.1128898

Heckman, J. J., Pinto, R., & Savelyev, P. (2013, October). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, *103*(6), 2052–2086. doi: 10.1257/aer.103.6.2052

Heckman, J. J., Stixrud, J., & Urzua, S. (2006, February). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. , w12006. doi: 10.3386/w12006

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

Hennegan, J., & Montgomery, P. (2016, February). Do Menstrual Hygiene Management Interventions Improve Education and Psychosocial Outcomes for Women and Girls in Low and Middle Income Countries? A Systematic Review. *PLOS ONE*, *11*(2), e0146985. doi: 10.1371/journal.pone.0146985

Hill, S. J., Lo, J., Vavreck, L., & Zaller, J. (2013, October). How Quickly We Forget: The Duration of Persuasion Effects From Mass Communication. *Political Communication*, *30*(4), 521–547. doi: 10.1080/10584609.2013.828143

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

Hochberg, Z., & Belsky, J. (2013, April). Evo-devo of human adolescence: Beyond disease models of early puberty. *BMC medicine*, *11*, 113. doi: 10.1186/1741-7015-11-113

Horn, J. L. (1965, June). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi: 10.1007/BF02289447

Hoynes, H., Schanzenbach, D. W., & Almond, D. (2016, April). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review*, *106*(4), 903–934. doi: 10.1257/aer.20130375

Hu, Q. (2023). Breaking Down Bullying: Empathy, Social Networks, and Adolescents.

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

Isaacs, E., Oates, J., & ILSI Europe a.i.s.b.l. (2008, August). Nutrition and cognition: Assessing cognitive abilities in children and young people. *European Journal of Nutrition*, *47*(S3), 4–24. doi: 10.1007/s00394-008-3002-y

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. (2021, September). Social Norms as a Barrier to Women's Employment in Developing Countries. *IMF Economic Review*, *69*(3), 576–595. doi: 10.1057/s41308-021-00140-w

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

Johnson, R. C., & Jackson, C. K. (2019, November). Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal: Economic Policy*, *11*(4), 310–349. doi: 10.1257/pol.20180510

Kaiser, H. F. (1958, September). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. doi: 10.1007/BF02289233

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

Katz, D., Allport, F. H., & Jenness, M. B. (1931). *Students' attitudes; a report of the Syracuse University reaction study*. Oxford, England: Craftsman Press.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (in press). Do financial concerns make workers less productive? *The Quarterly Journal of Economics*.

Kennedy, G., Ballard, T., Dop, M. C., & European Union. (2011). *Guidelines for measuring household and individual dietary diversity*. Rome: Food and Agriculture Organization of the United Nations.

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

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007, January). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119. doi: 10.1111/j.1468-0262.2007.00733.x

Knudsen, E. I. (2004, October). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, *16*(8), 1412–1425. doi: 10.1162/0898929042304796

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*(27), 10155–10162. doi: 10.1073/pnas.0600888103

Kudamatsu, M., Persson, T., & Strömberg, D. (2012). Weather and Infant Mortality in Africa. , 77.

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

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., ... Vakis, R. (2019, July). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, *5*(7), eaaw5226. doi: 10.1126/sciadv.aaw5226

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

Ledesma, R. D., Valero-Mora, P., & de Valencia, U. (n.d.). Determining the Number of Factors to Retain in EFA: An easy-to- use computer program for carrying out Parallel Analysis. *Exploratory Factor Analysis*, *12*(2), 11.

Lee, C.-T., Zhang, G., & Edwards, M. C. (2012, March). Ordinary Least Squares Estimation of Parameters in Exploratory Factor Analysis With Ordinal Data. *Multivariate Behavioral Research*, *47*(2), 314–339. doi: 10.1080/00273171.2012.658340

Levine, D., & Yang, D. (2014, July). *The Impact of Rainfall on Rice Output in Indonesia* (Tech. Rep. No. w20302). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w20302

Little, D. R., Lewandowsky, S., & Craig, S. (2014, March). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00239

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

Maccini, S., & Yang, D. (2009, May). Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review*, *99*(3), 1006–1026. doi: 10.1257/aer.99.3.1006

Macours, K., Premand, P., & Vakis, R. (2022, September). Transfers, Diversification and Household Risk Strategies: Can Productive Safety Nets Help Households Manage Climatic Variability? *The Economic Journal*, *132*(647), 2438–2470. doi: 10.1093/ej/ueac018

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

Malamud, O., Pop-Eleches, C., & Urquiola, M. (2016, March). *Interactions Between Family and School Environments: Evidence on Dynamic Complementarities?* (Tech. Rep. No. w22112). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w22112

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

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., & Stein, A. D. (2009, April). The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults. *The Economic Journal*, *119*(537), 734–763. doi: 10.1111/j.1468-0297.2009.02220.x

Marsh, D. R., Schroeder, D. G., Dearden, K. A., Sternin, J., & Sternin, M. (2004, November). The power of positive deviance. *BMJ*, *329*(7475), 1177–1179. doi: 10.1136/bmj.329.7475.1177

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

May, W. (2004, March). Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment. *Climate Dynamics*, *22*(2-3), 183–204. doi: 10.1007/s00382-003-0373-x

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

Minnesota Population Center. (2018). *Integrated Public Use Microdata Series, International: Version 7.1. Indonesia Spatially Harmonized Second-Level Geography (1970-2011).* https://international.ipums.org/international/gis_harmonized_2nd.shtml. doi: 10.18128/D020.V7.1

Minnesota Population Center. (2020). *Integrated Public Use Microdata Series, International: Version 7.1. Indonesia Census Data (1980, 1985, 1990, 1995, 2000, 2005, 2010).* https://international.ipums.org/international/gis_harmonized_2nd.shtml. doi: 10.18128/D020.V7.1

Monfreda, C., Ramankutty, N., & Foley, J. A. (2008, March). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000: GLOBAL CROP AREAS AND YIELDS IN 2000. *Global Biogeochemical Cycles*, *22*(1), n/a-n/a. doi: 10.1029/2007GB002947

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

Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, *12*(2), 125–135. doi: 10.1037/h0027568

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. Henry Holt and Company.

Muralidharan, K. (2017). Field Experiments in Education in Developing Countries. In A. Banerjee & E. Duflo (Eds.), *Handbook of Field Experiments* (pp. 323–385). doi: 10.1016/bs.hefe.2016.09.004

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

Muralidharan, K., & Sundararaman, V. (2015, August). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India *. *The Quarterly Journal of Economics*, *130*(3), 1011–1066. doi: 10.1093/qje/qjv013

Myers, D. G., & Bishop, G. D. (1970, August). Discussion Effects on Racial Attitudes. *Science*, *169*(3947), 778–779. doi: 10.1126/science.169.3947.778

Myers, D. G., & Lamm, H. (1976, July). The group polarization phenomenon. *Psychological Bulletin*, *83*(4), 602–627. doi: 10.1037/0033-2909.83.4.602

National Sample Survey Office, India. (2012). *HCE: Monthly per Capita Consumer Expenditure: Average: Tamil Nadu: Urban: Food | Economic Indicators | CEIC*.

Naylor, R. L., Battisti, D. S., Vimont, D. J., Falcon, W. P., & Burke, M. B. (2007, May). Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proceedings of the National Academy of Sciences*, *104*(19), 7752–7757. doi: 10.1073/pnas.0701825104

Nelson, C. A., & Gabard-Durnam, L. J. (2020, March). Early Adversity and Critical Periods: Neurodevelopmental Consequences of Violating the Expectable Environment. *Trends in Neurosciences*, *43*(3), 133–143. doi: 10.1016/j.tins.2020.01.002

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

Orr, L., Ferris, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., ... Kling, J. R. (2003). *Moving to opportunity: Interim impacts evaluation* (Report). Washington D.C: U.S. Dept. of HUD.

Oster, E., & Thornton, R. (2011, January). Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics*, *3*(1), 91–100. doi: 10.1257/app.3.1.91

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

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021, January). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology*, *72*(1), 533–560. doi: 10.1146/annurev-psych-071620-030619

Paradkar, N., & Chowdhury, S. R. (2015). Primary study for detection of arterial blood pressure waveform components. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2015*, 1959–1962. doi: 10.1109/EMBC.2015.7318768

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

Perkins, J. M., Kim, R., Krishna, A., McGovern, M., Aguayo, V. M., & Subramanian, S. (2017, November). Understanding the association between stunting and child development in low- and middle-income countries: Next steps for research and intervention. *Social Science & Medicine*, *193*, 101–109. doi: 10.1016/j.socscimed.2017.09.039

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.

Phillips-Howard, P. A., Nyothach, E., Ter Kuile, F. O., Omoto, J., Wang, D., Zeh, C., ... Laserson, K. F. (2016, November). Menstrual cups and sanitary pads to reduce school attrition, and sexually transmitted and reproductive tract infections: A cluster randomised controlled feasibility study in rural Western Kenya. *BMJ open*, *6*(11), e013229. doi: 10.1136/bmjopen-2016-013229

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

Prado, E. L., & Dewey, K. G. (2014, April). Nutrition and brain development in early life. *Nutrition Reviews*, *72*(4), 267–284. doi: 10.1111/nure.12102

Prentice, A. M., Ward, K. A., Goldberg, G. R., Jarjou, L. M., Moore, S. E., Fulford, A. J., & Prentice, A. (2013, May). Critical windows for nutritional interventions against stunting. *The American Journal of Clinical Nutrition*, *97*(5), 911–918. doi: 10.3945/ajcn.112.052332

Radloff, L. S. (1977, June). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. doi: 10.1177/014662167700100306

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

Rammstedt, B., & John, O. P. (2007, February). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. doi: 10.1016/j.jrp.2006.02.001

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

Raven, J. (2000, August). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. doi: 10.1006/cogp.1999.0735

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

Reece, J. D., Bunn, J. A., Choi, M., & Navalta, J. W. (2021, June). Assessing Heart Rate Using Consumer Technology Association Standards. *Technologies*, *9*(3), 46. doi: 10.3390/technologies9030046

Reynolds, A. J., Temple, J. A., White, B. A. B., Ou, S.-R., & Robertson, D. L. (2011, January). Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program: Cost-Benefit Analysis. *Child Development*, *82*(1), 379–404. doi: 10.1111/j.1467-8624.2010.01563.x

Rietveld, C. A., van Kippersluis, H., & Thurik, A. R. (2015, October). Self-Employment and Health:

Barriers or Benefits?: SELF-EMPLOYMENT AND HEALTH. *Health Economics*, *24*(10), 1302–1313. doi: 10.1002/hec.3087

Ritchie, J. T., & Nesmith, D. S. (2015, October). Temperature and Crop Development. In J. Hanks & J. T. Ritchie (Eds.), *Agronomy Monographs* (pp. 5–29). Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. doi: 10.2134/agronmonogr31.c2

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

Roser, M., & Ritchie, H. (2019). Hunger and undernourishment. *Our World in Data*.

Rossin-Slater, M., & Wüst, M. (2016). *Are Different Early Investments Complements or Substitutes? Long-Run and Intergenerational Evidence from Denmark* (Tech. Rep.).

Rubio-Codina, M., Attanasio, O., Meghir, C., Varela, N., & Grantham-McGregor, S. (2015). The Socioeconomic Gradient of Child Development: Cross-Sectional Evidence from Children 6–42 Months in Bogota. *Journal of Human Resources*, *50*(2), 464–483. doi: 10.3368/jhr.50.2.464

Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010, June). Crop planting dates: An analysis of global patterns: Global crop planting dates. *Global Ecology and Biogeography*, no-no. doi: 10.1111/j.1466-8238.2010.00551.x

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

Sarzosa, M., & Urzúa, S. (2021). Bullying among adolescents: The role of skills. *Quantitative Economics*, *12*(3), 945–980. doi: 10.3982/QE1215

Schilbach, F., Schofield, H., & Mullainathan, S. (2016, May). The Psychological Lives of the Poor. *American Economic Review*, *106*(5), 435–440. doi: 10.1257/aer.p20161101

Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006, February). The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions. *Review of Economics and Statistics*, *88*(1), 113–125. doi: 10.1162/rest.2006.88.1.113

Schlenker, W., & Roberts, M. J. (2006, September). Nonlinear Effects of Weather on Corn Yields*. *Review of Agricultural Economics*, *28*(3), 391–398. doi: 10.1111/j.1467-9353.2006.00304.x

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

Schweinhart, L. J., & Weikart, D. P. (1981, October). Effects of the Perry Preschool Program on Youths Through Age 15. *Journal of the Division for Early Childhood*, *4*(1), 29–39. doi: 10.1177/105381518100400105

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

Shah, M., Seager, J., Montalvao, J., & Goldstein, M. (2023, August). *Sex, Power, and Adolescence: Intimate Partner Violence and Sexual Behaviors* (Working Paper No. 31624). National Bureau of Economic Research. doi: 10.3386/w31624

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

Shiller, R. J. (2017, April). Narrative Economics. *American Economic Review*, *107*(4), 967–1004. doi: 10.1257/aer.107.4.967

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.

Smith, L. G. E., & Postmes, T. (2011). The power of talk: Developing discriminatory group norms through discussion. *British Journal of Social Psychology*, *50*(2), 193–215. doi: 10.1348/014466610X504805

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

Statistik, B. P., & SIAK, K. (2019). Indonesia (IDN) Administrative Boundary Common Operational Database (COD-AB).

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (1998). The second wave of the indonesia family life survey (IFLS3): Overview and field report. *NIA/NICHD*.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (2004). The third wave of the indonesia family life survey (IFLS3): Overview and field report. *NIA/NICHD*.

Strauss, J., Witoelar, F., & Sikoki, B. (2016). *The fifth wave of the indonesia family life survey: Overview and field report; RAND labor and population.* WR-1143/1-NIA/NICHD.

Strauss, J., Witoelar, F., Sikoki, B., & Wattie, A. M. (2009). *The fourth wave of the Indonesia Family Life Survey: Overview and field report.* RAND Labor and Population Working Paper WR-675/1-NIA/NICHD. Santa Monica, CA ....

Subramanian, T., Gupte, M., Dorairaj, V., Periannan, V., & Mathai, A. (2009, April). Psycho-social 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

Tebaldi, C., Hayhoe, K., Arblaster, J. M., & Meehl, G. A. (2006, November). Going to the Extremes: An Intercomparison of Model-Simulated Historical and Future Changes in Extreme Events. *Climatic Change*, *79*(3-4), 185–211. doi: 10.1007/s10584-006-9051-4

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

van den Besselaar, E. J. M., van der Schrier, G., Cornes, R. C., Iqbal, A. S., & Klein Tank, A. M. G. (2017, July). SA-OBS: A Daily Gridded Surface Temperature and Precipitation Dataset for Southeast Asia. *Journal of Climate*, *30*(14), 5151–5165. doi: 10.1175/JCLI-D-16-0575.1

Velicer, W. F. (1976, September). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. doi: 10.1007/BF02293557

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010, April). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. doi: 10.1175/2009JCLI2909.1

Victora, C. G., de Onis, M., Hallal, P. C., Blossner, M., & Shrimpton, R. (2010, March). Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *PEDIATRICS*, *125*(3), e473-e480. doi: 10.1542/peds.2009-1519

Wachs, T. D., Georgieff, M., Cusick, S., & McEwen, B. S. (2014, January). Issues in the timing of integrated early interventions: Contributions from nutrition, neuroscience, and psychological research: Timing of integrated early interventions. *Annals of the New York Academy of Sciences*, *1308*(1), 89–106. doi: 10.1111/nyas.12314

Walker, S. P., Chang, S. M., Powell, C. A., & Grantham-McGregor, S. M. (2005, November). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The Lancet*, *366*(9499), 1804–1807. doi: 10.1016/S0140-6736(05)67574-5

Walker, S. P., Chang, S. M., Wright, A. S., Pinto, R., Heckman, J. J., & Grantham-McGregor, S. M. (2021, August). Cognitive, psychosocial, and behaviour gains at age 31 years from the Jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry*, jcpp.13499. doi: 10.1111/jcpp.13499

WaterAid. (2013). *We can't wait: A report on sanitation and hygiene for women and girls.* WaterAid International. Retrieved from: http://www.zaragoza.es/contenidos/medioambiente/onu/1325-eng_We_cant_wait_sanitation_and_hygiene_for_women_and%20girls.pdf.

Webb, D. (2023). Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India. *mimeo, Paris School of Economics*.

Welz, M., Alfons, A., Demirer, M., & Chernozhukov, V. (2022). GenericML: Generic machine learning inference. *R package version*.

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

White, M. D. (2018). The Problems with Measuring and Using Happiness for Policy Purposes. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3191385

WHO, & UNICEF. (2009). WHO child growth standards and the identification of severe acute malnutrition in infants and children. , 1–12. doi: http://www.who.int/nutrition/publications/severemalnutrition/9789241598163/en/

Wolke, D., Copeland, W. E., Angold, A., & Costello, E. J. (2013). Impact of Bullying in Childhood on Adult Health, Wealth, Crime, and Social Outcomes. *Psychological Science*, *24*(10), 1958–1970.

World Bank. (2018). *Rural Population in Indonesia*.

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

# Chapter 2

# Menstrual Stigma, Hygiene, and Human Capital: Experimental Evidence from Madagascar

**Joint with Karen Macours and Julieta Vera Rueda**

## 2.1 Introduction

Human capital development is strongly affected by psychosocial factors: for example, performance at school can be determined by socioemotional skills (Ashraf et al., 2020; Edmonds et al., 2020; Dinarte & Egana del Sol, 2018), social interactions between students (Rao, 2019; Hu, 2023), or parents' aspirations for their child (Bernard et al., 2019). One potentially important psychosocial barrier to girls' human capital is *menstrual stigma* – the social stigma surrounding menstruation that makes it a taboo topic to discuss, and leads to the marginalization of menstruating women. Such stigma is widespread, with one study estimating that 1 in 3 women across the world who menstruate risk shame and harassment (WaterAid, 2013). Stigma could negatively impact (i) health, by inhibiting demand for menstrual products or inhibiting the flow of information about optimal menstrual hygiene (El-Gilany et al., 2005; Ali & Rizvi, 2010); (ii) school attendance, by discouraging girls from attending school during menstruation; or (iii) learning, by increasing stress or anxiety when menstruating and thereby inhibiting concentration while at school or by affecting the overall school social environment. More broadly, interventions in school – even those targeted on specific behaviors like pedagogy or hygiene – might have important impacts on other aspects of the school psychosocial environment that can have knock-on effects on human capital (Alan et al., 2021; Alan & Kubilay, 2024). For example, hygiene programs might reduce the stress associated with illness, create a collaborative environment that motivates teachers, or change students' gender attitudes by focusing on reproductive health.

Relatively few studies have examined interventions specifically designed to reduce menstrual stigma for girls in school. And while several previous studies have examined interventions aiming to directly improve menstrual *hygiene*, or a broader set of hygiene behaviors, they have generally seen mixed or small impacts on human capital outcomes – perhaps because they do not focus on the psychosocial effects of stigma or on the social environment in school (Oster & Thornton, 2011; Hennegan & Montgomery, 2016; Benshaul-Tolonen et al., 2020; A. Benshaul-Tolonen et al., 2021; Chirgwin et

al., 2021).

In this paper, we present results from a randomized controlled trial in 140 schools in rural Madagascar evaluating a program run by our NGO partner, CARE. The program aims to reduce menstrual stigma and improve menstrual hygiene, while also improving a broader set of hygiene behaviors (such as hand-washing and latrine usage) – thereby improving girls' human capital outcomes.

In this context, menstruation is highly stigmatized and hygiene behaviors are typically poor. For example, the taboo that surrounds menstruation means that only 45% of girls in our baseline survey had heard a classmate speak about menstruation (compared to 70% about hand-washing), and 39% say that one *should not* discuss menstruation openly. At baseline, only 55% of girls had received information from their mother (or other caregiver) about menstruation. And only 11% of girls use sanitary pads, while the remainder typically improvise with old cloth or fabric. Knowledge of good hygiene practices is generally mixed at baseline (e.g., 98% know that the place to defecate is not outside, but only 63% can name 2 important moments in the day for hand-washing). But this knowledge does not always translate to correct behavior (e.g., only 32% washed their hands 3 times the last day they were at school, and only 14% cleaned themselves with soap the last time they menstruated at school).

Our experimental design evaluates the effects of two components of the program. First, we evaluate a bundle of interventions that simultaneously (i) engages and trains school teachers to promote hygiene and reduce menstrual stigma in schools, and (ii) relaxes hard constraints to good hygiene practices by building hygiene infrastructure (hand-washing basins and latrines) and distributing free menstrual pads in schools.

Second, we evaluate the additional effect of an innovative approach to addressing the harmful social norm that underlies menstrual stigma, namely, identifying and amplifying the voice of *positive deviants* (Marsh et al., 2004) – people embedded within a social network who are willing engage in prosocial behaviors in defiance of a harmful social norm. Building on an earlier approach piloted by CARE, we identify around 4 "*Young Girl Leaders*" (YGLs) in each school. Using data from teacher and student surveys, these girls are selected based on their willingness to actively and openly speak about menstruation, despite the taboo nature of the topic, and on their broader leadership skills. YGLs are trained on a curriculum focusing on key hygiene behaviors, the reduction of menstrual stigma, and leadership skills, and are asked to promote hygiene and reduce stigma among their peers and classmates at school. The YGLs are thus intended to act as a prominent example of someone engaging in behaviors that undermine the norm of stigmatization from *within* the social network.

Working with positive deviants in this way may be more likely to lead to norm or behavior change than *individualized* attempts to change stigmatizing attitudes for at least two reasons (Paluck et al., 2021). First, driving norm change through positive deviants leverages a *social multiplier* effect. Rather than changing the attitudes of each person in the network individually, which may be costly and hard to scale, amplifying the voice of a positive deviant can allow them to spread attitude change through peer networks.[1] Second, by leveraging the peer network, it may be possible to not only change attitudes, but also people's beliefs about others' attitudes, thereby coordinating a change in norms.

To evaluate the effects of the program, each of the 140 schools is randomized into one of three

[1]This builds on the insight from the social learning literature that shows that an effective way of spreading *information* through a social network is to "seed" it with influential network members, and applies this insight to norm change (Banerjee et al., 2013; Beaman et al., 2021; Banerjee, Chandrasekhar, et al., 2019).

conditions: (i) *Control* (35 schools); (ii) *Base only* (35 schools), which receive teacher and parental sensitization, distribution of sanitary pads, and sanitation infrastructure construction; and (iii) *Base + YGL* (70 schools), which receive all the elements of the base program and also have an average of 4 girls nominated to be Young Girl Leaders (YGLs). After approximately 1.5 years of the program, we evaluate the program's effects on human capital, hygiene knowledge and behaviors, and menstrual stigma using an endline that includes an average of 16 girls in each school ($N$=2,250).

We find that the base program on its own leads to important impacts on human capital outcomes for girls. First, we document that the base program of hygiene infrastructure and school-level interventions leads to substantial improvements in girls' learning, as measured by standardized math and language achievement tests at endline and official school grades (0.1–0.2 SD). It also increased the probability of progressing to the next grade-level by 8 p.p. (16%). These improvements are comparable to the effects of interventions specifically designed to improve learning in low-income schools (Muralidharan, 2017), a notable result given that the program did not directly target improvements in learning. Second, there are no measurable effects on school attendance, as measured by the number of students present at school during a series of 5 unannounced spot checks carried out throughout the school year, or by official school enrolment registers and girls' self-reports. This suggests that the improvements in learning were driven by improvements in *quality* of time at school, rather than *quantity* of time at school. Third, we find no measurable effects of the program on self-reported health or psychosocial well-being. A generalized health index, an index of symptoms of urinary tract infections, an index of mental health, and an index of self-esteem all yield null effects for both the *Base only* schools and the *Base+YGL* schools.

Fourth, we find evidence of changes in objectively measured stress levels. Strikingly, girls' *heart rate* during the endline survey is 0.12 SD lower in treatment schools. This reduction in stress is directly welfare-relevant, but also points to an important mechanism that could underlie the effects on learning. Given existing evidence that stress acts as a "bandwidth constraint" that can inhibit cognitive function (Haushofer & Fehr, 2014; Schilbach et al., 2016; Kaur et al., in press), the stress reduction may have directly improved concentration and learning while at school.

The program also generated broader improvements in the schools' psychosocial environment that could drive learning effects. Building on qualitative reports that treatment schools had an improved social environment with more solidarity among students, we find (i) an increase of 0.23 SD in a network connection index, based on questions focusing on girls' network connections with others at school (e.g., their friends, who they do projects with, who they share with); and (ii) a reduction in the girls' reported severity of bullying at school (−0.08 SD). These improvements could enhance learning either by reducing stress, by increasing collaboration and peer learning, or by increasing students' (in line with evidence in Alan et al., 2021 and Alan & Kubilay, 2024). We do not find strong evidence in favor of learning improvements being driven by other channels, such as through better hygiene management or through increased teacher motivation.

These results speak to the importance of evaluating and including psychosocial components in school programs – even ones that do not explicitly target psychosocial outcomes.[2] The pattern of results also indicates that the human capital gains in our context operate through psychosocial channels other than the effects on *health* and *attendance* that are often emphasized in the literature on hygiene. Both qualitative and quantitative evidence suggest that other economic constraints are overriding

[2]The results on bullying also add to the relatively scarce literature in economics on the causes and effects of bullying at school, with evidence largely coming from high-income settings, e.g., Brown & Taylor (2008); Wolke et al. (2013); Eriksen et al. (2014); Sarzosa & Urzúa (2021); Hu (2023).

constraints to attendance in this low-income, rural context: for example, the main reason for school dropout is having to work or not being able to pay school fees. Even in such a context, however, an intervention targeting hygiene, menstrual hygiene and stigma has a substantial impact on learning.

Next, we show that the *Young Girl Leader* intervention combined with the base program is significantly more effective than just the base program at improving several measures along the theory of change. While both the base program with and without the *YGL* program lead to substantial improvements in girls' *knowledge* of hygiene (0.3-0.4 SD), these improvements appear *more quickly* when the *YGL* component is included, showing up as significant improvements even at midline. The *YGL* component also generates significant additional improvements in hygiene *behavior*, suggesting that YGLs are successful at partially closing the knowledge-behavior gap that exists at baseline.

In addition, results show a large reduction in menstrual stigma in schools with the *Base+YGL* compared to the control. A reduction in stigma is an outcome that is both important for welfare in its own right, but could also generate knock-on improvements in hygiene behaviors and human capital. A combined index of stigma indicates a 0.7 SD improvement in *Base+YGL* schools (compared to 0.4 SD in *Base only* schools). For example, girls report being more willing to speak about menstruation openly, having more progressive attitudes about behaviors related to menstruation, believing that others in their network have more progressive attitudes, and being less likely to feel shame in response to vignettes involving menstruation. And when using lab-in-the-field exercises to elicit revealed-preference measures of stigma, we find some indication that these changes are more likely to translate to changes in behavior when the YGL program is included. For example, girls are more likely to select to explain a menstruation-related topic in front of their class in *Base+YGL* schools. Broadly, these results suggest that the YGLs are effective at generating behavior change related to hygiene and menstruation while simultaneously changing norms of behavior among their peers.

The motivating idea behind the YGL program was that *positive deviants* who are willing to engage in prosocial behaviors in defiance of a harmful social norm would be particularly effective. We explore this hypothesis further using a heterogeneity analysis, examining whether schools with YGLs who score particularly highly on an index of positive deviance achieve better outcomes. Here, the results paint a mixed picture. Schools with positively deviant YGLs have better learning test scores, lower rates of severe bullying, and better reported mental health. At the same time, this heterogeneity appears to be strongly linked to the baseline school environment, suggesting that positively deviant YGLs may only be *found* in schools that are already more progressive, or may only be *effective* at changing behavior in more favorable school environments.

Finally, we document suggestive evidence that the base program may generate negative unintended consequences, but that the YGLs may offset these effects. First, there is some evidence that pregnancies increase in *Base only* schools (albeit from a very low baseline, i.e., from 2% to 4% of girls who have been pregnant in the last year at endline). Second, mothers report worse psychosocial behavior for their daughters in this same arm, possibly driven by backlash. Such negative effects don't materialize in the *Base+YGL* schools. On the other hand, there is some indication that the YGL program crowds out some of the learning effects of the base program, with generally lower point estimates on learning tests in *Base+YGL* schools compared to *Base only*, and suggestive evidence that the academic performance of girls selected as YGLs worsens. Given this, the relative merits of the base program and the YGL program depend on the importance of each outcome to overall welfare.

Our study contributes by showing that a program focused on reducing menstrual stigma and improving hygiene behaviors can generate substantial learning effects for adolescent girls. We thereby

add to the literature on menstrual hygiene (Oster & Thornton, 2011; Hennegan & Montgomery, 2016; Benshaul-Tolonen et al., 2020; A. Benshaul-Tolonen et al., 2021) by studying an intervention that specifically focuses on reducing stigma, by using standardized academic and cognitive tests to document improvements in learning, as well as an objective measure of anxiety/stress. Our learning results directly speak to ongoing policy debates on the learning crisis, and the related academic literature on the effect of various school-based interventions on learning outcomes (Akyeampong et al., 2023) and gender differences in such impacts (Evans & Yuan, 2022). We show that hygiene-focused programs can lead to improvements in learning in schools in low-income settings, a result that appears to be driven by changes in social dynamics at schools. This implies that policies should therefore seek to leverage these social dynamics, crowding in effort and motivation, and increasing between-peer solidarity as part of programs focused on health.

Second, we show that working with positive deviants can be an effective tactic for addressing harmful social norms. This builds on the literature examining the role of positive deviants or "trendsetters" (Bicchieri, 2017; Bicchieri & Funcke, 2018), and more broadly on the literature examining ways of addressing harmful social norms (Jayachandran, 2021; Dhar et al., 2022; Gulesci, Jindani, et al., 2023; Bursztyn, González, & Yanagizawa-Drott, 2020; Banerjee, La Ferrara, & Orozco-Olvera, 2019; Webb, 2023). Our results could be applied in other contexts where there are positive deviants whose voices can be amplified, such as against anti-minority discrimination, or against harmful practices like female genital mutilation.

Beyond these domains, the study contributes to the broader literature on water, sanitation and hygiene interventions (see Chirgwin et al., 2021 for a review of systematic reviews), the literatures on interventions targeting adolescent girls in Sub-Saharan Africa (Buehren et al., 2017; Bandiera et al., 2020; Bergstrom & Ozler, 2021; Hamory et al., 2023; Shah et al., 2023), and to the evidence of the effectiveness of using schools as entry points for health interventions (Dupas & Miguel, 2017).

## 2.2 Study design

### 2.2.1 Interventions

The study evaluated the effect of a bundle of interventions designed by our partner NGO, CARE Madagascar. The programs aimed to spread hygienic behaviors and reduce menstrual stigma, and to thereby improve human capital outcomes for girls in the later grade-levels of primary school, and all grade-levels of secondary schools. The region's secondary schools are mostly *lower* secondary schools, although our sample also includes 6 *upper* secondary schools.

Each of the 140 schools in our sample was randomly allocated into one of the following treatment conditions:

1. *Control* (35 schools), which received no school-level interventions.
2. *Base program only* (35 schools), which received a package of interventions that included teacher training, the construction of sanitation infrastructure, and vouchers for free menstrual pads for female students.
3. *Base program + Young Girl Leaders* (70 schools), which received all the elements of the base program, plus the Young Girl Leaders element of the program.

Randomization was stratified by three school characteristics: (i) whether it was primary or secondary, (ii) whether there was existing functional water access before the program, and (iii) the

number of girls above the age of 12.

*Base program*

The base program aimed to: (i) relax hard constraints that would otherwise prevent girls from engaging in hygienic behaviors, and (ii) promote hygienic behaviors by sensitizing teachers and parents.

*Infrastructure*. To relax hard constraints related to hygiene, the program financed and built hygiene and sanitation infrastructure in all the base program schools. Specifically, our partner built between 2 and 10 (mean: 3.9) new latrines on the school site, and in some cases rehabilitated existing latrines so that they were functional.[3] The infrastructure also included basins for students to wash their hands.

*Vouchers for sanitary pads*. To relax hard constraints related to *menstrual* hygiene, the NGO distributed vouchers which could be exchanged for a set of 6 free reusable sanitary pads from local tailors who were trained by the program. The vouchers were distributed in schools to all girls from the targeted grade-levels. Take-up of the vouchers among targeted girls was 91%.

*Teacher training & hygiene committee*. To sensitize teachers and parents to the hygienic behaviors being promoted, personnel from the local government ministry of education and our NGO partners trained 2 teachers in every school. The curriculum focused on the key hygiene behaviors described above. There were 2 days of training at the start of the program, and shorter refresher trainings that took place every semester. The trained teachers were asked to organize a short (5-minute) weekly session at school to talk about hygiene to the student body. In addition, an 8-member "hygiene committee" composed of teachers, parents, and students was formed in every treatment school. The committee was responsible for promoting hygienic behaviors in school and for maintaining the hygiene infrastructure. Adult members received an initial training session about the roles and responsibilities of the committee.[4]

*Young Girl Leaders intervention*

The Young Girl Leaders intervention involved selecting girls in each school and designating them as *Young Girl Leaders* (YGLs) who were responsible for peer-to-peer promotion of hygiene and menstrual hygiene behaviors.

An average of 4 YGLs were selected in each treated school.[5] Treatment assignment was fully respected: the targeted number of YGLs were identified and trained in all 70 treatment schools, and there were no YGLs in control or *Base only* schools.

YGLs were selected based on a survey of teachers and of the students whom those teachers nominated. We selected girls in the same way across all YGL schools, using a predetermined weighting of survey responses that incorporated (i) leadership qualities, (ii) a form of "positive deviance" (i.e., willingness to take prosocial actions in defiance of others, such as speaking openly about menstruation), and (iii) having started menstruating themselves. We also restricted to girls with a sufficiently good academic track record, to avoid hurting the performance of girls who were already struggling academically.

[3] In all schools, infrastructure was finalized by endline data collection, but was only at very early stages by midline.

[4] Together, the teacher training and the hygiene committee meant that sensitization on hygiene behaviors was not *only* carried out by the young girl leaders (YGLs), described below. This also meant that the activities of YGLs were more likely to be perceived to be endorsed by adults and figures of authority in the school, reducing the risk of backlash against YGLs.

[5] The number selected depended on the number of pupils in the school. There was one YGL for every 60 students in the school, apart from schools with fewer than 120 students who always had 2 YGLs.

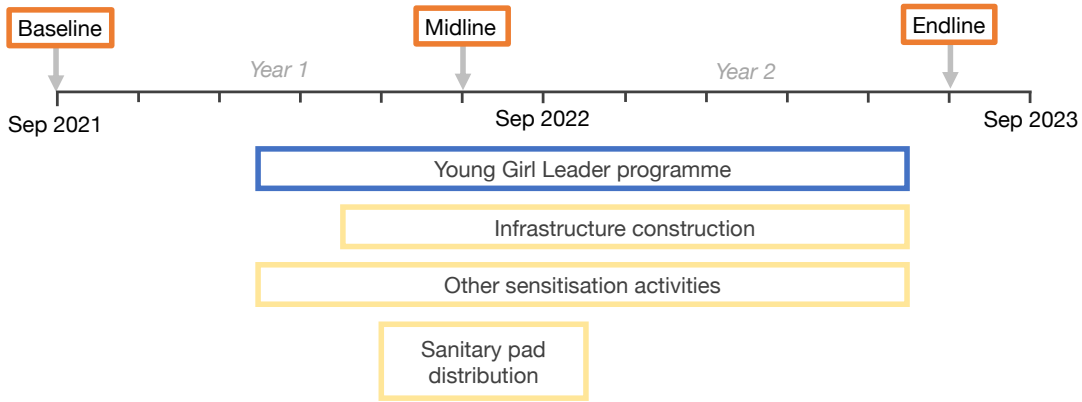
YGLs received an initial intensive training with weekly sessions of 1.5 hours for approximately 3 months. These training sessions were delivered using a cascade model, where technicians from local NGOs were trained by our main partner NGO, and then in turn trained YGLs. The sessions initially focused on enhancing the leadership and sensitization skills of the YGLs, and then moved onto a curriculum of topics on general hygiene and menstrual hygiene and stigma. The key messages and hygiene behaviors being targeted were as follows: (i) menstrual hygiene, including how to properly wash during menstruation, how to use and wash reusable sanitary pads, and basic information about menstruation and the body; (ii) using latrines; (iii) hand-washing with soap; (iv) using and keeping clean water; and (v) food hygiene. After the initial 3-month training period, technicians followed up with YGLs every two weeks to give additional guidance and encouragement. At the start of the second school year, a comprehensive refresher training covering all topics was provided to the YGLs. YGLs who graduated from a school were replaced with new YGLs, who were selected using the same criteria as the first school year and also received the same training.

YGLs were asked to spread positive messages and encourage hygienic behavior among their peers at school, and to openly discuss menstruation in a way that could reduce menstrual stigma. To test whether peer-to-peer sensitization was more effective if it was "formal" or "informal", we randomized the method through which YGLs were asked to sensitize their peers. In a *formal* condition (35 schools), they were told to give formal classroom sessions about hygiene and menstruation. In an *informal* condition (35 schools) they were told to simply talk to their classmates and friends about these topics, but were not instructed to hold classroom sessions. Results from the endline reveal, however, that there was only modest de facto variation in how the two modalities were implemented (Table B.1.1). For example, YGLs in the informal condition often also organized formal sensitization sessions. Possibly as a result, we find few differences in the effects of the two modalities (Table B.1.2), and therefore present the pooled results throughout the paper.

*Timeline*

The timeline of the program activities and main data collection exercises are summarized in Figure 2.1.

**Figure 2.1:** *Study timeline*




### 2.2.2 Data collection

*Baseline survey and sample selection*

140 primary and secondary schools in the rural areas of the Amoron'i Mania department of Madagascar were selected to be part of the study. We selected schools based on three criteria. First, we excluded schools that were deemed be to inaccessible by the operations team (e.g., because they

could not be accessed at all by vehicle). Then, we prioritized schools that had the most girls above the age of 12, for whom the intervention would be most relevant. Finally, we used a coarse measure of location (based on the distance to the subdistrict capital) to attempt to exclude schools that were too close together, to minimize spillover effects.

For the baseline survey (September to October 2021), we sampled an average of 17 girls per school to be surveyed, totaling $N$=2,393. This sampling was based on a listing exercise using official school enrolment registers. 90% of students were between 4th and 9th grade, but the sample included students between 3rd grade and 12th grade.[6] Girls' average age in the sample at baseline was 14 years (Table B.1.3), and about half of the girls had not reached menarche at baseline (with this share naturally declining by endline). The sample is characterized by high levels of poverty, and relatively large accumulated school delays (2.2 years on average).[7] Girls take on average 30 minutes to walk 2.5km to school, while only 2% of their households own a vehicle. Girls live in large households (mean size: 6.5), and 81% live in the same household as their mother. At baseline, we also interviewed the mother or primary female caregiver of all sample girls during the household interview, and, in cases where they existed, a brother in the same household as the girl.

For all relevant treatment comparisons, the treatment groups are well-balanced across key characteristics (Tables B.1.3, B.1.4), with only one statistically significant difference across treatment groups (girls in *Base only* schools are significantly less likely to have a brother in the household, $p$=0.03). We use LASSO to select all controls that predict both treatment status and outcomes (as per Belloni et al., 2014), to avoid this imbalance affecting the results.

*Midline survey*

A midline survey took place in the summer of 2022, after approximately half a year of program implementation. Attrition at midline was 11%, and slightly higher in the treatment villages than in the control (Table B.1.5, column 3).

*Endline survey: girls' survey*

Our main endline survey took place between May and July 2023 (i.e., towards the end of the second school year and at the start of the summer vacation). This was approximately 1.5 years after the start of the program's implementation. We surveyed 2255 out of 2393 girls at endline, corresponding to an attrition rate of 5.8%. Attrition at endline is balanced across treatment arms (Table B.1.5, column 1). The majority of girls were interviewed at their homestead. However, many students had moved to a different school, or left school and moved to a different city. For a subset of the 15% of girls who were not found after several attempts in their original location or nearby areas, we carried out additional intensive tracking in October 2023 (see Appendix Section B.2.5 for more details). This resulted in further in-person interviews at migration destinations, such as the capital Antananarivo and regions next to Amoron'i Mania. A small share of girls (3.7%) were interviewed by phone because they could not be reached in person. The phone survey used a shorter survey instrument that did not include standardized achievement tests. Participants in the base program arm were slightly more likely to be surveyed by phone ($p$=0.03).

[6] The baseline took place at the end of the summer vacation in 2021, using a sample based on school registers from the previous school year. We therefore avoided sampling grades where girls would be likely to drop out of school. In lower secondary schools, we sampled equally from girls who had just finished 6th, 7th, and 8th grade, and in high schools we sampled equally from girls who had just finished 10th and 11th grade. In primary schools, we sampled girls who were aged 12 and above and prioritized 5th-grade girls who were registered to repeat next year, filling out the remaining sample from other grades and younger ages in smaller schools where necessary.

[7] School dropout at early ages is also frequent: while we cannot observe this at baseline, 17% of girls in control schools drop out between year 1 and year 2 of the study.

*Endline survey: other surveys*
In addition to the girls' survey, during the endline data collection we also carried out 4 other surveys:

1. *Mothers*. We did a home-based survey of the mother or the same female caregiver that had been surveyed in the baseline. 2,287 were found, corresponding to an attrition rate of 4 % that was balanced across treatments (Table B.1.5, column 5).[8]
2. *Teachers and directors*. We surveyed all school directors and at least one teacher per grade-level for 3rd grade and above. During this survey, we also elicited the official marks received by all the children in at least one class for every grade-level.
3. *Male classmates*. Using up-to-date school registers, we randomly selected 10 boys in each school to do a short survey centered around a dictator game related to menstruation.
4. *"Willingness to speak" lab-in-the-field*. Using school registers, we randomly sampled 12 girls and 4 boys in each school to take part in an interactive exercise designed to measure girls' willingness to speak about menstruation in front of others. This exercise is described in more detail below.

*Spot-check absenteeism and school register data*
To measure school absenteeism without having to depend on unreliable official school registers, we carried out 5 waves of unannounced spot checks in all schools. Enumerators visited the schools without notifying beforehand, and took attendance calls for all classes above 3rd grade. They also digitized all the official school enrolment and attendance records from the last 3 months. The first spot check took place in February 2022, before the program had started, acting as a baseline. Subsequent waves took place in June 2022, October 2022, December 2022, January 2023, and June 2023.

### 2.2.3 Outcomes

Here, we describe the main outcomes used in the paper (see Appendix Section B.2.1 for more details).

Because respondents likely knew about the program's objectives to improve hygiene and menstrual hygiene, an important concern when measuring outcomes was the risk of *social desirability bias*, i.e. respondents biasing their responses to be favorable (possibly particularly in the treatment schools). Because of this, a key motivation for the data collection was to create *objective* or behavioral measures of the outcomes of interests. We place most weight on the results on such measures.

Where objective measures were not possible to elicit, we also used self-reported measures. For these, to evaluate the risk of social desirability bias driving our treatment effects, we use a shortened version of the Crowne & Marlowe (1960) social desirability index elicited at baseline, and test whether treatment effects are driven by girls with high social desirability scores (Table B.1.7; see Appendix Section B.2.4 for more details).

*Human capital outcomes*
We measure the effect of the base program and the YGL intervention on a set of human capital outcomes.

*Learning*. We measure how much children have learned in three complementary ways. First, at endline, we administer a set of standardized tests that evaluates academic skills learnt at school and

[8]For the other samples used, there was (i) no differential attrition for teachers (Table B.1.5, column 7); (ii) a lower probability of successfully eliciting school marks among *Base only* schools (column 8); (iii) no difference in the number of boys successfully interviewed or the probability of having to take boys from a secondary sample list across arms (columns 9 and 10); and (iv) no differential sample selection for girls in the lab-in-the-field (columns 12 and 13).

cognitive ability. It includes modules on (i) Malagasy listening; (ii) Malagasy reading fluency; (iii) Malagasy text comprehension; (iv) a digit-span test that measures working memory (forward and backward); and (v) math fluency, based on a timed written maths test involving simple arithmetic calculations. Most modules had been tested and validated earlier by local survey teams on other projects in Madagascar, and all of them elsewhere in Sub-Saharan Africa. In addition, we used pilot data collected before the deadline to validate all modules on a population of the same age and region as our study population. For the main analysis, we combine all 5 modules into a single index using factor analysis. We also consider the results of each test separately, to explore the differences between the academic achievement tests (reading and math fluency) and the other tests that capture attention, memory and cognition.

Second, we ask teachers and directors to give us the official marks received by students for their most recent exams. These are recorded by schools on student-specific paper "bulletins" and on class-specific registers, allowing enumerators to directly verify reported school marks. Students are given an overall mark out of 20 for the preceding trimester. We use this as an observed measure of academic achievement.

Third, girls report the grade-level they were in at baseline and endline (which were one academic year apart). We can therefore deduce whether a girl progressed to the next grade-level, or did not because they repeated a grade-level.

*School attendance*. We measure school attendance using the spot check data, first by measuring the number of students physically present at school when the spot check is carried out, and second by recording the number of students enrolled and present according to official school registers. For these outcomes, we include fixed effects for the 4 waves of the spot checks, or day-level fixed effects for school register data, and control for the value of the outcome in the first baseline wave. We also ask girls at endline (i) whether they are currently enrolled in school, and (ii) whether they were absent from school in the last 3 months.

*Health*. First, we measure generalized health by asking girls the physical functioning sub-scale of the Pediatric Quality of Life Inventory (PedsQL), following A. Benshaul-Tolonen et al. (2021). Girls are asked whether they had difficulty carrying out any daily activities, such as running or completing tasks at home. Second, since improvements in menstrual hygiene are associated with reductions in urinary tract infections (UTIs) (Phillips-Howard et al., 2016), we also asked girls to report symptoms of UTIs, following (Czura et al., 2019).

*Psychosocial well-being and mental health*. We measure psychosocial well-being by first asking an adapted version of the CESD depression index, in which girls report how many days over the last week they have experienced certain (positive and negative) emotions. Our preferred specification uses an index that combines both positive and negative emotions and is corrected for acquiescence bias, but we also show positive and negative emotions separately.[9] We also elicit a self-esteem module, asking girls a series of questions indicating whether they are confident in themselves and their abilities or not.

*Heart rate during endline survey*. To act as an objective measure of stress or anxiety, we measured

[9]Also known as "yay-saying", acquiescence bias is when respondents tend to agree with a statement from the enumerator, even if doing so results in contradictory responses to positive and negative statements intended to measure the same trait. We correct for this bias, following the psychometric literature (Rammstedt & Farmer, 2013; Laajaj & Macours, 2019). For each girl, we calculate an acquiescence score by taking the average difference between the positively-coded questions and the reversed negatively-coded questions, and dividing it by two to retain centering. This score is then added to the negatively-coded questions, and subtracted from the positively-coded questions.

girls' heart rates during the endline survey. This was incorporated to measure a combination of (i) general anxiety levels experienced by girls, and (ii) the anxiety levels experienced specifically when discussing topics related to hygiene and menstruation. We used wristband monitors that included optical heart rate sensors that recorded the girls' heart rate in beats per minute on a second-by-second basis. The monitors we used have been demonstrated to have similar accuracy to industry-standard chest strap monitors (Reece et al., 2021), but were significantly less invasive and more comfortable for girls to wear.[10]

*Hygiene knowledge and behavior*

To understand intermediate outcomes, we elicit a series of measures of knowledge and behavior related to hygiene and menstruation.

*Knowledge*. We ask girls questions about their knowledge of good hygienic practices. We create one index based on knowledge of overall hygiene practices (e.g., methods for cleaning water and how many times a day they should wash their hands), and another on their knowledge of menstruation and menstrual hygiene (e.g., naming hygienic menstrual products, and naming average length of a period). The sum of correct answers to the questions, normalized to have a mean of 0 and standard deviation of 1 in the control group, is used as the outcome.

*Behavior*. We ask girls questions related to whether they engage in the types of hygienic behaviors recommended by the YGL curriculum. One index is again related to overall hygiene behaviors (e.g., did they use soap when washing their hands at home last week?; did they use a latrine or defecate openly the last time?), while another is related specifically to menstruation (e.g., did they use a hygienic menstrual product?; did they wash at least 3 times per day the last time they menstruated?).

*Observed hygiene*. To alleviate concerns about social desirability bias leading to girls *overreporting* hygienic behaviors in treatment schools, enumerators also directly observed levels of cleanliness in three contexts. First, they marked the visible hygiene levels of girls themselves during the endline survey (e.g., was there visible dirt on the girl?; was the girl wearing shoes?). Second, they noted the hygiene levels at the original household of the girl, where the mother resided at endline (e.g., were animals separated from the people's living spaces?; were the dishes in the kitchen area clean?). Finally, in each school, when carrying out the teacher and director survey, enumerators noted features of the environment that indicate engagement in hygienic behaviors, such as whether there was soap available at the basins, whether there was a dedicated dispenser for trash, and whether any latrine had a cover.

*Menstrual stigma*

To understand whether the constraints imposed by menstrual stigma are relaxed by the program, we elicit an array of complementary measures of stigma. For all measures of stigma, a reduction in stigma is coded positively.

*Spread of information*. First, we seek to understand whether the taboo around speaking about menstruation is lifted by the program. We measure whether girls have received information about menstruation from others at school or at home using an index of questions focusing on how many times they heard classmates, teachers, and parents speak about menstruation in the last 3 months, and how many times they asked people questions about menstruation. A second series of questions measures how willing girls are to themselves talk about menstruation in front of others.

[10] In some cases, heart rate monitors were not fixed properly to girls' wrists, or the battery of monitors ran out on the field. Because of this, heart rate data is available for only 1904 girls. However, the proportion of data available is equal across treatment arms (Table B.1.5, column 6).

*Attitudes and perceived norms*. We measure how progressive girls' attitudes are towards menstruation. We use a series of Likert scale questions asking the level of agreement with statements like "Girls should be allowed to come to school when they menstruate" and "Girls should have to hide the fact that they are on their periods when they are at school". We also measure perceived norms by asking girls to estimate how much people in their social network agree with a subset of these attitude questions. They report how much they think other girls, boys, teachers in their school, and their mothers agree with the statements about menstruation. All four second-order belief questions are combined into a single index of norms.

*Shame*. To measure shame, we describe hypothetical vignettes about embarrassing scenarios related to menstruation to girls. We then ask them to report how much they feel shame-related emotions in response. To validate these measures, we ask enumerators to observe how visibly uncomfortable girls appear to be when talking about menstruation during the endline survey (e.g., did they look at the floor, did they hesitate, did they seem uncomfortable, did they speak excessively quietly).

*"Willingness to speak" lab-in-the-field*. This was an interactive exercise designed to measure girls' *revealed-preference* willingness to speak about menstruation in front of others. In each school, this activity was carried out twice, with 6 girls and 2 boys participating in each group. All 8 students sat in a classroom. Each girl was taken out one by one, and had two new topics explained to them by enumerators: one topic related broadly to hygiene/health (e.g., what is pneumonia?), and the other related to menstruation (e.g., what are menstrual cramps?).

Girls then faced two binary choices. In one choice ("*anonymous*"), they were told that a video would be shown to the rest of the class about one of the topics, and they could choose which one — the menstruation-related topic, or the hygiene-related topic. No one would be told which video the girl picked. Since this choice was anonymous, it was designed to minimize concerns about taboo and stigma, and to measure the desire to supply the information to the rest of the class. In the other binary choice ("*explanation*"), by contrast, girls were told that they would have to *themselves* explain the topic to the rest of the class.[11] Since this involved talking openly about a topic in front of others, the choice was likely to be strongly affected by social image concerns and the taboo around menstruation, allowing us to test for changes in the level of taboo driven by the treatments.

*Dictator game*. A standard prediction of theories of taboo is that those who *break* the taboo should face social sanctions. To test this, we ask endline girls and the 10 randomly selected endline boys to play two dictator games. Students are shown two 1-minute videos in a random order, both depicting a teenage girl from the region explaining a topic. One explains what menstruation is and how to maintain menstrual hygiene, and the other talks about the use of soap in hand-washing and the spread of germs. After each video, the student is shown 5 cookies, and asked how many they want to share with the person in the video. We interpret a decrease in the number of cookies as social sanctioning that may be incurred when breaking a taboo. Just as with the lab-in-the-field, we can test whether the taboo around menstruation has decreased by looking at whether students are more likely to share cookies with the girl speaking about menstruation (relative to the girl speaking about hand-washing).

*Demand for information in envelope*. The final revealed-preference measure of stigma we use is whether girls choose to find out information about menstruation or another topic, intended as a measure of the demand for information. In the endline and midline survey, girls were shown a series of pairs of envelopes, each of which showed a question on the outside, some of which related

[11] We randomly varied the order of explanation and anonymous choices, and the topics being proposed to each girl.

to menstruation (e.g., "Can stress affect the menstrual cycle?") and some which related to broader hygiene and health (e.g., "Are mosquitos more attracted by certain blood types?"). Girls make binary choices over whether they would prefer to open the menstruation-related envelope or the hygiene-related envelope, and get to see the answer to the question inside the envelope.

*Psychosocial environment at school*
To evaluate the social dynamics at school, we ask girls and teachers to report the severity and frequency of *bullying at school*. We ask them to report separately on cases of light-hearted teasing, severe intimidation or harassment, and incidences of bullying specifically related to menstruation. Girls reported on bullying they experienced themselves, and on a randomly selected subset of 2-3 girls in the same grade-level as them; these are combined into one index. Teachers report on a randomly selected subset of girls.

We also create a *network integration index* by asking girls about their social connections with their peers. For example, they are asked about how many friends they have, how many children they play with, how many they speak with at break, and how often they do homework together. For a subset of these questions, we ask separately about other girls and other boys.

*Effects on mothers*
To evaluate whether any changes in knowledge and behavior among girls are transmitted vertically to their mothers (e.g., through interactions in the household), we elicit from mothers at endline the same indexes of knowledge and behavior related to general hygiene and menstrual hygiene. We also ask mothers the same set of attitude questions about menstruation as girls. We further ask mothers a set of questions about their perception of the behavior of their daughter, using the standardized Strength and Difficulties (SDQ) instrument. This was intended to evaluate girls' psychosocial well-being and also to understand if mothers had a negative or positive reaction to the changes in girls' behavior induced by the program. Some questions focused on emotional behavior (e.g., whether they are often worried, or often angry), while others focused on their relations with others (e.g., are they often alone, do they have at least one friend?). Finally, we ask all mothers about the fertility status of their daughters (specifically, were they pregnant in the last year or are they pregnant now) to examine potential impacts on fertility.

### 2.2.4 Empirical specification

In our main specification, we estimate the following regression:

$$Y_{ij} = \beta_0 + \beta_1 BaseOnly_j + \beta_2 BaseYGL_j + \mathbf{X}'_{ij}\Gamma + \varepsilon_{ij} \tag{2.1}$$

$Y_{ij}$ is the outcome variable measured at endline for student $i$ in school $j$. $BaseOnly_j$ is a binary variable, equalling 1 when school $j$ is in the *Base only* treatment and 0 otherwise. $BaseYGL_j$ equals 1 when school $j$ is in the *Base program + YGL* treatment and is 0 otherwise. The main average treatment effects of interest are therefore given by $\beta_1$ and $\beta_2$. $\mathbf{X}_{ij}$ is a vector of controls, which always includes stratum fixed effects. In cases where it exists, it also includes $Y^0_{ij}$, the baseline analog of the outcome. We use double LASSO (Belloni et al., 2014) to select an additional set of controls that predict both the treatment and outcome variables. When further controls are included (e.g., wave fixed effects for spot check outcomes), we specify directly in the table notes. Standard errors are, unless otherwise specified, clustered at the school level.

## 2.3 Results

### 2.3.1 First stage: program implementation

The program was implemented with high fidelity to the assigned treatment conditions. First, in *Base only* and *Base+YGL* schools, there were improvements in the hygiene infrastructure at endline according to enumerator observations. On average, these treatment schools had 4.5 functional latrines (vs 2.3 in control schools), 85% of them had at least one separate latrine for girls (vs 23% in control schools), and 69% had a functional water basin (vs 37% in control schools). There were no statistically significant differences between *Base only* and *Base+YGL* schools in these outcomes. Second, in *Base only* and *Base+YGL* schools, menstrual pads were successfully distributed, with 100% of directors in these schools reporting that there had been a distribution (vs 0% in control schools), and 72% of girls at endline reporting having received pads (vs 2% in control schools)[12]. Third, in *Base only* and *Base+YGL* schools, directors reported that teachers had received training about hygiene and sanitation in the last 5 years in 99% of cases (vs 31% in control), and teachers sensitized children on hygiene and sanitation issues every week in 93% of schools (vs 46% in control). Finally, the *Young Girl Leader* program was carried out successfully: when asking girls at endline whether certain girls in their school carried out sensitization activities about menstruation and health, in 97% of schools (vs 9% in control schools) at least 20% of the girls said yes.

### 2.3.2 Human capital outcomes

In this section, we document the effects of the program on human capital outcomes, including learning, school attendance, health, mental health, and anxiety/stress (as measured by heart rate).

*Learning*. The base program leads to substantial improvements in girls' academic and cognitive skills, measured through a set of standardized learning tests administered at endline (Table 2.1). The pooled effect of the *Base only* and *Base+YGL* arms on the combined index (capturing tests of academic skills, memory, and attention) is 0.15 SD ($p$=0.006, column 1). The effects are strongest for academic achievement tests (language and math fluency) (0.13-0.25 SD, pooled $p$=0.001, column 2), while effects are smaller and less consistently significant for tests capturing memory, attention and other cognitive skills (0.08-0.14 SD, pooled $p$=0.07, column 3).[13] The magnitudes of the effects are comparable to the effects of interventions specifically designed to improve learning in schools in low-income settings (Muralidharan, 2017). They are thus substantial given that the program elements did not focus directly on learning.

The improvement in learning translates to improvements in official marks in treatment schools for girls (pooled effect: 0.11 SD, $p$=0.07; Table 2.1, column 4). The effect on boys' official marks is positive but not significant (pooled $p$=0.20, column 5), suggesting that there may be a gender-specific component to the improvements, although we cannot reject that the effect on girls and boys is the same ($p$=0.71). In addition, the improved marks for girls lead to a substantially increased probability of progressing to the next grade-level, i.e., of not repeating a grade-level (column 6). While in the control group only 51% of girls progressed to the next grade-level, this increases to

[12] This figure is lower than the 91% overall take-up rate, because some girls in the sample were too young to be in the grades who received pads.

[13] See Table B.1.8 for results on each individual test.

59% in the pooled treatment schools (effect: 8p.p., 16%, $p$=0.02).[14]

The improvements in learning across the different measures appear to be driven by the base program rather than the *YGL* intervention. For all measures, the point estimate is larger for *Base only* schools than *Base+YGL* schools, although the difference between these two arms is only significant for the subset of academic achievement tests ($p$=0.02). This suggests that relaxing hard constraints by building infrastructure and distributing sanitary pads, along with teacher training and parental involvement together drive the learning effects, while the *YGL* program does not generate additional learning benefits.

**Table 2.1:** *Education - Learning*

| VARIABLES | (1) Achievement + cognitive tests (combined) (Z) | (2) Achievement tests only (Z) | (3) Cognitive tests only (Z) | (4) Official mark (Z): girls | (5) Official mark (Z): boys | (6) Progressed one class (=1) |
|---|---|---|---|---|---|---|
| Base Only | 0.204*** | 0.248*** | 0.138* | 0.163** | 0.077 | 0.106*** |
| | (0.072) | (0.064) | (0.075) | (0.080) | (0.071) | (0.039) |
| Base + YGL | 0.113** | 0.131** | 0.080 | 0.090 | 0.086 | 0.068* |
| | (0.053) | (0.052) | (0.058) | (0.065) | (0.069) | (0.035) |
| Observations | 2,162 | 2,162 | 2,162 | 7,586 | 6,993 | 2,251 |
| Data source | Girls | Girls | Girls | Marks | Marks | Girls |
| Control mean | 0 | 0 | 0 | 0 | 0 | .515 |
| p: Treated = 0 | .0056 | .001 | .0694 | .0728 | .1954 | .019 |
| p: Base Only = Base + YGL | .1655 | .0195 | .4252 | .2809 | .8831 | .1887 |
| p: Base Only [Girls] = Base Only [Boys] | | | | .191 | .191 | |
| p: Base + YGL [Girls] = Base + YGL [Boys] | | | | .866 | .866 | |
| p: Treated [Girls] = Treated [Boys] | | | | .707 | .707 | |

*Notes:* Standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. Columns (1)-(3) also include the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. In column (1), all learning tests elicited at endline are combined into one index using factor analysis. In column (2), only reading and math fluency tests are included and are combined using an unweighted sum of Z-scores. In column (3), only listening, comprehension, digit span, and reverse digit span tests are used, and are combined using an unweighted sum of Z-scores. The official school marks (columns 4 and 5) were collected from grades notebooks provided by teachers and include all students in the class. *Progressed one class* = 1 when the girl moved up one grade-level in between the baseline and endline survey, and is coded as 0 if the girl is in the same grade-level at endline or is not enrolled at school at endline.

*School attendance*. There are no improvements in school attendance driven by the interventions (Table 2.2). The program has no significant effect on (i) the number of girls physically present at school during the unannounced spot checks (pooled $p$=0.99, column 1); (ii) the number of girls physically present in the 2 months preceding the spot checks, according to digitized official school registers (pooled $p$=0.31, column 2); or the number of students enrolled according to those school registers (pooled $p$=0.39, column 3).[15] When asking girls at endline, we find no effect on whether girls report being enrolled in school (pooled $p$=0.29, column 4), or whether they were absent in the last 3 months, conditional on being enrolled (pooled $p$=0.36, column 5).[16] Together with the

[14] Several additional results suggest the learning effects are robust. The randomization inference $p$-value of the pooled effect is significant for both the learning tests and grade progression (Table B.1.9). The distributions of the learning test results shift fairly uniformly, indicating that the effects are not driven by outliers (Figure B.1.1). We find no evidence of spillovers on learning tests (Table B.3.1), or any other primary human capital outcome. Learning results are also robust to including attrition weights (Table B.1.10 and Table B.1.11). And while we did not elicit learning tests at baseline, the number of years of school delay (a proxy for academic performance) is balanced across arms at baseline (Table B.1.3).

[15] Our standard errors imply a minimum detectable effect of approximately 8 girls on a given day, corresponding to an absenteeism rate of approximately 12% relative to the control mean. We therefore cannot rule out small effects on school attendance.

[16] There is also no effect on boys' attendance (Table B.1.12).

**Table 2.2:** *Education - Attendance*

| VARIABLES | (1) # girls present (spot-checks) | (2) # girls present (registers) | (3) # students enrolled (registers) | (4) Enrolled (=1) | (5) Absent in last 3 months (=1) |
|---|---|---|---|---|---|
| Base Only | -1.996 | -0.849 | 4.540 | -0.021 | -0.006 |
| | (3.458) | (2.854) | (4.569) | (0.024) | (0.030) |
| Base + YGL | 0.981 | -3.080 | 2.633 | -0.023 | -0.034 |
| | (2.715) | (2.715) | (3.929) | (0.022) | (0.029) |
| Observations | 657 | 34,455 | 693 | 2,251 | 1,820 |
| Data source | Spot checks | Registers | Registers | Girls | Girls |
| Control mean | 56.4 | 64.8 | 146. | .826 | .442 |
| Control SD | 49.1 | 95.1 | 103. | .38 | .497 |
| p: Treated = 0 | .9911 | .3323 | .3934 | .28 | .3587 |
| p: Base Only = Base + YGL | .3637 | .4206 | .5768 | .9129 | .2648 |

*Notes:* The table displays school attendance results from the spot-checks in column (1) and official school registers in columns (2)-(3), collected in June 2022, October 2022, December 2022, January 2023, and June 2023, and the girls survey in columns (4)-(5) at endline. Standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed-effects. Column 1: *# girls present* denotes the number of girls physically present at the school. The data is at the *wave* × *school* level and includes wave fixed-effects and controls for the outcome at baseline. Column 2 is at the *day* × *school* level and includes date (day) fixed effects and controls for the outcome at baseline. *# girls present* denotes how many girls were present according to the registers on that day. Column 3: *# students enrolled* is the number of students (girls and boys) enrolled according to the official registers in that wave of the spot-check. For the last two outcomes, girls were asked about their school attendance at endline: *Enrolled (=1)* refers to whether the girl was enrolled in the 2022-2023 academic year, and *Absent in last 3 months (=1)* is an indicator of whether the girl had been absent in the last 3 months, conditioned on being enrolled. LASSO-selected controls are included for columns 4 and 5. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

results above, these results suggest that the *quality* of learning increases while at school without increasing the *quantity* of time at school.

A plausible explanation for the lack of effects on school attendance is that absenteeism and dropout in this low-income context is driven by other hard economic factors that act as binding constraints. The overall rate of dropout is high (19% over the two-year period between baseline and endline). And in 63% of schools, directors reported that the main reasons for female dropout were economic factors, including (i) girls having to find a job (38% of schools), (ii) not being able to pay school fees (44% of schools), or (iii) not having enough food to eat to go to school (13% of schools).

*Health*. There are no measurable impacts on self-reported health outcomes (Table 2.3). Our measure of general health based on the Pediatric Quality of Life index shows no significant effect (pooled $p$=0.99, column 1). There are also no clear decreases in the symptoms of urinary tract infections (UTIs) (pooled $p$=0.49, column 2). The null effects seen here may be partly driven by measurement error on self-reported health outcomes, given that we do find impacts on objective measures of observed hygiene and heart rate during the endline survey (discussed below).

**Table 2.3:** *Health and psychosocial wellbeing*

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | General health (Z) | UTI health (Z) | Mental health (Reversed CESD) Pos. and neg. emotions (Z) | Mental health Pos. emotions only (Z) | Mental health Neg. emotions only (Z) | Self-esteem (Z) |
| Base Only | -0.033 | -0.085 | 0.017 | 0.043 | 0.031 | 0.021 |
| | (0.060) | (0.054) | (0.062) | (0.071) | (0.060) | (0.067) |
| Base + YGL | 0.014 | 0.002 | 0.057 | 0.089 | 0.046 | 0.014 |
| | (0.049) | (0.043) | (0.061) | (0.067) | (0.051) | (0.052) |
| Observations | 2,251 | 2,251 | 2,251 | 2,251 | 2,251 | 2,251 |
| Data source | Girls | Girls | Girls | Girls | Girls | Girls |
| p: Treated = 0 | .9912 | .4898 | .4808 | .258 | .3424 | .7176 |
| p: Base Only = Base + YGL | .3844 | .0857 | .3998 | .3828 | .7621 | .9071 |

*Notes:* The table displays health index results from the girls' surveys at endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$* p < 0.1$, *$** p < 0.05$, *$*** p < 0.01$. *General health* is an index constructed using questions about experiencing symptoms of physical pain and general health during the last month. A higher value indicates *better* health. *UTI health* is constructed using questions specific to experiencing urinary symptoms during the last month (a higher value indicates fewer symptoms, i.e., better health). *Mental health* (column 3) was constructed from questions aimed at measuring depression symptoms during the last week, combining both positive and negative emotions and correcting for acquiescence bias, with a higher value indicating better mental health (i.e., less depression). *Positive emotions* (column 4) and *Negative emotions* (column 5) include only the positive and negative symptoms from this scale, respectively. A higher value for positive emotions indicates more common experiences of positive emotions; a higher value for negative emotions indicates more commonly experiencing negative emotions. *Self-esteem* is constructed from positive and negative questions about how the girl perceived her own abilities.

*Self-reported psychosocial well-being*. There are no significant effects on self-reported measures of psychosocial wellbeing (Table 2.3, columns 3-7). The program does not affect a summary index of depression based on the CESD index ($p$=0.99, column 3), nor does it affect reports of the positive or negative emotions that compose the index when separated out (columns 4 and 5). A related measure of self-esteem, that asks girls questions about their confidence in their abilities and personality, also shows no significant effects ($p$=0.70, column 6).[17]

[17] At midline, there is some weak evidence of an increased incidence of negative emotions in *YGL* schools (Table B.1.13, column 4), which may be suggestive of an initial "backlash" effect against the program. But this fades out by endline, suggesting that any initially negative effect is mitigated with further program involvement.

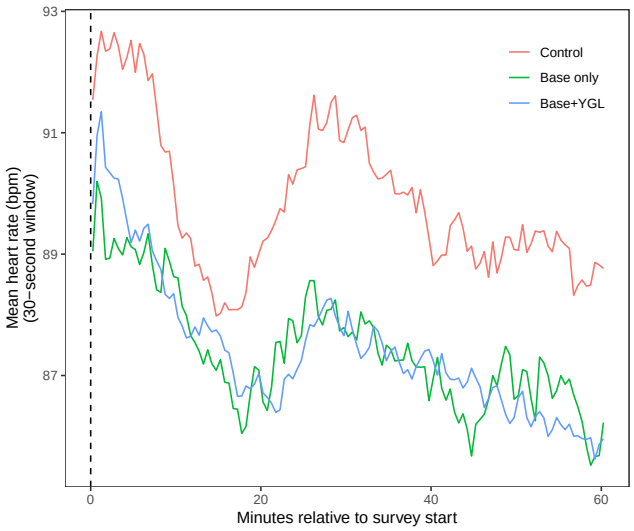
*Heart rate*. The interventions do, however, lead to a significant *decrease* on an objective biological measure of anxiety or stress: girls' heart rate during the endline survey. While heart rate naturally fluctuates in both treatment and control during the survey, heart rate is on average 2.4 bpm lower in treatment schools (Figure 2.2; $p$=0.02; effect size: 0.12 SD when using the between-girl standard deviation). The effects are very similar across *Base only* and *Base+YGL* schools ($p$ of difference: 0.99).[18]

**Figure 2.2:** *Stress: heart rate during endline survey*

**(a)** *Regression*

| | Dep var: HR bpm |
|---|---|
| Base only | −2.358** |
| | (1.140) |
| Base+YGL | −2.352** |
| | (1.008) |
| Num.Obs. | 9 105 420 |
| N girls | 1904 |
| Control mean | 87.6 |
| Control SD (between) | 19.0 |
| p : Treat = 0 | 0.014 |
| p : Base=Base+YGL | 0.995 |

**(b)** *Dynamics*




*Notes*: Panel (a) shows the effect of the intervention on the heart rate beats per minute (bpm) for girls in the endline survey. All girls who were included in the baseline sample and for whom heart rate data is available are included. *Time window*: observations are included until up to 120 minutes after the start of the survey. The unit of observation is the *girl* × *second (relative to the start of the survey)*. Results are robust to using a girl-level regression (Table B.1.14, column 1). The regression controls for 30-second window fixed effects, stratum fixed effects, and variables selected by double LASSO. Standard errors are clustered at the school level and are in parentheses. The bottom of the table denotes the p-values on the test of being treated Treated=0 and the equality of the coefficients Base Only = Base + YGL. * p < 0.1, ** p < 0.05, *** p <0.01. Panel (b) shows the average heart rate in bpm for each 30-second window relative to the start of the survey. The graph ends at 60 minutes, when almost all girls still have not yet finished the survey (see Appendix Section B.2.2 for more details).

We see two main interpretations of this reduction in heart rate (which are not mutually exclusive). First, it could capture a *general* improvement in stress or anxiety levels among the girls. Given the evidence that reductions in stress and anxiety can relax cognitive bandwidth constraints and thereby improve decision-making in a variety of domains (Haushofer & Fehr, 2014; Schilbach et al., 2016; Kaur et al., in press), this therefore points to an important potential mechanism that could underly the effects on learning. A reduction in stress or anxiety would also represent an important improvement in welfare in and of itself.

Second, the reduction in heart rate could be driven by a reduction in stress or anxiety *specifically* when talking about menstruation, thus capturing a reduction in menstrual stigma. We find mixed

[18]The distribution of heart rates shows an overall shift to the left in treatment groups (Figure B.1.2). However, these distributions also show a grouping of large heart rate values at approximately double the magnitude of the mode. These could reflect real episodes of high heart rates (heart rate can typically jump to 150-200 bpm in situations of intense stress). Alternatively, they could reflect measurement error, e.g., driven by the heart rate monitors erroneously counting *both* of the two main peaks in arterial pressure (called the systolic and diastolic peaks) that occur within one heartbeat (see e.g. Paradkar & Chowdhury (2015) for discussion of peak detection issues in arterial pressure). As robustness tests, we drop or halve any observations above 175 bpm, and show that this still yields significant pooled effects (Table B.1.14; $p$=0.10 and 0.08).

evidence regarding this channel, since heart rate does not seem to be differentially affected when actually speaking about menstruation during the endline (Table B.1.15, columns 1 and 2), although there is some evidence that the treatment effect grows larger as the survey goes on (Table B.1.15, column 4). On the other hand, girls may simply be correctly anticipating questions about menstruation even before they occur, since they were aware that the survey was about menstruation – so this evidence is not dispositive.

*Multiple hypothesis testing.* Given that we evaluate multiple primary outcomes, we check robustness to multiple hypothesis testing. We calculate Anderson (2008) sharpened $q$-values that control for the false discovery rate for our set of primary outcomes that are directly welfare-relevant: learning test scores, school attendance (as measured by spot-checks), general health, heart rate, and menstrual stigma (Table B.1.6). Our substantive conclusions are not affected by this adjustment: the pooled coefficient on learning has a $q$-value of 0.11 (compared to a $p$-value of 0.10); the pooled coefficient on heart rate is still significant at the 5% level; and the pooled coefficient on the stigma index is still significant at the 1% level.

### 2.3.3 Hygiene knowledge and behavior

In Table 2.4 we show that the base program led to substantial improvements in hygiene knowledge and behavior, and that the *YGL* intervention led to significant *additional* improvements in these dimensions.

*Knowledge*. The program led to large improvements of 0.3-0.4 SD in girls' knowledge of both general hygiene (pooled effect: 0.40 SD, $p<0.001$, column 1), and of menstruation and menstrual hygiene (pooled effect 0.34 SD, $p<0.001$, column 2). For both of these knowledge indexes, the point estimates on the *Base+YGL* arm are larger than for the *Base only* arm, hinting that the *YGLs* may have been effective at spreading knowledge about the hygiene-based curriculum. However, the two treatment arms are not significantly different from each other in either case ($p$ of difference: 0.12 and 0.35).

*Reported hygiene behavior*. Girls in *Base only* schools report engaging in more hygienic behaviors related to menstruation ($p<0.001$, column 4), but not more general hygiene behaviors ($p=0.23$, column 3).[19] By contrast, girls in *Base+YGL* schools report large improvements in both general (0.3 SD, $p<0.001$, column 3) and menstrual hygiene behaviors (0.5 SD, $p<0.001$, column 4). Notably, the coefficients are significantly larger in *Base+YGL* schools than *Base only* schools, suggesting that the *YGLs* play an important role in correcting the knowledge-behavior gap by embedding and promoting correct hygienic behaviors in their social networks.

*Observed hygiene*. Here we examine whether the reported improvements in hygienic behaviors translate to measures of hygiene that are observed directly by enumerators, and are therefore likely to be less subject to reporting biases like social desirability. We find that girls' *households* in *Base+YGL* schools are significantly cleaner ($p<0.001$, column 6), and treatment schools show very large improvements in observable measures of hygiene in the environment, such as the availability of soap and separation of trash ($p<0.001$, column 7). The *YGL* intervention leads to additional improvements in this measure ($p$ of difference $= 0.03$). We do not find significant improvements in the observed hygiene of the girls themselves during the endline survey (pooled $p=0.80$, column 5). Overall the significant results on observed measures corroborate the results on reported behavior, namely that the *YGLs* lead to significant improvements in hygienic behaviors in their schools.

[19] While self-reported hygiene behaviors are potentially subject to social desirability bias, we find no evidence that the effects are stronger for girls who are more likely to report socially desirable answers (Table B.1.7).

*Transmission to mothers*. We find mixed results on whether the changes in knowledge and behavior seen for the girls are transmitted vertically from girls to their mothers (Table B.1.16). While mothers of girls in *Base + YGL* schools show improvements of 0.2 SD in menstruation-related knowledge (column 2, $p<0.01$), there are no significant improvements in self-reported behaviors or general hygiene knowledge (columns 1, 3, and 4). There is, however, a large improvement in mothers' progressive *attitudes* towards menstruation (column 5, $p=0.01$).

**Table 2.4:** *Knowledge and behavior indexes*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Hygiene knowl. | Menstr. knowl. | Hygiene behavior | Menstr. behavior | Observed hygiene: Girls | Observed hygiene: Home | Observed hygiene: School |
| VARIABLES | (Z) | (Z) | (Z) | (Z) | (Z) | (Z) | (Z) |
| Base Only | 0.340*** | 0.314*** | 0.102 | 0.358*** | 0.0639 | 0.0205 | 1.826*** |
| | (0.0772) | (0.0572) | (0.0860) | (0.0689) | (0.0632) | (0.0234) | (0.200) |
| Base + YGL | 0.436*** | 0.344*** | 0.318*** | 0.558*** | -0.0517 | 0.0620*** | 2.261*** |
| | (0.0625) | (0.0492) | (0.0728) | (0.0544) | (0.0552) | (0.0172) | (0.178) |
| Observations | 2,162 | 2,162 | 2,162 | 2,251 | 2,162 | 2,282 | 140 |
| Data source | Girls | Girls | Girls | Girls | Girls | Mothers | Schools |
| p: Treated = 0 | <0.001 | <0.001 | <0.001 | <0.001 | .7995 | .0045 | <0.001 |
| p: Base Only = Base + YGL | .1208 | .4987 | .0024 | .0018 | .0254 | .0489 | .0131 |

*Notes:* The table displays knowledge and behavior index results from the girl (1)-(6), mother (7), and school (8) surveys in endline. In columns (1)-(7) standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and the baseline controls selected by double LASSO. In column (8), standard errors are robust and controls include stratum fixed effects. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$* p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*
*Hygiene knowledge index* combines questions about WASH knowledge related to the program. *Menstruation knowledge index* is constructed from questions that test girls' knowledge about menstruation. *Hygiene behavior index* compiles questions about hygiene behavior during the last week. *Menstruation behavior index* includes questions about girls' behavior during the last period, such as the use of serviettes. *Observed hygiene: Girls* is measured by enumerator observations about the cleanliness of the girl at endline. *Observed hygiene at home* is constructed from enumerators' observations of home status during the mother's interview. *Observed hygiene: School* is constructed from enumerators' observations of the hygiene of the school during the school-level interview of teachers and directors. All outcomes are in control group standard deviations.

### 2.3.4 Menstrual stigma

Here, we document the program's effects on measures of menstrual stigma, first concentrating on the effect on survey measures of stigma (Table 2.5), then on revealed preference lab-in-the-field measures (Table 2.6).

*Survey measures*. The base program led to significant improvements in a wide variety of reported measures of menstrual stigma, and the *YGL* intervention led to substantial additional improvements (Table 2.5). In *Base only* schools, girls reported a 0.4 SD improvement in a combined index of stigma (column 1). This results from 0.19–0.36 SD improvements (all $p$-values below 0.05) in indexes of (i) how much they had heard people speaking about menstruation (column 2); (ii) how willing they were to speak about menstruation (column 3); (iii) how progressive their attitudes were regarding menstruation (column 4); (iv) how progressive they believed others' attitudes were (i.e., a measure of norms, column 5); and (v) they reported a reduction in emotions of shame and guilt in response to vignettes related to menstruation (column 6).[20] For all of these same measures, the *YGL* intervention generated significantly larger improvements of between 0.44 and

[20] Girls typically *underestimate* how progressive others' attitudes towards menstruation are in both control and treatment schools (Figure B.1.3), suggesting a form of "pluralistic ignorance" (Katz et al., 1931; Bursztyn, González, & Yanagizawa-Drott, 2020) that could be driven by a reluctance to speak openly about menstruation. Their misperceptions about teachers' attitudes are significantly reduced by the program, in line with a reduction in stigma.

0.72 SD, resulting in an improvement of 0.73 SD on the combined index. These large changes suggest that the program reduced the stigma surrounding menstruation, at least according to girls' perceptions, and did so particularly effectively when *YGLs* were involved in challenging the stigma. This conclusion is reinforced by enumerators' observations of girls during the endline survey (column 7), which indicate that girls in the treatment schools were substantially more comfortable talking about menstruation than girls in control schools (e.g., they were less likely to look at the floor when discussing the subject). In addition, while the self-reported measures of stigma are in principle vulnerable to social desirability bias, we find no evidence that the treatment effects are driven by such bias (Table B.1.7).

**Table 2.5:** *Menstrual stigma - Survey measures*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| VARIABLES | Combined index (Z) | Received info. (Z) | Willingness to speak (Z) | Attitudes (Z) | Norms (Z) | Shame response to vignettes (rev.) (Z) | Observed shame (rev.) (Z) |
| Base Only | 0.396*** | 0.193** | 0.243*** | 0.360*** | 0.274*** | 0.303*** | 0.288*** |
| | (0.065) | (0.078) | (0.076) | (0.065) | (0.060) | (0.059) | (0.061) |
| Base + YGL | 0.730*** | 0.527*** | 0.720*** | 0.559*** | 0.467*** | 0.440*** | 0.220*** |
| | (0.055) | (0.064) | (0.065) | (0.058) | (0.049) | (0.055) | (0.053) |
| Observations | 2,251 | 2,251 | 2,245 | 2,249 | 2,162 | 2,160 | 2,251 |
| Data source | Girls | Girls | Girls | Girls | Girls | Girls | Girls |
| p: Treated = 0 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| p: Base Only = Base + YGL | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | .0085 | .1775 |

*Notes:* The table displays results from the girls (1)-(7) and the teachers surveys (8) in endline. In all regressions, standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. In addition, columns (1)-(7) include the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*. *Combined index* is an index that combines all indexes from (2) to (7) into one summary index. *Received info* is a factor analysis index from questions that asks girls how many times they have heard peers, family or teachers speak about menstruation in the last 3 months. *Willingness to speak* is an index constructed from questions that ask girls how likely they are to discuss topics related to menstruation. *Attitudes* is an index of how progressive girls' attitudes are when asked about questions related to menstruation. *Norms* is constructed from questions about the menstruation-related attitudes girls think their teachers and mothers have. *Shame response to vignettes (rev.)* index is constructed from the emotions that girls reported when prompted with situations where others hear them talk about menstruation or realize they had their period. A higher value indicates less shame. *Observed shame (rev.)* index is constructed from enumerator observations about the girls' behaviors during the menstruation survey questions. A higher value indicates less shame in their behavior. All outcomes are in control group standard deviations.

*Revealed-preference lab-in-the-field*. A series of revealed-preference measures show suggestive evidence that the reported reductions in menstrual stigma translate into changes in behavior (Table 2.6). When girls were informed they'd be asked to explain a topic in front of their schoolmates, and were asked whether to choose a menstruation-related topic or a broader hygiene-related topic, they were 4.4 p.p. (21%, 0.27 SD) more likely to choose the menstruation topic in the *Base+YGL* schools ($p$=0.04, column 1). By contrast, there is no effect on *Base only* schools ($p$=0.91, column 1; $p$ of difference = 0.06). The *YGLs* in schools are therefore effective at making girls more willing to speak in front of others about menstruation, indicating a reduction in the taboo on the topic. In line with this, the additional effect of *YGLs* is specific to a choice that involves speaking *in front of other people* (when social image concerns are more salient). When girls instead choose *anonymously* which video should be shown to their schoolmates (column 2), the pooled treatment effect is significant (+3.8 p.p., 13%, 0.08 SD, $p$=0.05), and *Base+YGL* girls do not choose the menstruation topics

more often than *Base only* girls ($p$ of difference = 0.52, column 2).[21]

There are no detectable effects on the probability of selecting a menstruation-related topic when presented with envelopes with information about various topics contained inside (pooled $p$=0.20) at endline. However, at *midline*, we found a stronger pooled effect of 4.1 p.p. on the probability of choosing a menstruation-related envelope (8.6%, 0.14 SD, $p$=0.003), with no difference between *Base only* and *Base+YGL* schools ($p$=0.88), suggesting that information demand does go up at least early on in the program.

Since undoing menstrual stigma also requires that girls face fewer social sanctions from an "out-group" when menstruating or discussing menstruation, we examine whether the program affects the behavior of *boys*. We do not find evidence of reduced menstrual stigma when boys play a dictator game with a girl who speaks about either menstruation or hand-washing. There is no effect on how many cookies boys shared with the girl speaking about menstruation (Table 2.6, column 5, pooled $p$=0.68), while boys in treatment schools share approximately 6% fewer cookies with the girl who spoke about hand-washing (column 4, pooled $p$=0.06).[22]

Overall, the set of revealed-preference measures provides suggestive evidence in favor of a reduction in menstrual stigma driven by the program, although the effects are more mixed and generally smaller in magnitude than the large effects on self-reported measures. This suggests that while attitudes and girls' perceptions of stigma changed substantially, this did not consistently translate into shifts in observed behavior. The set of menstrual stigma results also together indicate that the *YGL* component helped improve some of the outcomes more than the base intervention, while for other outcomes the base intervention alone was effective.

### 2.3.5 Psychosocial environment at school

We find evidence of generalized improvements in the psychosocial environment at schools driven by the base program (Table 2.7). Teachers, parents, and students reported changes of these kinds during qualitative work that took place around 9 months after the start of the program, and these findings are corroborated by two outcomes in the endline data.

First, girls report reductions in the *severity* of bullying in schools: while there is no effect on overall bullying levels or light teasing (Table 2.7, pooled $p$: 0.82 and 0.42; columns 1 and 2), there is a significant reduction of 0.08 SD in girls' reports of severe intimidation or harassment (pooled $p$=0.05, column 3).[23] This appears to represent a general improvement in relationships between students, rather than a reduction in menstrual stigma, since it is not driven by reduced bullying *about menstruation*. In fact, such teasing *increases* in treatment schools (Table 2.7, column 4, $p$<0.001), an effect that is driven solely by teasing *from boys* (Table B.1.19, column 5). This suggests that the program was not successful at reducing stigmatizing behaviors among boys, even while attitudes and behaviors strongly improved among girls, and overall severity of bullying decreased.

Second, the program led to a significant increase in a measure of network integration among students (Table 2.7, column 6, pooled $p$=0.001), based on asking girls how many social connections

[21] The double difference (comparing the difference between *Base only* and *Base+YGL* across the *explanation* and *anonymous* outcomes) has $p$-value of 0.06, in line with the claim that the *YGL* intervention's impact is specific to the *explanation* outcome.

[22] There is also no significant effect of any treatment on whether girls themselves share cookies with the other girls (Table B.1.17).

[23] We do not find significant effects on teachers' reports of bullying, although the point estimate on severe intimidation and harassment is negative and a similar magnitude to the girls' reports (Table B.1.18).

**Table 2.6:** *Menstrual stigma - lab measures*

| VARIABLES | (1) Lab in field Explanation Chose menstr. (=1) | (2) Lab in field Anonymous Chose menstr. (=1) | (3) Informative envelope: P(chose menstr.) | (4) Prop. cookies shared w/ handwashing girl | (5) Prop. cookies shared w/ menstruation girl |
|---|---|---|---|---|---|
| Base Only | -0.003 | 0.046** | 0.011 | -0.033 | -0.006 |
| | (0.026) | (0.021) | (0.017) | (0.020) | (0.018) |
| Base + YGL | 0.044** | 0.034 | 0.022 | -0.030* | -0.007 |
| | (0.021) | (0.021) | (0.016) | (0.017) | (0.017) |
| Observations | 1,634 | 1,634 | 2,156 | 1,325 | 1,325 |
| Data source | Lab in the field | Lab in the field | Girls | Boys | Boys |
| Control mean | .208 | .291 | .571 | .454 | .44 |
| Control SD | .406 | .455 | .274 | .206 | .199 |
| p: Treated = 0 | .1649 | .0495 | .2026 | .0567 | .6775 |
| p: Base Only = Base + YGL | .0599 | .5228 | .4219 | .8471 | .9525 |

*Notes:* The table displays results from the revealed-preference lab-in-the-field experiments. In all regressions, standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. Controls selected by double LASSO are included for column (3), where we have baseline measures. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*.

Columns (1)-(2) present results from the *lab-in-the-field* experiment where girls were informed they would need to explain or show (anonymously) a video on a topic in front of their schoolmates. They were asked to choose between a menstruation-related topic and a broader hygiene-related topic. *Explanation, chose menstr.* (column 1) indicates whether the girl chose to provide an in-person explanation about a menstruation-related topic to the group rather than another hygiene topic. *Anonymous, chose menstr.* (column 2) indicates that the girl chose to anonymously show a video about a menstruation-related topic instead of a hygiene-related topic. Controls in columns (1)-(2) include fixed effects for choice order (*Explanation first=1*). Column 3 displays results from the envelopes experiment where girls had to choose between envelopes containing multiple topics, including menstruation-related topics. *Informative envelope: P(chose menstr.)* is the proportion of times out of 4 that girls chose to learn about a menstruation-related topic instead of a non-stigmatizing topic (such as whether mosquitoes are equally attracted to all blood types). Columns (4)-(5) show results from a dictator game implemented during the boys' survey. Boys had to decide how many cookies to share with a girl after watching a video of the girl providing a menstruation or hand-washing-related explanation. The outcomes in these columns represent the proportion of cookies (out of 5) that boys agreed to share with the girl in the video. Controls include order of videos fixed effects.

**Table 2.7:** *Impacts on the school psychosocial environment*

| VARIABLES | (1) Bullying Girls report Combined (Z) | (2) Bullying Girls report Teasing only (Z) | (3) Bullying Girls report Intimidation/ harassment only (Z) | (4) Bullying Girls report About menstr. only (Z) | (5) Bullying Teacher-report Combined (Z) | (6) Network connection index (Z) | (7) Network connection w. girls only (Z) | (8) Network connection w. boys only (Z) |
|---|---|---|---|---|---|---|---|---|
| Base Only | 0.019 | 0.098 | -0.085* | 0.131* | -0.071 | 0.167* | 0.065 | 0.199** |
| | (0.061) | (0.067) | (0.046) | (0.070) | (0.057) | (0.086) | (0.076) | (0.084) |
| Base + YGL | -0.027 | 0.018 | -0.072* | 0.264*** | -0.032 | 0.259*** | 0.175*** | 0.259*** |
| | (0.054) | (0.059) | (0.041) | (0.062) | (0.053) | (0.075) | (0.063) | (0.076) |
| Observations | 2,162 | 2,162 | 2,162 | 2,162 | 3,637 | 2,162 | 2,162 | 2,162 |
| Data source | Girls | Girls | Girls | Girls | Teachers | Girls | Girls | Girls |
| p: Treated = 0 | .8211 | .4254 | .0474 | <0.001 | .386 | .0013 | .0212 | <0.001 |
| p: Base Only = Base + YGL | .3842 | .1519 | .7458 | .0458 | .3923 | .1908 | .0776 | .3817 |
| p: Base Only [Girls] = Base Only [Boys] | | | | | | | | .204 |
| p: Base + YGL [Girls] = Base + YGL [Boys] | | | | | | | | .156 |
| p: Treated [Girls] = Treated [Boys] | | | | | | | | .129 |

*Notes:* The table displays effects on bullying and network connection between students at endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects. All columns except column 5 have baseline controls selected by double LASSO. *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*.

Columns (1)-(4) use girls' reports about bullying in their school, combining peer reports (reporting about other girls being bullied, using a randomly selected subset of 2 or 3 girls in the same grade-level and removing cases where the respondent did not know the other girl) and self-reports (about themselves being bullied). Column 1 combines both light teasing and severe intimidation and harassment. Column 2 includes only teasing. Column 3 includes only severe intimidation and harassment. Column 4 includes only questions about bullying specifically related to menstruation. Column 5 is analogous to column (1) but uses reports from teachers about bullying in their class. Each teacher reports on a randomly selected sample of 3 girls from their class; the observations are at the *teacher × girl* level. *Network connection index (Z)* combines questions about the girls' awareness, friendship, interaction, cooperation, and communication with school peers. Column 6 refers to interactions with all peers. Column 7 restricts to the subset of questions that were asked specifically about female peers; column 8 is specifically about male peers.

they have of different types at endline (e.g., how many friends, people they play with, people they share with, people they do projects with). Social connections with both girls and boys increase significantly.[24]

### 2.3.6 Dynamics: comparing midline and endline

To understand the dynamic process through which the interventions affected the outcomes, Figure 2.3 shows the effect sizes of the two treatment arms at both midline (after approximately 4 months of the program) and endline (after approximately 1.5 years of the program).[25] Four key patterns stand out.

First, the null results seen on school attendance, and self-reported health and mental health also hold at midline, with no significant effects at the 5% level across all measures and treatments (Figure 2.3, panels 1-7).[26] This suggests that these human capital outcomes are not on average affected by the program at any point. We did not administer learning tests or measure heart rate at midline, so we cannot evaluate the dynamics of those outcomes.

Second, the program reduced *menstrual stigma* relatively rapidly, showing significant effects at midline (using a combined index of all menstrual stigma measures available at both midline and endline; Figure 2.3, panel 8; $p$=0.05 for *Base only* and $p$<0.001 for *Base+YGL*). These effects were consolidated by endline, with slightly larger point estimates (although the endline effects do not differ significantly from the midline ($p \in [0.14, 0.33]$).

Third, *knowledge* about menstruation and hygiene improved more rapidly in the *Base+YGL* schools compared to the *Base only* schools at midline ($p$ of difference = 0.08 for hygiene knowledge; 0.002 for menstruation knowledge). But by endline the differences had diminished, suggesting that *Base only* schools caught up (Figure 2.3, panels 9-10; $p$ of [*Base only=Base+YGL*] at endline = 0.17 and 0.92). There are also significant improvements in both types of knowledge between the midline and endline ($p$<0.05 for all comparisons for both treatment groups).

Finally, reported *behavior* was slow to move, with no significant effects at midline except a small effect on menstruation behavior in the *Base+YGL* schools ($p$=0.08). Only by endline do we see strong and significant effects (Figure 2.3, panels 11-12). Combined with the results above, this suggests that behavior takes longer to change than knowledge and reported stigma, in line with a literature showing that generating behavior change can be challenging, especially in the field of hygiene and health (Aboud & Singla, 2012).

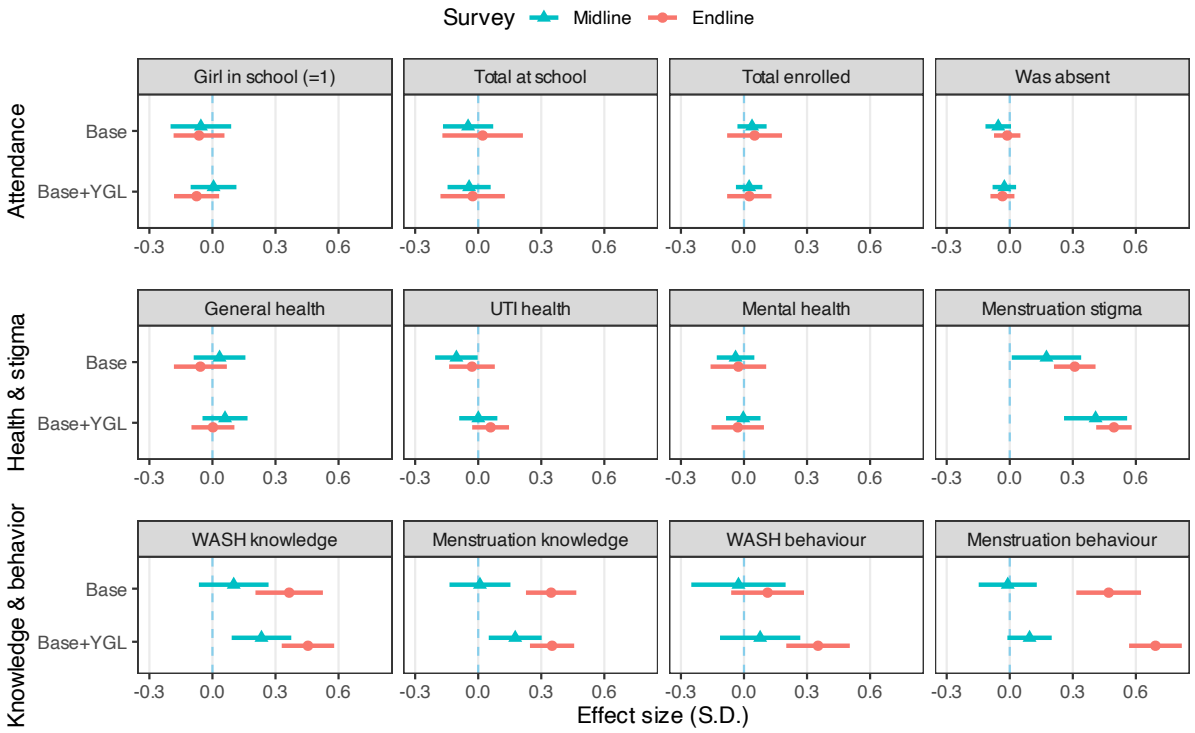
Broadly, the results confirm that outcomes earlier in the theory of change (knowledge and perceptions of stigma) are more responsive at midline, while behavior takes longer to change. The dynamic results further show that the YGLs were instrumental in shifting knowledge, stigma and menstrual behavior more quickly than the base program alone.

[24] Parts of the measured effects are driven by large increases in the number of peers girls speak with specifically about menstruation (Figure B.1.4). But even after removing this question from the index, there are still 0.13-0.16 SD increases in the network integration index for boys (pooled $p$-value = 0.04), suggesting a more general improvement in social dynamics between girls and boys.

[25] To ensure comparability across waves, we (i) only include individuals found at both midline and endline waves in the samples used for estimates in this particular figures; (ii) construct indexes so that they are only composed of questions that were common across midline and endline waves, and use the same factor loadings for factor analysis-based indexes; and (iii) normalize all outcomes using the control mean and standard deviation *at endline*.

[26] One exception, discussed above in section 2.3.2, is that we find suggestive evidence of an increase in the incidence of negative emotions in *YGL* schools at midline (Table B.1.13).

**Figure 2.3:** *Dynamics: comparing effects at midline and endline*




*Notes*: This figure shows the effects on each outcome at both midline and endline. Indices are re-calculated so that they only use the questions that were asked in both midline and endline. Only girls in both the midline and endline samples are included. All outcomes are normalized using the control mean and standard deviation from the endline. All index outcomes are combined using factor analysis, based on the factor loadings from the endline survey.

## 2.4 Mechanisms

In this section, we further investigate the mechanisms behind the program's impacts and the relationships between them. We first note that the relatively large changes in hygiene knowledge, behavior and social stigma in both treatment arms confirm that the program led to meaningful changes along the theory of change. It is thus plausible that they contributed to the observed improvements in final human capital outcomes, in particular in learning. We further argue that the improvements in learning are likely driven by improvements in the psychosocial environment at school, reductions in stress, and associated improvements in quality of learning while at school. We then investigate why the YGL program did not lead to additional impacts on learning or stress, despite generating large additional improvements in hygiene behavior and menstrual stigma. To do so, we explore heterogeneous impacts of the YGL program to understand how they relate to the effects of the base program. Finally, we examine possible negative consequences of the program by examining effects on fertility and backlash.

### 2.4.1 Mechanisms behind learning effects

Why does learning improve in treatment schools? Since the effects on learning are never larger in *Base+YGL* schools than *Base only* schools (Table 2.1), they are driven by base program. Note that in principle, any component of the base program could have led to reduced stress and improved the psychosocial environment (e.g., sanitary pads could reduce stress associated with menstruating at school; teacher training could increase their motivation and improve classroom dynamics). We cannot experimentally evaluate the relative importance of each component, and instead focus on understanding the key mechanisms.

*Health and attendance*. The null results on other human capital outcomes suggest that the effect on learning does not occur through measurable improvements in physical health outcomes or through spending an increased *quantity* of time at school. Instead, the *quality* of time at school seems to have improved.

*Stress*. The reduction in heart rate discussed above points to an important potential channel for this improvement in quality: reducing girls' anxiety and stress could allow them to concentrate more effectively while at school. This is in line with a literature on "bandwidth constraints" (Haushofer & Fehr, 2014; Mullainathan & Shafir, 2013; Kaur et al., in press) that suggests that stress can inhibit decision-making ability and productivity.

*Psychosocial environment*. The reductions in severe bullying in treatment schools and increases in network integration could also underly the learning effects. This channel could operate through the reductions in stress measured as reductions in heart rate (e.g., by reducing the stress associated with bullying or with social isolation). Additionally, the increased social integration could directly improve learning by encouraging peer-to-peer motivation or improving students' motivation.

*Hygiene behaviors and time in class*. The interventions focused on menstrual management and hygiene behaviors could have increased learning while at school by directly impacting how much time girls spent in class during a school day, even conditional on attending school.[27] For example, access to private toilets, sanitary pads, or improved knowledge of good menstrual hygiene may have reduced the likelihood of leaving class due to leaks during menstruation. We find some evidence

[27] While the unannounced spot checks at school could in theory pick up this intensive margin of school attendance, we are likely underpowered to detect such changes, especially since girls may be marked as attending school even if they are (for example) in the toilet during the rollcall.

that the program may indeed affect this "intensive margin" of school attendance. Of the 1053 girls who had started menstruating at baseline, girls in treatment schools were by endline 3.3 p.p. (28%) less likely to report having been absent from school because of menstruation (Table B.1.20, column 2, $p$=0.09); 3.9 p.p. (36%) less likely to have left school during the day because of menstruation (column 4, $p$=0.05); and 4.0 p.p. (18%) less likely to report not being able to fully participate in school activities because of menstruation (column 5, $p$=0.09).[28]

However, it is not clear that improved menstrual management allowing more time in the classroom is a key mechanism behind the learning effects. First, these mechanisms only affect those who have started menstruating by endline, i.e., around 67% of the sample. Second, the effects of menstruation on classroom time seem relatively small, even in the control group. For example, only 11% of menstruating girls in control schools report *ever* having had to leave class because of menstruation. Third, learning effects are not larger for girls who had reached menarche at baseline (Table B.1.21) or for girls who started menstruating between baseline and endline (Table B.1.22). The treatment effects are therefore not driven by girls who we would expect to be able to change their menstrual management behaviors the most.[29] These results together suggest that the learning effects are unlikely to be *primarily* driven by menstrual management behaviors or associated changes in time spent in class. However, this does not rule out that they could contribute to the program's effects, or that the "hard" components of the program could drive *psychosocial* changes that generate learning effects. For example, access to sanitary pads and good menstrual management may reduce the stress associated with girls fearing leaks or smell, enabling better concentration in class.

*Teacher behavior*. We do not find evidence that the learning effects are driven by increases in teacher effort in response to the program, since teacher absenteeism rates are unaffected by the program (Table B.1.23, column 4, $p$=0.62).[30] However, the program generates large improvements in teachers' progressive attitudes towards menstruation (Table B.1.23, column 1, pooled effect: 0.7 SD; $p$<0.001), and smaller but significant improvements in knowledge about menstruation (column 2, pooled $p$=0.05). These changes may make teachers more attentive to *girls* in classrooms, or more responsive when they are being bullied about menstruation, possibly leading to improved psychosocial outcomes or learning improvements for girls.

*Crowd out in the YGL program*. The effects on learning (Table 2.1) suggest that the base program drives the improvements, while including Young Girl Leaders in the intervention generates no further improvements. There may even be a negative effect of the Young Girl Leaders: the point estimates are lower for *Base+YGL* schools than for *Base Only* schools across all measures, and is significantly lower for the *achievement tests* (column 2; $p$=0.02). There is some suggestive evidence that this could be because the time and effort YGLs dedicate to sensitizing others about hygiene and stigma comes at the cost of academic performance, perhaps by crowding out traditional study time. In particular, we find a negative trend in grades of girls nominated as YGLs: between baseline and

[28] We use the baseline measure of menarche, because at endline the reduction in stigma in treatment schools appears to have made girls *more* likely to reveal that they had started menstruating than in control schools, making it an endogenous outcome. Note that we cannot rule out that lower menstruation-driven absenteeism is simply due to fewer girls being *enrolled* in treatment schools (given the negative, although insignificant, point estimate on enrollment in Table 2.2).

[29] In addition, the timing of impacts suggests that infrastructure access is unlikely to be a key driver of learning. Girls were more likely to progress to the next grade-level during the study period (Table 2.1, column 6), implying that learning had improved *by the end of the first school year*, when infrastructure construction had not been completed in the majority of treatment schools.

[30] There is also no effect on *female* teacher absenteeism, that could plausibly have been driven by improved access to hygiene infrastructure ($p$=0.64).

endline, YGLs' average academic score decreases by 1.2 (on a 20-point scale) and their rank within the class worsens by 6.4 places (Table B.1.24). These results should be interpreted with caution, since (i) YGLs were not chosen randomly, preventing us from making causal inferences about their performance compared to non-YGLs, and (ii) some of the time-trend could partly capture regression to the mean, since YGLs were selected to have above-average academic scores (and at endline they still have higher scores than non-YGLs; +0.34 points, $p<0.01$). Nevertheless, this pattern may explain overall weaker learning results in YGL schools. If YGLs influence others' behavior by acting as exemplars, their focus on hygiene and stigma may shift other students' attention away from traditional academic learning.[31] There is also a mechanical effect on school-level learning, since YGLs in the sample will reduce the measured mean.

### 2.4.2 Which YGLs were effective?

The design of the YGL program was motivated by the idea that one could identify *positively deviant* girls in each school, i.e., those willing to speak out in defiance of a harmful norm, and that those girls would be especially effective at changing others' behavior. To examine this hypothesis, we analyze heterogeneity with respect to the selected YGLs' positive deviance in each school using data from the YGL selection process.[32] We find that schools with more positively deviant YGLs indeed exhibit better human capital and behavior outcomes, suggesting that selecting the "right" YGLs can improve outcomes significantly. However, these improvements are strongly conditional on the baseline school environment, either because positively deviant YGLs can only be *found* in already-progressive school environments, or because positively deviant YGLs can only have positive effects if they are working within a progressive school environment.

We first motivate our analysis by using machine learning to test whether there is heterogeneity in the effect of the YGL program, based on a large set of baseline characteristics (following Chernozhukov et al., 2018). We find evidence of significant heterogeneity on the *Base + YGL* treatment for 3 outcomes (Table B.1.25): (i) whether girls are enrolled in school at endline; (ii) menstruation knowledge; and (iii) the index of diffusion of information about menstruation. The heterogeneity parameter also has a $p$-value of 0.105 for the reading fluency measure.

Motivated by this evidence, we then analyze heterogeneity with respect to positive deviance. To do this, we construct an index of each YGLs' *positive deviance*, using data collected during the YGL selection process on their their willingness to engage in prosocial behaviors in defiance of existing norms (see Appendix Section B.2.4 for the questions used).[33] We then average across the YGL of a school to obtain a school-level indicator of positive deviance of the selected YGLs. Importantly, the variation in this index does not only capture individual YGLs' characteristics, but is also likely to

[31] Girls in *Base+YGL* schools are significantly less likely to progress to the next grade-level when the *YGLs* were asked to carry out *formal* classroom sessions, compared to when they were asked to carry out *informal* sensitization, e.g., during recreation ($p$=0.05, Table B.1.2). While a similar difference is not observed in the learning tests ($p$=0.71), it nevertheless suggests that using classroom time to learn about hygiene and menstrual hygiene may crowd out academic learning.

[32] In the design phase, we considered randomizing the YGL selection criteria - selecting positive deviants in some schools and *popular* (network-central) girls in others. We stepped away from this design after piloting, because in many schools there were not enough candidates who satisfied other key criteria (having started menarche and having good school grades) to generate variation in the type of selected YGL *within* a school. Given that there was, however, variation *between* schools in positive deviance, we instead decided at the design stage to test for heterogeneity with respect to the selected YGLs' positive deviance.

[33] Our measure of positive deviance does not simply pick up the overall "quality" of the YGL, e.g., in terms of leaderships skills, confidence, popularity, or intelligence. When we construct a quality measure based on the first factor from an exploratory factor analysis of the YGL selection data, we do not find similar heterogeneity results.

capture features of the baseline school environment, since the probability of finding more positive deviant girls does not vary randomly across schools.

We then parametrically analyze heterogeneity using this school-level indicator of YGL positive deviance (Table B.1.26). Schools with more positively deviant YGLs have significantly better learning test scores, lower rates of severe bullying, and better reported mental health. They also generate strong improvements in UTI health. This suggests that schools with more positively deviant YGLs were more effective at generating learning, health, and psychosocial improvements. Plausibly, girls who are willing and experienced in speaking out in defiance of social norms are better able to speak openly about menstruation without damaging the overall psychosocial environment of the school. Conversely, schools where these positively deviant YGLs were *not* selected have significantly weaker human capital outcomes, and are likely to drive the overall lower point estimates on learning found in *Base+YGL* schools compared to *Base only* schools.

That said, the improved outcomes in schools with positively deviant YGLs appear to be strongly linked to the baseline school environment. First, the positive deviance of selected YGLs in a school at baseline is positively correlated with mothers' education and being a secondary school, and negatively correlated with girls having a brother at baseline (all $p$-values below 0.01). This suggests that the ability to find positively deviant YGLs is strongly related to the school environment.[34] Second, when examining the positive deviance of selected YGLs *relative* to the wider pool of girls who were considered as candidates for the YGL position in a given school, the pattern of heterogeneity does not persist: for example, relatively positively deviant YGLs are not associated with higher learning scores, and are even associated with less progression to the next grade (Table B.1.27). This implies that the earlier heterogeneity results are primarily driven by some schools having more positive deviants among their possible candidate YGLs. Together these results suggest that either (i) positively deviant girls who are willing to defy a harmful norm tend to be mainly *found* in school environments that are already more educated or progresive, or that (ii) positively deviant girls are only *effective* when the school environment is already educated or progressive. In either case, the results re-emphasize the importance of the school environment for girls' human capital outcomes.

### 2.4.3 Possible negative consequences of the program

*Fertility.* We find some evidence that the base program increased the probability of pregnancy (Table B.1.28, column 2). It is difficult to assess the robustness of this result, because the overall rates of pregnancy in the sample are very low: they move from 2% in the control group to 3.5% in the *Base only* group. However, the $p$-value is robust to using randomization inference (Table B.1.9), the effects are not driven by a few outlier schools with many pregnancies (Figure B.1.5), and the results are corroborated by more reports of girls dropping out of school due to pregnancy (Table B.1.28, column 3). This result therefore suggests an important cautionary note for programs that focus on reducing stigma regarding menstrual and reproductive health. Opening up discussion about such topics may also decrease stigma regarding sexual interactions and could lead to early pregnancies for a subset of girls.

*Backlash*. The program encourages girls to speak out against the harmful taboo surrounding menstruation, which could expose them to backlash from teachers, parents, or other students. The overall pattern of results suggests that such backlash was not common, for two reasons. First, the program generated improvements in psychosocial environment and large reductions in menstrual

[34]This could explain the negative heterogeneity result on "Received info.": in places where girls already knew more at baseline, YGLs plausibly focused less on spreading information.

stigma. Second, *YGLs*, who are plausibly most vulnerable to backlash, appear to be well supported by the school community members: 92% of teachers and directors say that YGLs had a positive impact on the school atmosphere, and 80% of girls said that they *admire* the YGLs in an open-ended question. Nevertheless, as discussed above, teasing about menstruation does increase in treatment schools (Table 2.7, column 4, $p<0.001$), which could have negative effects on a subset of girls. Since this was driven by teasing from boys, it suggests that a program that more explicitly targeted boys' attitudes may be helpful in mitigating backlash.

We also find some evidence that mothers report worsened psychosocial behavior for their daughters in *Base only* schools (effect: -0.15 SD, Table B.1.16, column 5, $p=0.05$), although not for *Base+YGL schools* ($p$ of difference: 0.02). The fact that mothers are unhappy with their daughters' behavior may be an indication that they do not agree with the messages of the program. It may also be related to the increases in pregnancy rates in *Base only* schools.

### 2.4.4 Welfare implications of the YGL program

Despite potential costs to academic learning, the overall welfare implications of the YGL program can still be positive, given the countervailing improvements in hygiene and menstrual hygiene-related knowledge, behavior and stigma documented in Sections 2.3.3 and 2.3.4, and the fact that the YGLs offset the negative impacts of the base program on fertility and mothers' perceptions of behavior. Broadly, girls in YGL schools may gain less in academic skills like reading and math, but they learn other skills that can improve their health and well-being. We cannot adjudicate which is more beneficial over the long run. The results do show, however, that these welfare trade-offs are weaker in schools with a more favorable initial environment.

## 2.5 Discussion

Motivated by the evident social stigma that surrounds menstruation in many contexts, policymakers have developed interventions that target menstrual hygiene in schools in low- and middle-income countries. While some interventions focus mainly on relaxing hard constraints to menstrual hygiene management, others also target the menstrual stigma that limits the flow of information and worsens the psychosocial environment for menstruating girls more broadly. This paper provides empirical evidence on the effects of such a broader intervention, which combined sanitary infrastructure and access to sanitary pads with efforts to destigmatize menstruation and its discussion. A unique feature of the approach was to identify and train "Young Girl Leaders" in each school to spread positive messages about hygiene and menstruation, and to encourage healthy behaviors.

After 1.5 years, the program led to large shifts in knowledge and reported hygiene and menstrual hygiene behaviors, as well as improvements in several measures of menstrual stigma, including impacts on objective and observed measures. The results demonstrate that targeted interventions in school can tackle social taboos. While we observe important impacts along the causal chain for the base program, impacts on behavior and stigma were larger (and started earlier) in schools with *Young Girl Leaders* - suggesting that this component was effective. This confirms that identifying and supporting positive deviants has the potential to be an effective way of addressing harmful social norms in school environments. At the same time, the attitudes and behaviors of girls' mothers and male classmates do not seem to be shifted by the program, and positively deviant YGLs only seem to be effective in already-favorable school environments. This points to the need for future research on interventions that target a broader set of actors (e.g., teachers, parents, and boys) to help YGLs be effective in more challenging school environments.

The base intervention itself led to substantial improvements in learning, despite being designed to primarily target hygiene and menstrual hygiene, and having no apparent impact on physical health outcomes and school attendance. It also led to reductions in stress (as measured by reduced heart rate at endline) that could generate improved learning and concentration, along with improvements in the school environment (as measured by increases in social ties students and reductions in severe bullying). This suggests that the focus on health and attendance (in both policy discussions and the academic literature) may be ignoring effects on important welfare-relevant outcomes like learning and stress, and may be neglecting the psychosocial channels that may underlie them. More broadly, the results are therefore highly relevant for discussions on girls' education and the learning crisis in low- and middle-income countries, and point to the importance of the school environment in affecting girls' human capital outcomes.

# 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.

Aboud, F. E., & Singla, D. R. (2012, August). Challenges to changing health behaviours in developing countries: A critical overview. *Social Science & Medicine (1982)*, *75*(4), 589–594. doi: 10.1016/j.socscimed.2012.04.009

Adhvaryu, A., Fenske, J., & Nyshadham, A. (2019, August). Early Life Circumstance and Adult Mental Health. *Journal of Political Economy*, *127*(4), 1516–1549. doi: 10.1086/701606

Adhvaryu, A., Nyshadham, A., Molina, T., & Tamayo, J. (2018). Helping Children Catch Up: Early Life Shocks and the PROGRESA Experiment. *NBER Working Paper Series*, 1–75.

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

Aiyar, A., & Cummins, J. R. (2017). Age-Profile Estimates of the Relationship Between Economic Growth and Child Health. , 46.

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

Akyeampong, K., Andrabi, T., Banerjee, A., Banerji, R., Dynarski, S., Glennerster, R., … Yoshikawa, H. (2023). *Cost-Effective Approaches to Improve Global Learning - What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries?* London, Washington D.C., New York: FCDO, the World Bank, UNICEF, and USAID.

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021, November). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics*, *136*(4), 2147–2194. doi: 10.1093/qje/qjab009

Alan, S., & Kubilay, E. (2024). Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting. *European University Institute, mimeo*.

Ali, T. S., & Rizvi, S. N. (2010, August). Menstrual knowledge and practices of female adolescents in urban Karachi, Pakistan. *Journal of Adolescence*, *33*(4), 531–541. doi: 10.1016/j.adolescence.2009.05.013

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

Almond, D., Currie, J., & Duque, V. (2018, December). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. doi: 10.1257/jel.20171164

Almond, D., & Mazumder, B. (2013). Fetal Origins and Parental Responses. *Annual Review of Economics*, *5*(1), 37–56. doi: 10.1146/annurev-economics-082912-110145

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

Ashraf, N., Bau, N., Low, C., & McGinn, K. (2020). Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment*. *The Quarterly Journal of Economics*. doi: 10.1093/qje/qjz039

Attanasio, O. (2015, December). The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies. *Journal of the European Economic Association*, *13*(6), 949–997. doi: 10.1111/jeea.12159

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., & Rubio-Codina, M. (2020, January). Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia. *American Economic Review*, *110*(1), 48–85. doi: 10.1257/aer.20150183

Attanasio, O., Meghir, C., & Nix, E. (2020, November). Human Capital Development and Parental Investment in India. *The Review of Economic Studies*, *87*(6), 2511–2541. doi: 10.1093/restud/rdaa026

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

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2020, January). Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. *American Economic Journal: Applied Economics*, *12*(1), 210–259. doi: 10.1257/app.20170416

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013, July). The Diffusion of Microfinance. *Science*, *341*(6144), 1236498–1236498. doi: 10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019, November). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *The Review of Economic Studies*, *86*(6), 2453–2490. doi: 10.1093/restud/rdz008

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

Barham, T., Macours, K., & Maluccio, J. A. (2013). Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *American Economic Review*, *103*(3), 467–471. doi: 10.1257/aer.103.3.467

Barham, T., Macours, K., & Maluccio, J. A. (2018). Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes. (August).

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., & Shapiro, J. S. (2016, February). Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy*, *124*(1), 105–159. doi: 10.1086/684582

Barreca, A. I. (2012, January). Climate change, humidity, and mortality in the United States. *Journal of Environmental Economics and Management*, *63*(1), 19–34. doi: 10.1016/j.jeem.2011.07.004

Bazzi, S., & Gudgeon, M. (2016). Local Government Proliferation , Diversity, and Conflict.

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021, June). Can Network Theory-Based Targeting Increase Technology Adoption? *American Economic Review*, *111*(6), 1918–1943. doi: 10.1257/aer.20200295

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.

Beguería, S., Vicente-Serrano, S. M., Reig, F., & Latorre, B. (2014, August). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, *34*(10), 3001–3023. doi: 10.1002/joc.3887

Beguería, S., & Vicente Serrano, S. M. (2017). *SPEIbase v.2.5 Global 04-month 1901-2015 SPEI, 2017*. http://digital.csic.es/handle/10261/153475?locale=en. doi: 10.20350/digitalCSIC/8508

Belachew, T., Hadley, C., Lindstrom, D., Getachew, Y., Duchateau, L., & Kolsteren, P. (2011, September). Food insecurity and age at menarche among adolescent girls in Jimma Zone Southwest Ethiopia: A longitudinal study. *Reproductive biology and endocrinology: RB&E*, *9*, 125. doi: 10.1186/1477-7827-9-125

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

Benjamin, D. J., Debnam, J., Fleurbaey, M., Heffetz, O., & Kimball, M. (2020). What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents. , 41.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Benshaul-Tolonen, Anja, Aja, Zulaika, G., Sommer, M., & Phillips-Howard, P. A. (2020). Measuring menstruation-related absenteeism among adolescents in low-income countries. In *The Palgrave Handbook of Critical Menstruation Studies* (pp. 705–723). Palgrave Macmillan.

Benshaul-Tolonen, A., Zulaika, G., Nyothach, E., Odour, C., Mason, L., Obor, D., & Laserson, K. F. (2021). *Sanitary products, absenteeism and psychosocial well-being: Evidence from a three-arm cluster randomized controlled feasibility study in Western Kenya* (Working Paper No. WP No.93). Columbia University, SIPA, CDEP-CGEG.

Bergstrom, K., & Ozler, B. (2021). *Improving the Well-Being of Adolescent Girls in Developing Countries*. The World Bank. doi: 10.1596/1813-9450-9827

Bernard, T., Dercon, S., Orkin, K., & Taffesse, A. S. (2019, May). Parental Aspirations for Children's Education: Is There a "Girl Effect"? Experimental Evidence from Rural Ethiopia. *AEA Papers and Proceedings*, *109*, 127–132. doi: 10.1257/pandp.20191072

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

Bharadwaj, P., Løken, K. V., & Neilson, C. (2013, August). Early Life Health Interventions and Academic Achievement. *American Economic Review*, *103*(5), 1862–1891. doi: 10.1257/aer.103.5.1862

Bharati, T., & Chin, S. (2016). Does Education Affect Time Preference? *SSRN Electronic Journal*. doi: 10.2139/ssrn.2880879

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. doi: 10.1093/acprof:oso/9780190622046.001.0001

Bicchieri, C., & Funcke, A. (2018, March). Norm Change: Trendsetters and Social Structure. *Social Research: An International Quarterly*, *85*(1), 1–21. doi: 10.1353/sor.2018.0002

Binswanger, H. P. (1980, August). Attitudes Toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics*, *62*(3), 395–407. doi: 10.2307/1240194

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

Breza, E. (2016, April). Field Experiments, Social Networks, and Development. In Y. Bramoullé, A. Galeotti, & B. W. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks* (pp. 411–439). Oxford University Press. doi: 10.1093/oxfordhb/9780199948277.013.25

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

Brown, S., & Taylor, K. (2008). Bullying, education and earnings: Evidence from the National Child Development Study. *Economics of Education Review*, *27*(4), 387–401.

Buehren, N., Goldstein, M. P., Gulesci, S., Sulaiman, M., & Yam, V. (2017, February). Evaluation of an adolescent development program for girls in Tanzania. *Policy Research Working Paper Series*(7961).

Burgess, R., Deschênes, O., Donaldson, D., & Greenstone, M. (2017). Weather, climate change and death in India. *Unpublished Manuscript*, 71.

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

Buss, A. T., Fox, N., Boas, D. A., & Spencer, J. P. (2014, January). Probing the early development of visual working memory capacity with functional near-infrared spectroscopy. *NeuroImage*, *85*, 314–325. doi: 10.1016/j.neuroimage.2013.05.034

Buss, A. T., & Spencer, J. P. (2014, June). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii, 1–103. doi: 10.1002/mono.12096

Buss, A. T., & Spencer, J. P. (2018, July). Changes in frontal and posterior cortical activity underlie the early emergence of executive function. *Developmental Science*, *21*(4), e12602. doi: 10.1111/desc.12602

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

Carneiro, P., Kraftman, L., Mason, G., Moore, L., Rasul, I., & Scott, M. (2021, August). The Impacts of a Multifaceted Prenatal Intervention on Human Capital Accumulation in Early Life. *American Economic Review*, *111*(8), 2506–2549. doi: 10.1257/aer.20191726

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

Cattell, R. B. (1966, April). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. doi: 10.1207/s15327906mbr0102_10

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

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2018, June). *Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India* (Working Paper No. 24678). National Bureau of Economic Research. doi: 10.3386/w24678

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

Chirgwin, H., Cairncross, S., Zehra, D., & Sharma Waddington, H. (2021, December). Interventions promoting uptake of water, sanitation and hygiene (WASH) technologies in low- and middle-income countries: An evidence and gap map of effectiveness studies. *Campbell Systematic Reviews*, *17*(4), e1194. doi: 10.1002/cl2.1194

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

Chuang, Y., & Schechter, L. (2015, November). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, *117*, 151–170. doi: 10.1016/j.jdeveco.2015.07.008

Clark, L. A., & Watson, D. (1995, September). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. doi: 10.1037/1040-3590.7.3.309

Cordero, M. E., D'Acuña, E., Benveniste, S., Prado, R., Nuñez, J. A., & Colombo, M. (1993). Dendritic development in neocortex of infants with early postnatal life undernutrition. *Pediatric Neurology*, *9*(6), 457–464. doi: 10.1016/0887-8994(93)90025-8

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*.

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, *97*(2), 25.

Cunha, F., & Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, *43*(4), 738–782. doi: 10.3368/jhr.43.4.738

Cunha, F., & Heckman, J. J. (2009, April). The Economics and Psychology of Inequality and Human DEvelopment. *Journal of the European Economic Association*, *7*(2-3), 320–364. doi: 10.1162/JEEA.2009.7.2-3.320

Cunha, F., Heckman, J. J., & Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931. doi: 10.3982/ECTA6551

Currie, J., & Almond, D. (2011). Human capital development before age five. In *Handbook of Labor Economics* (Vol. 4, pp. 1315–1486). Elsevier B.V. doi: 10.1016/S0169-7218(11)02413-0

Czura, K., Menzel, A., & Miotto, M. (2019). Menstrual Health, Worker Productivity and Well-being among Female Bangladeshi Garment Workers. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3519895

Dal Bó, E., Finan, F., Li, N. Y., & Schechter, L. (2018). Government Decentralization Under Changing State Capacity: Experimental Evidence From Paraguay. , 66.

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

Das, J., Do, Q. T., Friedman, J., & McKenzie, D. (2009). Mental health patterns and consequences: Results from survey data in five developing countries. *World Bank Economic Review*, *23*(1), 31–55. doi: 10.1093/wber/lhn010

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 Chaisemartin, C., & D'Haultfœuille, X. (2018, April). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, *85*(2), 999–1028. doi: 10.1093/restud/rdx049

de Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way Fixed Effects Regressions with Several Treatments. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3751060

de Quidt, J., & Haushofer, J. (2019). Depression through the Lens of Economics. In *The Economics of Poverty Traps* (pp. 127–152). doi: 10.7208/chicago/9780226574448.003.0003

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

Dean, J. T., & Jayachandran, S. (2020). Attending kindergarten improves cognitive development in India, but all kindergartens are not equal. , 64.

De Datta, S. K. (1981). *Principles and practices of rice production*. New York: Wiley.

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

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–798. doi: 10.1257/jel.52.3.740

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

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, *1*(3), 111–34. doi: 10.1257/app.1.3.111

Deoni, S., Dean, D., Joelson, S., O'Regan, J., & Schneider, N. (2018, September). Early nutrition influences developmental myelination and cognition in infants and young children. *NeuroImage*, *178*, 649–659. doi: 10.1016/j.neuroimage.2017.12.056

Department, U. S. S. (2022). *2022 Country Reports on Human Rights Practices*.

Deschênes, O., & Greenstone, M. (2011, October). Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US. *American Economic Journal: Applied Economics*, *3*(4), 152–185. doi: 10.1257/app.3.4.152

Deschênes, O., Greenstone, M., & Guryan, J. (2009, April). Climate Change and Birth Weight. *American Economic Review*, *99*(2), 211–217. doi: 10.1257/aer.99.2.211

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, March). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, *112*(3), 899–927. doi: 10.1257/aer.20201112

Dinarte, L., & Egana del Sol, P. (2018). Preventing Violence in the Most Violent Contexts. Behavioral and Neurophysiological Evidence. (May).

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

Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. In *Handbook of economic field experiments* (Vol. 2, pp. 3–93). Elsevier.

Edmonds, E., Feigenberg, B., & Leight, J. (2020, July). *Advancing the Agency of Adolescent Girls* (Tech. Rep. No. w27513). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w27513

El-Gilany, A.-H., Badawi, K., & El-Fedawy, S. (2005, January). Menstrual Hygiene among Adolescent Schoolgirls in Mansoura, Egypt. *Reproductive Health Matters*, *13*(26), 147–152. doi: 10.1016/S0968-8080(05)26191-8

Eriksen, T. L. M., Nielsen, H. S., & Simonsen, M. (2014). Bullying in Elementary School. *Journal of Human Resources*, *49*(4), 839–871.

Evans, D. K., & Yuan, F. (2022, February). What We Learn about Girls' Education from Interventions That Do Not Focus on Girls. *The World Bank Economic Review*, *36*(1), 244–267. doi: 10.1093/wber/lhab007

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

Fernández, R., Parsa, S., & Viarengo, M. (2019). Coming out in America: AIDS, Politics, and Cultural Change. , 65.

Ferrer, E. (2009, May). Fluid reasoning and the developing brain. *Frontiers in Neuroscience*, *3*(1). doi: 10.3389/neuro.01.003.2009

Fishman, R. (2016, February). More uneven distributions overturn benefits of higher precipitation for crop yields. *Environmental Research Letters*, *11*(2), 024004. doi: 10.1088/1748-9326/11/2/024004

Fishman, R. (2018, March). Groundwater depletion limits the scope for adaptation to increased rainfall variability in India. *Climatic Change*, *147*(1-2), 195–209. doi: 10.1007/s10584-018-2146-x

Fiske, A., & Holmboe, K. (2019, June). Neural substrates of early executive function development. *Developmental Review*, *52*, 42–62. doi: 10.1016/j.dr.2019.100866

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

Food and Agriculture Organisation. (2005). *Fertilizer use by crop in Indonesia.*

Frankenberg, E., & Karoly, L. (1995). *The 1993 indonesian family life survey: Overview and field report, publication no* (Tech. Rep.). DRU-1195/1-NICHD/AID, RAND, Santa Monica, CA.

Frederick, W. H., Worden, R. L., & of Congress. Federal Research Division, L. (2011). *Indonesia: A Country Study*. Federal Research Division, Library of Congress.

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

Garces, E., Thomas, D., & Currie, J. (2002). Longer-Term Effects of Head Start. *The American Economic Review*, *92*(4), 32.

Gertler, P., Heckman, J. J., Pinto, R., Chang, S. M., Grantham-McGregor, S., Vermeersch, C., ... Wright, A. (2021). Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31. , 50.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018, March). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, *19*(3), 123–137. doi: 10.1038/nrn.2018.1

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

Goodman-Bacon, A. (2018, September). *Difference-in-Differences with Variation in Treatment Timing* (Tech. Rep. No. w25018). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w25018

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

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007,

January). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60–70. doi: 10.1016/S0140-6736(07)60032-4

Grantham-McGregor, S., Powell, C., Walker, S., & Himes, J. (1991, July). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican Study. *The Lancet*, *338*(8758), 1–5. doi: 10.1016/0140-6736(91)90001-6

Groh, M., McKenzie, D., & Vishwanath, T. (2015). Reducing Information Asymmetries in the Youth Labor Market of Jordan with Psychometrics and Skill Based Tests. *The World Bank Economic Review*, *29*(suppl 1), S106-S117. doi: 10.1093/wber/lhv005

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., & Young, H. (2023). A Stepping Stone Approach to Norm Transitions. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4425503

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

Gunnsteinsson, S., Adhvaryu, A., Christian, P., Labrique, A., Sugimoto, J., Shamim, A. A., & West, K. P. (2014). Resilience to early life shocks: Evidence from the Interaction of a Randomized Controlled Trial and a Natural Experiment. *Working Paper*, 1–48.

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

Hamory, J., Baird, S., Das, S., Jones, N., Woldehanna, T., & Yadete, W. (2023). Do layered adolescent-centric interventions improve girls' capabilities? Evidence from a mixed-methods cluster randomised controlled trial in Ethiopia.

Hanson, J. L., Hair, N., Shen, D. G., Shi, F., Gilmore, J. H., Wolfe, B. L., & Pollak, S. D. (2013, December). Family Poverty Affects the Rate of Human Infant Brain Growth. *PLoS ONE*, *8*(12), e80954. doi: 10.1371/journal.pone.0080954

Harari, M., & Ferrara, E. L. (2018, October). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics*, *100*(4), 594–608. doi: 10.1162/rest_a_00730

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

Haushofer, J., & Fehr, E. (2014, May). On the psychology of poverty. *Science*, *344*(6186), 862–867. doi: 10.1126/science.1232491

Heckman, J. J. (2006, June). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(5782), 1900–1902. doi: 10.1126/science.1128898

Heckman, J. J., Pinto, R., & Savelyev, P. (2013, October). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, *103*(6), 2052–2086. doi: 10.1257/aer.103.6.2052

Heckman, J. J., Stixrud, J., & Urzua, S. (2006, February). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. , w12006. doi: 10.3386/w12006

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

Hennegan, J., & Montgomery, P. (2016, February). Do Menstrual Hygiene Management Interventions Improve Education and Psychosocial Outcomes for Women and Girls in Low and Middle Income Countries? A Systematic Review. *PLOS ONE*, *11*(2), e0146985. doi: 10.1371/journal.pone.0146985

Hill, S. J., Lo, J., Vavreck, L., & Zaller, J. (2013, October). How Quickly We Forget: The Duration of Persuasion Effects From Mass Communication. *Political Communication*, *30*(4), 521–547. doi: 10.1080/10584609.2013.828143

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

Hochberg, Z., & Belsky, J. (2013, April). Evo-devo of human adolescence: Beyond disease models of early puberty. *BMC medicine*, *11*, 113. doi: 10.1186/1741-7015-11-113

Horn, J. L. (1965, June). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi: 10.1007/BF02289447

Hoynes, H., Schanzenbach, D. W., & Almond, D. (2016, April). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review*, *106*(4), 903–934. doi: 10.1257/aer.20130375

Hu, Q. (2023). Breaking Down Bullying: Empathy, Social Networks, and Adolescents.

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

Isaacs, E., Oates, J., & ILSI Europe a.i.s.b.l. (2008, August). Nutrition and cognition: Assessing cognitive abilities in children and young people. *European Journal of Nutrition*, *47*(S3), 4–24. doi: 10.1007/s00394-008-3002-y

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. (2021, September). Social Norms as a Barrier to Women's Employment in Developing Countries. *IMF Economic Review*, *69*(3), 576–595. doi: 10.1057/s41308-021-00140-w

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

Johnson, R. C., & Jackson, C. K. (2019, November). Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal: Economic Policy*, *11*(4), 310–349. doi: 10.1257/pol.20180510

Kaiser, H. F. (1958, September). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. doi: 10.1007/BF02289233

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

Katz, D., Allport, F. H., & Jenness, M. B. (1931). *Students' attitudes; a report of the Syracuse University reaction study*. Oxford, England: Craftsman Press.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (in press). Do financial concerns make workers less productive? *The Quarterly Journal of Economics*.

Kennedy, G., Ballard, T., Dop, M. C., & European Union. (2011). *Guidelines for measuring household and individual dietary diversity*. Rome: Food and Agriculture Organization of the United Nations.

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

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007, January). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119. doi: 10.1111/j.1468-0262.2007.00733.x

Knudsen, E. I. (2004, October). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, *16*(8), 1412–1425. doi: 10.1162/0898929042304796

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*(27), 10155–10162. doi: 10.1073/pnas.0600888103

Kudamatsu, M., Persson, T., & Strömberg, D. (2012). Weather and Infant Mortality in Africa. , 77.

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

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., ... Vakis, R. (2019, July). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, *5*(7), eaaw5226. doi: 10.1126/sciadv.aaw5226

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

Ledesma, R. D., Valero-Mora, P., & de Valencia, U. (n.d.). Determining the Number of Factors to Retain in EFA: An easy-to- use computer program for carrying out Parallel Analysis. *Exploratory Factor Analysis*, *12*(2), 11.

Lee, C.-T., Zhang, G., & Edwards, M. C. (2012, March). Ordinary Least Squares Estimation of Parameters in Exploratory Factor Analysis With Ordinal Data. *Multivariate Behavioral Research*, *47*(2), 314–339. doi: 10.1080/00273171.2012.658340

Levine, D., & Yang, D. (2014, July). *The Impact of Rainfall on Rice Output in Indonesia* (Tech. Rep. No. w20302). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w20302

Little, D. R., Lewandowsky, S., & Craig, S. (2014, March). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00239

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

Maccini, S., & Yang, D. (2009, May). Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review*, *99*(3), 1006–1026. doi: 10.1257/aer.99.3.1006

Macours, K., Premand, P., & Vakis, R. (2022, September). Transfers, Diversification and Household Risk Strategies: Can Productive Safety Nets Help Households Manage Climatic Variability? *The Economic Journal*, *132*(647), 2438–2470. doi: 10.1093/ej/ueac018

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

Malamud, O., Pop-Eleches, C., & Urquiola, M. (2016, March). *Interactions Between Family and School Environments: Evidence on Dynamic Complementarities?* (Tech. Rep. No. w22112). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w22112

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

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., & Stein, A. D. (2009, April). The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults. *The Economic Journal*, *119*(537), 734–763. doi: 10.1111/j.1468-0297.2009.02220.x

Marsh, D. R., Schroeder, D. G., Dearden, K. A., Sternin, J., & Sternin, M. (2004, November). The power of positive deviance. *BMJ*, *329*(7475), 1177–1179. doi: 10.1136/bmj.329.7475.1177

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

May, W. (2004, March). Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment. *Climate Dynamics*, *22*(2-3), 183–204. doi: 10.1007/s00382-003-0373-x

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

Minnesota Population Center. (2018). *Integrated Public Use Microdata Series, International: Version 7.1. Indonesia Spatially Harmonized Second-Level Geography (1970-2011).* https://international.ipums.org/international/gis_harmonized_2nd.shtml. doi: 10.18128/D020.V7.1

Minnesota Population Center. (2020). *Integrated Public Use Microdata Series, International: Version 7.1. Indonesia Census Data (1980, 1985, 1990, 1995, 2000, 2005, 2010).* https://international.ipums.org/international/gis_harmonized_2nd.shtml. doi: 10.18128/D020.V7.1

Monfreda, C., Ramankutty, N., & Foley, J. A. (2008, March). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000: GLOBAL CROP AREAS AND YIELDS IN 2000. *Global Biogeochemical Cycles*, *22*(1), n/a-n/a. doi: 10.1029/2007GB002947

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

Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, *12*(2), 125–135. doi: 10.1037/h0027568

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. Henry Holt and Company.

Muralidharan, K. (2017). Field Experiments in Education in Developing Countries. In A. Banerjee & E. Duflo (Eds.), *Handbook of Field Experiments* (pp. 323–385). doi: 10.1016/bs.hefe.2016.09.004

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

Muralidharan, K., & Sundararaman, V. (2015, August). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India *. *The Quarterly Journal of Economics*, *130*(3), 1011–1066. doi: 10.1093/qje/qjv013

Myers, D. G., & Bishop, G. D. (1970, August). Discussion Effects on Racial Attitudes. *Science*, *169*(3947), 778–779. doi: 10.1126/science.169.3947.778

Myers, D. G., & Lamm, H. (1976, July). The group polarization phenomenon. *Psychological Bulletin*, *83*(4), 602–627. doi: 10.1037/0033-2909.83.4.602

National Sample Survey Office, India. (2012). *HCE: Monthly per Capita Consumer Expenditure: Average: Tamil Nadu: Urban: Food | Economic Indicators | CEIC*.

Naylor, R. L., Battisti, D. S., Vimont, D. J., Falcon, W. P., & Burke, M. B. (2007, May). Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proceedings of the National Academy of Sciences*, *104*(19), 7752–7757. doi: 10.1073/pnas.0701825104

Nelson, C. A., & Gabard-Durnam, L. J. (2020, March). Early Adversity and Critical Periods: Neurodevelopmental Consequences of Violating the Expectable Environment. *Trends in Neurosciences*, *43*(3), 133–143. doi: 10.1016/j.tins.2020.01.002

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

Orr, L., Ferris, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., ... Kling, J. R. (2003). *Moving to opportunity: Interim impacts evaluation* (Report). Washington D.C: U.S. Dept. of HUD.

Oster, E., & Thornton, R. (2011, January). Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics*, *3*(1), 91–100. doi: 10.1257/app.3.1.91

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

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021, January). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology*, *72*(1), 533–560. doi: 10.1146/annurev-psych-071620-030619

Paradkar, N., & Chowdhury, S. R. (2015). Primary study for detection of arterial blood pressure waveform components. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2015*, 1959–1962. doi: 10.1109/EMBC.2015.7318768

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

Perkins, J. M., Kim, R., Krishna, A., McGovern, M., Aguayo, V. M., & Subramanian, S. (2017, November). Understanding the association between stunting and child development in low- and middle-income countries: Next steps for research and intervention. *Social Science & Medicine*, *193*, 101–109. doi: 10.1016/j.socscimed.2017.09.039

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.

Phillips-Howard, P. A., Nyothach, E., Ter Kuile, F. O., Omoto, J., Wang, D., Zeh, C., ... Laserson, K. F. (2016, November). Menstrual cups and sanitary pads to reduce school attrition, and sexually transmitted and reproductive tract infections: A cluster randomised controlled feasibility study in rural Western Kenya. *BMJ open*, *6*(11), e013229. doi: 10.1136/bmjopen-2016-013229

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

Prado, E. L., & Dewey, K. G. (2014, April). Nutrition and brain development in early life. *Nutrition Reviews*, *72*(4), 267–284. doi: 10.1111/nure.12102

Prentice, A. M., Ward, K. A., Goldberg, G. R., Jarjou, L. M., Moore, S. E., Fulford, A. J., & Prentice, A. (2013, May). Critical windows for nutritional interventions against stunting. *The American Journal of Clinical Nutrition*, *97*(5), 911–918. doi: 10.3945/ajcn.112.052332

Radloff, L. S. (1977, June). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. doi: 10.1177/014662167700100306

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

Rammstedt, B., & John, O. P. (2007, February). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. doi: 10.1016/j.jrp.2006.02.001

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

Raven, J. (2000, August). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. doi: 10.1006/cogp.1999.0735

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

Reece, J. D., Bunn, J. A., Choi, M., & Navalta, J. W. (2021, June). Assessing Heart Rate Using Consumer Technology Association Standards. *Technologies*, *9*(3), 46. doi: 10.3390/technologies9030046

Reynolds, A. J., Temple, J. A., White, B. A. B., Ou, S.-R., & Robertson, D. L. (2011, January). Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program: Cost-Benefit Analysis. *Child Development*, *82*(1), 379–404. doi: 10.1111/j.1467-8624.2010.01563.x

Rietveld, C. A., van Kippersluis, H., & Thurik, A. R. (2015, October). Self-Employment and Health:

Barriers or Benefits?: SELF-EMPLOYMENT AND HEALTH. *Health Economics*, *24*(10), 1302–1313. doi: 10.1002/hec.3087

Ritchie, J. T., & Nesmith, D. S. (2015, October). Temperature and Crop Development. In J. Hanks & J. T. Ritchie (Eds.), *Agronomy Monographs* (pp. 5–29). Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. doi: 10.2134/agronmonogr31.c2

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

Roser, M., & Ritchie, H. (2019). Hunger and undernourishment. *Our World in Data*.

Rossin-Slater, M., & Wüst, M. (2016). *Are Different Early Investments Complements or Substitutes? Long-Run and Intergenerational Evidence from Denmark* (Tech. Rep.).

Rubio-Codina, M., Attanasio, O., Meghir, C., Varela, N., & Grantham-McGregor, S. (2015). The Socioeconomic Gradient of Child Development: Cross-Sectional Evidence from Children 6–42 Months in Bogota. *Journal of Human Resources*, *50*(2), 464–483. doi: 10.3368/jhr.50.2.464

Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010, June). Crop planting dates: An analysis of global patterns: Global crop planting dates. *Global Ecology and Biogeography*, no-no. doi: 10.1111/j.1466-8238.2010.00551.x

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

Sarzosa, M., & Urzúa, S. (2021). Bullying among adolescents: The role of skills. *Quantitative Economics*, *12*(3), 945–980. doi: 10.3982/QE1215

Schilbach, F., Schofield, H., & Mullainathan, S. (2016, May). The Psychological Lives of the Poor. *American Economic Review*, *106*(5), 435–440. doi: 10.1257/aer.p20161101

Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006, February). The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions. *Review of Economics and Statistics*, *88*(1), 113–125. doi: 10.1162/rest.2006.88.1.113

Schlenker, W., & Roberts, M. J. (2006, September). Nonlinear Effects of Weather on Corn Yields*. *Review of Agricultural Economics*, *28*(3), 391–398. doi: 10.1111/j.1467-9353.2006.00304.x

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

Schweinhart, L. J., & Weikart, D. P. (1981, October). Effects of the Perry Preschool Program on Youths Through Age 15. *Journal of the Division for Early Childhood*, *4*(1), 29–39. doi: 10.1177/105381518100400105

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

Shah, M., Seager, J., Montalvao, J., & Goldstein, M. (2023, August). *Sex, Power, and Adolescence: Intimate Partner Violence and Sexual Behaviors* (Working Paper No. 31624). National Bureau of Economic Research. doi: 10.3386/w31624

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

Shiller, R. J. (2017, April). Narrative Economics. *American Economic Review*, *107*(4), 967–1004. doi: 10.1257/aer.107.4.967

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.

Smith, L. G. E., & Postmes, T. (2011). The power of talk: Developing discriminatory group norms through discussion. *British Journal of Social Psychology*, *50*(2), 193–215. doi: 10.1348/014466610X504805

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

Statistik, B. P., & SIAK, K. (2019). Indonesia (IDN) Administrative Boundary Common Operational Database (COD-AB).

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (1998). The second wave of the indonesia family life survey (IFLS3): Overview and field report. *NIA/NICHD*.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (2004). The third wave of the indonesia family life survey (IFLS3): Overview and field report. *NIA/NICHD*.

Strauss, J., Witoelar, F., & Sikoki, B. (2016). *The fifth wave of the indonesia family life survey: Overview and field report; RAND labor and population.* WR-1143/1-NIA/NICHD.

Strauss, J., Witoelar, F., Sikoki, B., & Wattie, A. M. (2009). *The fourth wave of the Indonesia Family Life Survey: Overview and field report.* RAND Labor and Population Working Paper WR-675/1-NIA/NICHD. Santa Monica, CA ....

Subramanian, T., Gupte, M., Dorairaj, V., Periannan, V., & Mathai, A. (2009, April). Psycho-social 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

Tebaldi, C., Hayhoe, K., Arblaster, J. M., & Meehl, G. A. (2006, November). Going to the Extremes: An Intercomparison of Model-Simulated Historical and Future Changes in Extreme Events. *Climatic Change*, *79*(3-4), 185–211. doi: 10.1007/s10584-006-9051-4

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

van den Besselaar, E. J. M., van der Schrier, G., Cornes, R. C., Iqbal, A. S., & Klein Tank, A. M. G. (2017, July). SA-OBS: A Daily Gridded Surface Temperature and Precipitation Dataset for Southeast Asia. *Journal of Climate*, *30*(14), 5151–5165. doi: 10.1175/JCLI-D-16-0575.1

Velicer, W. F. (1976, September). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. doi: 10.1007/BF02293557

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010, April). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. doi: 10.1175/2009JCLI2909.1

Victora, C. G., de Onis, M., Hallal, P. C., Blossner, M., & Shrimpton, R. (2010, March). Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *PEDIATRICS*, *125*(3), e473-e480. doi: 10.1542/peds.2009-1519

Wachs, T. D., Georgieff, M., Cusick, S., & McEwen, B. S. (2014, January). Issues in the timing of integrated early interventions: Contributions from nutrition, neuroscience, and psychological research: Timing of integrated early interventions. *Annals of the New York Academy of Sciences*, *1308*(1), 89–106. doi: 10.1111/nyas.12314

Walker, S. P., Chang, S. M., Powell, C. A., & Grantham-McGregor, S. M. (2005, November). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The Lancet*, *366*(9499), 1804–1807. doi: 10.1016/S0140-6736(05)67574-5

Walker, S. P., Chang, S. M., Wright, A. S., Pinto, R., Heckman, J. J., & Grantham-McGregor, S. M. (2021, August). Cognitive, psychosocial, and behaviour gains at age 31 years from the Jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry*, jcpp.13499. doi: 10.1111/jcpp.13499

WaterAid. (2013). *We can't wait: A report on sanitation and hygiene for women and girls.* WaterAid International. Retrieved from: http://www.zaragoza.es/contenidos/medioambiente/onu/1325-eng_We_cant_wait_sanitation_and_hygiene_for_women_and%20girls.pdf.

Webb, D. (2023). Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India. *mimeo, Paris School of Economics*.

Welz, M., Alfons, A., Demirer, M., & Chernozhukov, V. (2022). GenericML: Generic machine learning inference. *R package version*.

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

White, M. D. (2018). The Problems with Measuring and Using Happiness for Policy Purposes. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3191385

WHO, & UNICEF. (2009). WHO child growth standards and the identification of severe acute malnutrition in infants and children. , 1–12. doi: http://www.who.int/nutrition/publications/severemalnutrition/9789241598163/en/

Wolke, D., Copeland, W. E., Angold, A., & Costello, E. J. (2013). Impact of Bullying in Childhood on Adult Health, Wealth, Crime, and Social Outcomes. *Psychological Science*, *24*(10), 1958–1970.

World Bank. (2018). *Rural Population in Indonesia*.

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

# Chapter 3

# Critical Periods in Cognitive and Socioemotional Development: Evidence from Weather Shocks in Indonesia

## 3.1 Introduction

Cognitive and socioemotional skills are fundamental measures of human capital, and are predictive of adult labour market outcomes and well-being (Cunha & Heckman, 2007, 2009; Grantham-McGregor et al., 2007; Heckman et al., 2006; Attanasio, 2015). An extensive literature shows that an individual's circumstances in early life can have a major impact on their cognitive and socioemotional development, and therefore on their economic welfare (Bharadwaj et al., 2013; Almond et al., 2018; Carneiro et al., 2021; Attanasio, Cattan, et al., 2020; Walker et al., 2021). A key claim in this literature is that *timing matters*: the same shock or intervention may have different impacts at different stages of life. Which stages are particularly important for determining adult human capital? Some commentators suggest that circumstances before age 5 are most important, with younger ages being increasingly sensitive within that window (Currie & Almond, 2011; Almond et al., 2018). Others cite a narrower range, claiming that the first 1000 days of life are a "critical period" for cognitive development, thereby justifying a focus on the earliest years for policy intervention (Heckman, 2006).

There is, however, little causal evidence on how much timing matters for the effects of early life circumstance on adult human capital. Few studies directly compare the effects of policies or shocks across ages (Almond et al., 2018). The focus on the broad window (to age 5) is a convention sometimes justified by evidence that adult outcomes improve linearly with increasing exposure to beneficial policies only in this window (Hoynes et al., 2016).

In this paper, I use a setting in which individuals are subject to repeated, uncorrelated, exogenous shocks throughout their childhood and carry out a fine-grained test for the existence of critical periods in both cognitive and socioemotional development. I examine the effect of weather variation at every age from in utero to age 15 on adult cognitive and socioemotional outcomes. The data come from individuals in the Indonesia Family Life Survey (IFLS) born in rural areas between 1988 and 2000. Such individuals experience the weather variation as a shock to aggregate household welfare that translate to changes in nutritional investments for children. My approach provides

causal evidence on whether the timing of such shocks matters for adult human capital.

I show evidence of a critical period at age 2 for the effect of weather variation on adult cognitive ability. A 1 standard deviation improvement in my primary weather measure at age 2 leads to a significant increase in adult cognitive ability of 0.063 standard deviations. Similar weather variation does not have a statistically significant effect at any other age (including in utero). The age 2 effect is highly robust to different empirical specifications, and I use placebo tests to demonstrate that such a strongly significant result is extremely unlikely to be the result of artefacts in the data. By contrast, there is no positive evidence of a critical period for socioemotional development: effects are either null or cannot be isolated to a single critical period.

I show suggestive evidence that weather variation has an effect on adult cognitive ability by affecting nutritional investments during childhood. Households in rural areas experience weather variation as a shock to aggregate household welfare, affecting their per capita expenditures and investment in the nutrition of young children. The effects can be isolated to 3-8 months after harvest, allowing me to pin down the timing of the shocks' impacts. Using this result, I show that adult cognitive skills are particularly sensitive to weather variation in the second half of the year after the 2nd birthday. I corroborate these results using a secondary weather measure that is constructed to be highly predictive of rice harvests in the IFLS sample. At earlier ages, parents appear to engage in compensatory behaviours that mitigate the effects of shocks. In particular, parents offset negative weather shocks by prolonging breastfeeding and avoiding a nutritional cost to children (and vice versa).

These parental reactions interact with biological factors to produce critical periods in this setting. Numerous neural mechanisms could drive an impact of nutrition on cognitive and socioemotional function. Brain and cortical growth is fast in the first 2 years of life, and is associated with sensitivity to nutritional deficiency (Gilmore et al., 2018). Malnourished children have less well-developed dendrites (the neuronal branches that receive signals from other neurons) (Cordero et al., 1993), while breastfeeding can promote myelination, in which neurons grow a fatty sheath that promotes signal transmission (Deoni et al., 2018). Other neurological processes may be sensitive to nutrition, including neuron proliferation, axonal growth, synapse formation and pruning, and the degree of cortical gyration (folding) (Prado & Dewey, 2014; Gilmore et al., 2018). All of these processes occur over different age windows, so shocks at different ages are likely to have different impacts (Nelson & Gabard-Durnam, 2020). But we currently have little direct evidence on how nutrition impacts each of the mechanisms, and how this impact can vary by age.

My empirical setting presents a number of advantages when testing for critical periods. A key advantage of my strategy is that I use serially uncorrelated shocks as an exogenous source of variation. Each child is subject to multiple weather shocks over their life, and the correlation of these shocks across periods is close to 0. Each shock therefore provides independent variation that can be used to compare effects of similar shocks at different ages of childhood, thus testing for critical periods. Previous literature, by contrast, has typically examined single shocks or interventions (e.g. Heckman et al., 2013), or shocks that are correlated over time (e.g. Adhvaryu et al., 2019). Such settings present greater challenges when trying to test for the timing of effects.

A second advantage is the richness of the IFLS data. It is a large 5-wave panel survey that includes a total of over 30,000 individuals, and is representative of 83% of the Indonesian population. All adult respondents are asked for a detailed migration history. This, along with very low attrition rates on migrants, provides an unusual opportunity to reconstruct the location of individuals for all years

of their life since birth in order to examine the effects of weather shocks up until 15 years of age. The IFLS contains data to test critical periods on a range of cognitive and socioemotional measures, including Raven's matrices, depression, personality, wellbeing. The long-term panel also provides contemporaneous data on many individuals from their childhood years. I use these data to examine the mechanisms through which weather shocks may be operating, including via early-childhood investments and household consumption.

The final advantage is that Indonesia is an ideal setting for testing the timing of the effects of weather shocks. It is a geographically large country: east to west, it spans 5,100km. There are diverse patterns of weather across provinces and districts, meaning that weather varies significantly across individuals within a given year and across years (Frederick et al., 2011; Maccini & Yang, 2009). Indonesia is a lower-middle-income country with a large rural population and large agricultural sector (World Bank, 2018). This population experiences weather shocks as agricultural income shocks that affect children through changes in nutrition and other investment channels.[1] And children are vulnerable to the malnutrition that results, with 48% of children under 5 having a height-for-age that is 2 standard deviations below the norm. This high stunting rate is comparable to other lower-middle income countries, such as India, Bangladesh, and Kenya (Roser & Ritchie, 2019). The critical periods I examine are therefore highly welfare-relevant for this and other similar lower income populations, but are less likely to generalise to less vulnerable populations.

My paper contributes in a variety of ways to the literature on human capital accumulation. I test many of the claims of the Cunha-Heckman framework (Cunha & Heckman, 2007, 2008; Cunha et al., 2010) and see how they apply to the rural Indonesian context.[2]

First, the paper tests whether the timing of parental investment matters for adult outcomes. The neurodevelopmental and nutrition literatures that justify the focus on the first 1000 days are based on correlations between brain growth, stunting, and cognitive development, rather than the effects of causally identified shocks on cognitive development (Isaacs et al., 2008; Gilmore et al., 2018; Victora et al., 2010; Aiyar & Cummins, 2017; Perkins et al., 2017; Prentice et al., 2013). Another branch of studies use structural models to test when development is most malleable Cunha et al. (2010).

We therefore have little direct causal evidence on the question of timing. A notable exception is Barham et al. (2013), who show that children exposed to conditional cash transfers in Nicaragua at an earlier age (between in utero and age 2) have higher cognitive scores by age 10 than those who were exposed later (between age 2 and 5). My work contributes relative to this paper by providing a more disaggregated test of timing, by testing a wider age-range and broader set of outcomes, and by testing longer-term effects into adulthood.

Second, I contribute to the literature on dynamic complementarity that investigates whether early investments can increase the productivity of later investments (Almond & Mazumder, 2013; Adhvaryu et al., 2018; Gunnsteinsson et al., 2014; Rossin-Slater & Wüst, 2016; Malamud et al., 2016; Johnson & Jackson, 2019). The literature typically focuses on the interaction between shocks of different types, whereas my results show evidence of positive interactions for repeated shocks of the

[1] For example, Levine & Yang (2014) find that rainfall shocks are positively associated with deviations in district-level rice output from district-level mean, and rice is the main harvested crop across much of Indonesia (see Appendix Figure C.1.3). They interpret this as justification for interpreting higher rainfall as a positive contemporaneous shock to local economic conditions in Indonesia.

[2] I thereby complement a recent literature that adapts the Cunha-Heckman framework to a developing-country setting (Attanasio, Cattan, et al., 2020; Attanasio, Meghir, & Nix, 2020).

same type.

Third, I inform the debate on the differences between cognitive and socioemotional development (Cunha et al., 2010; Attanasio, Cattan, et al., 2020). My results provide some of the first causal estimates that suggest a strong critical period for weather shocks on cognitive development, but no evidence for such critical period for socioemotional development, adding to the evidence that human capital cannot be considered as a single "unidimensional object" (Attanasio, 2015).

Fourth, I provide evidence on cognitive development between the ages of 7 and 14, a poorly-understood period described as the "missing middle" (Almond et al., 2018). I show that the effects of weather shocks on adult cognitive skills are latent during middle childhood and only appear after age 15. Many studies of early-childhood interventions in the US find that treatment effects fade out by ages 8 or 9 (Currie & Almond, 2011; Garces et al., 2002; Deming, 2009). I provide evidence for an alternative explanation of apparent fade-out: longer-term effects are hard to observe during this middle childhood period and so can only be detected in adulthood. This is consistent with other papers showing very long term effects of early childhood interventions (Grantham-McGregor et al., 1991; Walker et al., 2005; Gertler et al., 2021; Walker et al., 2021).

My results also add evidence to the extensive literature on the effects of early-life weather shocks on economically relevant dimensions. Previous papers have shown that weather shocks can impact a wide range of outcomes (Dell et al., 2014), most relevantly on multiple measures of human capital, including adult height, health, socio-economic status, schooling, mortality, and birth weight (Maccini & Yang, 2009; Deschênes & Greenstone, 2011; Deschênes et al., 2009; Burgess et al., 2017; A. Barreca et al., 2016; A. I. Barreca, 2012). My results contribute by focusing specifically on how the timing of these shocks can affect the development of cognitive and socioemotional skills.

Identifying critical periods is important for policy. In this context, the impact of weather shocks on cognitive development is isolated to the year between the second and third birthdays. By contrast, there is no strong evidence for such a critical period for socioemotional development. Further research should examine whether similar critical periods exist for the effects of policies that act through similar channels, for example by affecting household income or nutritional investments. If critical periods do exist, this would imply that targeting these ages could greatly increase the cost-effectiveness of such programmes.

## 3.2 Data and measurement system

### 3.2.1 IFLS data

The main source of outcome data is the Indonesia Family Life Survey (IFLS) (Frankenberg & Karoly, 1995; Strauss et al., 1998, 2004, 2009, 2016). It is a longitudinal survey with 5 waves between 1993 and 2015, containing information on over 30,000 individuals. Individuals come from 13 of the 27 provinces in Indonesia, and are representative of 83% of the Indonesian population. Later waves include questions that assess cognitive ability and other socioemotional skills. Cognitive questions include Raven's progressive matrices (Raven, 2000) and number series tests, some of the most widely used measures in the literature. Subjective wellbeing is measured using cognitive evaluation questions such as the Cantrill Ladder. Positive and negative affect are measured using questions about the emotions the respondent felt in the last day. Personality questions are based on the "Big 5" personality traits (Rammstedt & John, 2007), and depression is measured using the CESD-10 scale (Radloff, 1977). These indexes are extensively used in psychology and economics and have

been validated in developing-country contexts (e.g., Das et al., 2009; de Quidt & Haushofer, 2019; Rietveld et al., 2015; Dal Bó et al., 2018; Groh et al., 2015). More details are found in Appendix Section C.1.1. In order to combine skill measures into a low-dimensional set of factors, I use a latent factor model (Heckman et al., 2013) to generate factor scores (see Section 3.2.3). These are then used as the dependent variables in my main specification.

I use a number of additional outcome variables to test for different mechanisms, including household expenditure, childhood nutritional and parental-time investment, child height-for-age, maternal nutrition, years of schooling, and whether children are being breastfed.

I construct a year-by-year history of each individual's location using the IFLS. First, I use information from retrospective data on each person's birth location, location at age 12, and subsequent migrations. I combine this with contemporaneous data on respondents' location at each of the five IFLS waves. The sample is restricted to individuals born in rural areas, since these households are likely to be sensitive to weather shocks that change their household welfare via nutrition and agricultural income. More detail on the measures used in the IFLS data can be found in Appendix Section C.1.1.

### 3.2.2 District, weather, and crop data

The primary weather variable is the Standardised Precipitation Evapotranspiration Index (SPEI), a state-of-the-art drought index Vicente-Serrano et al. (2010). The impact of rainfall on crop cycles depends on *potential evapotranspiration* (the ability of soil to retain water), which is determined by multiple environmental features, such as temperature, latitude, wind-speed, and the number of sunlight hours. The SPEI combines information on all these features. It has been shown to be a better predictor of crop yields than other climate indexes based solely on temperature or precipitation (Beguería et al., 2014), and has recently been used in the conflict and economics literature (e.g., see Harari & Ferrara, 2018).

I use a monthly gridded dataset for SPEI from the Global SPEI database (Beguería & Vicente Serrano, 2017) and aggregate it geographically to the level of the Indonesian district. The main weather variable $SPEI_{i(t+j)}$ (or *SPEI growing season*) is constructed by averaging the district-level SPEI over the growing season of the main crop of that district in that period. The index is normalised to have a mean of 0 and a standard deviation of 1, so that a value of 1 represents weather conditions during the growing season that are 1 standard deviation *better* for crops compared to the local long-run average (1961-2015).

I match this shock to individuals using the constructed year-by-year record of each individual's location. For example, the variable $SPEI_{i(t+2)}$ for an individual $i$ born in $t$ denotes the normalised average SPEI during the growing season of the main crop of the district where $i$ spent the year after her 2nd birthday. See Appendix Section C.1.3 for more details.

To corroborate the results with the SPEI, I construct a "predicted harvest" weather variable. This variable is a composite index of weather features that are most predictive of rice yields in the IFLS data (including, among them, the SPEI index). More details are in Appendix Section C.4.

### 3.2.3 Measurement of skills

I use the latent factor model seen in Heckman et al. (2013) to combine individual skill measures into indexes. This technique, common in the psychometric literature, helps to reduce measurement

error by explicitly accounting for classical error.[3] It also reduces the dimensionality of the set of outcomes by combining correlated proxy variables that measure a single psychological trait.

The process used to generate the indices is described in Appendix Section C.2, and the results are summarised in Table 3.1. There is a single cognitive factor that uses all the available cognitive measures. I detect 3 socioemotional latent factors, which I interpret as *affect/wellbeing*, *personality*[4], and *depression*. The factor scores for each individual are used as dependent variables in the main results. Factor scores should be interpreted positively: a higher depression factor score will mean *better* mental health or *less* depression. Scores are normalised with a mean of 0 and standard deviation of 1.

**Table 3.1:** *Latent factors and their corresponding measures*

| *Cognitive* | *Socioemotional* | | |
|---|---|---|---|
| | "Affect / Wellbeing" | "Personality" | "Depression" |
| Raven's Matrices | Affect: Frustrated | Big 5: Agreeableness - Forgiving | CES-D: I was bothered by things |
| Maths - Written | Affect: Sad | Big 5: Agreeableness - Considerate | CES-D: I had trouble concentrating |
| Word Recall | Affect: Enthusiastic | Big 5: Agreeableness - Rude | CES-D: I felt depressed |
| Maths - Oral | Affect: Lonely | Big 5: Conscientiousness - Thorough | CES-D: I felt hopeful about the future |
| Number Series | Affect: Content | Big 5: Conscientiousness - Lazy | CES-D: I felt fearful |
| | Affect: Worried | Big 5: Extroversion - Outgoing | CES-D: My sleep was restless |
| | Affect: Bored | Big 5: Openness - Original | CES-D: I was happy |
| | Affect: Happy | Big 5: Openness - Artistic | CES-D: I felt lonely |
| | Affect: Angry | | CES-D: I could not get going |
| | Affect: Tired | | |
| | Affect: Stressed | | |
| | Wellbeing: Life Satisfaction | | |
| | Wellbeing: Assessment of situation | | |
| | CES-D: I was happy | | |

*Notes*: Each column represents a single factor, containing a list of all the measures which load on that factor. The questions for each measure are in Appendix Table C.1.1. The full process used to generate the factors is described in Appendix Section C.2.

### 3.2.4 Samples and summary statistics

Table 3.2 summarises the samples used in the paper. To focus on households that are most likely to be impacted by weather variation, I restrict all samples to rural areas, and contemporaneous mechanism samples to farming households.

The main sample used in Sections 3.4.1 and 3.4.2 includes all individuals born after 1988 for whom socioemotional or cognitive outcome data is available.[5] Individuals were at most 5 years old at the time of the first wave of the IFLS data in 1993. The analysis of potential mechanisms using the earliest IFLS data is therefore still relevant for the early childhood period of my main sample. The primary weather variable is based on definitions of the growing season from 2000, and is also relevant for this sample.

[3]Psychological or skill measures tend to be "riddled with measurement error" (Cunha & Heckman, 2009; Heckman et al., 2013). Such error may be especially severe in low- and middle-income country contexts (Laajaj & Macours, 2019; Laajaj et al., 2019).

[4]The interpretation of the personality factor is unclear, since it combines questions that are supposed to identify distinct personality traits (e.g., extraversion and conscientiousness). This difficulty is common in studies in developing countries (Laajaj & Macours, 2019; Laajaj et al., 2019), where factor analyses do not separate the "Big 5" traits, and instead identify a single "personality" factor. I show results on the Big 5 personality traits when they are treated separately in Appendix Figure C.14.1.

[5]The main sample does not include individuals born after 2000 because the last wave of the IFLS survey was in 2015 and individuals were only asked questions on adult cognitive and socioemotional outcomes if they were at least 15 years old.

The mechanism samples include individuals outside the main sample to increase power. Analyses of early life investments include all children aged 0-5, and analyses of breastfeeding behaviour focuses on children ages 0-3.

Appendix Table C.3.1 presents summary statistics for individuals in the main sample. Factor scores and SPEI growing season variables are normalised with a mean of 0 and standard deviation of 1. The within-sample standard deviation is slightly lower than 1, since the target sample only covers a subset of the years (1961-2015) and districts used to standardise. 53% of the main sample are female. For the majority of the sample (52%), the highest level of education reached by a parent is elementary.

**Table 3.2:** *Guide to different IFLS samples used*

| *Sample* | | *N* | *Level of observation* | *Outcomes measured in* | *Sample restrictions* |
|---|---|---|---|---|---|
| **Main Sample** | Cognitive & Socioemotional | 5572 | Individual | IFLS3-5 | Born in rural area between 1988 and 2000 |
| **Mechanism Samples** | Investment Indexes | 4247 | Individual x Wave | IFLS3-5 | Living in rural farming households, ages 0-5 |
| | Household Expenditure | 11571 | Household x Wave | IFLS1-4 | Rural farming households |
| | Breastfeeding | 11656 | Individual x Wave | IFLS1-5 | Living in rural farming households, ages 0-3 |

*Notes*: This describes the samples used in the results, including how many observations are in each sample, the unit of observation of the sample, which waves the outcomes are measured in, and any additional sample restrictions on the sample. IFLS1 was carried out in 1993, IFLS2 in 1997, IFLS3 in 2000, IFLS4 in 2007 and IFLS5 in 2015. IFLS5 is not used for the (main) household expenditure results because there was no price adjustment data available for this wave at the district × year level, as in the other waves.

## 3.3 Empirical strategy

The main specification uses the following equation:

$$Y_{irt} = \sum_{j=-2}^{J=15} \beta_j SPEI_{i(t+j)} + \boldsymbol{X}'_{ir}\gamma + \delta_r + \mu_t + \varepsilon_{irt} \tag{3.1}$$

Here, $Y_{irt}$ denotes the adult outcome variable of interest for an individual $i$ born in region $r$ (Indonesian *kabupaten*) in year $t$. The weather variable, $SPEI_{i(t+j)}$, describes the average SPEI over the growing season of the main crop in the district where $i$ was located in the year after their $j$th birthday. $\boldsymbol{X}_{ir}$ is a vector of individual-level controls that includes dummies for the highest level of education achieved by a parent of $i$, religion, whether the the $i$ was born during the growing season of their birth district, and gender. $\delta_r$ and $\mu_t$ are birth-region and birth-year fixed effects.

The coefficients of interest in this specification are the $\beta_j$s. These measure the effects of weather variation on adult outcomes for every year from 2 years before birth all the way to age 15. I can disentangle effects at different ages because the serial correlation of weather variation across years is very low.[6]

The birth-year fixed effects control for any unobserved factors that are common to all individuals in the sample born in the same year, such as macroeconomic shocks. The birth-region fixed effects control for any cohort-invariant unobserved factors that are common across all individuals born in the same region. My identification strategy then relies on the claim that the variation in SPEI is exogenous, conditional on these fixed effects.

To test for the existence of critical periods in early childhood development, I carry out two types of hypothesis tests. First, I test whether there is an overall effect of weather variation (regardless of timing) on adult outcomes using an F-test on the joint null hypothesis $H_0 : \beta_{-2} = ... = \beta_{15} = 0$. Second, I test whether the effect is concentrated on specific periods by looking at individual coefficients, i.e., testing the null $H_j : \beta_j = 0$. To account for multiple hypothesis testing, I report sharpened q-values (Anderson, 2008) that control for the false discovery rate (FDR), i.e., the expected proportion of rejections that are falsely rejected (see Appendix Section C.10 for more details).

The coefficients $\beta_j$ incorporate all subsequent changes in parental investments, so should be interpreted as the *reduced-form* impact of weather variation on adult outcomes. For example, if parents react to early life shocks by investing more in a child's education, this will be incorporated into the estimates. If such parental responses are large, these estimates will differ from the kind of estimates used in the biomedical literature, which detect biological critical periods by testing whether the direct impact of shocks differs at different ages (and do not incorporate subsequent investment responses) (Knudsen, 2004; Knudsen et al., 2006).

Empirically, the direct impact and reduced-form impact of shocks may not differ substantially. I show that good weather only increases nutritional investment in a window between 3 and 8 months after harvest, and doesn't increase it between 1 and 2 years after harvest. This suggests that later parental responses may be empirically close to 0, although I cannot rule out investment responses along other unmeasured dimensions. In any case, the reduced-form impacts I measure are still welfare-relevant. They may even be welfare-relevant in a wider range of scenarios than direct or biological

[6]The empirical correlation in the main sample between $SPEI_{i(t+s)}$ and $SPEI_{i(t+s-1)}$ is only 0.076 (95% confidence interval: [0.071, 0.081]). Appendix Table C.3.2 and Figure C.3.3 show the full correlation matrix.

effects, since the welfare impacts of a shock depend on whether parents respond by compensating or reinforcing.[7]

Below, I outline four threats to my identification strategy and provide evidence that they do not invalidate the main results.

First, endogenous decisions to migrate in response to weather shocks could confound the causal effect of weather shocks on adult outcomes. For example, if intelligent individuals are more likely to migrate (within a given year) to areas with good weather, this selection could positively bias the causal estimate of good weather on cognitive ability. Migration threatens my estimation when people migrate into areas based on (i) current weather patterns (in the same year as the migration), or (ii) future weather patterns.[8] The migrant must be either rapid or have foresight, in addition to being selective. Although migration is relatively common in the sample, Column (2) of Appendix Table C.11.2 shows that this type of selective migration does not occur in the main sample. The probability that an individual migrates to a specific destination district in a given year is unaffected by the weather experienced in that district, alleviating the concerns about selective migration in my context.

Second, selective mortality could bias my estimates. For example, if negative weather shocks (e.g., drought) selectively kill children with poor cognitive development, then a positive relationship between weather and adult outcomes would be biased downwards. I test for such an effect in Appendix Section C.11. While there may be a positive impact of SPEI variation in utero on infant mortality, the magnitude is small and does not affect the probability of inclusion in my main sample.[9] This implies that selection bias is close to 0, although it cannot be ruled out entirely.

Third, two-way fixed effect specifications can yield biased estimates of the true treatment effect in the presence of heterogeneous treatment effects (Goodman-Bacon, 2018; de Chaisemartin & D'Haultfœuille, 2018; de Chaisemartin & d'Haultfoeuille, 2020). The concern arises when the fixed-effect estimator gives a higher weight to district-cohort groups with a higher average treatment effect on the treated (ATT), leading to an upward bias on the overall estimate (and vice versa). The weight assigned to each district-cohort group depends on their treatment status. In the current context, the concern is unlikely to undermine the results: the weather shocks I examine are plausibly as-good-as-random, so they are likely to be uncorrelated with the (heterogeneous) ATTs across districts and cohorts. In Appendix Section C.9 I discuss this issue further, and show evidence that the assigned weights for each group are unlikely to be correlated with the ATT in that group.

Fourth, imperfect recall may lead to significant measurement error on retrospective reports of location in each year of childhood. A key advantage of the IFLS data is that the majority of my sample were children in IFLS households during early waves of the survey. I therefore infer individuals' locations throughout their childhood by combining retrospective questions with contemporaneous data on the individuals' locations in survey years. This is likely to mitigate much of the measurement

[7] A simple model (Appendix Section C.6, and particularly Appendix Section C.6.1) shows that the direct effect determines welfare only in cases that are unlikely to hold in practice – namely when shocks are marginal and when parents have perfect knowledge of the human capital production function. The reduced form effect has welfare implications in more general settings.

[8] By contrast, migration responses based on *past* weather patterns should be thought of as one of the potential channels through which weather impacts adult outcomes. The results in Column (1) of Appendix Table C.11.2 suggest that migration is not a key channel through which weather variation affects adult outcomes, since the SPEI growing season in the pre-migration district does not affect the probability of migrating.

[9] Moreover, shocks during early life (after age 1) have no detectable effect on infant mortality, suggesting that the large effect on cognitive skills seen at age 2 is unlikely to be driven by selective mortality.

error stemming from retrospective migration reports.

## 3.4 Main results

### 3.4.1 Cognitive outcomes

Table 3.3 and Figure 3.1 show the results of running the main specification with adult cognitive factor z-score as the outcome variable $Y_{irt}$. In column (1), there is a strong and statistically significant positive coefficient (0.063) on SPEI growing season at age 2. The q-value of 0.002 implies that the probability that the observed coefficient is a false positive is below 1%, even after adjusting for multiple hypothesis testing. The p-value on the F-test in column (1) is 0.001, implying that we can also reject the joint null across all ages. None of the other coefficients has a q-value below 0.1, suggesting that SPEI growing season has a particularly strong effect at age 2. I do not have sufficient power to rule out positive coefficients of less than 0.05 in utero, at age 1 or age 3.[10] The coefficient at age 2 is larger for males (0.071, column (2)) than for females (0.047, column (3)), but this difference is not statistically significant (p=0.667).

Given that the weather shock is mild, and outcomes are measured at least 13 years after the second birthday, the magnitude of the coefficient at age 2 is relatively large. Increasing the SPEI growing season index by one standard deviation results in an average increase of 0.063 standard deviations in adult cognitive score.[11]

The results presented are strong evidence in favour of a critical period in early childhood for cognitive development. The large estimates on age 2 suggest that this age is particularly sensitive, while there is no positive evidence of impacts on cognitive development at other ages, although I cannot rule out small positive impacts.

There are reasons to expect cognitive development to be especially sensitive to shocks in utero and in the first year of life. Brain growth, which is usually associated with vulnerability to nutritional deficiency, is very fast during this period (Gilmore et al., 2018). I find no evidence of an effect of weather variation in these early stages. There is some evidence that this is due to compensatory breastfeeding behaviours by parents (Section 3.5.4), and I discuss other possible explanations in the Discussion (Section 3.6).

To check the robustness of the critical period for cognitive development at age 2, I present a number of additional results. First, I rerun the main results with a number of different specifications (Appendix Table C.12.1). The coefficient at age 2 remains strongly positive, with a q-value below 0.05, when I: (i) expand my sample to include older individuals (column 1), (ii) use siblings fixed effects and birth order fixed effects instead of birth district fixed effects (column 2), (iii) use a different version of the SPEI weather variable (column 3), and (iv) use alternative constructions for the cognitive index based on Anderson (2008) and the equally-weighted indices of Kling et al. (2007).The coefficients are also highly stable when including age fixed effects as controls in the regression (Appendix Table C.12.4). None of these specifications yields a q-value below 0.05 for SPEI growing season at any other age. When using a sample of people born in urban areas (columns 4 and 5), the coefficient at age 2 is close to 0, consistent with later evidence that weather variation is acting by changing nutritional investment and agricultural incomes.

[10] In keeping with a possible effect at age 3, examining effects over 2-year intervals suggests that ages 2-3 have strong impacts on cognitive scores (Appendix Table C.12.8).

[11] This effect is likely driven by an improvement in nutritional investment lasting approximately 6 months (see Section 3.5.1). The magnitude is therefore comparable to Maluccio et al. (2009), who estimate that a nutritional intervention targeting children for the first 36 months of their life led to 0.25SD increases in cognitive ability.

Second, Appendix Table C.12.2 shows that coefficients on early life shocks are not biased by the inclusion of later shocks in the specification. A priori, omitting interaction terms between shock variables at different ages (e.g., resulting from dynamic complementarity) could lead to omitted variable bias on earlier shocks. [12] However, any bias is empirically close to 0, since the coefficient at age 2 remains stable when removing later SPEI variables from the regression.

Third, I relax the assumption of linear effects that is built into my main specification (Equation 3.1). Using a non-parametric residual plot, I show that we can't reject that the effect of the shock at age 2 is linear (Appendix Figure C.13.1). Appendix Section C.13 shows additional results using alternative specifications, including binary shocks with different thresholds, different linear effects for positive and negative values, and squared term specification.

Fourth, Appendix Section C.7 presents the results of two types of placebo tests. I make use of (i) shocks from before conception, and (ii) synthetic shocks that are the result of randomly reshuffling each individual's shocks within their lifetime. I show that observing a coefficient with a q-value of less than 0.05 is highly unlikely when using these placebo shocks. This verifies that my main result is picking up true variation, not just random noise.

### 3.4.2 Socioemotional outcomes

Table 3.4 and Figure 3.2 show the results of estimating Equation 3.1 with each of the the 3 socioemotional indexes as outcome variables. In contrast to the results on adult cognitive score, there is no clear evidence of a single critical period for socioemotional development. Across all three indexes, there are no coefficients with a q-value of less than 0.05, and the majority of coefficients are not significantly different from 0 even before adjusting for multiple hypothesis testing. The p-value on the F-test for affect/wellbeing is 0.080, denoting weak evidence that the shocks together are causing variation in the adult factor score. This may be driven by a positive effect at age 10, or a negative effect at age 14, but neither coefficient is significant after adjusting for multiple hypotheses. The F-test on the personality index is insignificant; we cannot reject the null that the index is unaffected by *SPEI growing season* across all ages.[13] For depression, the F-test is highly significant. This is likely driven by a positive effect in utero (with a coefficient of 0.043) and a negative effect at age 11 (with a coefficient at -0.050). Both of these coefficients have a q-value of 0.086; they are significant after adjusting for multiple hypothesis testing, but only at the 10% level.

In Appendix Section C.8, I examine whether insufficient power could explain these null results. I use simulated data that uses before-conception weather to proxy the null distribution of estimated treatment effects, and show that cognitive, depression, and personality indexes all yield a power of 80% when true treatment effects are between 0.06 and 0.08SD. By contrast, the affect/wellbeing index is significantly more noisy, only yielding 80% power when the treatment effect is 0.093SD. I cannot confidently rule out smaller effect sizes, so the results should not be interpreted as strong

[12] The omitted variable bias would be 0 when (i) shocks across ages are independent, and (ii) the mean value of all the shocks is 0. To see this, denote $D$ as the value of the shock and consider that $Cov(D_{i(t+j)}, D_{i(t+j)} \times D_{i(t+k)}) = E[D^2_{i(t+j)}]E[D_{i(t+k)}] - E[D_{i(t+j)}]^2 E[D_{i(t+k)}]$. A sufficient condition for this to be 0 is (i) $E[D_{i(t+j)}] = 0$ for all $j$, and (ii) $D_{i(t+j)} \perp D_{i(t+k)}$. The correlation between shocks at different ages is weak but not 0 (Appendix Table C.3.2), and the mean value of all the shocks is slightly positive in most cases, so missing interaction effects may cause some bias on my coefficients.

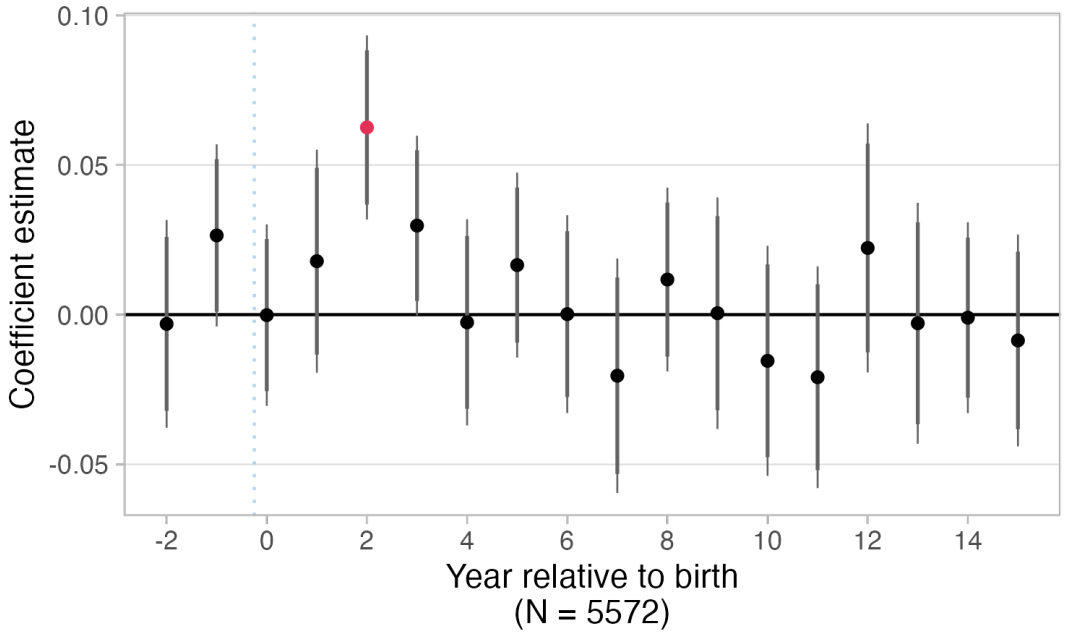
[13] Treating the Big 5 personality traits separately yields similar results, with no coefficient with a q-value below 0.05, and no significant F-tests.

**Table 3.3:** *Effect of SPEI growing season on adult cognitive factor score*

| | Dep Var: Adult Cognitive Z-score | | | | | |
|---|---|---|---|---|---|---|
| | All (1) | q value (1) | Male (2) | q value (2) | Female (3) | q value (3) |
| SPEI growing season, age = -2 | −0.003 (0.018) | [1] | −0.019 (0.021) | [1] | 0.017 (0.026) | [1] |
| SPEI growing season, age = -1 | 0.026 (0.015)* | [0.884] | 0.037 (0.021)* | [0.682] | 0.021 (0.022) | [1] |
| SPEI growing season, age = 0 | −0.000 (0.015) | [1] | 0.010 (0.024) | [1] | −0.008 (0.021) | [1] |
| SPEI growing season, age = 1 | 0.018 (0.019) | [1] | 0.014 (0.028) | [1] | 0.030 (0.022) | [0.9] |
| SPEI growing season, age = 2 | 0.063 (0.016)*** | [0.002] | 0.071 (0.027)*** | [0.182] | 0.047 (0.020)** | [0.705] |
| SPEI growing season, age = 3 | 0.030 (0.015)* | [0.796] | 0.048 (0.023)** | [0.512] | 0.013 (0.025) | [1] |
| SPEI growing season, age = 4 | −0.003 (0.017) | [1] | −0.009 (0.025) | [1] | 0.008 (0.029) | [1] |
| SPEI growing season, age = 5 | 0.017 (0.016) | [1] | 0.001 (0.024) | [1] | 0.031 (0.023) | [0.9] |
| SPEI growing season, age = 6 | 0.000 (0.017) | [1] | 0.006 (0.024) | [1] | 0.002 (0.022) | [1] |
| SPEI growing season, age = 7 | −0.020 (0.020) | [1] | −0.007 (0.030) | [1] | −0.032 (0.027) | [0.921] |
| SPEI growing season, age = 8 | 0.012 (0.016) | [1] | 0.021 (0.027) | [1] | 0.013 (0.025) | [1] |
| SPEI growing season, age = 9 | 0.000 (0.020) | [1] | −0.007 (0.027) | [1] | 0.006 (0.026) | [1] |
| SPEI growing season, age = 10 | −0.015 (0.019) | [1] | 0.015 (0.027) | [1] | −0.035 (0.022) | [0.9] |
| SPEI growing season, age = 11 | −0.021 (0.019) | [1] | 0.005 (0.025) | [1] | −0.038 (0.025) | [0.9] |
| SPEI growing season, age = 12 | 0.022 (0.021) | [1] | 0.002 (0.029) | [1] | 0.037 (0.029) | [0.9] |
| SPEI growing season, age = 13 | −0.003 (0.020) | [1] | −0.008 (0.026) | [1] | −0.004 (0.032) | [1] |
| SPEI growing season, age = 14 | −0.001 (0.016) | [1] | 0.019 (0.027) | [1] | −0.015 (0.025) | [1] |
| SPEI growing season, age = 15 | −0.009 (0.018) | [1] | 0.006 (0.021) | [1] | −0.033 (0.025) | [0.9] |
| Birth year fixed effects | Yes | | Yes | | Yes | |
| Birth district fixed effects | Yes | | Yes | | Yes | |
| F statistic | 2.490 | | 1.704 | | 2.003 | |
| p-value for F-test | 0.001 | | 0.044 | | 0.012 | |
| $R^2$ | 0.212 | | 0.229 | | 0.256 | |
| N | 5572 | | 2601 | | 2971 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are in parentheses and are clustered by birth district. The table uses Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Column (1) includes all observations, Column (2) includes only males, and Column (3) includes only females. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Figure 3.1:** *Effect of SPEI shock at different stages of childhood on cognitive factor scores*




*Notes*: Each point corresponds to a coefficient from column (1) of Table 3.3, so represents the effect of SPEI growing season at each age on cognitive factor z-score in adulthood. The model controls for birth year fixed effects, birth district fixed effects, and individual-level controls (dummies for highest level of parental education, sex, and religion). Error bars are 90% and 95% confidence intervals based on standard errors clustered at the district level (not adjusted for multiple testing). Points are highlighted in red when the q-value is less than 0.05.

evidence against other sensitive periods with smaller treatment magnitudes.[14]

The positive effect on depression in utero is in keeping with recent evidence on the effect of economic shocks before birth on adult mental health (Adhvaryu et al., 2019). The possible negative effect on depression at age 11 and on affect/wellbeing at age 14 is more surprising. I offer two potential explanations. First, an additional set of channels may be operating for girls and not boys, since the negative effects are more pronounced among girls (Appendix Table C.14.1). For example, better nutrition in the early teenage years has been associated with *earlier* fertility or menarche for females (Belachew et al., 2011; Hochberg & Belsky, 2013; Barham et al., 2018), which could have a negative effect on wellbeing.[15] Second, the effects on socioemotional scores might be driven by noise on lower power tests. The coefficients are only weakly significant, and they are not robust to alternative specifications. When I include siblings fixed effects (Appendix Table C.14.2), none of the coefficients has a q-value below 0.1, and the previously significant coefficients for depression are close to 0 and are not significant (even before multiple hypothesis testing). The coefficients in the main specification are unlikely to be picking up true critical periods in socioemotional development.

[14]To further evaluate whether the null effects are due to measurement error, Appendix Table C.2.7 analyzes two indicators of statistical noise in the outcome variables. Serial correlation is high for cognitive skills (0.52) but much lower for depression (0.15). The 7-year gap between measurements makes this lower correlation unsurprising, but it may indicate higher measurement error. The average inter-item correlation is in the standard recommended range of 0.15-0.5 (Clark & Watson, 1995) for all indexes apart from personality, whose value is only 0.13, suggesting a more noisy dependent variable.

[15]Appendix Table C.14.3 shows that fertility is unlikely to be a key mechanism for the negative impacts on wellbeing, because shocks in adolescence do not affect the probability of giving birth. I am unable to test whether the age of menarche is an explanatory factor.

Overall, I find no strong evidence for the existence of a single critical period in the socioemotional development in rural Indonesia. Weather shocks in early life appear to have no significant effect on adult personality outcomes. Similar weather shocks may have some influence on adult affect/wellbeing and depression outcomes, but it is difficult to isolate the timing of this influence. There is no strong evidence for a single critical period for these outcomes, and instead there may be multiple ages in which affect/wellbeing and depression outcomes are weakly malleable, or malleable only for girls.

**Table 3.4:** *Effect of SPEI growing season on adult socioemotional factor scores*

| | Affect/Wellbeing (1) | q value (1) | Personality (2) | q value (2) | Depression (3) | q value (3) |
|---|---|---|---|---|---|---|
| SPEI growing season, age = -2 | 0.007 (0.022) | [1] | −0.020 (0.020) | [1] | 0.014 (0.019) | [1] |
| SPEI growing season, age = -1 | 0.036 (0.023) | [1] | 0.002 (0.021) | [1] | 0.043 (0.015)*** | [0.084] |
| SPEI growing season, age = 0 | 0.017 (0.021) | [1] | −0.006 (0.025) | [1] | −0.012 (0.017) | [1] |
| SPEI growing season, age = 1 | −0.013 (0.027) | [1] | 0.014 (0.024) | [1] | −0.007 (0.017) | [1] |
| SPEI growing season, age = 2 | 0.020 (0.027) | [1] | −0.023 (0.023) | [1] | −0.018 (0.017) | [1] |
| SPEI growing season, age = 3 | −0.016 (0.022) | [1] | −0.014 (0.024) | [1] | 0.008 (0.016) | [1] |
| SPEI growing season, age = 4 | 0.025 (0.026) | [1] | 0.026 (0.025) | [1] | 0.012 (0.021) | [1] |
| SPEI growing season, age = 5 | 0.016 (0.023) | [1] | −0.066 (0.023)*** | [0.098] | −0.021 (0.019) | [1] |
| SPEI growing season, age = 6 | 0.009 (0.020) | [1] | −0.028 (0.024) | [1] | −0.001 (0.019) | [1] |
| SPEI growing season, age = 7 | 0.030 (0.023) | [1] | −0.040 (0.022)* | [1] | 0.029 (0.019) | [0.923] |
| SPEI growing season, age = 8 | 0.013 (0.023) | [1] | −0.016 (0.023) | [1] | −0.008 (0.018) | [1] |
| SPEI growing season, age = 9 | −0.013 (0.026) | [1] | −0.026 (0.023) | [1] | −0.005 (0.020) | [1] |
| SPEI growing season, age = 10 | 0.045 (0.022)** | [0.563] | −0.021 (0.026) | [1] | −0.009 (0.016) | [1] |
| SPEI growing season, age = 11 | −0.025 (0.020) | [1] | −0.028 (0.026) | [1] | −0.050 (0.019)*** | [0.086] |
| SPEI growing season, age = 12 | 0.035 (0.027) | [1] | −0.007 (0.026) | [1] | 0.015 (0.023) | [1] |
| SPEI growing season, age = 13 | 0.018 (0.021) | [1] | −0.031 (0.025) | [1] | −0.005 (0.020) | [1] |
| SPEI growing season, age = 14 | −0.060 (0.024)** | [0.343] | −0.010 (0.020) | [1] | 0.021 (0.024) | [1] |
| SPEI growing season, age = 15 | 0.025 (0.026) | [1] | −0.041 (0.026) | [1] | 0.032 (0.018)* | [0.744] |
| Birth year FEs | Yes | | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | | Yes | |
| F statistic | 1.542 | | 0.998 | | 2.750 | |
| p-value for F-test | 0.080 | | 0.464 | | 0.000 | |
| $R^2$ | 0.069 | | 0.071 | | 0.089 | |
| N | 4956 | | 4956 | | 5565 | |

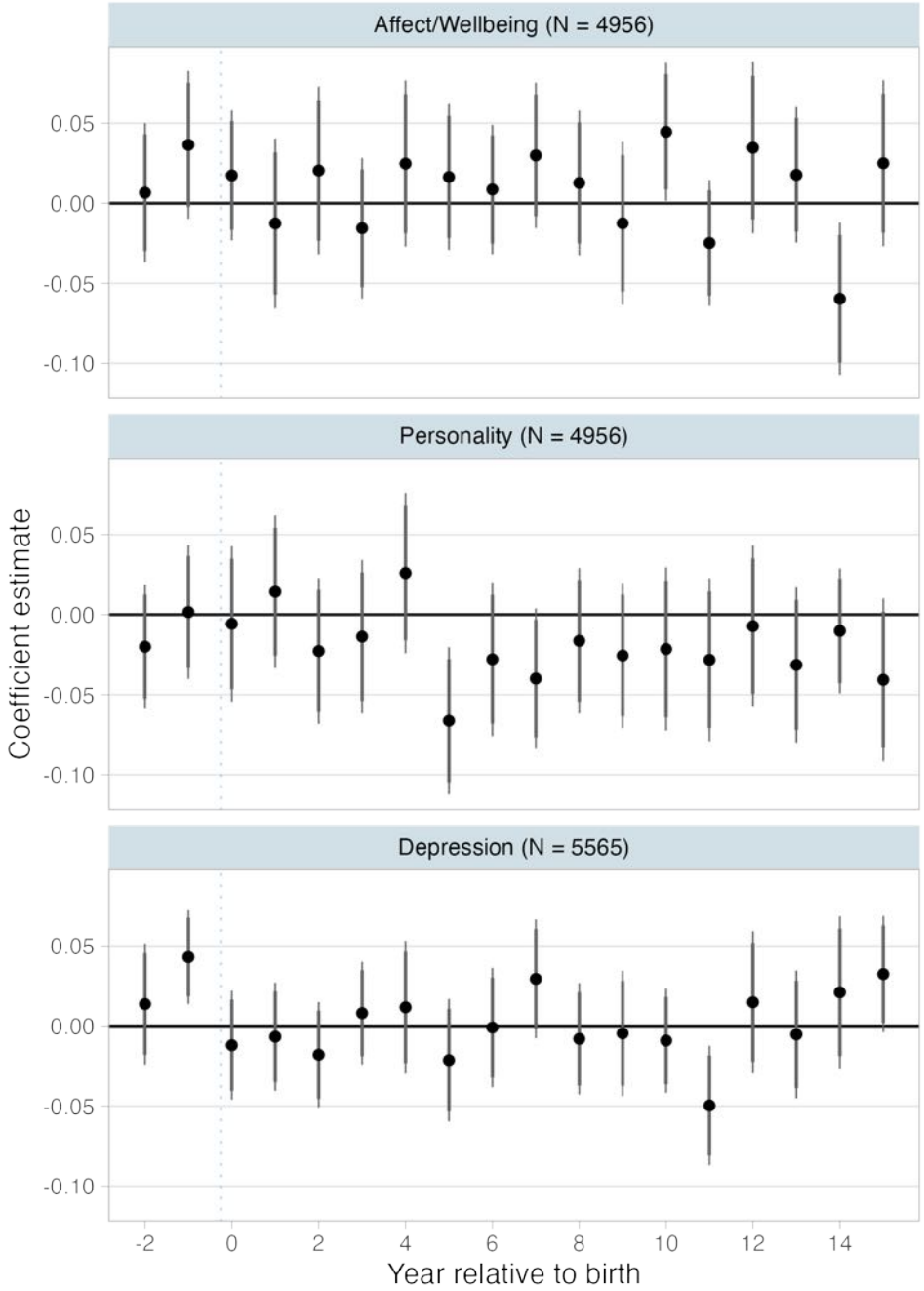
*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table uses Equation 3.1 with the three socioemotional factors measured in adulthood (above aged 15) as the dependent variable. All socioemotional scores are internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

### 3.4.3 Cognitive: ages 7-14

Section 3.4.1 showed significant effects of early life weather shocks on *adult* cognitive ability. To understand whether it is possible to detect these long-term effects earlier, and inform the question of the "missing middle" (Almond et al., 2018), this section examines middle-childhood cognitive ability (measured between 7 and 14 years) as an outcome.

Table 3.5 shows that the impact of early life weather shocks on cognitive score cannot be detected in middle childhood. I use the 75% of the main sample for whom cognitive outcomes are measured

**Figure 3.2:** *Effect of SPEI growing season on adult socioemotional outcomes*




*Notes*: Each point corresponds to a coefficient from Table 3.4, so represents the effect of an SPEI shock in the specific year relative to birth on the respective socioemotional Z-score in adulthood. All coefficients control for birth year fixed effects, birth district fixed effects, and individual-level controls (dummies for highest level of parental education, sex, and religion). Error bars are 90% and 95% confidence intervals based on standard errors clustered at the district level (not adjusted for multiple testing). Points are highlighted in red when the q-value is less than 0.05.

at between the ages of 7 and 14.[16] Columns (1) and (2) show no evidence of any impact of SPEI growing season on cognitive scores at ages 7-14. The F-tests on both specifications are insignificant, and the strong effect at age 2 seen in adulthood is not detectable at this earlier age. Columns (3) and (4) demonstrate that the effect of weather at age 2 on adult outcomes is not *mediated* by the cognitive score at age 7-14. With adult cognitive score as the outcome, adding a control for middle-childhood cognitive score has almost no impact on the coefficients for SPEI growing season. This is true despite a very strong correlation between cognitive score in middle-childhood and adulthood (the last coefficient of column (4)). Together these results provide a warning against using the effect on cognitive scores in middle childhood as a sufficient statistic for long-term impacts on cognitive development.

### 3.4.4 Interaction effects (dynamic complementarity)

To measure whether there may be dynamic complementarity in investments over the course of childhood, I examine interaction effects of SPEI growing season at different ages. First, I focus on the earliest periods, and test for the interaction effect between the shocks at age 2 and other years below 4. Table 3.6, columns (1)-(3), show the results of these estimations. For females, there is no evidence of dynamic complementarity. For males, there is weak evidence of dynamic complementarity: a positive interaction between weather variation at ages 1 and 2, significant at the 5% level. The coefficient implies that experiencing weather that is 1 SD better at age 2 increases average adult cognitive score by 0.08 SD, while having 1 SD better weather at *both* ages 1 and 2 increases average adult cognitive score by 0.14 SD. In line with this, Appendix Table C.12.5 shows that shocks in early life have weakly cumulative effects.

Next, I analyse interaction effects between early childhood and adolescence (columns (4) to (6)). These may be of particular interest if early shocks to human capital change the returns to educational investment later in life. To increase power, I average the SPEI growing season across multiple ages. While there is no interaction effect between early-life SPEI (averaged from ages 0 to 3) and late-childhood SPEI (averaged from ages 10 to 12), there is a positive interaction between early-life SPEI and adolescent SPEI (averaged from 13-15). The complementarity is unlikely to be driven by biological changes during puberty, since the pattern is similar across genders (Appendix Table C.12.6), and puberty occurs earlier in girls. Alternatively, changes in educational investment during middle school could explain the interaction effect. Although there are no detectable effects on the number of years children stay at school (Appendix Table C.12.7), I cannot rule out changes along other margins of educational investment during middle school.

[16] Individuals who enter the sample after age 14 are not included. This pattern will occur if a new individual enters a household that includes an original IFLS member. It is possible that such individuals are systematically different to the main sample, but selection into IFLS households is unlikely to be driven by early life circumstance. The strong positive effect at age 2 in column (3) is similar to the main results, mitigating concerns of a selective sample.

**Table 3.5:** *Cognitive "missing middle"*

| | Dep var: Cog at 7-14 | | Dep var: Cog at 15+ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| SPEI growing season, age = -1 | 0.020 | 0.010 | 0.037** | 0.032* |
| | (0.019) | (0.026) | (0.018) | (0.018) |
| SPEI growing season, age = 0 | 0.003 | 0.040 | 0.004 | 0.003 |
| | (0.021) | (0.028) | (0.016) | (0.015) |
| SPEI growing season, age = 1 | 0.005 | 0.023 | 0.020 | 0.019 |
| | (0.021) | (0.028) | (0.019) | (0.018) |
| SPEI growing season, age = 2 | 0.028 | 0.035 | 0.074*** | 0.067*** |
| | (0.022) | (0.031) | (0.019) | (0.018) |
| SPEI growing season, age = 3 | 0.027 | 0.044* | 0.030* | 0.023 |
| | (0.021) | (0.027) | (0.015) | (0.014) |
| SPEI growing season, age = 4 | −0.022 | −0.004 | −0.015 | −0.009 |
| | (0.018) | (0.028) | (0.019) | (0.018) |
| SPEI growing season, age = 5 | −0.002 | −0.012 | 0.008 | 0.008 |
| | (0.017) | (0.026) | (0.016) | (0.015) |
| SPEI growing season, age = 6 | 0.013 | 0.043* | 0.002 | −0.001 |
| | (0.018) | (0.025) | (0.017) | (0.016) |
| Cog score (aged 7-14) | | | | 0.258*** |
| | | | | (0.017) |
| Birth year FEs | Yes | Yes | Yes | Yes |
| Birth district FEs | Yes | No | Yes | Yes |
| Siblings FEs | No | Yes | No | No |
| Birth order FEs | No | Yes | No | No |
| F statistic | 0.773 | 1.010 | 3.709 | 3.342 |
| p-value for F-test | 0.627 | 0.426 | 0.001 | 0.001 |
| $R^2$ | 0.170 | 0.851 | 0.225 | 0.270 |
| N | 4370 | 4370 | 4370 | 4370 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with only SPEI growing seasons from age -1 to age 6 as explanatory variables. The first two columns use the cognitive score measured between ages 7 and 14 as the dependent variable. Columns (3) and (4) use cognitive score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0. Appendix Table C.12.3 shows the same results broken down by gender.

**Table 3.6:** *Cognitive interaction (dynamic complementarity)*

| | Dep Var: Adult Cog Z-score | | | | | |
|---|---|---|---|---|---|---|
| | Both (1) | Male (2) | Female (3) | Both (4) | Both (5) | Both (6) |
| SPEI growing season, age 0 | −0.002 | 0.014 | −0.016 | | | |
| | (0.014) | (0.021) | (0.019) | | | |
| SPEI growing season, age 1 | 0.015 | 0.009 | 0.023 | | | |
| | (0.017) | (0.026) | (0.019) | | | |
| SPEI growing season, age 2 | 0.068*** | 0.075*** | 0.051** | | | |
| | (0.015) | (0.025) | (0.021) | | | |
| SPEI growing season, age 3 | 0.031** | 0.046** | 0.013 | | | |
| | (0.014) | (0.021) | (0.023) | | | |
| SPEI growing season, (age 0 x age 2) | 0.000 | −0.016 | 0.006 | | | |
| | (0.016) | (0.022) | (0.019) | | | |
| SPEI growing season, (age 1 x age 2) | 0.027* | 0.057** | 0.003 | | | |
| | (0.015) | (0.025) | (0.018) | | | |
| SPEI growing season, (age 3 x age 2) | −0.016 | 0.003 | −0.032 | | | |
| | (0.014) | (0.024) | (0.022) | | | |
| SPEI growing season, ages 0-3 | | | | 0.127*** | 0.080** | 0.091** |
| | | | | (0.035) | (0.037) | (0.041) |
| SPEI growing season, ages 10-12 | | | | −0.021 | | |
| | | | | (0.033) | | |
| SPEI growing season, (ages 0-3 x ages 10-12) | | | | −0.051 | | |
| | | | | (0.053) | | |
| SPEI growing season, ages 13-15 | | | | | 0.013 | |
| | | | | | (0.033) | |
| SPEI growing season, (ages 0-3 x ages 13-15) | | | | | 0.081** | |
| | | | | | (0.039) | |
| SPEI growing season, ages 10-15 | | | | | | −0.010 |
| | | | | | | (0.059) |
| SPEI growing season, (ages 0-3 x ages 10-15) | | | | | | 0.056 |
| | | | | | | (0.054) |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth district fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.211 | 0.229 | 0.253 | 0.209 | 0.209 | 0.209 |
| N | 5572 | 2601 | 2971 | 5572 | 5572 | 5572 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running versions of Equation 3.1. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. All columns use cognitive factor score measured in adulthood (above aged 15) as the dependent variable. In columns (1) to (3) the explanatory variables are SPEI growing season from age -1 to age 3 *and* the interactions between SPEI growing season at age 2 with all other weather variables. Column (2) is only males, and column (3) is only females. *SPEI growing season, ages 0-3* denotes the mean value of SPEI growing season for an individual across ages 0, 1, 2, and 3. The age ranges 13-15 and 10-15 are defined analogously. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

## 3.5 Mechanisms

In this section I examine a number of potential mechanisms that could explain the effects on cognitive development. The results suggest that weather has an impact on nutritional investment between 3 and 8 months after harvest. I use this to show that effects on cognitive development are particularly strong in the *second* half of the year between the second and third birthday. I then show evidence that breastfeeding behaviour may be insulating children from shocks at earlier ages.

Throughout this section, I corroborate the results on *SPEI growing season* with an additional weather variable, the *predicted harvest index*. To construct it, I first regress rice yields from the IFLS data on a number of weather measures from the climatology literature. The index is the predicted value from this regression, normalised to have a mean of 0 and standard deviation of 1. It is thus a summary index for the weather features that predict high rice yields in the IFLS sample. See Appendix Section C.4 for more details.

### 3.5.1 Investment indexes

This section provides evidence that weather variation has a positive effect on investment for under-5s in IFLS households, suggesting that weather shocks may be affecting cognitive development through nutritional channels.

I focus on two important channels through which increases in agricultural yields could affect early childhood development. I construct a *nutritional investment index* (composed of measures of food frequency and dietary diversity for young children) and a *parental time index* (composed of questions on whether and how much the father and mother work).[17]

First, I examine how weather from the last 2 years affects contemporaneous investments for children under 5 years old. The empirical specification is:

$$I_{iry} = \alpha + \beta_0 Weather_{ry} + \beta_1 Weather_{r,y-1} + \pi_r + \mu_y + \varepsilon_{iry} \tag{3.2}$$

where $I_{iry}$ is the investment index for an under-5 individual $i$ in region $r$ measured in year $y$. $Weather_{ry}$ is the weather that corresponds to the harvest in year $y$ (either SPEI growing season or predicted harvest index), and $\pi_r$ and $\mu_y$ are district and year fixed effects.

Second, I use data on *when* each household was measured relative to harvest to understand the timing of impacts in more detail. Motivated by results disaggregated by month (Appendix Figure C.15.1), I define two 6-month periods during the year: *around harvest* (3 months before harvest to 2 months after), and *after harvest* (3 to 8 months after). Using the interaction of weather variation with an indicator of which period investment was measured in, the results indicate when the shocks "bite" and affect investment. Appendix Section C.5 gives more details on the specification.

Table 3.7 shows the results. Both weather measures positively impact nutritional investment. SPEI growing season has a positive and statistically significant effect of 0.065 standard deviations in the year of harvest, which can be isolated to the period 3-8 months after harvest. The coefficient on predicted harvest across the whole year is not significantly different from 0, but this disguises a positive effect that is also isolated to between 3 and 8 months after harvest. Why might the effects on nutrition be focused on this period? Plausibly, the smaller yields of households that experience poor weather provide food stocks that run out after only 2-3 months. They may thus experience

[17] The survey questions used are described in Appendix Section C.1.1. The construction of the indexes is described in Appendix Section C.2.7.

earlier, longer, and more severe lean seasons than households with normal or favourable weather conditions.

For the parental time index, none of the coefficients is significant at the 10% level. The coefficients are not precisely estimated, but this does suggest that weather improvements do not lead to reductions in parental time investment (for example, due to the labour requirements of good harvests) that might offset the positive effects on nutritional investment.

### 3.5.2 Household expenditure

To complement the results showing that weather variation affects nutritional investment, I here examine whether this is driven by changes in aggregate household welfare that would result from changes in harvest yields for rural households.

Table 3.8 shows the effect of weather shocks in the last 2 years on contemporaneous log household per capita expenditure. A 1 SD improvement in the SPEI growing season increases per capita expenditures by approximately 6% in the period between 3 and 8 months after harvest. Similarly, the predicted harvest index is associated with a 10.4% increase in total per capita expenditure in this period (corresponding to the coefficient on the log outcome of 0.099). Good weather and accompanying good harvests thus appear to improve aggregate household expenditure, which then translates to improved nutritional investment, specifically in the window 3 to 8 months after harvest.[18] [19] These results are robust to different constructions of the household expenditure variable (Appendix Table C.15.1).

In Appendix Section C.17, I examine whether the changes in nutritional investment and household expenditure translate into improved height-for-age (as measured between ages 2 and 6, giving enough time for nutritional status to show up as changes in height). The predicted harvest index at age 1 has a positive effect on female height-for-age in early childhood, but there are no detectable effects of early-life SPEI growing season on height-for-age. One interpretation is that height-for-age is mostly determined by cumulative nutritional investments over the course of a longer window in early childhood (in contrast to the apparently strong critical period for cognitive development at age 2). This, along with noisy measures of height-for-age at ages 2-6 may make it harder to detect the effects of temporary changes in weather earlier in life.

### 3.5.3 Effects on cognitive development: 6 month intervals

The effects of weather on expenditure and nutritional investment are isolated to a window between 3 and 8 months after harvest, providing a clear demarcation of the timing of the effect of the variation. By assuming that shocks only "bite" in the window 3-8 months after harvest, I use this result to understand the timing of critical periods in cognitive development using more granular 6-month intervals.[20]

I assign every 6-month interval in an individual's lifetime *either* to the 6-month period when the

[18] There may be general equilibrium effects of weather shocks in local areas, for example, effects on the price of agricultural goods. This channel should be thought of as an extension of the main mechanism, since it will likely affect cognitive development primarily through changes in nutritional investment and economic welfare.

[19] Wetter weather may increase the risks of floods or may lead to a worse early-life disease environment. This is unlikely to be driving results, since (i) it would not explain a positive relationship between SPEI at age 2 and cognitive development, (ii) the results persist when using the predicted harvest index which includes multiple weather measures not based on rainfall, and (iii) large positive SPEI shocks (associated with heavy rainfall and flooding) do not have negative effects on cognitive development (Appendix Section C.13).

[20] The variation in harvest time across different regions implies that impacts will not be confounded by other cyclical shocks to nutrition, such as Ramadan (see Appendix Figure C.16.1).

**Table 3.7:** *Impact of weather shocks on indexes of investment for u5s*

| | | Dep Var: Nutrition Index | | | | Dep Var: Parental Time Index | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Outcome measured in: | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| SPEI growing season | Year 0 | 0.065*** | | | | −0.008 | | | |
| | | (0.024) | | | | (0.037) | | | |
| SPEI growing season | Year 1 | 0.009 | | | | −0.002 | | | |
| | | (0.021) | | | | (0.022) | | | |
| SPEI growing season | Months -3 to 2 after harvest | | 0.024 | | | | −0.006 | | |
| | | | (0.027) | | | | (0.039) | | |
| SPEI growing season | Months 3 to 8 after harvest | | 0.092*** | | | | 0.035 | | |
| | | | (0.034) | | | | (0.040) | | |
| SPEI growing season | Months 9 to 14 after harvest | | 0.016 | | | | −0.011 | | |
| | | | (0.025) | | | | (0.034) | | |
| SPEI growing season | Months 15 to 20 after harvest | | −0.012 | | | | −0.000 | | |
| | | | (0.026) | | | | (0.022) | | |
| Predicted harvest index | Year 0 | | | 0.028 | | | | −0.036 | |
| | | | | (0.028) | | | | (0.028) | |
| Predicted harvest index | Year 1 | | | 0.005 | | | | 0.003 | |
| | | | | (0.022) | | | | (0.022) | |
| Predicted harvest index | Months -3 to 2 after harvest | | | | −0.013 | | | | 0.013 |
| | | | | | (0.042) | | | | (0.037) |
| Predicted harvest index | Months 3 to 8 after harvest | | | | 0.058* | | | | −0.032 |
| | | | | | (0.031) | | | | (0.031) |
| Predicted harvest index | Months 9 to 14 after harvest | | | | 0.029 | | | | −0.015 |
| | | | | | (0.027) | | | | (0.028) |
| Predicted harvest index | Months 15 to 20 after harvest | | | | −0.013 | | | | 0.004 |
| | | | | | (0.028) | | | | (0.028) |
| Year FEs | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District FEs | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Age (month) FEs | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | | 0.139 | 0.139 | 0.145 | 0.129 | 0.131 | 0.135 | 0.148 | 0.124 |
| N | | 4247 | 4247 | 4247 | 4246 | 4247 | 4247 | 4247 | 4246 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses. When the dependent variable is *SPEI growing season*, the standard errors are clustered by district. When the dependent variable is *predicted harvest index*, the standard errors are bootstrapped to account for the construction of the regressor, as described in Appendix Section C.5.3. The unit of observation is (individual × wave). The dependent variable in columns (1)-(4) is nutritional investment index for children under 5 years old (composed of food frequency and dietary diversity measures). The dependent variable in columns (5)-(8) is parental time investment index for children under 5 years old (composed of measures of time spent working for both parents, and whether their primary activity is working). The survey questions used are described in Appendix Section C.1.1. The construction of the indexes is described fully in Appendix Section C.2.7. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. For the SPEI growing season, the harvest date is defined as the average harvest date of the main crop grown in that district (as according to aggregate data, see description in Section C.1.2). For the Predicted Harvest Index, the harvest date is defined as the average harvest date of rice in that district based on IFLS harvest data (see Section C.4.2). "Year 0" is the period between 3 months before and 8 months after harvest (inclusive). Analogously, "Year 1" is the period 9 months and 20 months after harvest (inclusive). For the 6-monthly effects, the coefficients reported come from a version of Equation C.4 (but with investment index as the dependent variable) and are $\beta_0$ for -3 to 2 months; $\beta_0 + \gamma_0$ for 3-8 months; $\beta_1$ for 9-14 months; $\beta_1 + \gamma_1$ for 15-20 months.

**Table 3.8:** *Effect of weather variation on log household per capita expenditure*

| | | SPEI growing season | | | | Predicted Harvest Index | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Outcome measured in: | Total (1) | Food (2) | Total (3) | Food (4) | Total (5) | Food (6) | Total (7) | Food (8) |
| SPEI growing season | Year 0 | 0.014 | 0.017 | | | | | | |
| | | (0.019) | (0.018) | | | | | | |
| SPEI growing season | Year 1 | 0.001 | 0.010 | | | | | | |
| | | (0.025) | (0.023) | | | | | | |
| SPEI growing season | Months -3 to 2 after harvest | | | −0.028 | −0.051* | | | | |
| | | | | (0.031) | (0.029) | | | | |
| SPEI growing season | Months 3 to 8 after harvest | | | 0.058** | 0.057** | | | | |
| | | | | (0.028) | (0.024) | | | | |
| SPEI growing season | Months 9 to 14 after harvest | | | −0.021 | 0.005 | | | | |
| | | | | (0.028) | (0.024) | | | | |
| SPEI growing season | Months 15 to 20 after harvest | | | −0.013 | −0.006 | | | | |
| | | | | (0.019) | (0.019) | | | | |
| Predicted harvest index | Year 0 | | | | | 0.068** | 0.066** | | |
| | | | | | | (0.031) | (0.030) | | |
| Predicted harvest index | Year 1 | | | | | −0.001 | −0.002 | | |
| | | | | | | (0.029) | (0.028) | | |
| Predicted harvest index | Months -3 to 2 after harvest | | | | | | | −0.021 | −0.016 |
| | | | | | | | | (0.054) | (0.050) |
| Predicted harvest index | Months 3 to 8 after harvest | | | | | | | 0.099** | 0.094** |
| | | | | | | | | (0.038) | (0.036) |
| Predicted harvest index | Months 9 to 14 after harvest | | | | | | | −0.009 | −0.017 |
| | | | | | | | | (0.034) | (0.034) |
| Predicted harvest index | Months 15 to 20 after harvest | | | | | | | −0.002 | 0.004 |
| | | | | | | | | (0.035) | (0.034) |
| Year FEs | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District FEs | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample mean | | 12.728 | 12.210 | 12.728 | 12.210 | 12.728 | 12.210 | 12.728 | 12.210 |
| $R^2$ | | 0.120 | 0.103 | 0.127 | 0.109 | 0.118 | 0.101 | 0.120 | 0.103 |
| N | | 11561 | 11571 | 11529 | 11539 | 11561 | 11571 | 11557 | 11567 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. When the dependent variable is *SPEI growing season*, the standard errors are clustered by district. When the dependent variable is *predicted harvest index*, the standard errors are bootstrapped to account for the construction of the regressor, as described in Appendix Section C.5.3. The unit of observation is (household × wave). The dependent variable is based on log household per capita expenditure adjusted for inflation at the district and year level, which is available for IFLS waves 1-4 (see more on this in Appendix Section C.1.1). For columns (1), (3), and (5), and (7), the dependent variable is total expenditure calculated in this way. For the remaining columns, it is food expenditure only. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. For the SPEI growing season, the harvest date is defined as the average harvest date of the main crop grown in that district (as according to aggregate data, see description in Section C.1.2). For the Predicted Harvest Index, the harvest date is defined as the average harvest date of rice in that district based on IFLS harvest data (see Section C.4.2). "Year 0" is the period between 3 months before and 8 months after harvest (inclusive). Analogously, "Year 1" is the period 9 months and 20 months after harvest (inclusive). For the 6-monthly effects, the coefficients reported come from Equation C.4 and are $\beta_0$ for -3 to 2 months; $\beta_0 + \gamma_0$ for 3-8 months; $\beta_1$ for 9-14 months; $\beta_1 + \gamma_1$ for 15-20 months.

weather shock will bite (3-8 months after harvest), *or* to the 6-month period when weather has no effect (3 months before to 2 months after harvest). This assignment will depend on when the individual is born relative to harvest. The shock will be more salient for some individuals than others, so I account for differences across individuals in shock exposure in a number of ways.[21]

Figure 3.3 shows the results.[22] For both SPEI growing season and predicted harvest index, we again see strong evidence of a critical period at age 2. The similarity of the results across both weather measures suggests that the critical period in cognitive development at age 2 is a robust finding. The 6-month interval specification shows the effects to be particularly concentrated in the *second* half of the year between the second and third birthdays. For both weather indexes, there are a strong effects of around 0.08 standard deviations at age 30-35 months, with p-values below 0.01. The q-value for SPEI growing season is below 0.05, while for the predicted harvest index I do not have the power to reject null effects after adjusting for multiple hypothesis testing.[23] At age 24-29 months, the evidence is more mixed. SPEI growing season yields a positive coefficient of 0.06 standard deviations that is only significant before adjusting for multiple hypothesis testing, and the coefficient on predicted harvest index is close to 0.

### 3.5.4 Breastfeeding

I find a surprising lack of evidence of an effect of weather variation *before* age 2. One potential explanation for this result is that children are protected from nutritional shocks at these ages because they are being breastfed. This is plausible, since the median weaning age for a child in the IFLS sample is 609 days old, shortly before the second birthday. Without a source of exogenous variation in breastfeeding choices, I cannot rigorously test whether breastfeeding mediates the effects of weather variation. However, two pieces of evidence suggest that it might play an offsetting role.

First, mothers change when they stop breastfeeding in a way that might offset the effects of shocks at age 1. Using data on whether children are still being breastfed at the time of an IFLS survey, I measure whether this is affected by the most recent SPEI growing season, and how this effect depends on the age of the child (Figure 3.4). Around the most common weaning time (500-700 days old), increases in SPEI growing season are associated with a *decrease* in the probability of the child being breastfed. When harvests are bad around age 1, mothers prolong breastfeeding. This is likely to protect children from shocks at this age.[24]

Second, for the small subset of children who are never breastfed, there is a very strong positive effect of SPEI growing season at age 1 on cognitive development (Appendix Table C.16.1), suggesting that

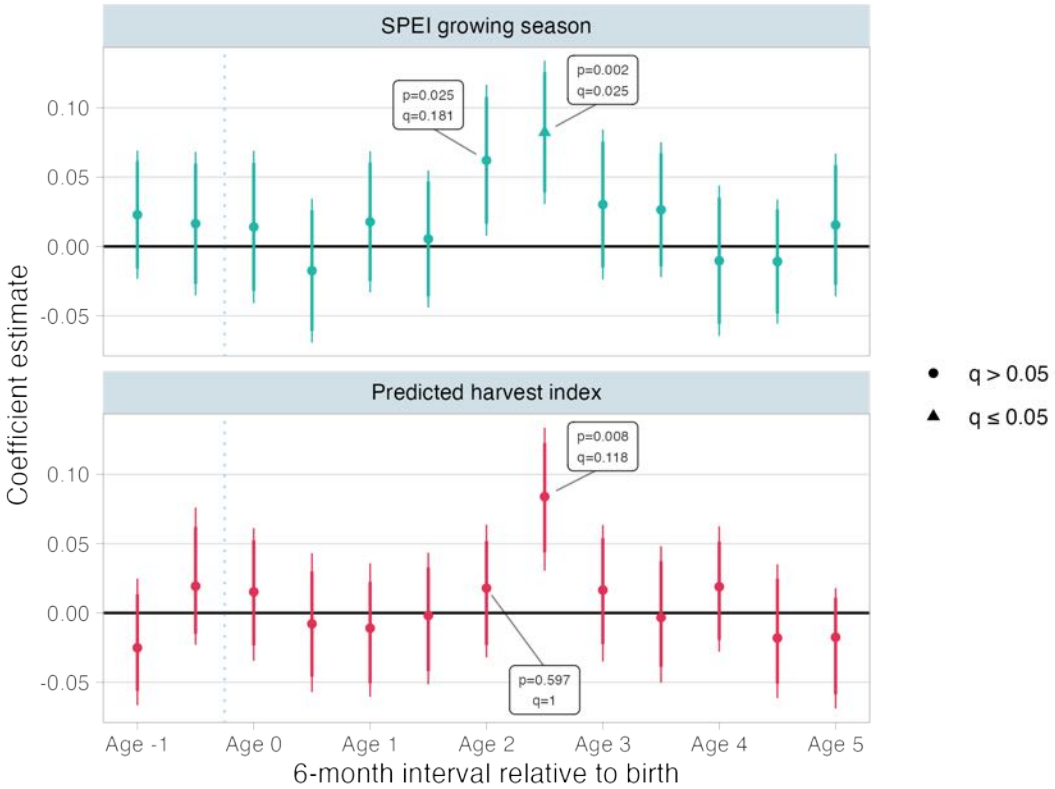
[21] If an individual is born 3 months after harvest, then the first 6-month period of their life (age 0-5 months) is assigned the value of the shock that bites in that period, and the second 6-month period of their life (age 6-11 months) is assigned a value of 0 (because the shock has no detectable effect in this period). A person born 2 months later in the same district is assigned the same values, but will in reality be less exposed to the assigned shock. To control for this differing exposure to the shock variable, I include (district × birth-month-of-year) fixed effects, and show robustness to weighting the shock variable by an individual's exposure to the shock (Appendix Table C.12.13). This methodology is described in detail in Appendix Section C.5.2.

[22] For these results, I increase power by focusing only on the first 5 years of life. Examining up to age 15 yields similar coefficients for age 2.5 across both indexes (Appendix Table C.12.9). However, testing 35 hypotheses simultaneously reduces the power to detect effects after adjusting for multiple hypothesis testing. Appendix Table C.12.10 shows the results by year for predicted harvest, which are qualitatively similar to the results with SPEI growing season. The results on the predicted harvest index are robust to different constructions of the index (Appendix Tables C.12.11 and C.12.12).

[23] Appendix Figure C.12.1 shows that the coefficient estimates are larger for males, as they were for the main results.

[24] The protective effect of prolonging breastfeeding may be weaker if maternal nutrition also suffers. Appendix Table C.16.2 shows some weak evidence that the predicted harvest index is positively correlated with maternal nutrition between 3 and 8 months after harvest, although the smaller sample of mothers makes it difficult to reject null effects.

**Figure 3.3:** *Effect of weather on cognitive development: 6-month intervals*




*Notes*: Each point represents the effect of weather variation in each 6-month age interval on adult cognitive z-score. The top chart uses SPEI growing season as the main explanatory variable, and the bottom chart uses the predicted harvest index. Error bars are 90% and 95% confidence intervals based on standard errors clustered at the district level (not adjusted for multiple testing). Error bars are 90% and 95% confidence intervals based on standard errors that are not adjusted for multiple testing. When the dependent variable is *SPEI growing season*, the standard errors are clustered by district. When the dependent variable is *predicted harvest index*, the standard errors are bootstrapped to account for the construction of the regressor, as described in Appendix Section C.5.3. A triangular point is used when the q-value is less than 0.05. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given specification. The dependent variable in all models is the adult cognitive factor score (measured after age 15), which is standardised to have 0 mean and standard deviation of 1. Individuals are matched to values of predicted harvest index using the process described in Section C.5.2. 6-monthly specifications control for (district × month-of-birth) fixed effects, to account for variation due to the timing of birth relative to the 6-month "bucket" individuals are assigned to (see Appendix Section C.5.2 for details). Both specifications also control for birth year fixed effects and individual-level controls (dummies for highest level of parental education, sex, and religion). All models use the individual-level attrition-corrected weights provided in the IFLS data.

they are particularly vulnerable to nutrition shocks due to changes in the weather. This second result should be interpreted with caution. Only 1.4% (N = 53) of the sample with available data are never breastfed, so the coefficient estimate is noisily estimated. The decision not to breastfeed is also endogenous, and may be affected by weather variation in a way that confounds the causal estimates. For example, the most common reason given for not breastfeeding is insufficient breastmilk, which could be caused by poor nutrition of the caretaker.

### 3.5.5 Discussion of mechanisms

The lack of effect before age 2 is surprising given the existing literature that proposes the first 1000 days as the critical period for cognitive development.[25] There are a number of factors that could explain this.

First, children may be protected from nutritional shocks when they are breastfeeding, as I discussed in Section 3.5.4. More generally, parental responses can offset the effects of negative shocks. This appeared to be the case for breastfeeding behaviour: parents responded to negative shocks by prolonging breastfeeding (and vice versa), thereby attenuating the effects of shocks before age 2. Parents may also be more willing to reduce other expenditures and maintain nutritional expenditures for younger children. Since the impacts I measure are *reduced-form*, measured at least 13 years after shocks at age 2, they incorporate the effects of any subsequent parental investments.

Another explanation for the discrepancy between my results and the literature could be that existing literature typically focuses on methods that do not make use of exogenous shocks. A neurodevelopmental literature shows a correlation between brain growth and cognitive development (Isaacs et al., 2008; Gilmore et al., 2018), while a literature in nutrition shows a correlation between stunting and cognitive development (Victora et al., 2010; Aiyar & Cummins, 2017; Perkins et al., 2017). While stunting and brain growth occur most rapidly before age 2, this does not directly imply that exogenous shocks at this time will have the biggest effect cognitive development in this period.

Indeed, there are a number of plausible mechanisms from the neurodevelopmental literature that could explain effects after age 2. First, even though brain volume growth is fastest before age 2, one study found it to be more sensitive to economic circumstance in the 24-36 month period than at earlier ages (Hanson et al., 2013).[26] Second, the prefrontal cortex is an important determinant of cognitive skills (Ferrer, 2009; Little et al., 2014), and this part of the brain follows a slower growth path, with important changes between ages 2 and 4 (Fiske & Holmboe, 2019). Third, connectivity between the frontal and parietal cortices may be important for cognitive ability, and later ages (after age 2) are key for the development of this connectivity (Buss & Spencer, 2014; Buss et al., 2014; Buss & Spencer, 2018; Fiske & Holmboe, 2019).[27] It remains an important avenue of research to explore the relative importance of such factors for explaining the effects of nutritional and economic shocks in low- and middle-income countries.

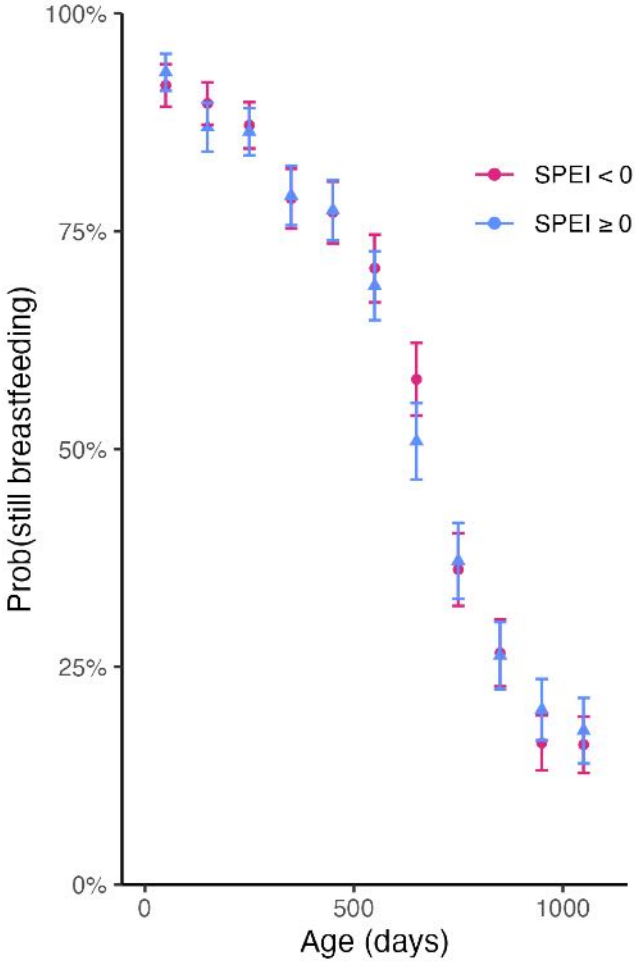
[25] It does not preclude later effects, such as the one at age 30-35 months observed here: a review concluded that the earliest years are not the sole sensitive time period (Wachs et al., 2014), and studies show effects of preschool programmes at ages 3-4 on cognitive development (Dean & Jayachandran, 2020; Schweinhart & Weikart, 1981; Reynolds et al., 2011), suggesting that cognitive development is not "set in stone" by 24 months after birth.

[26] (Hanson et al., 2013) show that the divergence in grey matter volume between US children in mid- and low-SES groups is somewhat starker in the 24–36 month period, suggesting heightened sensitivity.

[27] The studies indicate that the development of a stronger fronto-parietal network between the ages of 3 and 4.5 years underlies the development of visual working memory capacity, which has been shown to be correlated to performance on cognitive tests (Little et al., 2014).

**Figure 3.4:** *Breastfeeding timing*

**(a)** *Binary split*




**(b)** *Interaction specification*

| | (1) |
|---|---|
| SPEI | 0.008 |
| | (0.015) |
| SPEI x (Age 100-200 days) | −0.033* |
| | (0.019) |
| SPEI x (Age 200-300 days) | −0.013 |
| | (0.019) |
| SPEI x (Age 300-400 days) | −0.018 |
| | (0.022) |
| SPEI x (Age 400-500 days) | −0.011 |
| | (0.019) |
| SPEI x (Age 500-600 days) | −0.041* |
| | (0.021) |
| SPEI x (Age 600-700 days) | −0.060** |
| | (0.024) |
| SPEI x (Age 700-800 days) | −0.012 |
| | (0.025) |
| SPEI x (Age 800-900 days) | −0.010 |
| | (0.023) |
| SPEI x (Age 900-1000 days) | 0.014 |
| | (0.023) |
| SPEI x (Age 1000-1100 days) | 0.011 |
| | (0.020) |
| Year fixed effects | Yes |
| Birth district fixed effects | Yes |
| $R^2$ | 0.362 |
| N | 11656 |

*Notes*: Sample used is all children under the age of 3 years in rural farming households in the IFLS. For these children, mothers were asked whether they were still breastfeeding their child as of the date of interview. Panel **(a)** shows the raw probabilities of still being breastfed for each 100-day bin of age. The sample is split into those who experienced a positive SPEI growing season shock at the last harvest (SPEI $\geq$ 0) and those who experienced a negative SPEI growing season shock at the last harvest (SPEI $<$ 0). Panel **(b)** uses an interaction specification. SPEI growing season (as a linear variable, rather than a binary one) is interacted with each 100-day age bin. The outcome is a binary indicator of whether the child was still being breastfeed as of the date of interview. Standard errors are reported in parentheses and are clustered by birth district. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. All models use the individual-level attrition-corrected weights provided in the IFLS data. *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

## 3.6 Discussion & conclusion

In this paper I present new evidence for the existence of a critical period at age 2 for cognitive development as a result of weather shocks in early childhood. Weather shocks at this age on individuals born in rural Indonesia between 1988 and 2000 have a long lasting effect on their cognitive ability as an adult, whereas weather shocks at other ages have no detectable impact. When examining similar evidence on adult socioemotional outcomes, I find no such evidence of a critical period for these weather shocks, with either null effects or effects that cannot be clearly isolated to single periods. I cannot rule out smaller positive impacts of less than 0.06 standard deviations for these null effects, but the results on cognitive development at age 2 suggest a particularly strong impact in this period.

The observed effects on cognitive development are consistent with an economic channel in which weather affects rural households' expenditure and nutritional investment, leading to impacts on cognitive development that persist into adult life. Caretakers partially compensate for shocks before age 2, in particular by prolonging breastfeeding. If breastfeeding is playing a protective role, this suggests the importance of improving insurance against shocks for families with children who have recently been weaned. More generally, the way parents react to shocks and interventions can play an important role in determining long-run impacts. Further research should seek to understand which behavioural factors lead to crowd-in or crowd-out of caretaker investment.

The results I present are most relevant to other contexts in which agricultural harvests play an important role in determining nutritional investments in children, and in which malnutrition is sufficiently severe that it can hamper the cognitive development of young children. By contrast, critical periods are likely to manifest differently in contexts with less agricultural dependency and nutritional vulnerability. Further research should investigate whether the critical periods found here exist for the effects of *policy interventions* that target nutrition and cognitive development in contexts similar to rural Indonesia. Intervening in such periods may yield large gains in cost-effectiveness.

# 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.

Aboud, F. E., & Singla, D. R. (2012, August). Challenges to changing health behaviours in developing countries: A critical overview. *Social Science & Medicine (1982)*, *75*(4), 589–594. doi: 10.1016/j.socscimed.2012.04.009

Adhvaryu, A., Fenske, J., & Nyshadham, A. (2019, August). Early Life Circumstance and Adult Mental Health. *Journal of Political Economy*, *127*(4), 1516–1549. doi: 10.1086/701606

Adhvaryu, A., Nyshadham, A., Molina, T., & Tamayo, J. (2018). Helping Children Catch Up: Early Life Shocks and the PROGRESA Experiment. *NBER Working Paper Series*, 1–75.

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

Aiyar, A., & Cummins, J. R. (2017). Age-Profile Estimates of the Relationship Between Economic Growth and Child Health. , 46.

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

Akyeampong, K., Andrabi, T., Banerjee, A., Banerji, R., Dynarski, S., Glennerster, R., … Yoshikawa, H. (2023). *Cost-Effective Approaches to Improve Global Learning - What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries?* London, Washington D.C., New York: FCDO, the World Bank, UNICEF, and USAID.

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021, November). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics*, *136*(4), 2147–2194. doi: 10.1093/qje/qjab009

Alan, S., & Kubilay, E. (2024). Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting. *European University Institute, mimeo*.

Ali, T. S., & Rizvi, S. N. (2010, August). Menstrual knowledge and practices of female adolescents in urban Karachi, Pakistan. *Journal of Adolescence*, *33*(4), 531–541. doi: 10.1016/j.adolescence.2009.05.013

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

Almond, D., Currie, J., & Duque, V. (2018, December). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. doi: 10.1257/jel.20171164

Almond, D., & Mazumder, B. (2013). Fetal Origins and Parental Responses. *Annual Review of Economics*, *5*(1), 37–56. doi: 10.1146/annurev-economics-082912-110145

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

Ashraf, N., Bau, N., Low, C., & McGinn, K. (2020). Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment*. *The Quarterly Journal of Economics*. doi: 10.1093/qje/qjz039

Attanasio, O. (2015, December). The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies. *Journal of the European Economic Association*, *13*(6), 949–997. doi: 10.1111/jeea.12159

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., & Rubio-Codina, M. (2020, January). Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia. *American Economic Review*, *110*(1), 48–85. doi: 10.1257/aer.20150183

Attanasio, O., Meghir, C., & Nix, E. (2020, November). Human Capital Development and Parental Investment in India. *The Review of Economic Studies*, *87*(6), 2511–2541. doi: 10.1093/restud/rdaa026

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

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2020, January). Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. *American Economic Journal: Applied Economics*, *12*(1), 210–259. doi: 10.1257/app.20170416

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013, July). The Diffusion of Microfinance. *Science*, *341*(6144), 1236498–1236498. doi: 10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019, November). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *The Review of Economic Studies*, *86*(6), 2453–2490. doi: 10.1093/restud/rdz008

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

Barham, T., Macours, K., & Maluccio, J. A. (2013). Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *American Economic Review*, *103*(3), 467–471. doi: 10.1257/aer.103.3.467

Barham, T., Macours, K., & Maluccio, J. A. (2018). Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes. (August).

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., & Shapiro, J. S. (2016, February). Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy*, *124*(1), 105–159. doi: 10.1086/684582

Barreca, A. I. (2012, January). Climate change, humidity, and mortality in the United States. *Journal of Environmental Economics and Management*, *63*(1), 19–34. doi: 10.1016/j.jeem.2011.07.004

Bazzi, S., & Gudgeon, M. (2016). Local Government Proliferation , Diversity, and Conflict.

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021, June). Can Network Theory-Based Targeting Increase Technology Adoption? *American Economic Review*, *111*(6), 1918–1943. doi: 10.1257/aer.20200295

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.

Beguería, S., Vicente-Serrano, S. M., Reig, F., & Latorre, B. (2014, August). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, *34*(10), 3001–3023. doi: 10.1002/joc.3887

Beguería, S., & Vicente Serrano, S. M. (2017). *SPEIbase v.2.5 Global 04-month 1901-2015 SPEI, 2017*. http://digital.csic.es/handle/10261/153475?locale=en. doi: 10.20350/digitalCSIC/8508

Belachew, T., Hadley, C., Lindstrom, D., Getachew, Y., Duchateau, L., & Kolsteren, P. (2011, September). Food insecurity and age at menarche among adolescent girls in Jimma Zone Southwest Ethiopia: A longitudinal study. *Reproductive biology and endocrinology: RB&E*, *9*, 125. doi: 10.1186/1477-7827-9-125

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

Benjamin, D. J., Debnam, J., Fleurbaey, M., Heffetz, O., & Kimball, M. (2020). What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents. , 41.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Benshaul-Tolonen, Anja, Aja, Zulaika, G., Sommer, M., & Phillips-Howard, P. A. (2020). Measuring menstruation-related absenteeism among adolescents in low-income countries. In *The Palgrave Handbook of Critical Menstruation Studies* (pp. 705–723). Palgrave Macmillan.

Benshaul-Tolonen, A., Zulaika, G., Nyothach, E., Odour, C., Mason, L., Obor, D., & Laserson, K. F. (2021). *Sanitary products, absenteeism and psychosocial well-being: Evidence from a three-arm cluster randomized controlled feasibility study in Western Kenya* (Working Paper No. WP No.93). Columbia University, SIPA, CDEP-CGEG.

Bergstrom, K., & Ozler, B. (2021). *Improving the Well-Being of Adolescent Girls in Developing Countries*. The World Bank. doi: 10.1596/1813-9450-9827

Bernard, T., Dercon, S., Orkin, K., & Taffesse, A. S. (2019, May). Parental Aspirations for Children's Education: Is There a "Girl Effect"? Experimental Evidence from Rural Ethiopia. *AEA Papers and Proceedings*, *109*, 127–132. doi: 10.1257/pandp.20191072

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

Bharadwaj, P., Løken, K. V., & Neilson, C. (2013, August). Early Life Health Interventions and Academic Achievement. *American Economic Review*, *103*(5), 1862–1891. doi: 10.1257/aer.103.5.1862

Bharati, T., & Chin, S. (2016). Does Education Affect Time Preference? *SSRN Electronic Journal*. doi: 10.2139/ssrn.2880879

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. doi: 10.1093/acprof:oso/9780190622046.001.0001

Bicchieri, C., & Funcke, A. (2018, March). Norm Change: Trendsetters and Social Structure. *Social Research: An International Quarterly*, *85*(1), 1–21. doi: 10.1353/sor.2018.0002

Binswanger, H. P. (1980, August). Attitudes Toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics*, *62*(3), 395–407. doi: 10.2307/1240194

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

Breza, E. (2016, April). Field Experiments, Social Networks, and Development. In Y. Bramoullé, A. Galeotti, & B. W. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks* (pp. 411–439). Oxford University Press. doi: 10.1093/oxfordhb/9780199948277.013.25

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

Brown, S., & Taylor, K. (2008). Bullying, education and earnings: Evidence from the National Child Development Study. *Economics of Education Review*, *27*(4), 387–401.

Buehren, N., Goldstein, M. P., Gulesci, S., Sulaiman, M., & Yam, V. (2017, February). Evaluation of an adolescent development program for girls in Tanzania. *Policy Research Working Paper Series*(7961).

Burgess, R., Deschênes, O., Donaldson, D., & Greenstone, M. (2017). Weather, climate change and death in India. *Unpublished Manuscript*, 71.

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

Buss, A. T., Fox, N., Boas, D. A., & Spencer, J. P. (2014, January). Probing the early development of visual working memory capacity with functional near-infrared spectroscopy. *NeuroImage*, *85*, 314–325. doi: 10.1016/j.neuroimage.2013.05.034

Buss, A. T., & Spencer, J. P. (2014, June). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii, 1–103. doi: 10.1002/mono.12096

Buss, A. T., & Spencer, J. P. (2018, July). Changes in frontal and posterior cortical activity underlie the early emergence of executive function. *Developmental Science*, *21*(4), e12602. doi: 10.1111/desc.12602

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

Carneiro, P., Kraftman, L., Mason, G., Moore, L., Rasul, I., & Scott, M. (2021, August). The Impacts of a Multifaceted Prenatal Intervention on Human Capital Accumulation in Early Life. *American Economic Review*, *111*(8), 2506–2549. doi: 10.1257/aer.20191726

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

Cattell, R. B. (1966, April). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. doi: 10.1207/s15327906mbr0102_10

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

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2018, June). *Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India* (Working Paper No. 24678). National Bureau of Economic Research. doi: 10.3386/w24678

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

Chirgwin, H., Cairncross, S., Zehra, D., & Sharma Waddington, H. (2021, December). Interventions promoting uptake of water, sanitation and hygiene (WASH) technologies in low- and middle-income countries: An evidence and gap map of effectiveness studies. *Campbell Systematic Reviews*, *17*(4), e1194. doi: 10.1002/cl2.1194

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

Chuang, Y., & Schechter, L. (2015, November). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, *117*, 151–170. doi: 10.1016/j.jdeveco.2015.07.008

Clark, L. A., & Watson, D. (1995, September). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. doi: 10.1037/1040-3590.7.3.309

Cordero, M. E., D'Acuña, E., Benveniste, S., Prado, R., Nuñez, J. A., & Colombo, M. (1993). Dendritic development in neocortex of infants with early postnatal life undernutrition. *Pediatric Neurology*, *9*(6), 457–464. doi: 10.1016/0887-8994(93)90025-8

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*.

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, *97*(2), 25.

Cunha, F., & Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, *43*(4), 738–782. doi: 10.3368/jhr.43.4.738

Cunha, F., & Heckman, J. J. (2009, April). The Economics and Psychology of Inequality and Human DEvelopment. *Journal of the European Economic Association*, *7*(2-3), 320–364. doi: 10.1162/JEEA.2009.7.2-3.320

Cunha, F., Heckman, J. J., & Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931. doi: 10.3982/ECTA6551

Currie, J., & Almond, D. (2011). Human capital development before age five. In *Handbook of Labor Economics* (Vol. 4, pp. 1315–1486). Elsevier B.V. doi: 10.1016/S0169-7218(11)02413-0

Czura, K., Menzel, A., & Miotto, M. (2019). Menstrual Health, Worker Productivity and Well-being among Female Bangladeshi Garment Workers. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3519895

Dal Bó, E., Finan, F., Li, N. Y., & Schechter, L. (2018). Government Decentralization Under Changing State Capacity: Experimental Evidence From Paraguay. , 66.

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

Das, J., Do, Q. T., Friedman, J., & McKenzie, D. (2009). Mental health patterns and consequences: Results from survey data in five developing countries. *World Bank Economic Review*, *23*(1), 31–55. doi: 10.1093/wber/lhn010

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 Chaisemartin, C., & D'Haultfœuille, X. (2018, April). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, *85*(2), 999–1028. doi: 10.1093/restud/rdx049

de Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way Fixed Effects Regressions with Several Treatments. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3751060

de Quidt, J., & Haushofer, J. (2019). Depression through the Lens of Economics. In *The Economics of Poverty Traps* (pp. 127–152). doi: 10.7208/chicago/9780226574448.003.0003

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

Dean, J. T., & Jayachandran, S. (2020). Attending kindergarten improves cognitive development in India, but all kindergartens are not equal. , 64.

De Datta, S. K. (1981). *Principles and practices of rice production*. New York: Wiley.

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

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–798. doi: 10.1257/jel.52.3.740

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

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, *1*(3), 111–34. doi: 10.1257/app.1.3.111

Deoni, S., Dean, D., Joelson, S., O'Regan, J., & Schneider, N. (2018, September). Early nutrition influences developmental myelination and cognition in infants and young children. *NeuroImage*, *178*, 649–659. doi: 10.1016/j.neuroimage.2017.12.056

Department, U. S. S. (2022). *2022 Country Reports on Human Rights Practices*.

Deschênes, O., & Greenstone, M. (2011, October). Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US. *American Economic Journal: Applied Economics*, *3*(4), 152–185. doi: 10.1257/app.3.4.152

Deschênes, O., Greenstone, M., & Guryan, J. (2009, April). Climate Change and Birth Weight. *American Economic Review*, *99*(2), 211–217. doi: 10.1257/aer.99.2.211

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, March). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, *112*(3), 899–927. doi: 10.1257/aer.20201112

Dinarte, L., & Egana del Sol, P. (2018). Preventing Violence in the Most Violent Contexts. Behavioral and Neurophysiological Evidence. (May).

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

Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. In *Handbook of economic field experiments* (Vol. 2, pp. 3–93). Elsevier.

Edmonds, E., Feigenberg, B., & Leight, J. (2020, July). *Advancing the Agency of Adolescent Girls* (Tech. Rep. No. w27513). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w27513

El-Gilany, A.-H., Badawi, K., & El-Fedawy, S. (2005, January). Menstrual Hygiene among Adolescent Schoolgirls in Mansoura, Egypt. *Reproductive Health Matters*, *13*(26), 147–152. doi: 10.1016/S0968-8080(05)26191-8

Eriksen, T. L. M., Nielsen, H. S., & Simonsen, M. (2014). Bullying in Elementary School. *Journal of Human Resources*, *49*(4), 839–871.

Evans, D. K., & Yuan, F. (2022, February). What We Learn about Girls' Education from Interventions That Do Not Focus on Girls. *The World Bank Economic Review*, *36*(1), 244–267. doi: 10.1093/wber/lhab007

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

Fernández, R., Parsa, S., & Viarengo, M. (2019). Coming out in America: AIDS, Politics, and Cultural Change. , 65.

Ferrer, E. (2009, May). Fluid reasoning and the developing brain. *Frontiers in Neuroscience*, *3*(1). doi: 10.3389/neuro.01.003.2009

Fishman, R. (2016, February). More uneven distributions overturn benefits of higher precipitation for crop yields. *Environmental Research Letters*, *11*(2), 024004. doi: 10.1088/1748-9326/11/2/024004

Fishman, R. (2018, March). Groundwater depletion limits the scope for adaptation to increased rainfall variability in India. *Climatic Change*, *147*(1-2), 195–209. doi: 10.1007/s10584-018-2146-x

Fiske, A., & Holmboe, K. (2019, June). Neural substrates of early executive function development. *Developmental Review*, *52*, 42–62. doi: 10.1016/j.dr.2019.100866

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

Food and Agriculture Organisation. (2005). *Fertilizer use by crop in Indonesia.*

Frankenberg, E., & Karoly, L. (1995). *The 1993 indonesian family life survey: Overview and field report, publication no* (Tech. Rep.). DRU-1195/1-NICHD/AID, RAND, Santa Monica, CA.

Frederick, W. H., Worden, R. L., & of Congress. Federal Research Division, L. (2011). *Indonesia: A Country Study*. Federal Research Division, Library of Congress.

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

Garces, E., Thomas, D., & Currie, J. (2002). Longer-Term Effects of Head Start. *The American Economic Review*, *92*(4), 32.

Gertler, P., Heckman, J. J., Pinto, R., Chang, S. M., Grantham-McGregor, S., Vermeersch, C., ... Wright, A. (2021). Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31. , 50.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018, March). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, *19*(3), 123–137. doi: 10.1038/nrn.2018.1

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

Goodman-Bacon, A. (2018, September). *Difference-in-Differences with Variation in Treatment Timing* (Tech. Rep. No. w25018). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w25018

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

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007,

January). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60–70. doi: 10.1016/S0140-6736(07)60032-4

Grantham-McGregor, S., Powell, C., Walker, S., & Himes, J. (1991, July). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican Study. *The Lancet*, *338*(8758), 1–5. doi: 10.1016/0140-6736(91)90001-6

Groh, M., McKenzie, D., & Vishwanath, T. (2015). Reducing Information Asymmetries in the Youth Labor Market of Jordan with Psychometrics and Skill Based Tests. *The World Bank Economic Review*, *29*(suppl 1), S106-S117. doi: 10.1093/wber/lhv005

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., & Young, H. (2023). A Stepping Stone Approach to Norm Transitions. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4425503

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

Gunnsteinsson, S., Adhvaryu, A., Christian, P., Labrique, A., Sugimoto, J., Shamim, A. A., & West, K. P. (2014). Resilience to early life shocks: Evidence from the Interaction of a Randomized Controlled Trial and a Natural Experiment. *Working Paper*, 1–48.

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

Hamory, J., Baird, S., Das, S., Jones, N., Woldehanna, T., & Yadete, W. (2023). Do layered adolescent-centric interventions improve girls' capabilities? Evidence from a mixed-methods cluster randomised controlled trial in Ethiopia.

Hanson, J. L., Hair, N., Shen, D. G., Shi, F., Gilmore, J. H., Wolfe, B. L., & Pollak, S. D. (2013, December). Family Poverty Affects the Rate of Human Infant Brain Growth. *PLoS ONE*, *8*(12), e80954. doi: 10.1371/journal.pone.0080954

Harari, M., & Ferrara, E. L. (2018, October). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics*, *100*(4), 594–608. doi: 10.1162/rest_a_00730

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

Haushofer, J., & Fehr, E. (2014, May). On the psychology of poverty. *Science*, *344*(6186), 862–867. doi: 10.1126/science.1232491

Heckman, J. J. (2006, June). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, *312*(5782), 1900–1902. doi: 10.1126/science.1128898

Heckman, J. J., Pinto, R., & Savelyev, P. (2013, October). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, *103*(6), 2052–2086. doi: 10.1257/aer.103.6.2052

Heckman, J. J., Stixrud, J., & Urzua, S. (2006, February). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. , w12006. doi: 10.3386/w12006

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

Hennegan, J., & Montgomery, P. (2016, February). Do Menstrual Hygiene Management Interventions Improve Education and Psychosocial Outcomes for Women and Girls in Low and Middle Income Countries? A Systematic Review. *PLOS ONE*, *11*(2), e0146985. doi: 10.1371/journal.pone.0146985

Hill, S. J., Lo, J., Vavreck, L., & Zaller, J. (2013, October). How Quickly We Forget: The Duration of Persuasion Effects From Mass Communication. *Political Communication*, *30*(4), 521–547. doi: 10.1080/10584609.2013.828143

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

Hochberg, Z., & Belsky, J. (2013, April). Evo-devo of human adolescence: Beyond disease models of early puberty. *BMC medicine*, *11*, 113. doi: 10.1186/1741-7015-11-113

Horn, J. L. (1965, June). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi: 10.1007/BF02289447

Hoynes, H., Schanzenbach, D. W., & Almond, D. (2016, April). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review*, *106*(4), 903–934. doi: 10.1257/aer.20130375

Hu, Q. (2023). Breaking Down Bullying: Empathy, Social Networks, and Adolescents.

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

Isaacs, E., Oates, J., & ILSI Europe a.i.s.b.l. (2008, August). Nutrition and cognition: Assessing cognitive abilities in children and young people. *European Journal of Nutrition*, *47*(S3), 4–24. doi: 10.1007/s00394-008-3002-y

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. (2021, September). Social Norms as a Barrier to Women's Employment in Developing Countries. *IMF Economic Review*, *69*(3), 576–595. doi: 10.1057/s41308-021-00140-w

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

Johnson, R. C., & Jackson, C. K. (2019, November). Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal: Economic Policy*, *11*(4), 310–349. doi: 10.1257/pol.20180510

Kaiser, H. F. (1958, September). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. doi: 10.1007/BF02289233

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

Katz, D., Allport, F. H., & Jenness, M. B. (1931). *Students' attitudes; a report of the Syracuse University reaction study*. Oxford, England: Craftsman Press.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (in press). Do financial concerns make workers less productive? *The Quarterly Journal of Economics*.

Kennedy, G., Ballard, T., Dop, M. C., & European Union. (2011). *Guidelines for measuring household and individual dietary diversity*. Rome: Food and Agriculture Organization of the United Nations.

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

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007, January). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119. doi: 10.1111/j.1468-0262.2007.00733.x

Knudsen, E. I. (2004, October). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, *16*(8), 1412–1425. doi: 10.1162/0898929042304796

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*(27), 10155–10162. doi: 10.1073/pnas.0600888103

Kudamatsu, M., Persson, T., & Strömberg, D. (2012). Weather and Infant Mortality in Africa. , 77.

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

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., ... Vakis, R. (2019, July). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, *5*(7), eaaw5226. doi: 10.1126/sciadv.aaw5226

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

Ledesma, R. D., Valero-Mora, P., & de Valencia, U. (n.d.). Determining the Number of Factors to Retain in EFA: An easy-to- use computer program for carrying out Parallel Analysis. *Exploratory Factor Analysis*, *12*(2), 11.

Lee, C.-T., Zhang, G., & Edwards, M. C. (2012, March). Ordinary Least Squares Estimation of Parameters in Exploratory Factor Analysis With Ordinal Data. *Multivariate Behavioral Research*, *47*(2), 314–339. doi: 10.1080/00273171.2012.658340

Levine, D., & Yang, D. (2014, July). *The Impact of Rainfall on Rice Output in Indonesia* (Tech. Rep. No. w20302). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w20302

Little, D. R., Lewandowsky, S., & Craig, S. (2014, March). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00239

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

Maccini, S., & Yang, D. (2009, May). Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *American Economic Review*, *99*(3), 1006–1026. doi: 10.1257/aer.99.3.1006

Macours, K., Premand, P., & Vakis, R. (2022, September). Transfers, Diversification and Household Risk Strategies: Can Productive Safety Nets Help Households Manage Climatic Variability? *The Economic Journal*, *132*(647), 2438–2470. doi: 10.1093/ej/ueac018

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

Malamud, O., Pop-Eleches, C., & Urquiola, M. (2016, March). *Interactions Between Family and School Environments: Evidence on Dynamic Complementarities?* (Tech. Rep. No. w22112). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w22112

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

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., & Stein, A. D. (2009, April). The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults. *The Economic Journal*, *119*(537), 734–763. doi: 10.1111/j.1468-0297.2009.02220.x

Marsh, D. R., Schroeder, D. G., Dearden, K. A., Sternin, J., & Sternin, M. (2004, November). The power of positive deviance. *BMJ*, *329*(7475), 1177–1179. doi: 10.1136/bmj.329.7475.1177

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

May, W. (2004, March). Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment. *Climate Dynamics*, *22*(2-3), 183–204. doi: 10.1007/s00382-003-0373-x

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

Minnesota Population Center. (2018). *Integrated Public Use Microdata Series, International: Version 7.1. Indonesia Spatially Harmonized Second-Level Geography (1970-2011).* https://international.ipums.org/international/gis_harmonized_2nd.shtml. doi: 10.18128/D020.V7.1

Minnesota Population Center. (2020). *Integrated Public Use Microdata Series, International: Version 7.1. Indonesia Census Data (1980, 1985, 1990, 1995, 2000, 2005, 2010).* https://international.ipums.org/international/gis_harmonized_2nd.shtml. doi: 10.18128/D020.V7.1

Monfreda, C., Ramankutty, N., & Foley, J. A. (2008, March). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000: GLOBAL CROP AREAS AND YIELDS IN 2000. *Global Biogeochemical Cycles*, *22*(1), n/a-n/a. doi: 10.1029/2007GB002947

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

Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, *12*(2), 125–135. doi: 10.1037/h0027568

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. Henry Holt and Company.

Muralidharan, K. (2017). Field Experiments in Education in Developing Countries. In A. Banerjee & E. Duflo (Eds.), *Handbook of Field Experiments* (pp. 323–385). doi: 10.1016/bs.hefe.2016.09.004

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

Muralidharan, K., & Sundararaman, V. (2015, August). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India *. *The Quarterly Journal of Economics*, *130*(3), 1011–1066. doi: 10.1093/qje/qjv013

Myers, D. G., & Bishop, G. D. (1970, August). Discussion Effects on Racial Attitudes. *Science*, *169*(3947), 778–779. doi: 10.1126/science.169.3947.778

Myers, D. G., & Lamm, H. (1976, July). The group polarization phenomenon. *Psychological Bulletin*, *83*(4), 602–627. doi: 10.1037/0033-2909.83.4.602

National Sample Survey Office, India. (2012). *HCE: Monthly per Capita Consumer Expenditure: Average: Tamil Nadu: Urban: Food | Economic Indicators | CEIC*.

Naylor, R. L., Battisti, D. S., Vimont, D. J., Falcon, W. P., & Burke, M. B. (2007, May). Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proceedings of the National Academy of Sciences*, *104*(19), 7752–7757. doi: 10.1073/pnas.0701825104

Nelson, C. A., & Gabard-Durnam, L. J. (2020, March). Early Adversity and Critical Periods: Neurodevelopmental Consequences of Violating the Expectable Environment. *Trends in Neurosciences*, *43*(3), 133–143. doi: 10.1016/j.tins.2020.01.002

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

Orr, L., Ferris, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., ... Kling, J. R. (2003). *Moving to opportunity: Interim impacts evaluation* (Report). Washington D.C: U.S. Dept. of HUD.

Oster, E., & Thornton, R. (2011, January). Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics*, *3*(1), 91–100. doi: 10.1257/app.3.1.91

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

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021, January). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology*, *72*(1), 533–560. doi: 10.1146/annurev-psych-071620-030619

Paradkar, N., & Chowdhury, S. R. (2015). Primary study for detection of arterial blood pressure waveform components. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2015*, 1959–1962. doi: 10.1109/EMBC.2015.7318768

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

Perkins, J. M., Kim, R., Krishna, A., McGovern, M., Aguayo, V. M., & Subramanian, S. (2017, November). Understanding the association between stunting and child development in low- and middle-income countries: Next steps for research and intervention. *Social Science & Medicine*, *193*, 101–109. doi: 10.1016/j.socscimed.2017.09.039

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.

Phillips-Howard, P. A., Nyothach, E., Ter Kuile, F. O., Omoto, J., Wang, D., Zeh, C., ... Laserson, K. F. (2016, November). Menstrual cups and sanitary pads to reduce school attrition, and sexually transmitted and reproductive tract infections: A cluster randomised controlled feasibility study in rural Western Kenya. *BMJ open*, *6*(11), e013229. doi: 10.1136/bmjopen-2016-013229

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

Prado, E. L., & Dewey, K. G. (2014, April). Nutrition and brain development in early life. *Nutrition Reviews*, *72*(4), 267–284. doi: 10.1111/nure.12102

Prentice, A. M., Ward, K. A., Goldberg, G. R., Jarjou, L. M., Moore, S. E., Fulford, A. J., & Prentice, A. (2013, May). Critical windows for nutritional interventions against stunting. *The American Journal of Clinical Nutrition*, *97*(5), 911–918. doi: 10.3945/ajcn.112.052332

Radloff, L. S. (1977, June). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. doi: 10.1177/014662167700100306

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

Rammstedt, B., & John, O. P. (2007, February). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. doi: 10.1016/j.jrp.2006.02.001

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

Raven, J. (2000, August). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41*(1), 1–48. doi: 10.1006/cogp.1999.0735

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

Reece, J. D., Bunn, J. A., Choi, M., & Navalta, J. W. (2021, June). Assessing Heart Rate Using Consumer Technology Association Standards. *Technologies*, *9*(3), 46. doi: 10.3390/technologies9030046

Reynolds, A. J., Temple, J. A., White, B. A. B., Ou, S.-R., & Robertson, D. L. (2011, January). Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program: Cost-Benefit Analysis. *Child Development*, *82*(1), 379–404. doi: 10.1111/j.1467-8624.2010.01563.x

Rietveld, C. A., van Kippersluis, H., & Thurik, A. R. (2015, October). Self-Employment and Health:

Barriers or Benefits?: SELF-EMPLOYMENT AND HEALTH. *Health Economics*, *24*(10), 1302–1313. doi: 10.1002/hec.3087

Ritchie, J. T., & Nesmith, D. S. (2015, October). Temperature and Crop Development. In J. Hanks & J. T. Ritchie (Eds.), *Agronomy Monographs* (pp. 5–29). Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. doi: 10.2134/agronmonogr31.c2

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

Roser, M., & Ritchie, H. (2019). Hunger and undernourishment. *Our World in Data*.

Rossin-Slater, M., & Wüst, M. (2016). *Are Different Early Investments Complements or Substitutes? Long-Run and Intergenerational Evidence from Denmark* (Tech. Rep.).

Rubio-Codina, M., Attanasio, O., Meghir, C., Varela, N., & Grantham-McGregor, S. (2015). The Socioeconomic Gradient of Child Development: Cross-Sectional Evidence from Children 6–42 Months in Bogota. *Journal of Human Resources*, *50*(2), 464–483. doi: 10.3368/jhr.50.2.464

Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010, June). Crop planting dates: An analysis of global patterns: Global crop planting dates. *Global Ecology and Biogeography*, no-no. doi: 10.1111/j.1466-8238.2010.00551.x

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

Sarzosa, M., & Urzúa, S. (2021). Bullying among adolescents: The role of skills. *Quantitative Economics*, *12*(3), 945–980. doi: 10.3982/QE1215

Schilbach, F., Schofield, H., & Mullainathan, S. (2016, May). The Psychological Lives of the Poor. *American Economic Review*, *106*(5), 435–440. doi: 10.1257/aer.p20161101

Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006, February). The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions. *Review of Economics and Statistics*, *88*(1), 113–125. doi: 10.1162/rest.2006.88.1.113

Schlenker, W., & Roberts, M. J. (2006, September). Nonlinear Effects of Weather on Corn Yields*. *Review of Agricultural Economics*, *28*(3), 391–398. doi: 10.1111/j.1467-9353.2006.00304.x

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

Schweinhart, L. J., & Weikart, D. P. (1981, October). Effects of the Perry Preschool Program on Youths Through Age 15. *Journal of the Division for Early Childhood*, *4*(1), 29–39. doi: 10.1177/105381518100400105

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

Shah, M., Seager, J., Montalvao, J., & Goldstein, M. (2023, August). *Sex, Power, and Adolescence: Intimate Partner Violence and Sexual Behaviors* (Working Paper No. 31624). National Bureau of Economic Research. doi: 10.3386/w31624

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

Shiller, R. J. (2017, April). Narrative Economics. *American Economic Review*, *107*(4), 967–1004. doi: 10.1257/aer.107.4.967

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.

Smith, L. G. E., & Postmes, T. (2011). The power of talk: Developing discriminatory group norms through discussion. *British Journal of Social Psychology*, *50*(2), 193–215. doi: 10.1348/014466610X504805

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

Statistik, B. P., & SIAK, K. (2019). Indonesia (IDN) Administrative Boundary Common Operational Database (COD-AB).

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (1998). The second wave of the indonesia family life survey (IFLS3): Overview and field report. *NIA/NICHD*.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (2004). The third wave of the indonesia family life survey (IFLS3): Overview and field report. *NIA/NICHD*.

Strauss, J., Witoelar, F., & Sikoki, B. (2016). *The fifth wave of the indonesia family life survey: Overview and field report; RAND labor and population.* WR-1143/1-NIA/NICHD.

Strauss, J., Witoelar, F., Sikoki, B., & Wattie, A. M. (2009). *The fourth wave of the Indonesia Family Life Survey: Overview and field report.* RAND Labor and Population Working Paper WR-675/1-NIA/NICHD. Santa Monica, CA ....

Subramanian, T., Gupte, M., Dorairaj, V., Periannan, V., & Mathai, A. (2009, April). Psycho-social 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

Tebaldi, C., Hayhoe, K., Arblaster, J. M., & Meehl, G. A. (2006, November). Going to the Extremes: An Intercomparison of Model-Simulated Historical and Future Changes in Extreme Events. *Climatic Change*, *79*(3-4), 185–211. doi: 10.1007/s10584-006-9051-4

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

van den Besselaar, E. J. M., van der Schrier, G., Cornes, R. C., Iqbal, A. S., & Klein Tank, A. M. G. (2017, July). SA-OBS: A Daily Gridded Surface Temperature and Precipitation Dataset for Southeast Asia. *Journal of Climate*, *30*(14), 5151–5165. doi: 10.1175/JCLI-D-16-0575.1

Velicer, W. F. (1976, September). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. doi: 10.1007/BF02293557

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010, April). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. doi: 10.1175/2009JCLI2909.1

Victora, C. G., de Onis, M., Hallal, P. C., Blossner, M., & Shrimpton, R. (2010, March). Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *PEDIATRICS*, *125*(3), e473-e480. doi: 10.1542/peds.2009-1519

Wachs, T. D., Georgieff, M., Cusick, S., & McEwen, B. S. (2014, January). Issues in the timing of integrated early interventions: Contributions from nutrition, neuroscience, and psychological research: Timing of integrated early interventions. *Annals of the New York Academy of Sciences*, *1308*(1), 89–106. doi: 10.1111/nyas.12314

Walker, S. P., Chang, S. M., Powell, C. A., & Grantham-McGregor, S. M. (2005, November). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The Lancet*, *366*(9499), 1804–1807. doi: 10.1016/S0140-6736(05)67574-5

Walker, S. P., Chang, S. M., Wright, A. S., Pinto, R., Heckman, J. J., & Grantham-McGregor, S. M. (2021, August). Cognitive, psychosocial, and behaviour gains at age 31 years from the Jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry*, jcpp.13499. doi: 10.1111/jcpp.13499

WaterAid. (2013). *We can't wait: A report on sanitation and hygiene for women and girls.* WaterAid International. Retrieved from: http://www.zaragoza.es/contenidos/medioambiente/onu/1325-eng_We_cant_wait_sanitation_and_hygiene_for_women_and%20girls.pdf.

Webb, D. (2023). Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India. *mimeo, Paris School of Economics*.

Welz, M., Alfons, A., Demirer, M., & Chernozhukov, V. (2022). GenericML: Generic machine learning inference. *R package version*.

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

White, M. D. (2018). The Problems with Measuring and Using Happiness for Policy Purposes. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3191385

WHO, & UNICEF. (2009). WHO child growth standards and the identification of severe acute malnutrition in infants and children. , 1–12. doi: http://www.who.int/nutrition/publications/severemalnutrition/9789241598163/en/

Wolke, D., Copeland, W. E., Angold, A., & Costello, E. J. (2013). Impact of Bullying in Childhood on Adult Health, Wealth, Crime, and Social Outcomes. *Psychological Science*, *24*(10), 1958–1970.

World Bank. (2018). *Rural Population in Indonesia*.

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

# Appendix A

# Appendix for Chapter 1

## A.1 Additional tables and figures

**Figure A.1.1:** *Experimental design (detailed)*

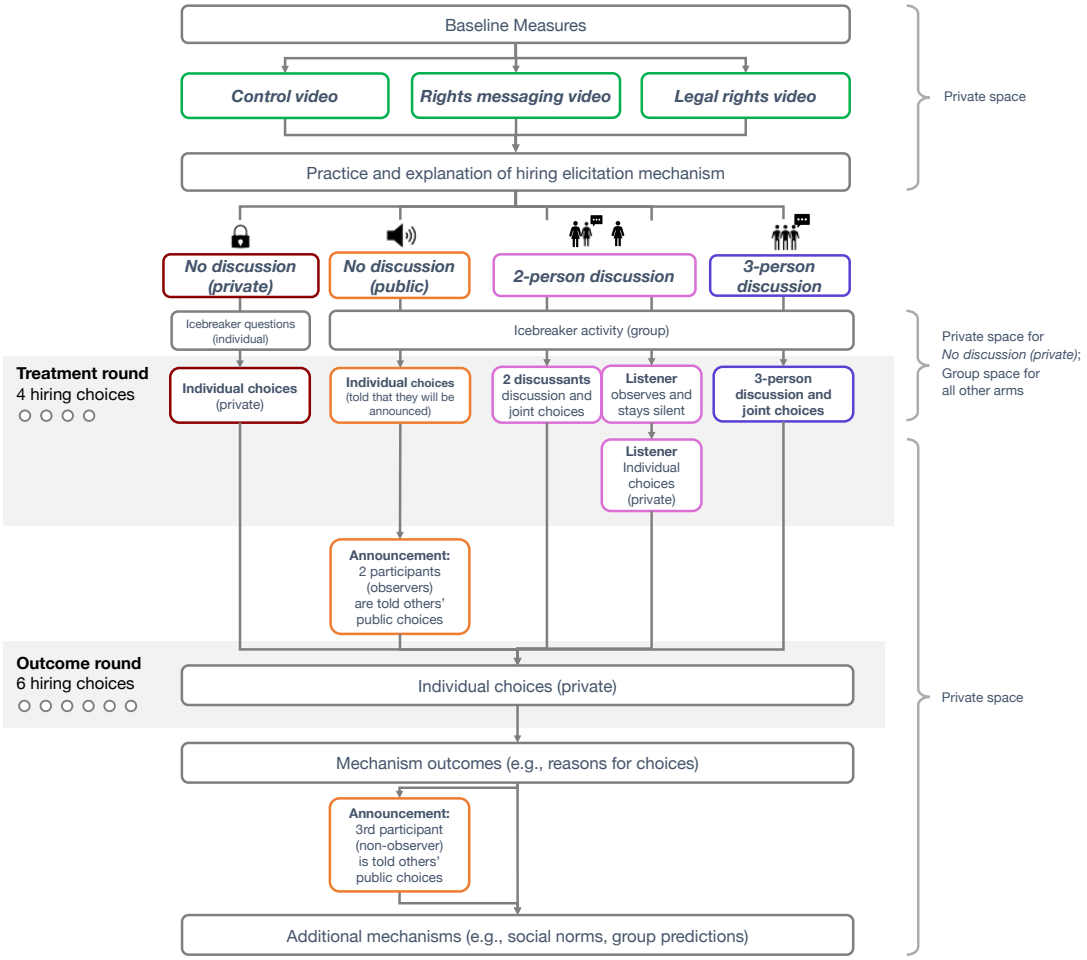
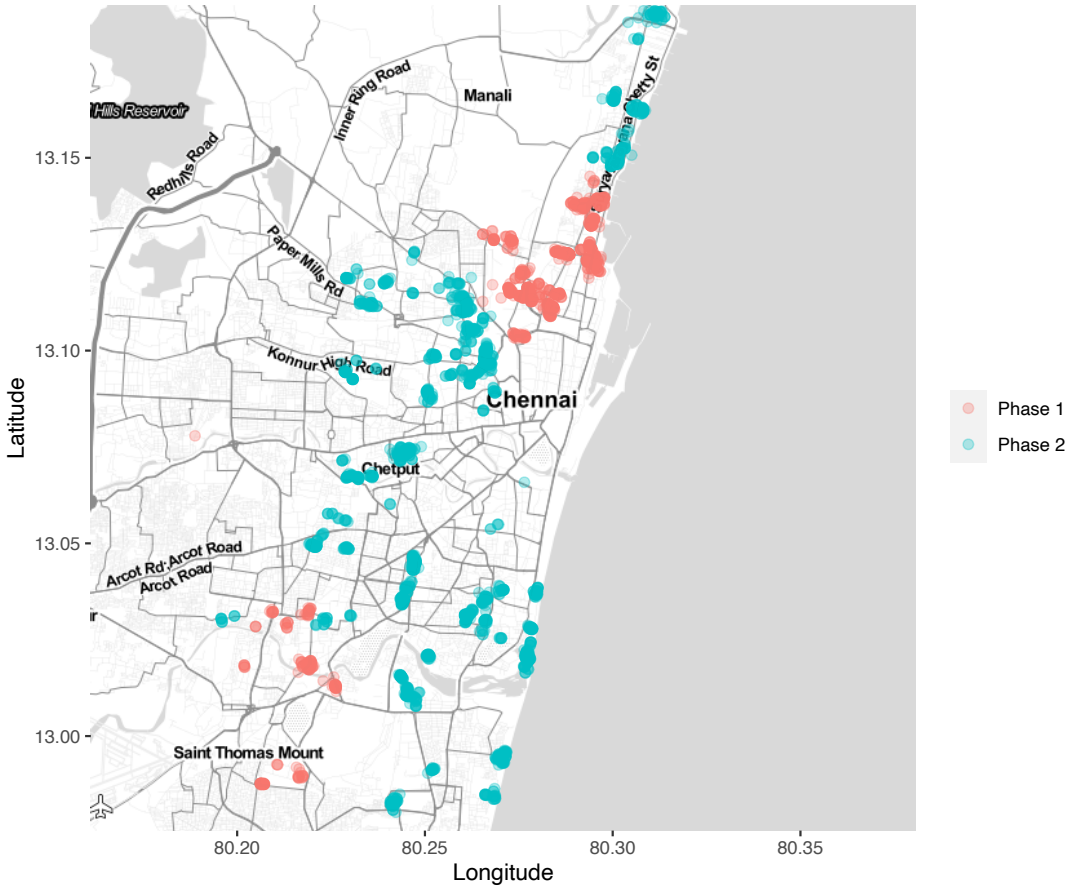


**Figure A.1.2:** *Survey locations*




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

**Table A.1.1:** *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 A.1.2:** *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 1.1. 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 A.10.9). 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 A.1.3:** *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 A.1.4:** *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 A.1.5:** *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 A.1.6:** *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 A.10.9).

**Table A.1.7:** *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 1.1. 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 A.10.9). 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 A.1.8:** *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 × 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 A.10.9).

**Table A.1.9:** *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 1.1. 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 A.10.9). 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 A.1.10:** *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** | 0.110*** | 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** | 0.119*** | 0.118*** |
| | (0.027) [0.025] | (0.027) [0.027] | (0.037) [0.001] | (0.037) [0.001] |
| 3-person discussion | 0.194*** | 0.193*** | 0.233*** | 0.231*** |
| | (0.035) [<0.001] | (0.035) [<0.001] | (0.058) [<0.001] | (0.057) [<0.001] |
| Rights messaging video × 3-person discussion | −0.082 | −0.081 | −0.146* | −0.144* |
| | (0.051) [0.105] | (0.050) [0.109] | (0.080) [0.069] | (0.080) [0.070] |
| Legal rights video × 3-person discussion | 0.002 | 0.003 | −0.002 | 0.000 |
| | (0.048) [0.959] | (0.048) [0.957] | (0.075) [0.982] | (0.075) [0.996] |
| No discussion (public) | | | 0.034 | 0.032 |
| | | | (0.040) [0.389] | (0.039) [0.422] |
| 2-person discussion (listener) | | | 0.193*** | 0.189*** |
| | | | (0.056) [<0.001] | (0.056) [<0.001] |
| 2-person discussion (speaker) | | | 0.159*** | 0.154*** |
| | | | (0.050) [0.002] | (0.050) [0.002] |
| Rights messaging video × No discussion (public) | | | 0.001 | 0.002 |
| | | | (0.059) [0.989] | (0.058) [0.971] |
| Rights messaging video × 2-person discussion (listener) | | | −0.105 | −0.110 |
| | | | (0.079) [0.184] | (0.079) [0.162] |
| Rights messaging video × 2-person discussion (speaker) | | | −0.085 | −0.083 |
| | | | (0.070) [0.226] | (0.070) [0.238] |
| Legal rights video × No discussion (public) | | | 0.015 | 0.017 |
| | | | (0.058) [0.789] | (0.058) [0.773] |
| Legal rights video × 2-person discussion (listener) | | | −0.081 | −0.075 |
| | | | (0.081) [0.321] | (0.082) [0.357] |
| Legal rights video × 2-person discussion (speaker) | | | 0.022 | 0.023 |
| | | | (0.067) [0.748] | (0.067) [0.727] |
| Num. observations | 4498 | 4498 | 4436 | 4436 |
| Num. participants | 2249 | 2249 | 2218 | 2218 |
| Num. groups | 751 | 751 | 741 | 741 |
| Controls | X | X | X | X |
| *p*-val: (Rights messaging video \| 3-person discussion) | 0.804 | 0.822 | 0.638 | 0.656 |
| *p*-val: (Rights messaging video \| No discussion (public)) | | | 0.014 | 0.012 |
| *p*-val: (Rights messaging video \| 2-person discussion (listener)) | | | 0.908 | 0.967 |
| *p*-val: (Rights messaging video \| 2-person discussion (speaker)) | | | 0.664 | 0.635 |
| *p*-val: (Legal rights video \| 3-person discussion) | 0.099 | 0.100 | 0.073 | 0.071 |
| *p*-val: (Legal rights video \| No discussion (public)) | | | 0.003 | 0.003 |
| *p*-val: (Legal rights video \| 2-person discussion (listener)) | | | 0.579 | 0.540 |
| *p*-val: (Legal rights video \| 2-person discussion (speaker)) | | | 0.013 | 0.012 |

*Notes*: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Outcome is whether a participant chose the transgender worker in the private outcome round (restricting analysis to only choices with transgender workers). Sample in columns (1) and (2) includes only the *3-person discussion* arm and the *No discussion (private)* arm, in both phases. *p-val: (Rights messaging video | 3-person discussion)* denotes the *p*-value on the test that the effect of the rights messaging video is 0 for participants in the 3-person discussion arm. Other p-values are defined analogously. Controls include stratum fixed effects; phase fixed effects (columns 1 and 2 only); whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; and the controls selected by double LASSO (see Section A.10.9).

**Table A.1.11:** *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 A.10.9).

**Table A.1.12:** *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 | 13494 | 13494 | 13494 | 13494 | 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 1.1. 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 A.10.9). Controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown.

**Table A.1.13:** *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 A.10.9). 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 A.1.14:** *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.

**Table A.1.15:** *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 A.10.9). *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 A.1.16:** *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 A.1.17:** *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 |

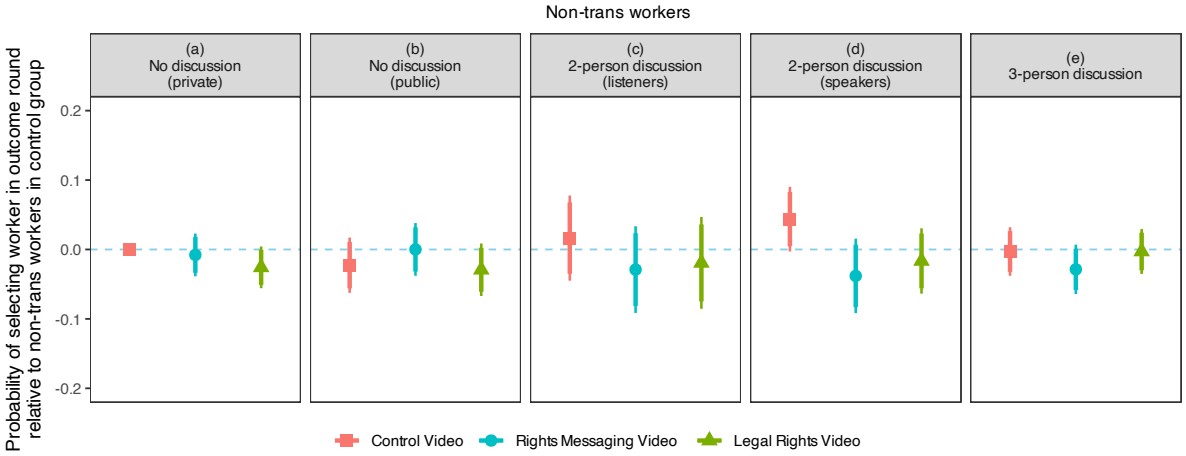
*Notes*: The table measures for spillover effects in the follow-up round 2-9 weeks after the main data collection. *No discussion (private) + Above med. number of discussion groups nearby* is 1 if the participant was in the *No discussion (private)* arm *and* the number of other groups within a certain 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 A.1.18:** *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 × 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 1.3.9 and 1.5. 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 A.10.9).

**Figure A.1.3:** *Placebo test: effect of all treatment arms on the probability of selecting non-transgender alternative workers in outcome round*



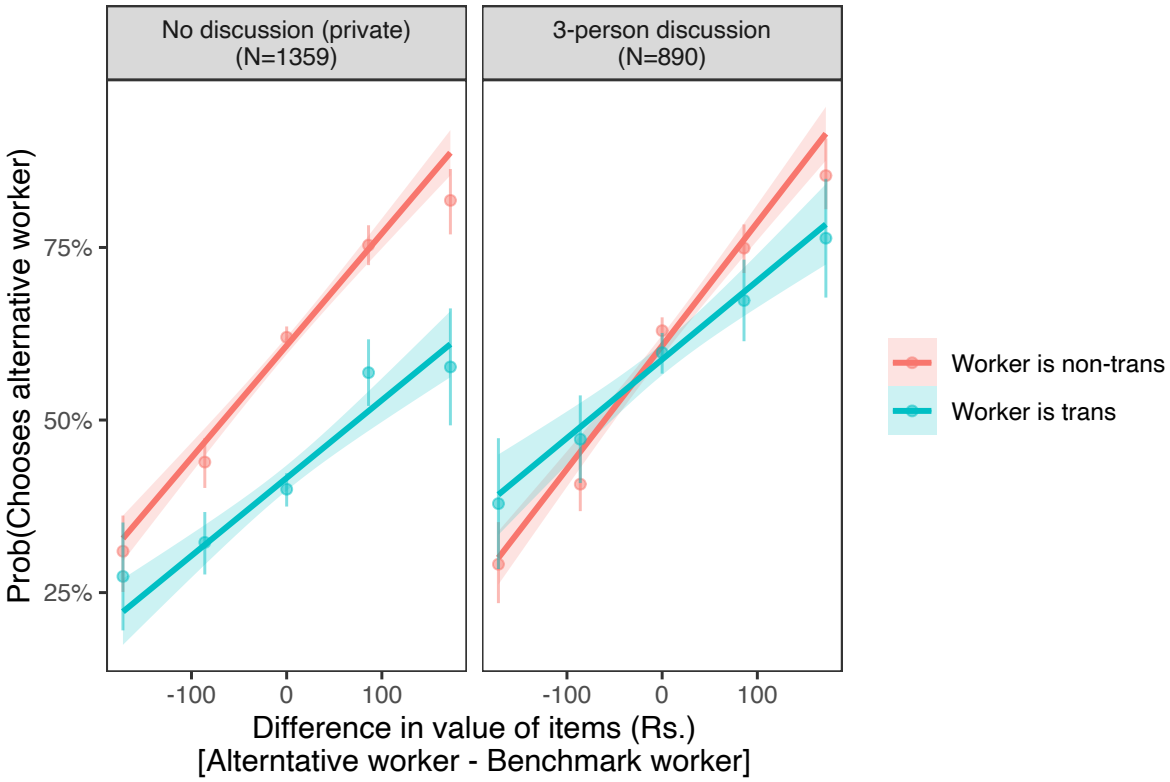

*Notes*: Shows the probability of selecting a non-transgender *alternative* worker in the outcome round, relative to the probability of selecting a non-transgender alternative worker in the *No discussion (private)*, *Control video* arm. 95% confidence intervals are base on standard errors clustered at the group-of-3 level. After adjusting for multiple hypothesis testing using the procedure in Anderson (2008), no q-value is below 0.84. Controls include stratum fixed effects; whether 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 number of items offered; relative reliability score; whether the relative reliability score was shown; and the controls selected by double LASSO (see Section A.10.9). Unit of observation is the participant × choice level.

**Table A.1.19:** *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 |

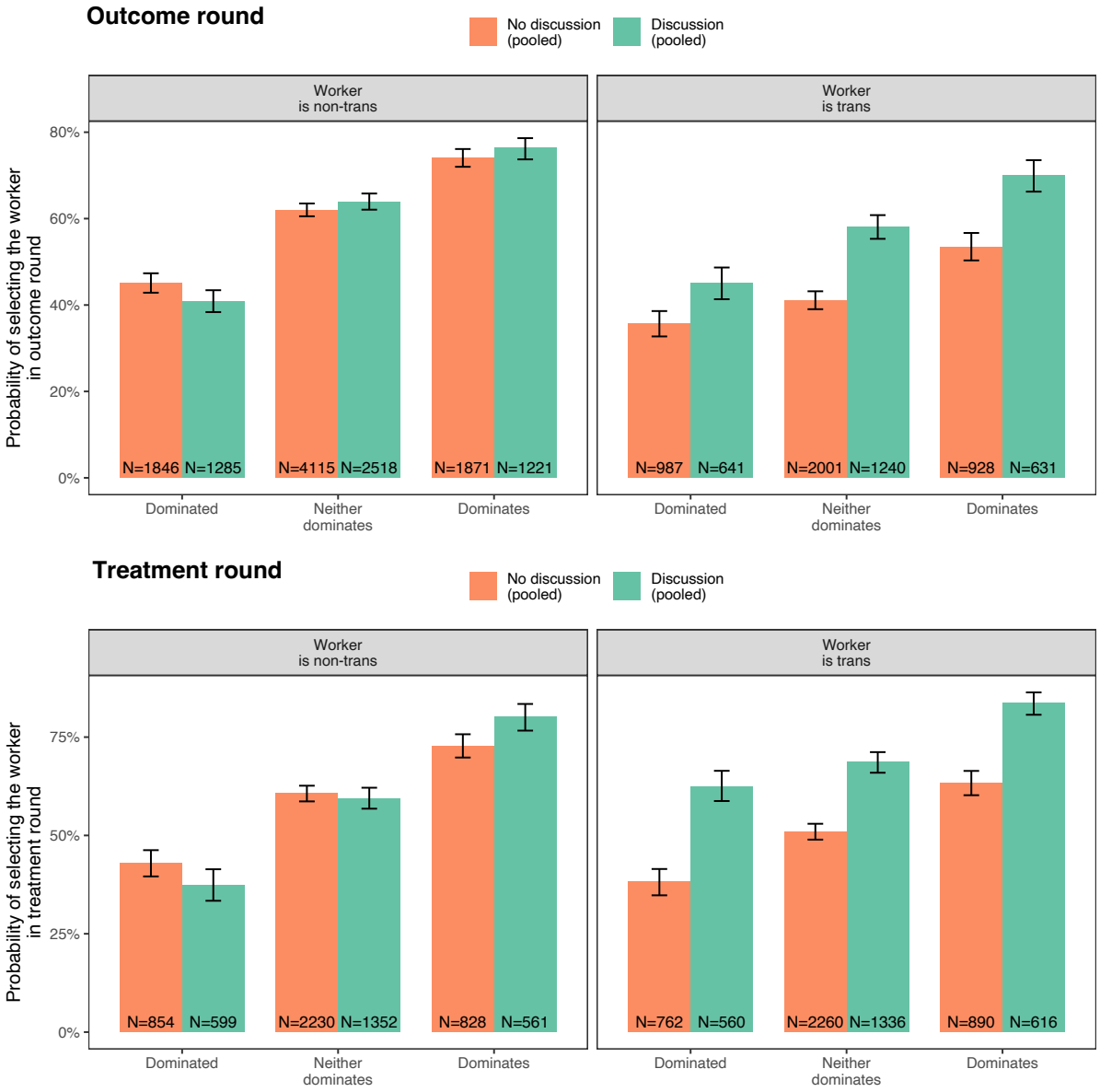
*Notes*: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Sample includes all participants in both phases.
Column (1): The unit of observation is the participant. The dependent variable is the incentivized 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 × 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 A.10.9).

**Figure A.1.4:** *Inferring WTP to avoid transgender workers*



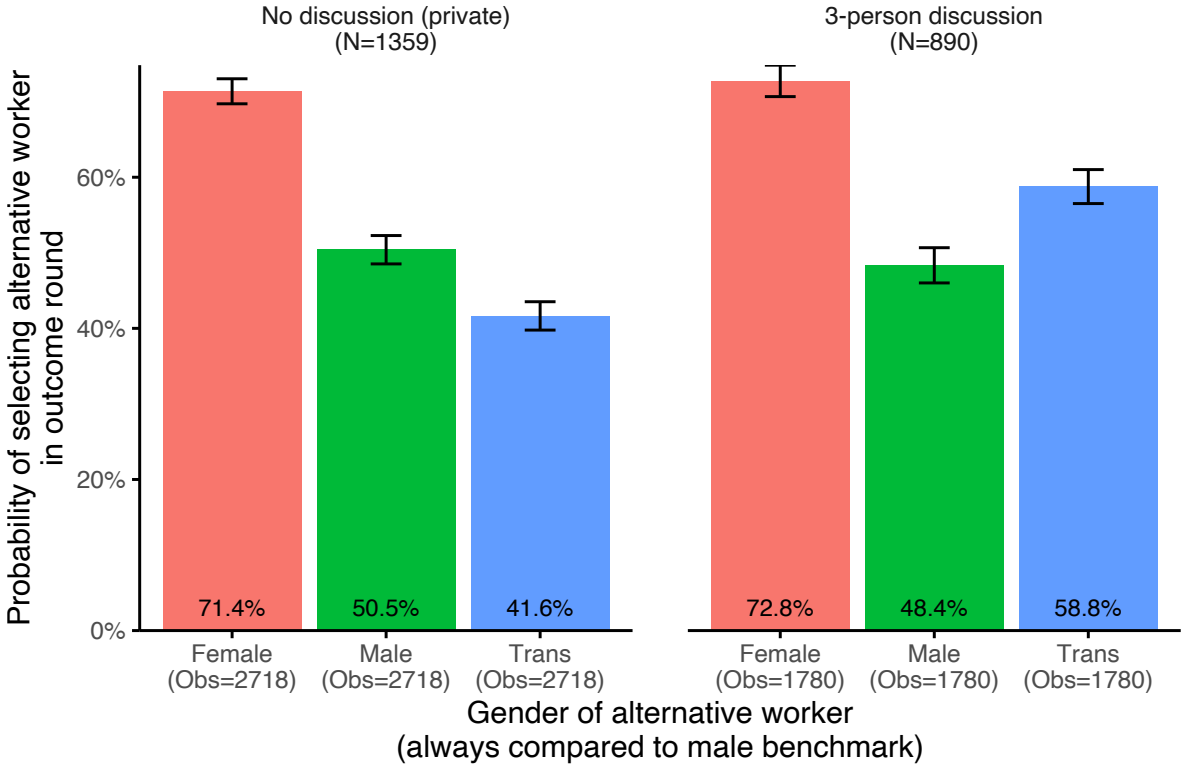

*Notes*: Points represent the probability of choosing the alternative worker at the given difference in value of items in Rs. Solid lines represent a linear fit. I take the reduction in probability that an option is chosen when a worker is transgender in each treatment group, and divide it by the gradient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

**Figure A.1.5:** *Dominated and dominating choices: negative discrimination decreases and positive discrimination increases*



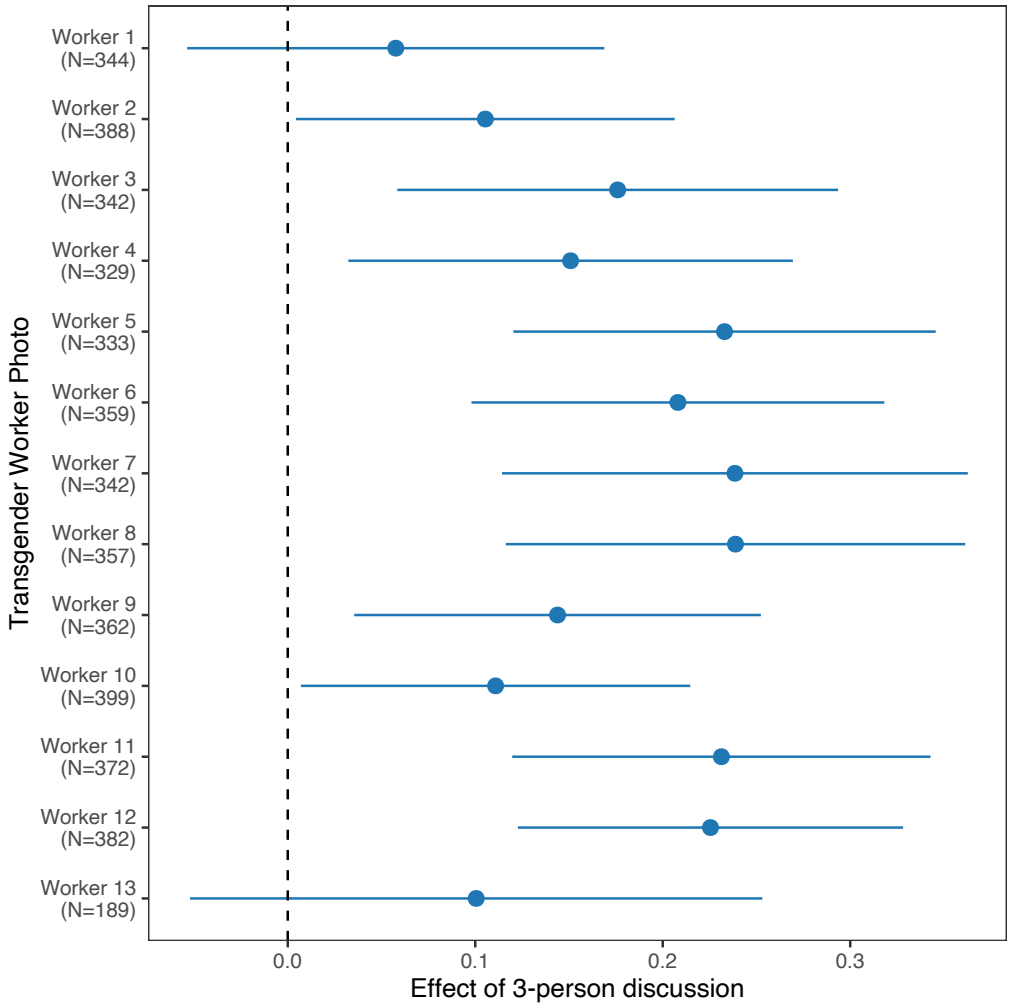

*Notes*: Includes all participants from both phase 1 and 2, apart from listeners. *Discussion (pooled)* = participants in *3-person discussion* arm or speakers in the *2-person discussion* arm. *No discussion (pooled)* = participants in *No discussion (public)* or *No discussion (private)* arm. Unit 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.

**Figure A.1.6:** *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 A.1.7:** *Treatment effect of discussion is positive for all transgender worker photos*




*Notes*: Each plot shows the coefficient from the regression equivalent to Table 1.1, column 3, but restricting the sample to *only* the choices that included the worker photo on the vertical axis.

**Table A.1.20:** *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.

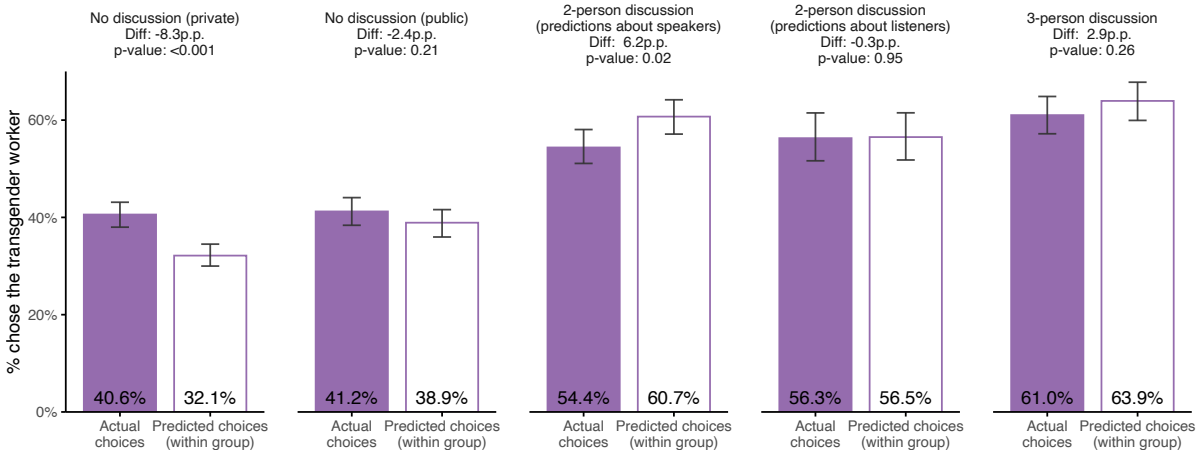
*Column (1)*: dependent variable is the number of statements the participant agreed with on each of the lists of statements shown to them. Each participant sees both List A and List B, and the anti-trans statement ("In general, if I see a transgender person, I walk away") is randomly included in either 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 A.10.9).

**Figure A.1.8:** *Predictions about others in group (Phase 2)*



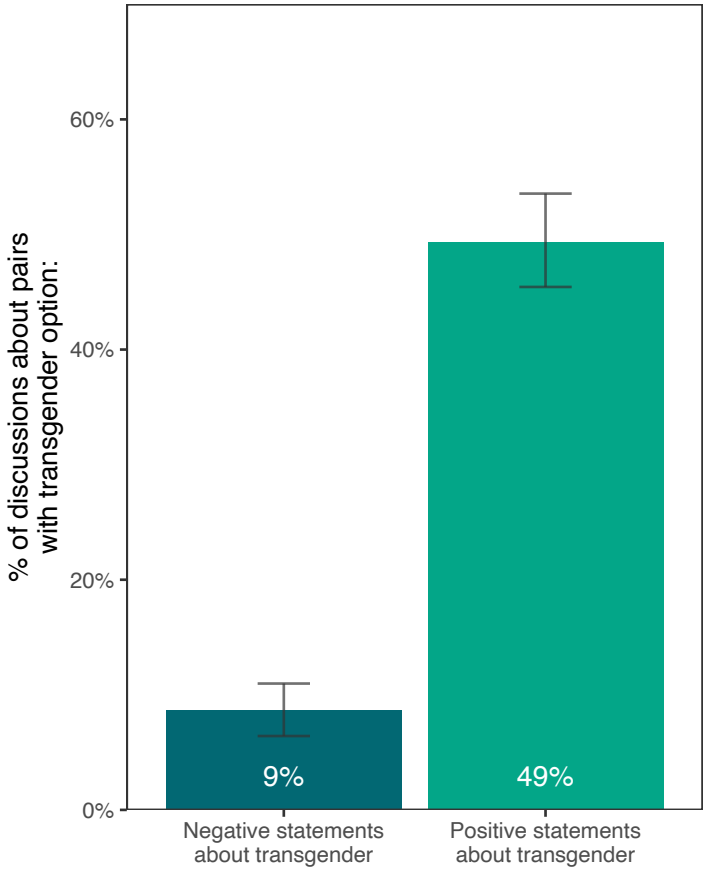

*Notes*: Sample includes all participants in phase 2. Unit of observation is participant × prediction. Only choices that include a transgender photo are included. Hollow bars represent the probability that a participant predicts that their group-member selects a transgender delivery worker. The prediction 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 A.1.21:** *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 | |

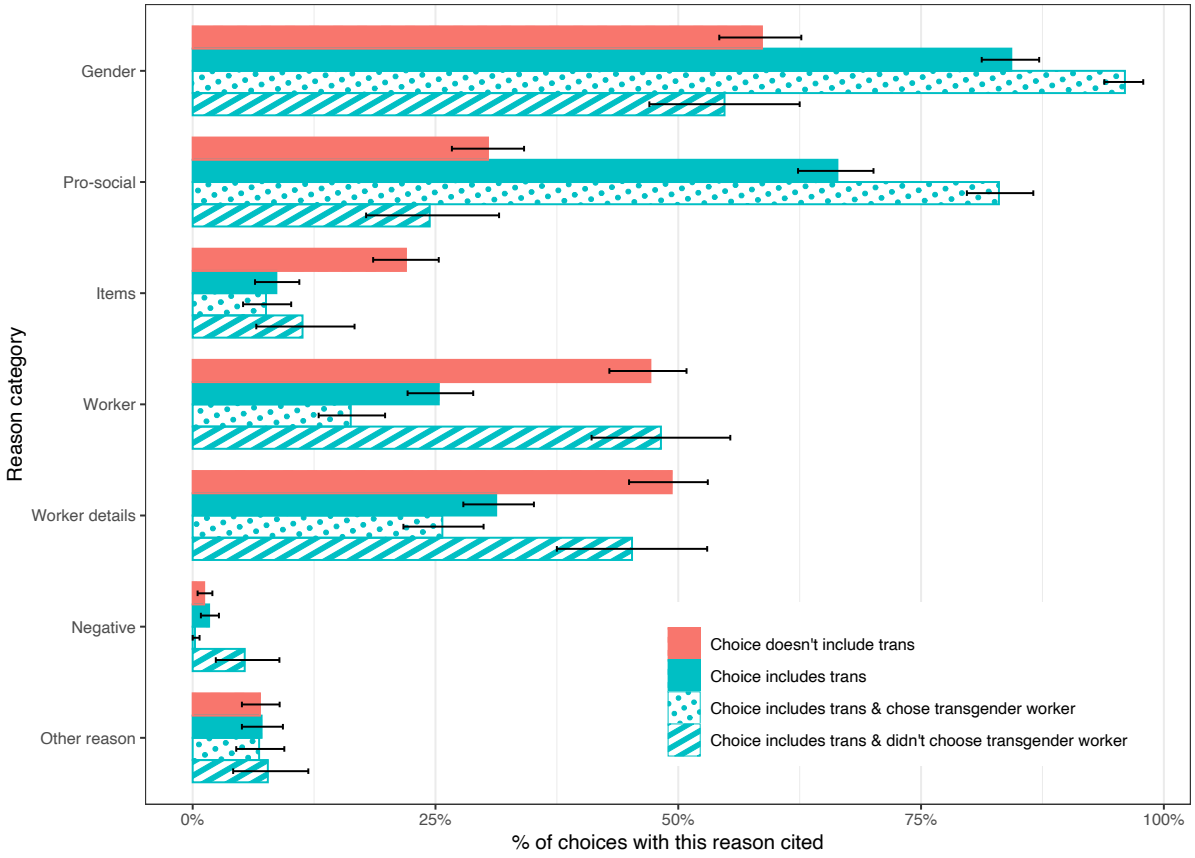
*Notes*: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. The outcomes are based on *treatment round choices*, i.e., during the discussion in the 3-person discussion arm. The specification used is seen in equation 1.1, and is otherwise the same as Tables 1.1 and 1.2.

**Figure A.1.9:** *Participants in the discussion are more likely to say positive statements about transgender workers than negative statements*



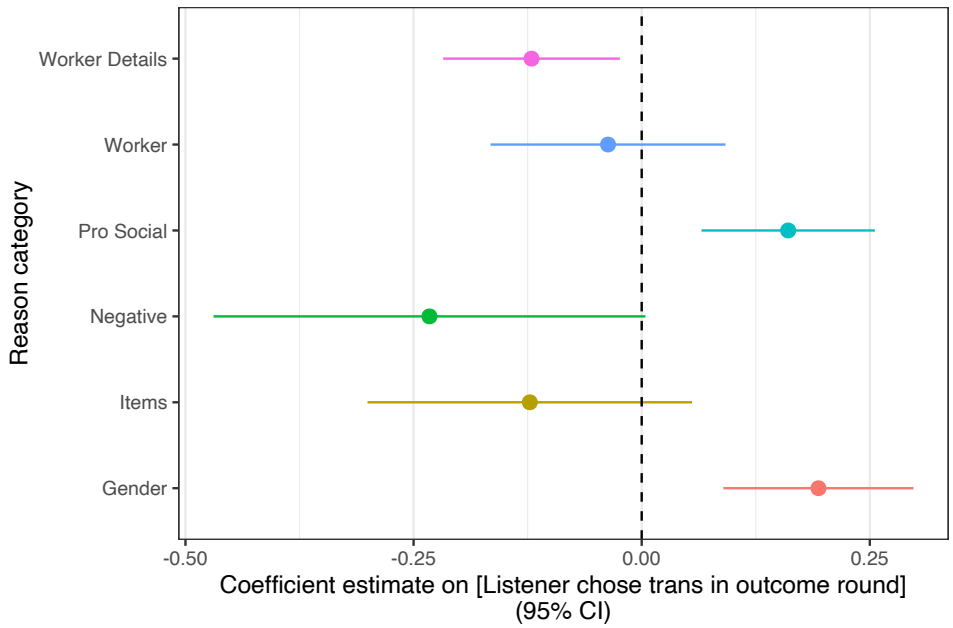

*Notes*: Unit of observation is the participant × choice level. Only choices that include a transgender worker are included. Enumerators coded for each choice whether each participant said a positive statement about the transgender worker, a negative statement about the worker, or both. Participants 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 A.1.10:** *Reasons cited in the 3-person discussions*



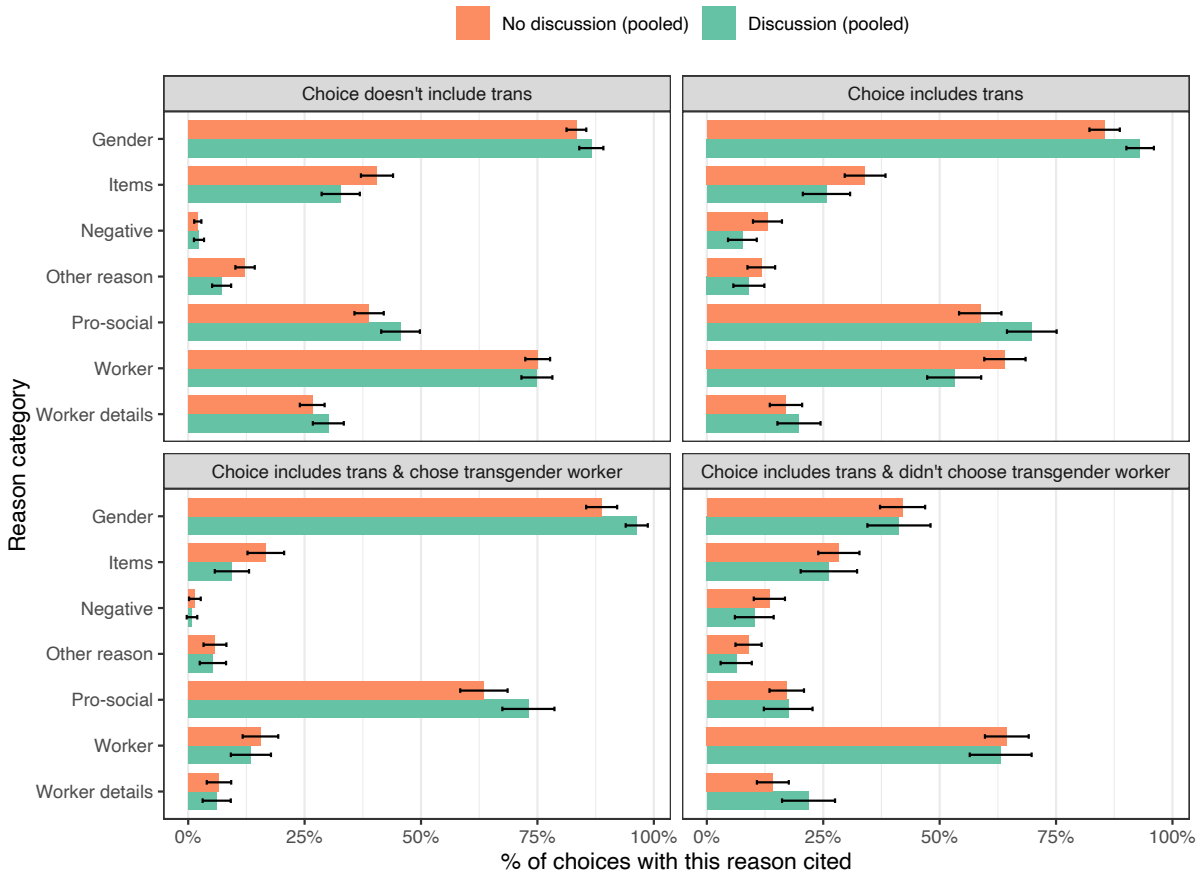

*Notes*: Unit of observation is a group × choice. Sample is the *3-person discussion* arm in both phase 1 and 2. Confidence intervals are based on a bootstrapped binomial distribution. One enumerator observed the discussion and marked the main reasons that the participants said they were selecting the 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 A.1.11:** *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 A.1.10. 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 × choice.

**Figure A.1.12:** *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 A.1.10.

**Table A.1.22:** *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 1.1. 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 A.10.9). 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 A.1.23:** *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 | |

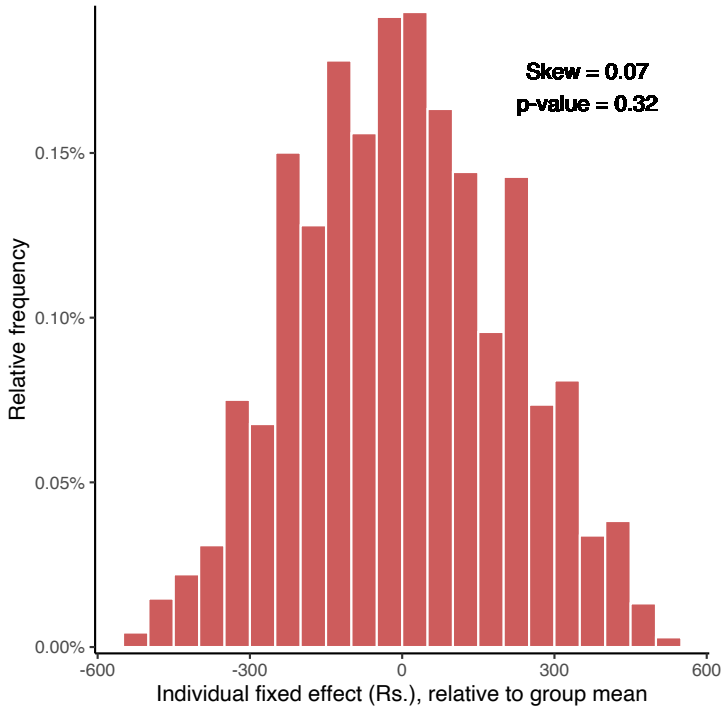
*Notes*: Unit of observation is the participant × choice. The variable of analysis is whether the participant selects a transgender worker, restricting to only choices where a transgender worker is shown. *ICC* denotes the intra-cluster correlation coefficient of this variable. The first two rows show 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 A.1.24:** *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 |

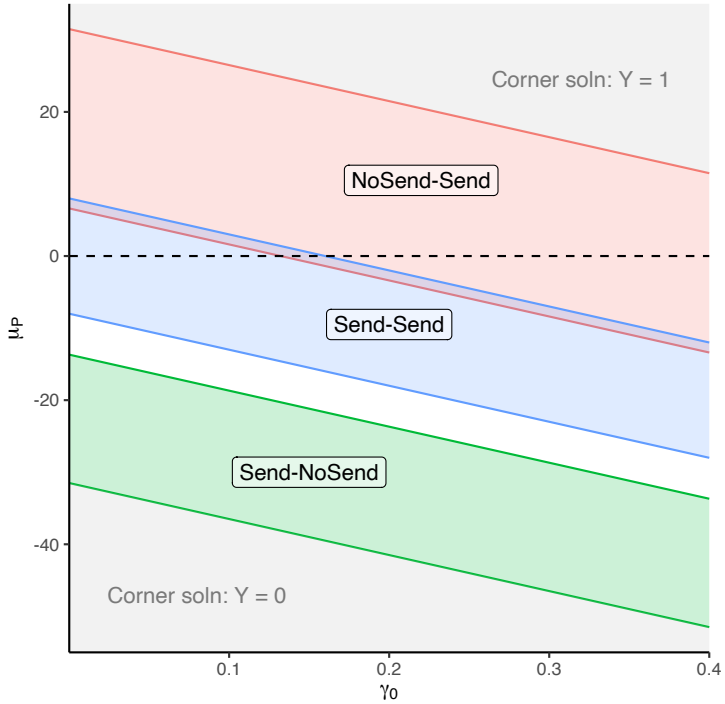
*Notes*: *P(spoke first)* is the probability that a participant spoke first in their group in the discussion of a choice, as marked by enumerator observations. The mean is 33%. *P(dominated)* is the probability that a participant dominated the discussion of a choice, as marked by enumerator observations. The mean is 55%. *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 A.1.13:** *The distribution of preferences for transgender workers within a group in the No discussion (private) arm is symmetric*



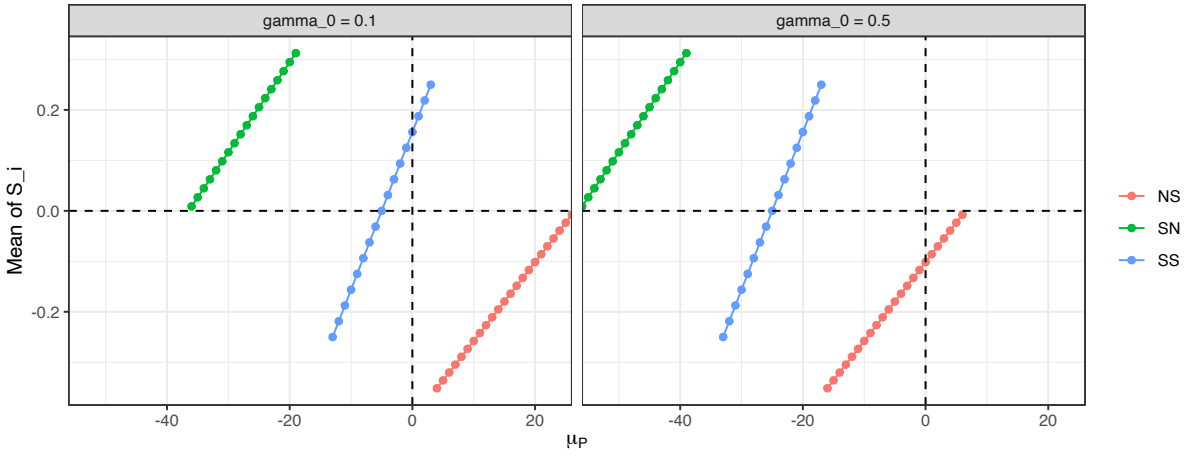

*Notes*: Sample is *No discussion (private)* only. I run a regression of whether the participant selects a transgender worker on a series of individual-specific fixed effects interacted with whether the choice included a transgender worker. The coefficient on these interaction terms gives an indication 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 A.1.14:** *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 $(\mu_P, \gamma_0)$ 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 A.1.15:** *Net persuasion in the model*



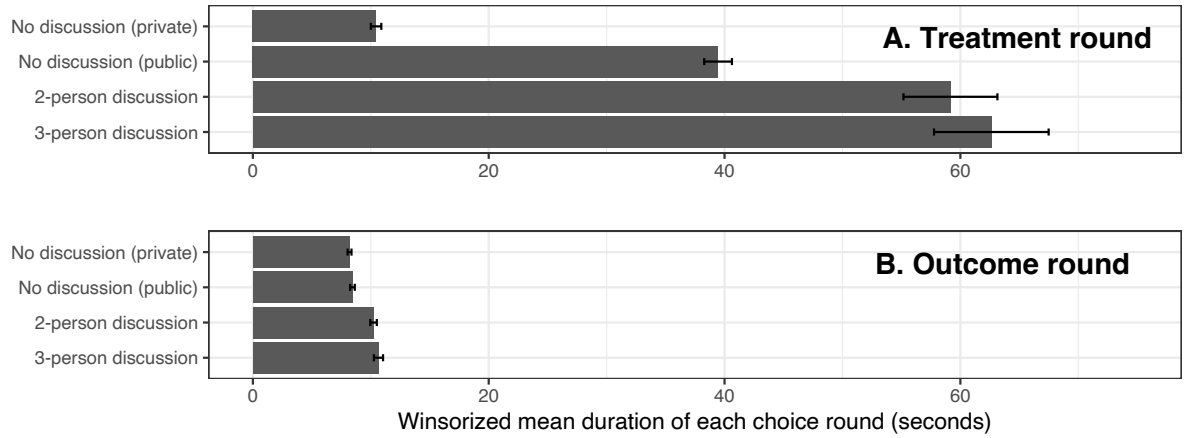

*Notes*: Parameter values: $\alpha = 10$, $\gamma_1 = 0.01$, $R = 50$, $c = 2.5$. Horizontal axis shows the value of $\mu_P$. Vertical axis shows the mean value of $S_i$, which depends on the type of equilibrium $Q \in \{SS, NS, SN\}$ and the proportion of people who choose $Y_i$ in that equilibrium.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

**Table A.1.25:** *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 1.1. 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 A.10.9). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown.

**Figure A.1.16:** *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 A.1.26:** *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 A.10.9). 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.

**Table A.1.27:** *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 A.1.28:** *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 A.1.29:** *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 A.10.4, 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 A.10.9).

**Table A.1.30:** *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 A.10.3. *Column (4)*. Participants were read two lists of words, described in Section A.10.4, 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 A.10.9).

**Table A.1.31:** *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 |

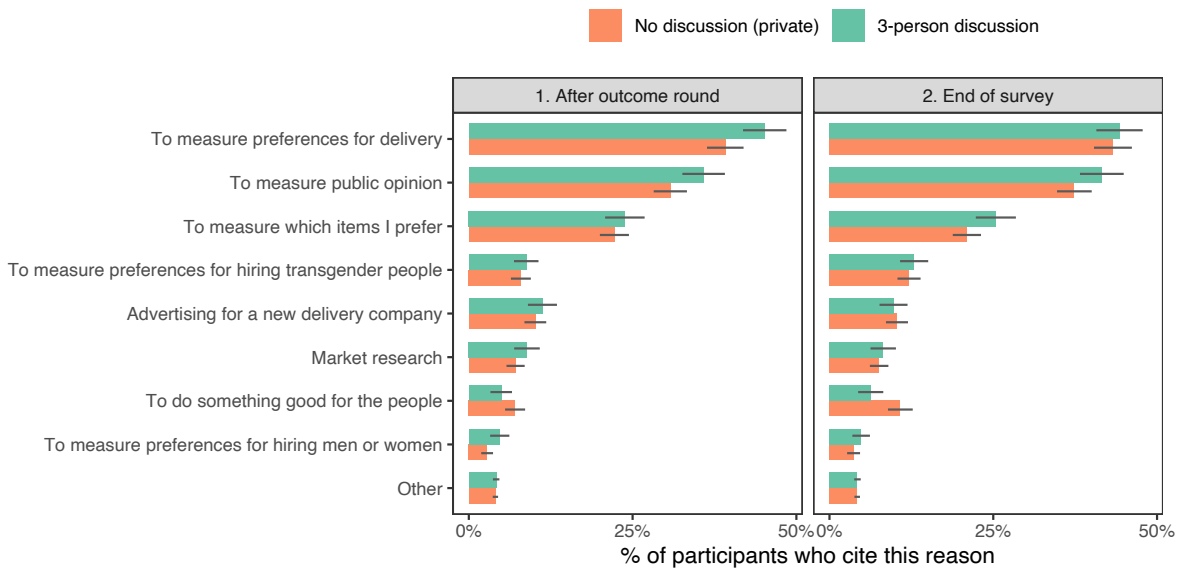
*Notes*: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets, and use randomization inference for the rights video coefficients. Unit of observation is the participant level. Sample includes participants from all 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 A.10.4, 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 A.10.9).

**Figure A.1.17:** *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 A.1.32:** *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 A.10.3. *Column (4)*. Participants were read two lists of words, described in Section A.10.4, 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 A.10.9).

**Table A.1.33:** *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 | |

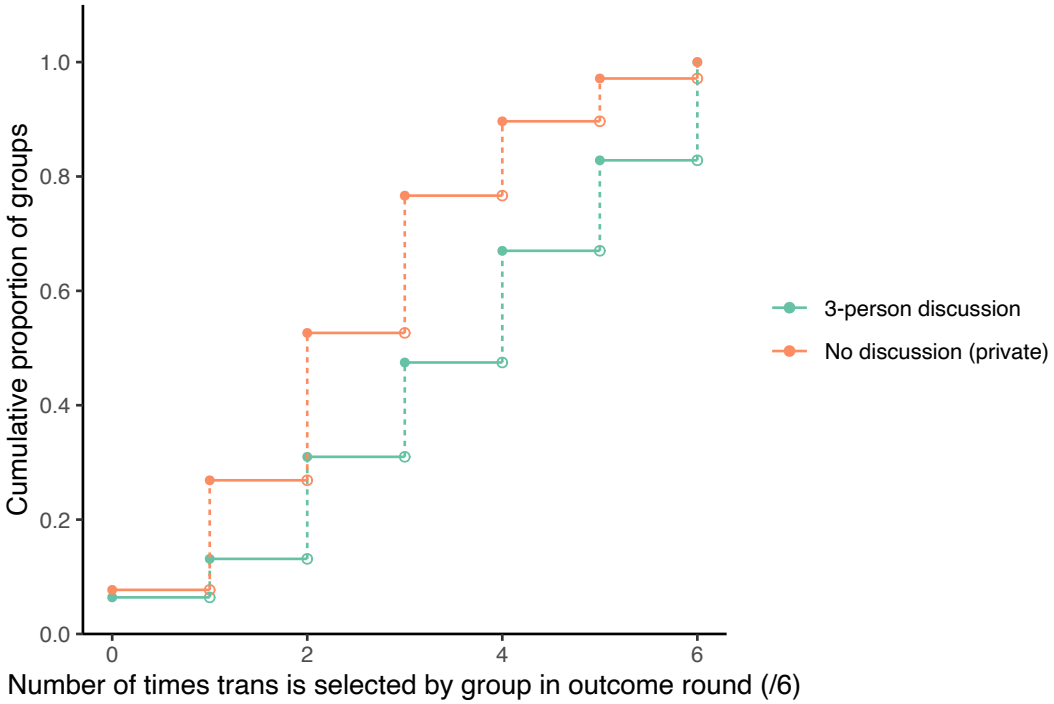
*Notes*: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. randomization inference p-values are in brackets. Unit of observation is the participant × choice level. Sample includes only the subsample of 582 individuals in phase 1 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 1.1. 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 A.10.9). 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 A.1.34:** *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 1.1. 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 A.10.9). 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 A.1.18:** *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 A.1.35:** *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) |
| $\pi_{-i}$ = 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] | | |
| $\pi_{-i}$ × 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] |
| $\pi_{-i}$ × 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)* ($\pi_{-i}$) 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 A.1.36:** *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) |
| $\pi_{-i}$ = 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] |
| $\pi_{-i}$ × 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] |
| $\pi_{-i}$ × 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: $\pi_{-i}$ x Listener = $\pi_{-i}$ 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)* ($\pi_{-i}$) 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 A.10.9).

**Table A.1.37:** *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 1.1. 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 A.10.9). 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 A.1.38:** *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 A.10.8 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 A.10.6 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.

## A.2 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:

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

- $\Delta V_{ijk}$ is observed.
- Assume that the preference shock term $\Delta u_{ijk}$ 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: $[\mu(Discuss), \mu(Control), \sigma_D^2(Discuss), \sigma_D^2(Control), \sigma_d^2, \sigma_\kappa^2, \sigma_\eta^2, \sigma_\varepsilon^2, \alpha]$

### A.2.1 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.

*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 $\Delta V_{ijk}$, 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.

*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):

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

where $\pi_{.1}$ denotes the marginal probability of $Y_A = Y_B = 1$, and $\pi_{11}$ 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 $\pi_{11}$ 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 A.1, the correlation between the outcome variables is:

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

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

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

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.

### A.2.2 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 $\theta$ 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 $\Delta V_{ijk}$ 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) $\sigma_\kappa^2$ is identified by the within-group correlation for non-transgender choices; (ii) $\sigma_D^2$ is identified by comparing the latter with the within-group correlation for transgender choices; (iii) $\sigma_\eta^2$ is identified by the additional within-individual correlation for non-transgender choices; (iv) $\sigma_d^2$ is identified by the additional within-individual correlation for transgender choices; (v) $\sigma_\varepsilon^2$ is the remaining unexplained variance.

### A.2.3 Results

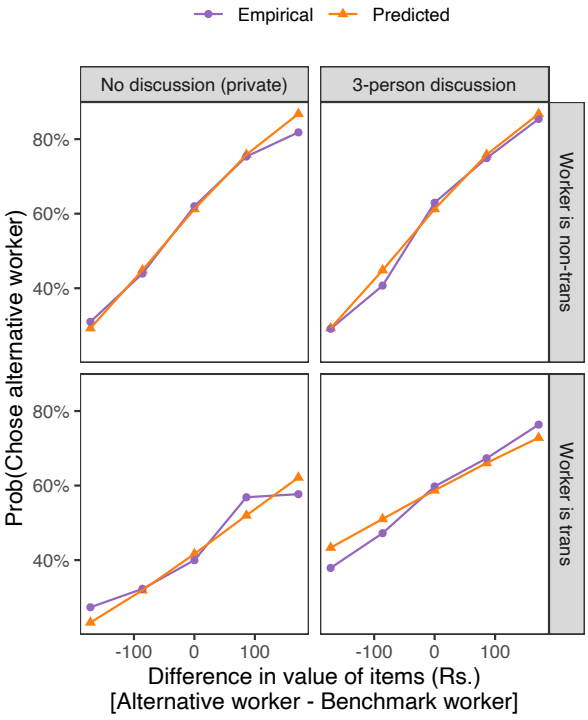
Table A.2.1 shows the estimated parameters. In the control group, I estimate a large distaste for transgender people of 129 Rs. In the 3-person discussion group, this distaste is estimated to be slightly negative at -38 Rs, implying a small positive preference for selecting transgender workers. Figure A.2.1 shows that the model performs well at predicting the empirical probabilities of selecting a transgender worker in each cell.

**Table A.2.1:** *Structural estimates*

| Parameter | Value |
|---|---|
| $\mu(Discuss)$ | -38.4 |
| $\mu(Control)$ | 128.7 |
| $\sigma_D^2(Discuss)$ | 123.8 |
| $\sigma_D^2(Control)$ | 236.4 |
| $\sigma_d^2$ | 0.0 |
| $\sigma_\varepsilon^2$ | 85.8 |
| $\sigma_\kappa^2$ | 56.5 |
| $\sigma_\eta^2$ | 64.7 |
| $\alpha$ | 59.1 |

*Notes*: Results from the structural estimation of the model in Section A.2.

**Figure A.2.1:** *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 $\Delta V_{ijk} \in \{-172, -86, 0, 86, 176\}$.

## A.3 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).[1] 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 A.1.11 and Figure A.1.5 show that the interactions between the main treatment effect on (Worker is trans × 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 A.1.11, column 2). The increase in positive discrimination is particularly strong, however: the interaction term (Worker is trans × Discussion (pooled) × 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 A.1.5 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 A.1.23). However, this does not translate to more extreme behavior between groups. In Figure A.1.18, 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 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.

[1] See the notes of Table A.1.11 for the precise definitions used.

## A.4 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.

## A.5 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.

## A.6 Proofs

### A.6.1 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 1.2, 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 $\sigma_{-i}$ by $Q$, since $\sigma_{-i}$ is fixed by the equilibrium type $Q$.

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

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

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.

*Expected $P_{-i}$*

Let $\pi_1$ 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**:

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

*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.:

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

*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 A.2 yields the full equilibrium conditions for each type of equilibrium. These can be rearranged to derive the value of $P_Q^*$ for each $Q$:

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

*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 $\mu_P \leq \eta_{Y=1}^{SS}$ where:

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

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 A.3 and the expectations from Section A.6.1 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 $\kappa^{SS}_{Y=1,S\neq 0} < \kappa^{SS}_{Y=1,S\neq -1}$, so only the constraint based on $\kappa^{SS}_{Y=1,S\neq 0}$ 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 $\mu_P \geq \kappa^{SS}_{Y=0,S\neq 0}$ where:

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

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

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

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 $\kappa^{SS}_{Y=0,S\neq 0} > \kappa^{SS}_{Y=0,S\neq 1}$, so only the constraint based on $\kappa^{SS}_{Y=0,S\neq 0}$ is binding.

A $SS$ equilibrium is feasible when there is some range of values $\mu_P$ 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:

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

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

The $SS$ equilibrium is therefore feasible when:

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

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 $\mu_P$, i.e., $\mu_P \in [\kappa^{SS}_{Y=0,S\neq 0}, \kappa^{SS}_{Y=1,S\neq 0}]$, 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:

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

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 $\alpha$ and $\gamma_1$, 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 $\mu_P$ 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 $\alpha$ and $\gamma_1$, 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.

### A.6.2 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 $\mu_P$, 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 $\mu_P$ 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:

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

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

$$\mathscr{M}^+(p) := E[P_i | P_i > p]$$
$$\mathscr{M}^-(p) := E[P_i | P_i < p]$$

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:

$$\mathscr{M}^+(p) = \frac{\mu_P + R + p}{2};$$
$$\mathscr{M}^-(p) = \frac{\mu_P - R + p}{2}$$

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 $\mu_P = \tilde{\mu}_P$.

## A.7 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.

### A.7.1 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.

### A.7.2 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.

### A.7.3 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.

### A.7.4 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 1.4 and Table A.1.10.
- *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 A.1.22). However, for reader clarity in the main text I did not pool any treatment arms when presenting the phase 2 outcome round results (Table 1.5). The corresponding pooled results are presented in Table A.1.37.
- *Heterogeneity with respect to round 1 observations*. Table A.1.35 shows the heterogeneous effects of *observers*'s choices with respect to the round 1 choices they observed. Table A.1.36 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 A.1.38 shows heterogeneous effects of the discussion with respect to persuasiveness and group relations. I find no detectable heterogeneity.

## A.8 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 A.8.1).

## A.8.1 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.]

## A.9 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.

## A.10 Data and measurement

### A.10.1 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.

### A.10.2 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.

### A.10.3 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 A.1.30.

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).

### A.10.4 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.

### A.10.5 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

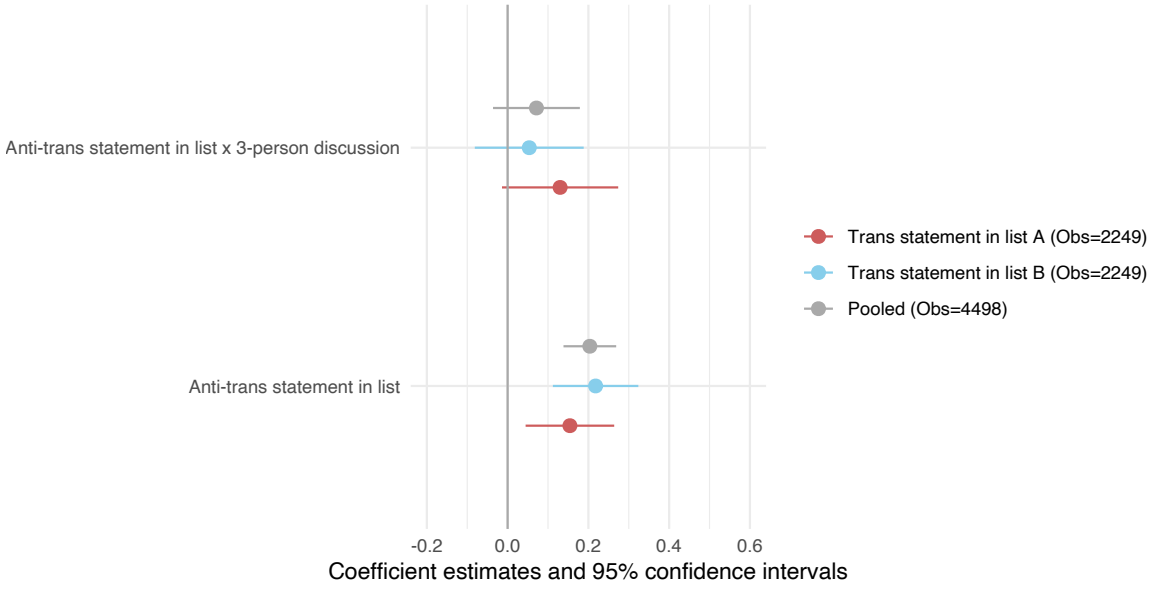
5 non-sensitive statements, shown in Appendix Table A.10.1. For each participant, either list A or list B is randomly selected to include one additional statement: "In general, if I see a transgender, I walk away." Enumerators read out each list and asked the participant how many statements they agreed 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 (E. 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 A.10.1 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 A.10.1:** *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 A.10.1:** *Treatment effect on list experiment does not depend on which list is used to estimate it*




*Notes*:

### A.10.6 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 A.10.2. 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 A.10.2:** *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 |

### A.10.7 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.[2] 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

[2]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.

select one of their choices to organize the pickup with. 2 of the 4 pairs included a transgender worker.

The enumerator giving the interview did not know what responses were given. We did not ask the respondent for their choice verbally, as in the main hiring rounds. Instead, we gave the tablet directly to the respondent, and they clicked their preferred answer. Upon clicking, the tablet would automatically skip to the next question and not reveal again the answer chosen before, making it impossible for the enumerator to know what was selected. We truthfully told respondents that enumerators wouldn't know what was selected, making the answers anonymous.[3] 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.

### A.10.8 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 A.10.3.
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.

[3] 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.

### A.10.9 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 A.10.3. In addition:

- 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 A.10.4.

**Table A.10.3:** *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 A.10.4:** *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.

# Appendix B

# Appendix for Chapter 2

## B.1 Supplementary tables and figures

**Table B.1.1:** *Implementation methods used by YGLs*

| | **Base + YGLs Informal** | **Base + YGLs Formal** |
|---|---|---|
| ***Informal implementation*** | 52% | 24% |
| Outside the classroom | 23% | 9% |
| Without specific slots | 20% | 10% |
| Unannounced | 7% | 4% |
| Without adult help | 2% | 1% |
| ***Formal implementation*** | 48% | 76% |
| With the help of teachers | 11% | 18% |
| In a classroom | 15% | 26% |
| With specific time slots | 22% | 32% |
| N | 35 | 35 |

*Notes:* The table shows results from the school survey. Directors were asked about the way the YGLs implemented their sensitization. Each row shows the proportion of each response. Column 1 includes the 35 *Base + YGL* schools where YGLs were asked to carry out informal discussions with their peers. Column 2 includes the 35 *Base + YGL* schools where YGLs were asked to carry out formal sensitization sessions.

**Table B.1.2:** *Effects of formal vs informal YGL program*

| *Outcome* | (1) Base + YGL (Informal) | (2) Base + YGL (Formal) | (3) *p*-val. of diff. | (4) N |
|---|---|---|---|---|
| Learning test (Z) | 0.102* | 0.124** | .7093 | 2162 |
| | (0.060) [0.089] | (0.063) [0.048] | | |
| Progressed one class (=1) | 0.101** | 0.036 | .0459 | 2251 |
| | (0.039) [0.011] | (0.037) [0.334] | | |
| Enrolled (=1) | –0.013 | –0.033 | .4373 | 2251 |
| | (0.026) [0.628] | (0.024) [0.172] | | |
| Absent in last 3 months (=1) | –0.037 | –0.032 | .8688 | 1820 |
| | (0.032) [0.252] | (0.033) [0.343] | | |
| General health (Z) | 0.020 | 0.008 | .8387 | 2251 |
| | (0.053) [0.709] | (0.059) [0.885] | | |
| UTI health (Z) | –0.043 | 0.044 | .1129 | 2251 |
| | (0.053) [0.424] | (0.048) [0.358] | | |
| Mental health (Z) | 0.036 | 0.076 | .4976 | 2251 |
| | (0.067) [0.593] | (0.068) [0.261] | | |
| Self-esteem (Z) | 0.050 | –0.020 | .2498 | 2251 |
| | (0.060) [0.405] | (0.062) [0.739] | | |
| Heart rate (Z) | –0.178*** | –0.076 | .0648 | 1899 |
| | (0.056) [0.002] | (0.058) [0.193] | | |
| Hygiene knowl. (Z) | 0.401*** | 0.470*** | .2511 | 2162 |
| | (0.063) [0.000] | (0.075) [0.000] | | |
| Menstr. knowl. (Z) | 0.353*** | 0.334*** | .7079 | 2162 |
| | (0.053) [0.000] | (0.057) [0.000] | | |
| Hygiene behavior (Z) | 0.368*** | 0.267*** | .1799 | 2162 |
| | (0.080) [0.000] | (0.083) [0.001] | | |
| Menstr. behavior (Z) | 0.566*** | 0.550*** | .8120 | 2251 |
| | (0.065) [0.000] | (0.063) [0.000] | | |
| Observed hygiene: Girls (Z) | 0.088 | 0.015 | .2156 | 2162 |
| | (0.064) [0.166] | (0.061) [0.808] | | |
| Combined stigma index (Z) | 0.672*** | 0.785*** | .0821 | 2251 |
| | (0.062) [0.000] | (0.066) [0.000] | | |
| Received info. (Z) | 0.494*** | 0.560*** | .3799 | 2251 |
| | (0.077) [0.000] | (0.071) [0.000] | | |
| Willingness to speak (Z) | 0.605*** | 0.832*** | .0025 | 2245 |
| | (0.072) [0.000] | (0.076) [0.000] | | |
| Attitudes (Z) | 0.531*** | 0.586*** | .4573 | 2249 |
| | (0.064) [0.000] | (0.073) [0.000] | | |
| Norms (Z) | 0.425*** | 0.510*** | .1296 | 2162 |
| | (0.053) [0.000] | (0.059) [0.000] | | |
| Shame response to vignettes (rev.) (Z) | 0.430*** | 0.449*** | .7617 | 2160 |
| | (0.060) [0.000] | (0.065) [0.000] | | |
| Observed shame (rev.) (Z) | 0.235*** | 0.206*** | .5979 | 2251 |
| | (0.060) [0.000] | (0.059) [0.000] | | |
| Bullying: teasing (Z) | 0.075 | –0.050 | .0438 | 2162 |
| | (0.059) [0.205] | (0.071) [0.486] | | |
| Bullying: intimidation/harassment (Z) | –0.025 | –0.119*** | .0345 | 2162 |
| | (0.045) [0.569] | (0.046) [0.010] | | |
| Network (Z) | 0.268*** | 0.250*** | .8311 | 2162 |
| | (0.083) [0.001] | (0.088) [0.004] | | |
| Official mark (Z) | 0.058 | 0.109 | .4425 | 14579 |
| | (0.075) [0.436] | (0.068) [0.111] | | |

*Notes:* The table compares the effects of the *YGL* informal and formal modalities using girls' outcomes at endline. Each row represents a regression of the outcome on *Base Only*, *Base + YGL (Informal)* and *Base + YGL (Formal)*. Column (1) shows the coefficient on *Base + YGL (Informal)*; column (2) shows the coefficient on *Base + YGL (Formal)*; column (3) shows the *p*-value for the test of *Base + YGL (Informal)* = textitBase + YGL (Formal). Standard errors are clustered at the school level and are in parentheses; coefficient p-values are in brackets; controls include stratum fixed effects and the baseline controls selected by double LASSO. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

**Table B.1.3:** *Baseline balance: demographics*

| Variable | (1) Total Mean/(SD) | (2) Control Mean/(SD) | (3) Base + YGL Mean/(SD) | (4) Base Only Mean/(SD) | (3)-(2) Pairwise t-test P-value | (4)-(2) Pairwise t-test P-value |
|---|---|---|---|---|---|---|
| Girl's age (years) | 14.000 | 13.741 | 14.104 | 14.033 | 0.292 | 0.483 |
| | (6.798) | (6.904) | (6.535) | (7.356) | | |
| Girl already had her period (=1) | 0.471 | 0.447 | 0.483 | 0.468 | 0.606 | 0.807 |
| | (1.387) | (1.392) | (1.340) | (1.509) | | |
| Girl has a brother (=1) | 0.548 | 0.574 | 0.556 | 0.508 | 0.512 | 0.033** |
| | (0.548) | (0.531) | (0.567) | (0.511) | | |
| Distance to school (km) | 2.320 | 2.260 | 2.379 | 2.259 | 0.518 | 0.996 |
| | (3.521) | (3.546) | (3.732) | (3.094) | | |
| Time to school (min) | 30.167 | 29.722 | 30.755 | 29.404 | 0.637 | 0.896 |
| | (40.650) | (44.587) | (40.191) | (38.511) | | |
| Was absent last 3 months (=1) | 0.552 | 0.546 | 0.573 | 0.518 | 0.497 | 0.517 |
| | (0.798) | (0.739) | (0.858) | (0.718) | | |
| Days absent last 3 months | 1.412 | 1.391 | 1.322 | 1.613 | 0.671 | 0.289 |
| | (3.633) | (2.988) | (3.637) | (4.106) | | |
| Father passed away (=1) | 0.085 | 0.077 | 0.083 | 0.096 | 0.687 | 0.256 |
| | (0.319) | (0.225) | (0.357) | (0.319) | | |
| Father in household (=1) | 0.696 | 0.711 | 0.696 | 0.682 | 0.592 | 0.401 |
| | (0.604) | (0.556) | (0.605) | (0.657) | | |
| Mother passed away (=1) | 0.049 | 0.040 | 0.050 | 0.053 | 0.331 | 0.282 |
| | (0.210) | (0.183) | (0.223) | (0.209) | | |
| Mother in household (=1) | 0.816 | 0.835 | 0.805 | 0.820 | 0.208 | 0.595 |
| | (0.450) | (0.469) | (0.439) | (0.459) | | |
| Guardian knows how to read and write (=1) | 0.828 | 0.829 | 0.836 | 0.810 | 0.799 | 0.474 |
| | (0.485) | (0.479) | (0.514) | (0.428) | | |
| Guardian has no education (=1) | 0.095 | 0.085 | 0.096 | 0.102 | 0.465 | 0.310 |
| | (0.365) | (0.255) | (0.418) | (0.342) | | |
| Guardian attended secondary or higher education (=1) | 0.303 | 0.308 | 0.293 | 0.318 | 0.648 | 0.794 |
| | (0.648) | (0.681) | (0.623) | (0.685) | | |
| Guardian occupation: Agriculture (=1) | 0.773 | 0.768 | 0.781 | 0.762 | 0.696 | 0.899 |
| | (0.666) | (0.716) | (0.632) | (0.697) | | |
| Observations | 2393 | 568 | 1787 | 606 | | |
| Schools | 140 | 35 | 70 | 35 | | |

*Notes:* Sample includes all girls in baseline. Columns (1)-(4) show means and standard deviations of demographic outcomes from the girls' baseline survey. Columns (5)-(6) show the p-value of a pairwise test comparing *Base Only* and *Base + YGL* with *control* , respectively. Standard errors cluster at the school level. *$* p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

**Table B.1.4:** *Baseline balance: outcome indexes*

| Variable | (1) Total Mean/(SD) | (2) Control Mean/(SD) | (3) Base + YGL Mean/(SD) | (4) Base Only Mean/(SD) | (3)-(2) Pairwise t-test P-value | (4)-(2) Pairwise t-test P-value |
|---|---|---|---|---|---|---|
| Wealth index (Z) | 0.000 | –0.136 | 0.083 | –0.039 | 0.208 | 0.632 |
| | (3.487) | (3.397) | (3.480) | (3.578) | | |
| Years of school delay | 2.224 | 2.160 | 2.233 | 2.264 | 0.628 | 0.598 |
| | (3.278) | (2.881) | (3.191) | (3.830) | | |
| WASH knowledge index (Z) | 0.093 | 0.000 | 0.107 | 0.151 | 0.317 | 0.206 |
| | (1.988) | (2.178) | (1.960) | (1.883) | | |
| WASH behaviour index (Z) | 0.008 | –0.000 | –0.011 | 0.054 | 0.906 | 0.606 |
| | (1.678) | (1.901) | (1.596) | (1.658) | | |
| Menstruation knowledge index (Z) | 0.047 | –0.000 | 0.099 | –0.014 | 0.504 | 0.937 |
| | (2.999) | (2.890) | (2.978) | (3.195) | | |
| Menstruation behaviour index (Z) | 0.070 | –0.000 | 0.094 | 0.086 | 0.475 | 0.591 |
| | (2.700) | (2.555) | (2.677) | (2.937) | | |
| Mental health index (Z) | 0.020 | –0.000 | 0.061 | –0.042 | 0.412 | 0.604 |
| | (1.373) | (1.488) | (1.383) | (1.234) | | |
| Menstrual health problems index (Z) | 0.018 | 0.000 | 0.042 | –0.013 | 0.699 | 0.919 |
| | (2.110) | (2.187) | (2.055) | (2.203) | | |
| Menstruation inhibiting school activity index (Z) | 0.032 | 0.000 | 0.042 | 0.041 | 0.316 | 0.330 |
| | (0.489) | (1.006) | (0.021) | (0.017) | | |
| Received info. on hygiene (Z) | 0.037 | –0.000 | 0.055 | 0.036 | 0.455 | 0.687 |
| | (1.311) | (1.604) | (1.107) | (1.418) | | |
| Received info. on menstruation (Z) | 0.021 | –0.000 | 0.066 | –0.049 | 0.520 | 0.694 |
| | (2.096) | (1.999) | (2.077) | (2.243) | | |
| Willingness to speak about hygiene (Z) | –0.006 | –0.000 | –0.011 | –0.001 | 0.899 | 0.989 |
| | (1.701) | (1.752) | (1.776) | (1.540) | | |
| Willingness to speak about menstruation (Z) | 0.040 | –0.000 | 0.058 | 0.043 | 0.415 | 0.648 |
| | (1.571) | (1.344) | (1.527) | (1.857) | | |
| Exposure to bullying (self-reported) (Z) | 0.032 | 0.000 | 0.021 | 0.082 | 0.757 | 0.357 |
| | (1.491) | (1.308) | (1.453) | (1.729) | | |
| Attitudes towards menstruation (Z) | –0.072 | 0.000 | –0.134 | –0.014 | 0.213 | 0.907 |
| | (2.057) | (2.163) | (1.997) | (2.078) | | |
| Shame response to vignettes (Z) | 0.020 | –0.000 | 0.015 | 0.048 | 0.880 | 0.692 |
| | (2.096) | (1.898) | (2.107) | (2.292) | | |
| Perceived norms related to menstruation (Z) | 0.076 | –0.000 | 0.089 | 0.120 | 0.277 | 0.222 |
| | (1.748) | (1.516) | (1.813) | (1.841) | | |
| Social desirability index (Z) | 0.006 | 0.000 | 0.037 | –0.051 | 0.538 | 0.456 |
| | (1.165) | (1.218) | (1.166) | (1.105) | | |
| Observed hygiene: home (Z) | –0.024 | 0.000 | –0.065 | 0.035 | 0.407 | 0.721 |
| | (1.606) | (1.571) | (1.515) | (1.789) | | |
| Observations | 2393 | 568 | 1787 | 606 | | |
| Schools | 140 | 35 | 70 | 35 | | |

*Notes:* Sample includes all girls in baseline. Columns (1)-(4) show means and standard deviations of indexes calculated using the girls' baseline survey. Columns (5)-(6) show the p-value of a pairwise test comparing *Base Only* and *Base + YGL* with *control*, respectively. Standard errors cluster at the school level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

**Table B.1.5:** *Probability of attrition*

| VARIABLES | (1) In Endline (=1) | (2) Endline in person (=1) | (3) In Midline (=1) | (4) In Midline and Endline (=1) | (5) Mother in Endline (=1) | (6) Heart rate available (=1) | (7) Class with marks (=1) | (8) Class with teacher int. (=1) | (9) Boy from main sample list (=1) | (10) Girl from main sample list (=1) | (11) P(register available) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Base Only | -0.0269 | -0.0438** | -0.0493* | -0.0620** | -0.0179 | -0.0488 | -0.0930** | -0.0185 | 0.0368 | -0.0209 | -0.0067 |
| | (0.0189) | (0.0199) | (0.0266) | (0.0289) | (0.0120) | (0.0307) | (0.0465) | (0.0479) | (0.0229) | (0.0252) | (0.0131) |
| Base + YGL | 0.0015 | -0.0193 | -0.0374* | -0.0386 | -0.0125 | -0.0265 | -0.0276 | -0.0037 | -0.0134 | -0.0124 | 0.0062 |
| | (0.0165) | (0.0170) | (0.0211) | (0.0240) | (0.0116) | (0.0262) | (0.0344) | (0.0402) | (0.0230) | (0.0188) | (0.0114) |
| Observations | 2,385 | 2,385 | 2,385 | 2,385 | 2,385 | 2,385 | 500 | 500 | 1,325 | 1,627 | 3,626 |
| Data source | Girls | Girls | Girls | Girls | Girls | Heart Rate | Teachers | Teachers | Boys | Lab in field | Registers |
| Control mean | .951 | .928 | .891 | .859 | .963 | .802 | .919 | .887 | .837 | .853 | .901 |
| p: Treated = 0 | .6072 | .0921 | .0444 | .0491 | .1962 | .1764 | .1487 | .8183 | .8722 | .4041 | .8592 |
| p: Base Only = Base + YGL | .0889 | .1557 | .6076 | .3113 | .5246 | .3876 | .1012 | .7259 | .0122 | .7112 | .2554 |

*Notes:* The table displays the probability of each observation from the sampling frame being in the data. *Outcomes:* Column (1): a girl from the baseline sample was interviewed at endline. Column (2): girl from the baseline sample was interviewed in person (rather than through phone). Column (3): girl from the baseline sample was interviewed at midline. Column (4): girl was interviewed both at midline and endline. Column (5): the girl's mother or caretaker from the baseline sample was also interviewed at endline. Column (6): heart rate data was available for the girl's interview. Columns (7)-(8): we aimed to digitize the official school marks and interview teachers from a sample of 500 classes that included all target grade-levels (from 3rd grade to 12th grade), randomly selecting one class per grade-level if there were multiple. Column (7) shows probability of collecting school marks for each class; column (8) shows probability of interviewing the teacher. Column (9): the proportion of boys interviewed from the main sampling frame of 10 randomly selected boys based on the school registers (rather than a randomly selected replacement list). Column (10): for the lab in the field, the proportion of girls interviewed based on a sampling frame of 12 randomly selected girls based on school registers (instead of a randomly selected replacement list). Column (11): at the *class* × *wave* level, the proportion of measured days for which official absenteeism data was available. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects.

**Table B.1.6:** *Multiple hypothesis testing adjustment*

| Outcome | Pooled coeff. | p-val. | q-val. |
|---|---|---|---|
| Learning test (Z) | 0.143 | 0.006 | 0.012 |
| # girls present (spot-checks) | -0.029 | 0.991 | 0.657 |
| General health (Z) | -0.001 | 0.991 | 0.657 |
| Heart rate (bpm) | -2.354 | 0.014 | 0.015 |
| Stigma index (Z) | 0.617 | <0.001 | <0.001 |

*Notes:* $q$-values are calculated using the Anderson (2008) sharpened process that controls for the false discovery rate across the 5 outcome variables shown. All specifications follow the corresponding specification in the main results, but with a pooled treatment that includes both *Base only* and *Base+YGL* schools. All models cluster at the school level. All models control for baseline analogues of the outcome variable, and all models apart from the spot checks (row 2) use LASSO-selected controls. The spot checks model controls for wave fixed effects; the heart rate model controls for time-window fixed effects and includes all observations less than 120 minutes after the start of the survey.

**Table B.1.7:** *Are effects driven by participants with high social desirability bias?*

| *Outcome* | Base Only | Base + YGL | High SDB | Base × High SDB | (Base+YGL) × High SDB | N |
|---|---|---|---|---|---|---|
| Learning test (Z) | 0.203*** | 0.162*** | 0.057 | 0.004 | –0.090 | 2162 |
| | (0.077) [0.009] | (0.063) [0.010] | (0.093) [0.539] | (0.090) [0.965] | (0.088) [0.310] | |
| Progressed one class (=1) | 0.125** | 0.098** | 0.001 | –0.040 | –0.060 | 2251 |
| | (0.049) [0.011] | (0.048) [0.043] | (0.057) [0.993] | (0.059) [0.492] | (0.056) [0.285] | |
| Enrolled (=1) | –0.001 | –0.017 | –0.001 | –0.041 | –0.013 | 2251 |
| | (0.034) [0.988] | (0.034) [0.625] | (0.050) [0.988] | (0.051) [0.416] | (0.047) [0.774] | |
| Absent in last 3 months (=1) | 0.066 | –0.058 | 0.031 | –0.150** | 0.040 | 1820 |
| | (0.044) [0.128] | (0.039) [0.140] | (0.062) [0.615] | (0.062) [0.015] | (0.059) [0.501] | |
| General health (Z) | 0.006 | –0.001 | 0.031 | –0.078 | 0.032 | 2251 |
| | (0.080) [0.939] | (0.076) [0.994] | (0.123) [0.802] | (0.145) [0.591] | (0.121) [0.790] | |
| UTI health (Z) | –0.115 | 0.019 | 0.087 | 0.056 | –0.039 | 2251 |
| | (0.087) [0.185] | (0.061) [0.755] | (0.098) [0.371] | (0.113) [0.617] | (0.085) [0.649] | |
| Mental health (Z) | 0.101 | 0.065 | 0.115 | –0.171 | –0.019 | 2251 |
| | (0.073) [0.167] | (0.070) [0.350] | (0.107) [0.281] | (0.122) [0.159] | (0.111) [0.865] | |
| Self-esteem (Z) | –0.003 | –0.038 | 0.036 | 0.053 | 0.108 | 2251 |
| | (0.088) [0.977] | (0.069) [0.584] | (0.094) [0.701] | (0.109) [0.628] | (0.096) [0.263] | |
| Heart rate (Z) | –0.092 | –0.151* | –0.032 | –0.078 | 0.046 | 1899 |
| | (0.100) [0.358] | (0.080) [0.061] | (0.121) [0.790] | (0.148) [0.598] | (0.117) [0.691] | |
| Hygiene knowl. (Z) | 0.374*** | 0.466*** | –0.097 | –0.067 | –0.052 | 2162 |
| | (0.085) [0.000] | (0.072) [0.000] | (0.099) [0.329] | (0.103) [0.512] | (0.089) [0.557] | |
| Menstr. knowl. (Z) | 0.329*** | 0.368*** | –0.009 | –0.030 | –0.042 | 2162 |
| | (0.067) [0.000] | (0.060) [0.000] | (0.086) [0.919] | (0.095) [0.753] | (0.081) [0.604] | |
| Hygiene behavior (Z) | 0.134 | 0.338*** | 0.064 | –0.065 | –0.044 | 2162 |
| | (0.094) [0.151] | (0.085) [0.000] | (0.121) [0.595] | (0.121) [0.591] | (0.116) [0.705] | |
| Menstr. behavior (Z) | 0.276*** | 0.600*** | –0.102 | 0.171* | –0.078 | 2251 |
| | (0.084) [0.001] | (0.062) [0.000] | (0.085) [0.227] | (0.104) [0.098] | (0.083) [0.348] | |
| Combined stigma index (Z) | 0.456*** | 0.796*** | –0.008 | –0.119 | –0.122 | 2251 |
| | (0.074) [0.000] | (0.062) [0.000] | (0.082) [0.920] | (0.083) [0.151] | (0.081) [0.132] | |
| Received info. (Z) | 0.299*** | 0.607*** | 0.100 | –0.215 | –0.157 | 2251 |
| | (0.103) [0.004] | (0.080) [0.000] | (0.119) [0.399] | (0.147) [0.142] | (0.110) [0.152] | |
| Willingness to speak (Z) | 0.295*** | 0.754*** | –0.085 | –0.104 | –0.061 | 2245 |
| | (0.092) [0.001] | (0.077) [0.000] | (0.113) [0.455] | (0.110) [0.349] | (0.107) [0.568] | |
| Attitudes (Z) | 0.406*** | 0.604*** | –0.051 | –0.085 | –0.077 | 2249 |
| | (0.075) [0.000] | (0.061) [0.000] | (0.081) [0.530] | (0.080) [0.289] | (0.081) [0.340] | |
| Norms (Z) | 0.239*** | 0.464*** | –0.065 | 0.079 | 0.021 | 2162 |
| | (0.064) [0.000] | (0.053) [0.000] | (0.100) [0.513] | (0.102) [0.443] | (0.091) [0.817] | |
| Shame response to vignettes (rev.) (Z) | 0.347*** | 0.503*** | 0.087 | –0.094 | –0.121 | 2160 |
| | (0.072) [0.000] | (0.065) [0.000] | (0.087) [0.317] | (0.097) [0.333] | (0.082) [0.139] | |
| Bullying: teasing (Z) | 0.069 | 0.056 | 0.044 | 0.061 | –0.097 | 2162 |
| | (0.085) [0.417] | (0.079) [0.482] | (0.101) [0.661] | (0.108) [0.574] | (0.099) [0.328] | |
| Bullying: intimidation/harassment (Z) | –0.095 | –0.065 | –0.074 | 0.026 | –0.019 | 2162 |
| | (0.077) [0.217] | (0.065) [0.317] | (0.108) [0.494] | (0.123) [0.835] | (0.104) [0.856] | |
| Network (Z) | 0.228** | 0.329*** | 0.048 | –0.122 | –0.130 | 2162 |
| | (0.113) [0.043] | (0.096) [0.001] | (0.121) [0.692] | (0.129) [0.345] | (0.102) [0.204] | |

*Notes:* Each line is one regression, and each column is a coefficient from that regression *High SDB* = 1 when a girl's Crowne-Marlowe social desirability score elicited at baseline is above the median, and is 0 otherwise. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Standard errors are clustered at the school level and are in parentheses; $p$-values are in brackets. Only self-reported outcomes from the girls' survey are included.

**Table B.1.8:** *Effects on each individual learning test*

| VARIABLES | (1) Math fluency | (2) Reading fluency | (3) Listening | (4) Comprehension | (5) Digit span | (6) Reverse digit span |
|---|---|---|---|---|---|---|
| Base Only | 0.253*** | 0.160*** | 0.103 | 0.0731 | 0.115* | 0.0482 |
| | (0.0822) | (0.0496) | (0.0799) | (0.0763) | (0.0628) | (0.0682) |
| Base + YGL | 0.150** | 0.0741* | 0.0351 | 0.0335 | 0.0348 | 0.107** |
| | (0.0626) | (0.0412) | (0.0625) | (0.0583) | (0.0519) | (0.0505) |
| Observations | 2,162 | 2,162 | 2,162 | 2,162 | 2,162 | 2,162 |
| Data source | Girls | Girls | Girls | Girls | Girls | Girls |
| p: Treated = 0 | .0036 | .0099 | .3362 | .4047 | .2123 | .0681 |
| p: Base Only = Base + YGL | .1089 | .0374 | .3526 | .5698 | .1546 | .3627 |

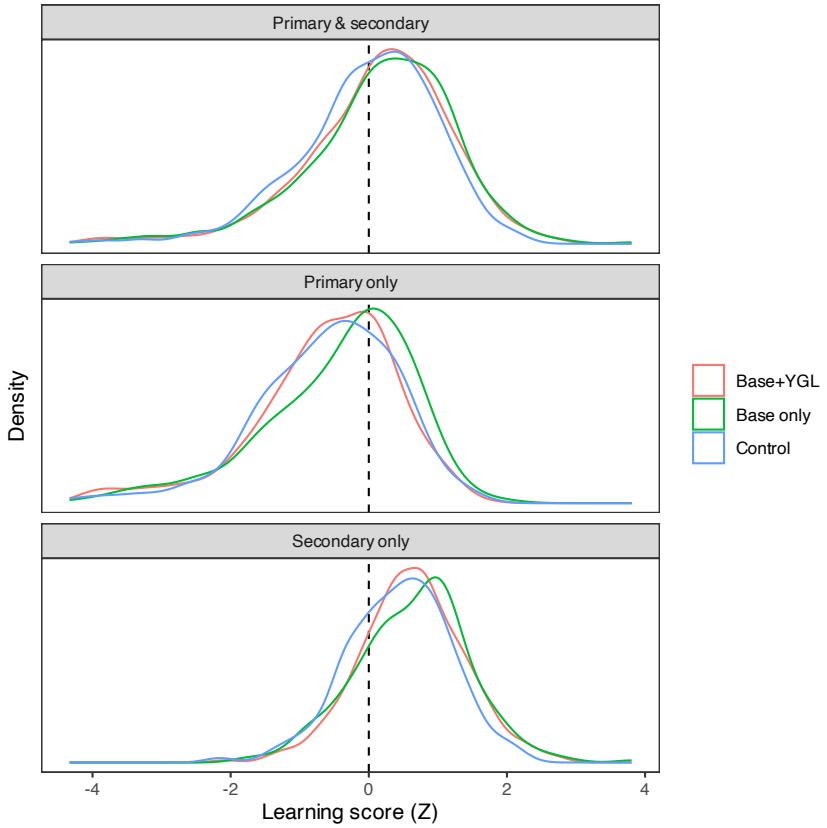
*Notes:* * p< 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the school level and are in parentheses. Sample includes all girls in endline interviewed in person. Controls include stratum fixed effects and the baseline controls selected by double LASSO. The outcomes are the individual components of the learning tests implemented during the girls' endline survey. All scores are in control group standard deviations.

**Table B.1.9:** *Randomization inference p-values*

| | Base+YGL | Base Only | Pooled | Base Only vs Base+YGL |
|---|---|---|---|---|
| Learning test (Z) | .191 | .107 | .024 | .193 |
| Progressed one class (=1) | .041 | .023 | .013 | .152 |
| Enrolled (=1) | .442 | .435 | .365 | .491 |
| Absent in last 3 months (=1) | .335 | 1 | .438 | .999 |
| General health (Z) | .805 | .651 | .992 | .376 |
| UTI health (Z) | .973 | .157 | .581 | .084 |
| Mental health (Z) | .419 | .821 | .461 | .805 |
| Heart rate (Z) | .034 | .074 | .018 | .504 |
| Self-esteem (Z) | .811 | .807 | .742 | .668 |
| Hygiene knowl. (Z) | <0.001 | .005 | <0.001 | .913 |
| Menstr. knowl. (Z) | <0.001 | <0.001 | <0.001 | .686 |
| Hygiene behavior (Z) | <0.001 | .315 | <0.001 | .998 |
| Menstr. behavior (Z) | <0.001 | <0.001 | <0.001 | .998 |
| Pregnant now or last year (=1) | .397 | .039 | .153 | .219 |

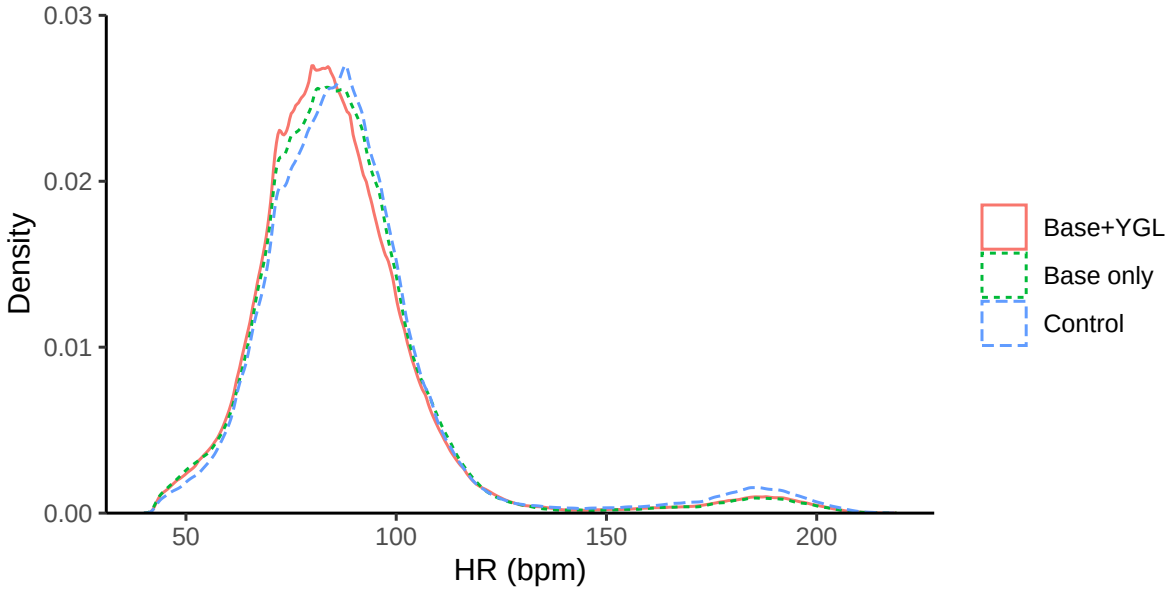
*Notes:* Table displays the *p*-values using randomization inference from 1000 replications. There is one row per outcome. Columns (1) and (2) result from a regression of the outcome on *Base Only* and *Base+YGL*. Column (3) results from a regression of the outcome on a pooled treatment variable. Column (4) tests if the effects of *Base Only* and *Base+YGL* are different. All regressions control for stratum fixed effects and LASSO controls.

**Figure B.1.1:** *Distribution of learning test scores at endline*




*Notes*: Density plot of learning test scores (academic achievement and cognitive tests combined) in each treatment arm. First panel includes both primary and secondary, second panel is only primary schools, and third panel is only secondary schools.

**Figure B.1.2:** *Distribution of heart rate at the girl × second level*




*Notes*: Shows the distribution of heart rate observations in each treatment group, measured in beats per minute. Unit of observation is the *girl* × *second*. Sample includes all girls with heart rate data at endline.

**Table B.1.10:** *Effects on girls with attrition weights*

| | (1) Base Only | (2) Base + YGL | (3) p: Treated=0 | (4) p:Base Only=Base+YGL | (5) N |
|---|---|---|---|---|---|
| Learning test (Z) | 0.209*** | 0.128** | .003 | .2119 | 2162 |
| | (0.072) [0.004] | (0.054) [0.017] | | | |
| Progressed one class (=1) | 0.096** | 0.056 | .0408 | .152 | 2251 |
| | (0.038) [0.013] | (0.035) [0.106] | | | |
| Enrolled (=1) | –0.027 | –0.029 | .1977 | .9294 | 2251 |
| | (0.024) [0.261] | (0.023) [0.209] | | | |
| Absent in last 3 months (=1) | 0.006 | –0.027 | .5553 | .1909 | 1820 |
| | (0.030) [0.840] | (0.029) [0.355] | | | |
| General health (Z) | –0.054 | –0.005 | .6637 | .3727 | 2251 |
| | (0.060) [0.371] | (0.048) [0.922] | | | |
| UTI health (Z) | –0.094* | –0.002 | .4113 | .0699 | 2251 |
| | (0.054) [0.080] | (0.044) [0.955] | | | |
| Mental health (Z) | 0.011 | 0.066 | .4387 | .2716 | 2251 |
| | (0.064) [0.859] | (0.061) [0.283] | | | |
| Self-esteem (Z) | –0.017 | –0.005 | .8892 | .8430 | 2251 |
| | (0.068) [0.803] | (0.052) [0.921] | | | |
| Heart rate (Z) | –0.107* | –0.110** | .0232 | .9515 | 1899 |
| | (0.057) [0.063] | (0.051) [0.031] | | | |
| Hygiene knowl. (Z) | 0.327*** | 0.438*** | <0.001 | .0763 | 2162 |
| | (0.077) [0.000] | (0.063) [0.000] | | | |
| Menstr. knowl. (Z) | 0.315*** | 0.338*** | <0.001 | .5942 | 2162 |
| | (0.057) [0.000] | (0.048) [0.000] | | | |
| Hygiene behavior (Z) | 0.097 | 0.306*** | <0.001 | .0031 | 2162 |
| | (0.084) [0.248] | (0.070) [0.000] | | | |
| Menstr. behavior (Z) | 0.358*** | 0.549*** | <0.001 | .0041 | 2251 |
| | (0.070) [0.000] | (0.053) [0.000] | | | |
| Observed hygiene: Girls (Z) | –0.055 | 0.061 | .7471 | .0303 | 2162 |
| | (0.064) [0.395] | (0.056) [0.276] | | | |
| Combined stigma index (Z) | 0.401*** | 0.738*** | <0.001 | <0.001 | 2251 |
| | (0.066) [0.000] | (0.058) [0.000] | | | |
| Received info. (Z) | 0.184** | 0.514*** | <0.001 | <0.001 | 2251 |
| | (0.078) [0.018] | (0.063) [0.000] | | | |
| Willingness to speak (Z) | 0.247*** | 0.737*** | <0.001 | <0.001 | 2245 |
| | (0.077) [0.001] | (0.065) [0.000] | | | |
| Attitudes (Z) | 0.348*** | 0.556*** | <0.001 | <0.001 | 2249 |
| | (0.065) [0.000] | (0.061) [0.000] | | | |
| Norms (Z) | 0.271*** | 0.458*** | <0.001 | <0.001 | 2162 |
| | (0.061) [0.000] | (0.051) [0.000] | | | |
| Shame response to vignettes (rev.) (Z) | 0.291*** | 0.419*** | <0.001 | .0171 | 2160 |
| | (0.059) [0.000] | (0.055) [0.000] | | | |
| Observed shame (rev.) (Z) | 0.270*** | 0.221*** | <0.001 | .3459 | 2251 |
| | (0.063) [0.000] | (0.055) [0.000] | | | |
| Bullying: teasing (Z) | 0.104 | 0.017 | .4192 | .1086 | 2162 |
| | (0.065) [0.110] | (0.058) [0.768] | | | |
| Bullying: intimidation/harassment (Z) | –0.062 | –0.062 | .1301 | .9854 | 2162 |
| | (0.049) [0.209] | (0.042) [0.135] | | | |
| Network (Z) | 0.172** | 0.262*** | <0.001 | .2023 | 2162 |
| | (0.086) [0.045] | (0.075) [0.000] | | | |

*Notes:* The table displays results for Base Only (1) and Base+YGL (2) from the girls survey in endline. The estimations are adjusted by attrition weights; standard errors are clustered at the school level and are in parentheses; coefficient p-values are in brackets; controls include stratum fixed effects and controls selected by LASSO. Columns (3) and (4) denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Column (5) shows the number of observations. ** p < 0.1, ** p < 0.05, *** p < 0.01.*

**Table B.1.11:** *Effects on school marks and teacher outcomes with attrition weights*

| | (1) Base Only | (2) Base + YGL | (3) p: Treated=0 | (4) p:Base Only=Base+YGL | (5) N |
|---|---|---|---|---|---|
| **Grades:** | | | | | |
| Official mark (Z) | 0.127* | 0.087 | .092 | .4902 | 14579 |
| | (0.069) [0.068] | (0.061) [0.158] | | | |
| Official mark (Z): girls | 0.171** | 0.087 | .059 | .2109 | 7586 |
| | (0.078) [0.029] | (0.061) [0.152] | | | |
| Official mark (Z): boys | 0.082 | 0.092 | .1696 | .8707 | 6993 |
| | (0.073) [0.260] | (0.070) [0.189] | | | |
| **Teachers:** | | | | | |
| Positive Menstruation Beliefs (Z) | 0.697*** | 0.688*** | <0.001 | .8871 | 743 |
| | (0.092) [0.000] | (0.089) [0.000] | | | |
| Mens. Knowledge Index (Z) | 0.187 | 0.249** | .0173 | .5898 | 743 |
| | (0.124) [0.133] | (0.101) [0.015] | | | |
| Positive Gender Views (Z) | 0.118 | 0.144 | .2267 | .8098 | 779 |
| | (0.130) [0.366] | (0.120) [0.231] | | | |
| Absent Last Month (=1) | –0.089 | –0.024 | .3214 | .2258 | 743 |
| | (0.056) [0.113] | (0.049) [0.617] | | | |

*Notes:* The table displays results for Base Only (1) and Base+YGL (2) from the teacher survey in endline. The estimations are adjusted by attrition weights; standard errors are clustered at the school level and are in parentheses; coefficient p-values are in brackets; controls include stratum fixed effects. Columns (3) and (4) denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Column (5) shows the number of observations. ** p < 0.1, ** p < 0.05, *** p < 0.01.*

**Table B.1.12:** *Effects on attendance for boys*

| VARIABLES | (1) # boys present (spot-checks) | (2) # boys present (registers) |
|---|---|---|
| Base Only | -2.443 | 1.906 |
| | (4.003) | (3.103) |
| Base + YGL | -1.720 | -0.0568 |
| | (3.070) | (2.615) |
| Observations | 657 | 34,455 |
| Data source | Spot checks | Spot checks |
| Control mean | 52.5 | 57.3 |
| Control SD | 42.583 | 78.712 |
| p: Treated = 0 | .5172 | .8138 |
| p: Base Only = Base + YGL | .8288 | .5323 |

*Notes:* Column 1: outcome is the number of boys physically present during the spot checks. Observations are at the school × wave level. Specification controls for spot check wave fixed effects. Column 2: outcome is the number of boys present on a given day according to official school registers. Observations are at the school × day level. Specification controls for day fixed effects. Both columns control for stratum fixed effects and cluster standard errors at the school level.

**Table B.1.13:** *Mental health (depression) at midline and endline*

| VARIABLES | (1) Midline Index Factor Analysis (Z) | (2) Midline Index Raw Sum (Z) | (3) Midline Positive Emotions (Z) | (4) Midline Negative Emotions (Z) | (5) Endline Index Factor Analysis (Z) | (6) Endline Index Raw Sum (Z) | (7) Endline Positive Emotions (Z) | (8) Endline Negative Emotions (Z) |
|---|---|---|---|---|---|---|---|---|
| Base Only | -0.0401 | -0.0243 | 0.0343 | -0.0196 | -0.0267 | 0.0076 | 0.0586 | 0.0918 |
| | (0.0456) | (0.0519) | (0.0493) | (0.0684) | (0.0675) | (0.0617) | (0.0731) | (0.0638) |
| Base + YGL | -0.0034 | 0.0599 | 0.0503 | 0.1381** | -0.0297 | -0.0011 | 0.0503 | 0.0641 |
| | (0.0418) | (0.0471) | (0.0416) | (0.0580) | (0.0636) | (0.0584) | (0.0669) | (0.0536) |
| Observations | 1,958 | 1,958 | 1,958 | 1,958 | 1,959 | 1,959 | 1,959 | 1,959 |
| Data source | Girls | Girls | Girls | Girls | Girls | Girls | Girls | Girls |
| p: Treated = 0 | .6857 | .4542 | .242 | .1133 | .6315 | .9749 | .4029 | .1533 |
| p: Base Only = Base + YGL | .3906 | .084 | .7334 | .0133 | .9545 | .8604 | .886 | .6082 |

*Notes:* The table shows the disaggregated results on mental health at midline and endline. Only girls who were interviewed at *both* midline and endline are included. All outcomes are normalized by the mean and standard deviation of the control group at endline. Only questions that were included in both midline and endline are included. *Index: Factor Analysis* is an index constructed using factor-analysis-based weightings on both "positive" and "negative" emotion questions, and correcting for acquiescence bias. *Index: Raw Sum* simply sums the values of the positive and (reverse-coded) negative emotion questions. For both of these measures, a higher index indicates improved mental health. *Positive emotions* is a sum of the value of only the positive emotions. *Negative emotions* is a sum of the value of only negative emotions (a higher score indicating more frequent negative emotions).

**Table B.1.14:** *Robustness of effects on heart rate to dealing with outliers*

| | HR (bpm) | | |
|---|---|---|---|
| | Individual-level | Drop outliers | Divide outliers |
| Base only | −2.206* | −0.981 | −1.047 |
| | (1.177) | (0.805) | (0.794) |
| Base+YGL | −2.254** | −1.171* | −1.238* |
| | (1.030) | (0.703) | (0.698) |
| Num.Obs. | 1904 | 8 877 622 | 9 105 420 |
| N girls | 1904 | 1904 | 1904 |
| Control mean | 87.9 | 85 | 85.3 |
| Control SD (between) | 19.0 | 14.6 | 12.2 |
| p : Treat = 0 | 0.023 | 0.099 | 0.078 |
| p : Base=Base+YGL | 0.961 | 0.775 | 0.772 |

*Notes:* Outcomes are heart rate measured in beats per minute. Column (1): Observations are amalgamated to the *girl* level by taking the mean heart rate throughout the first 120 minutes of the endline survey. Column (2): observations are at the *girl* × *second* level, with observations with heart rate above 175 bpm are dropped from the sample. Column (3): observations are at the *girl* × *second* level, with observations with heart rate above 175 bpm are divided by 2 (based on the assumption that the heart rate monitor may be erroneously double counting both the systolic and diastolic peaks within a single heartbeat). This division creates a unimodal distribution of heart rate observations. Observations are at the girl × second level. * p< 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the school level and are in parentheses. Sample includes all girls in endline interviewed with heart rate data. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Columns 2 and 3 also include 30-second time-window fixed effects.

**Table B.1.15:** *Heart rate: dynamics*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Treated (=1) | −1.694 | −1.706 | −1.361 | −1.363 |
| | (1.102) | (1.101) | (1.110) | (1.110) |
| After first question about menstr. (=1) | −1.217* | 1.526* | | |
| | (0.666) | (0.894) | | |
| Treated * After first question about menstr. | −0.028 | −0.019 | | |
| | (0.758) | (0.771) | | |
| Time since start of survey (m) | | | −0.038* | |
| | | | (0.020) | |
| Treated * Time since start of survey | | | −0.012 | −0.077*** |
| | | | (0.022) | (0.022) |
| FEs: 30-second window | | X | | X |
| Num. obs. | 6 678 438 | 6 678 438 | 6 678 438 | 6 678 438 |

*Notes:* All specifications use stratum fixed effects and LASSO-selected controls. Standard errors are clustered at the school level. Observations are at the second × girl level. Only observations in the first 60 minutes of the survey are included, since girls' interviews typically last between 60 and 90 minutes and so the number of observations begins to drop rapidly and become a selected sample. *Treated*=1 when a girl is in a *Base only* or *Base+YGL* school, and 0 otherwise. Columns (1) and (2) examine the role of the first question about menstruation ("Have you ever had your period?"). Column (1) does not control for 30-second time-window fixed effects, but column (2) does. Columns (3) and (4) examine a linear time trend (measured in minutes). Column (4) also controls for time-window fixed effects. Since *Time since start of survey* is perfectly correlated with the time-window fixed effects, it is not included in the model in column (4).

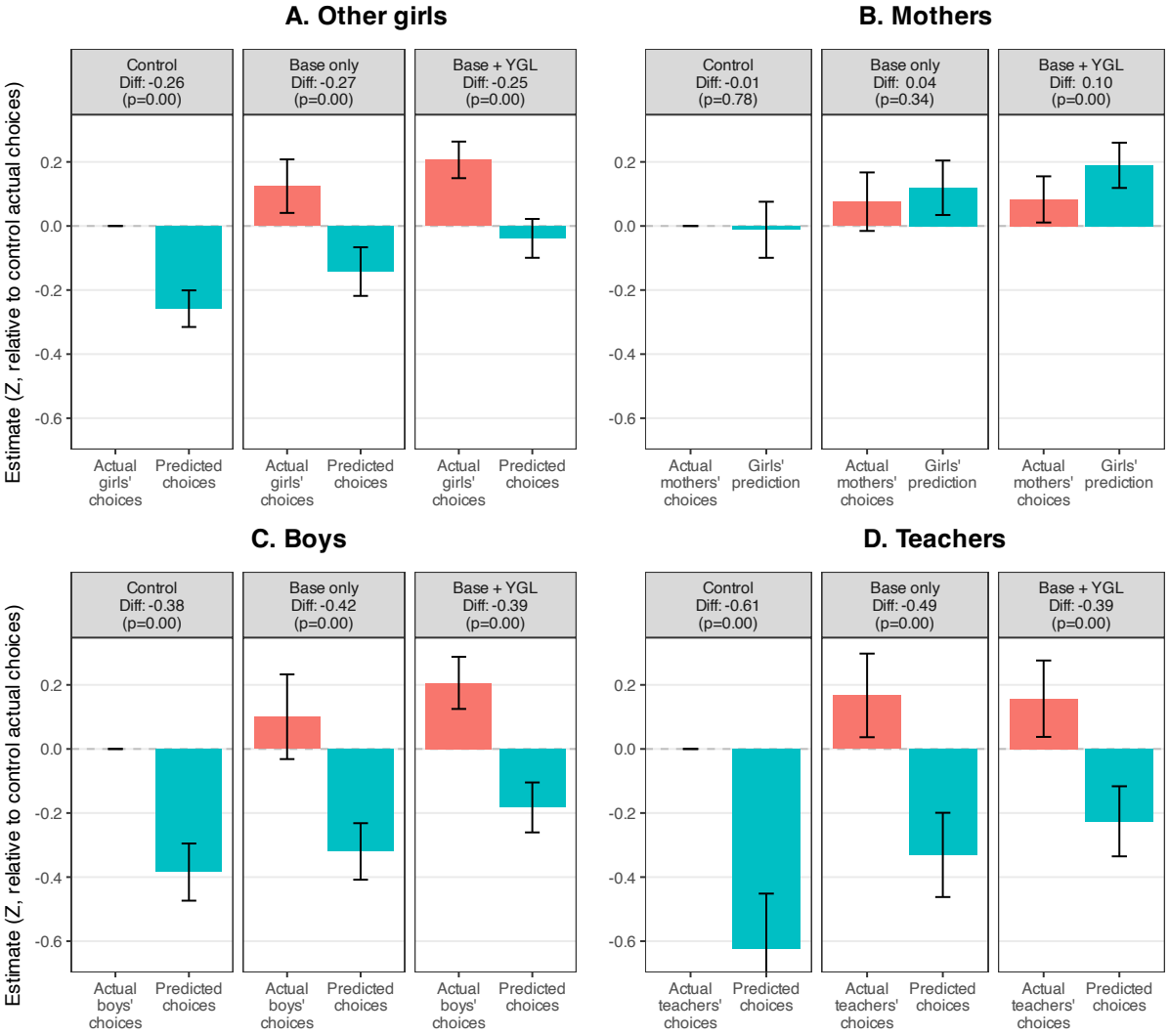
**Table B.1.16:** *Effects on mothers' knowledge, behavior, and perceptions of girls' behavior*

| VARIABLES | (1) Hygiene knowl. (Z) | (2) Menstr. knowl. (Z) | (3) Hygiene behavior (Z) | (4) Menstr. behavior (Z) | (5) Attitudes (Z) | (6) Girls' behavior (Z) |
|---|---|---|---|---|---|---|
| Base Only | 0.00614 | 0.0915 | -0.105 | 0.0294 | 0.327*** | -0.151** |
| | (0.0857) | (0.0736) | (0.0767) | (0.0619) | (0.0652) | (0.0769) |
| Base + YGL | 0.0240 | 0.218*** | -0.0473 | 0.00840 | 0.343*** | 0.00811 |
| | (0.0739) | (0.0599) | (0.0644) | (0.0580) | (0.0546) | (0.0607) |
| Observations | 2,282 | 2,281 | 2,282 | 2,282 | 2,282 | 2,255 |
| Data source | Mothers | Mothers | Mothers | Mothers | Mothers | Mothers |
| p: Treated = 0 | 0.800 | 0.0100 | 0.280 | 0.780 | 0.0100 | 0.480 |
| p: Base Only = Base + YGL | 0.790 | 0.0500 | 0.390 | 0.620 | 0.790 | 0.0200 |

*Notes:* The table displays knowledge and behavior index results from the mother survey in endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01.*

*Hygiene knowledge index* combines questions about WASH knowledge related to the program. *Menstruation knowledge index* is constructed from questions that tests mothers' knowledge about menstruation. *Hygiene behavior index* compiles questions about hygiene behavior during the last week. *Menstruation behavior index* includes questions about mothers' behavior during the last period, such as the use of serviettes. *Attitudes* includes questions about attitudes towards menstruation, where a higher value indicates more progressive attitudes. *Girls' behavior* includes mothers' reports about the psychosocial behavior of their daughters.

**Figure B.1.3:** *Misperceptions about others' attitudes*



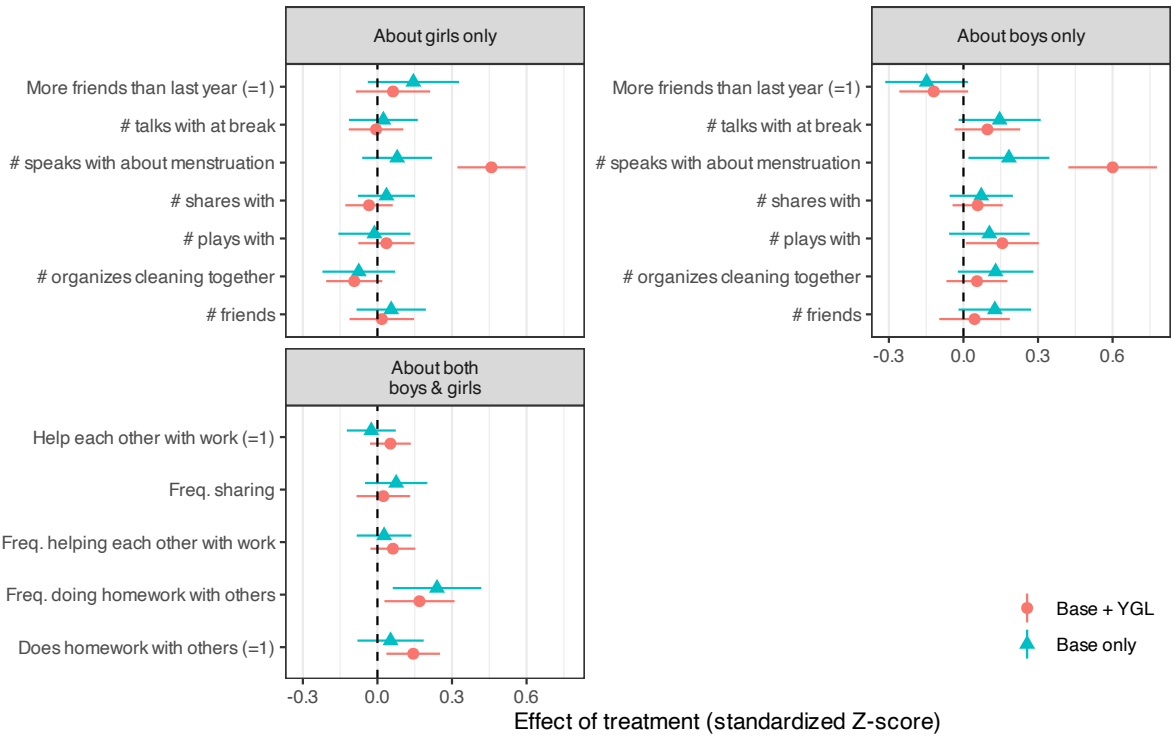

*Notes*: Panel **A** shows predictions about other girls at school. Panel **B** shows girls' predictions about their mother. Panel **C** shows girls' predictions about boys at school. Panel **D** shows predictions about teachers at school. All outcomes are always normalized to the actual choices in the control group. Pooled treatment effect of [*Base only* and *Base + YGL*] on the average *misperception* regarding teachers' attitudes (predicted - actual) is 0.20 SD ($p$=0.03). All other treatment effects on average misperceptions are not significant at the 10% level.

**Table B.1.17:** *Dictator game: girls*

| VARIABLES | (1) Prop. cookies shared w/ handwashing girl | (2) Prop. cookies shared w/ menstruation girl |
|---|---|---|
| Base Only | -0.0049 | 0.0177 |
| | (0.0143) | (0.0128) |
| Base + YGL | 0.0195 | 0.0074 |
| | (0.0127) | (0.0117) |
| Observations | 2,161 | 2,161 |
| Data source | Girls | Girls |
| Control mean | .42 | .422 |
| Control SD | .221 | .206 |
| p: Treated = 0 | .3431 | .315 |
| p: Base Only = Base + YGL | .0442 | .3872 |

*Notes:* This shows the results of the girls' dictator game. In all regressions, standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$* p < 0.1$, *$** p < 0.05$, *$*** p < 0.01$. Columns (4)-(5) show results from a dictator game implemented during the girls' survey. Respondents had to decide how many cookies to share with another girl after watching a video of that girl providing a menstruation- or hand-washing-related explanation. The outcomes in these columns represent the proportion of cookies (out of 5) that respondents agreed to share with the girl in the video. Controls include order of videos fixed effects.

**Figure B.1.4:** *Effect on individual network questions*




*Notes*: This figure shows the effects on each outcome that is used in the network index in Table 2.7. Only girls included in the baseline sample were used. All outcomes are normalized using the control mean and standard deviation from the endline. All regressions control for stratum fixed effects and confidence intervals are calculated based on standard errors clustered at the school level.

**Table B.1.18:** *Effects on bullying*

| VARIABLES | (1) Self-report Combined (Z) | (2) Self-report Teasing only (Z) | (3) Self-report Intimidation/ harassment only (Z) | (4) Peer-report Combined (Z) | (5) Peer-report Teasing only (Z) | (6) Peer-report Intimidation/ harassment only (Z) | (7) Teacher-report Combined (Z) | (8) Teacher-report Teasing only (Z) | (9) Teacher-report Intimidation/ harassment only (Z) |
|---|---|---|---|---|---|---|---|---|---|
| Base Only | 0.0228 | 0.101 | -0.0853* | -0.0809 | -0.0350 | -0.0841* | -0.0710 | -0.0294 | -0.0763 |
| | (0.0591) | (0.0655) | (0.0457) | (0.0534) | (0.0611) | (0.0464) | (0.0567) | (0.0674) | (0.0498) |
| Base + YGL | -0.0322 | 0.0127 | -0.0724* | -0.0559 | -0.0126 | -0.0736* | -0.0324 | 0.0260 | -0.0544 |
| | (0.0532) | (0.0583) | (0.0403) | (0.0490) | (0.0596) | (0.0444) | (0.0529) | (0.0612) | (0.0477) |
| Observations | 2,162 | 2,162 | 2,162 | 3,033 | 3,033 | 3,033 | 3,637 | 3,637 | 3,637 |
| Data source | Girls | Girls | Girls | Girls | Girls | Girls | Teachers | Teachers | Teachers |
| p: Treated = 0 | .7758 | .4487 | .043 | .1496 | .6921 | .0577 | .386 | .8636 | .1785 |
| p: Base Only = Base + YGL | .2831 | .1039 | .7424 | .5957 | .7033 | .7941 | .3923 | .3271 | .5811 |

*Notes:* Table shows girls' and teachers' reports of bullying. Columns (1)-(3) are girls' reports about bullying towards themselves. Columns (4)-(6) are girls' reports about bullying towards other girls, using a randomly selected subset of 2 or 3 girls in their grade-level (and removing cases where the respondent did not know the other girl). Columns (7)-(9) are teachers' reports about randomly selected girls in their class. Each teacher reports on a randomly selected sample of 3 girls from their class; the observations are at the *teacher* × *girl* level. Columns (1), (4) and (7) combine reports of both light teasing and severe intimidation/harassment. Columns (2), (5), and (8) include only light teasing. Columns (3), (6) and (9) include only severe intimidation/harassment. Combined indexes are constructed using loadings from exploratory factor analysis. The same relative loadings are used for teasing and intimidation/harassment. Standard errors are clustered at the school level and are in parentheses. Sample includes all girls in endline interviewed in person. Controls include stratum fixed effects and the baseline controls selected by double LASSO. All outcomes are in control group standard deviations.

**Table B.1.19:** *Effects on bullying by boys or by other girls*

| VARIABLES | (1) Teased by boy(s) (Self-reported) (=1) | (2) Teased by girls(s) (Self-reported) (=1) | (3) Teased by boy(s) (Peer-reported) (=1) | (4) Teased by girls(s) (Peer-reported) (=1) | (5) Teased about menstr. by boy(s) (=1) | (6) Teased about menstr. by girls(s) (=1) |
|---|---|---|---|---|---|---|
| Base Only | 0.0220 | 0.0192 | -0.0113 | -0.0152 | 0.0319* | -0.00268 |
| | (0.0257) | (0.0223) | (0.0256) | (0.0197) | (0.0165) | (0.0116) |
| Base + YGL | 0.00316 | 0.00643 | 0.0102 | 0.00468 | 0.0653*** | 0.00647 |
| | (0.0222) | (0.0198) | (0.0245) | (0.0176) | (0.0138) | (0.0111) |
| Observations | 2,162 | 2,162 | 3,033 | 3,033 | 2,251 | 2,251 |
| Survey | Girls | Girls | Girls | Girls | Girls | Girls |
| Outcome mean control | 0.240 | 0.220 | 0.250 | 0.160 | 0.0500 | 0.0500 |
| Outcome sd control | 0.428 | 0.416 | 0.435 | 0.364 | 0.218 | 0.214 |
| p: Treated = 0 | 0.650 | 0.560 | 0.920 | 0.860 | 0.0100 | 0.630 |
| p: Base Only = Base + YGL | 0.410 | 0.520 | 0.390 | 0.370 | 0.0300 | 0.370 |

*Notes:* Columns (1), (3), (5): outcome = 1 when girl was teased by a boy. Columns (2), (4), (6): outcome = 1 when girl was teased by a girl. Columns (1) and (2) are girls' reports about bullying towards themselves. Columns (3) and (4) are girls reports about randomly selected girls in their class. Columns (5) and (6) are self-reports about bullying specifically about menstruation. Sample includes all girls in endline interviewed in person. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Standard errors are clustered at the school level and are in parentheses.

**Table B.1.20:** *Effects on restrictions to girls' activities due to menstruation (among girls who reached menarche at baseline*

| VARIABLES | (1) Any (=1) | (2) Absent (=1) | (3) Absent because ill (=1) | (4) Left school during day (=1) | (5) Couldn't participate fully (=1) | (6) Missed social activities (=1) | (7) Other activities (=1) |
|---|---|---|---|---|---|---|---|
| Base Only | -0.0026 | -0.0228 | -0.0006 | -0.0475* | 0.0006 | 0.0002 | -0.0040 |
| | (0.0456) | (0.0273) | (0.0191) | (0.0252) | (0.0351) | (0.0251) | (0.0262) |
| Base + YGL | -0.0706** | -0.0385** | -0.0174 | -0.0339* | -0.0610*** | -0.0298 | 0.0103 |
| | (0.0339) | (0.0195) | (0.0177) | (0.0202) | (0.0232) | (0.0225) | (0.0248) |
| Observations | 1,053 | 1,053 | 1,050 | 1,053 | 1,053 | 1,053 | 1,053 |
| Data source | Girls | Girls | Girls | Girls | Girls | Girls | Girls |
| Control mean | .375 | .117 | .05 | .108 | .225 | .096 | .113 |
| Control SD | .485 | .322 | .219 | .311 | .418 | .295 | .317 |
| p: Treated = 0 | .1463 | .0899 | .4925 | .0513 | .0923 | .3745 | .8126 |
| p: Base Only = Base + YGL | .1276 | .5168 | .2325 | .4965 | .0628 | .1054 | .5291 |

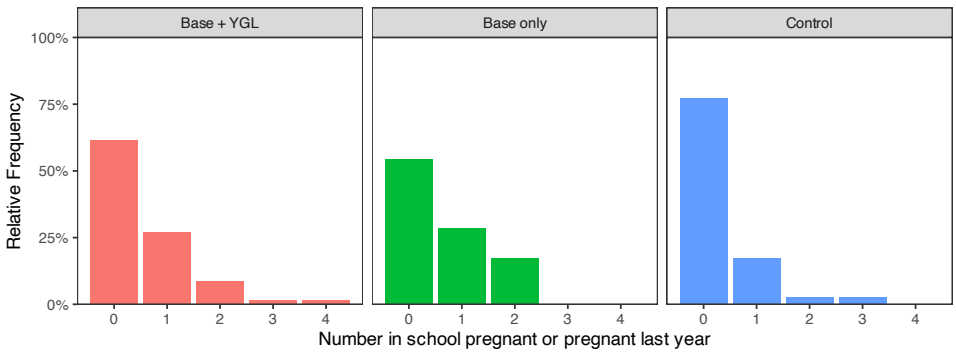
*Notes:* Shows girls' reports of restrictions to their activities because of menstruation at endline. Sample only includes 1,053 girls who reported having started menstruating at *baseline* and interviewed in person at endline. We use baseline menarche rather than endline because at endline there is a small (5%) but significant difference in the share of girls report having started menarche, possibly related to reduced menstrual stigma leading to less underreporting in treatment schools. Standard errors are clustered at the school level and are in parentheses. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Column (1): outcome is whether any of columns (2)-(7) are true. Column (2): has the girl ever been absent from school because of menstruation? Column (3): has the girl ever been absent because they were ill because of menstruation? Column (4): has the girl ever left school during the school day because of menstruation? Column (5): does the girl report not being able to participate fully in school activities because of menstruation? Column (6): has the girl missed any social activities because of menstruation? Column (7): has the girl been restricted from any other activities due to menstruation? The recall period used is whether these restrictions have *ever* occurred to them, so the questions were asked to all girls including those who had dropped out of school in the last year.

**Table B.1.21:** *Heterogeneity by whether girls were menstruating at baseline*

| | Base Only | Base + YGL | Menstr. at baseline | Base × Menstr. at base. | Base+YGL × Menstr. at base. | N |
|---|---|---|---|---|---|---|
| Learning test (Z) | 0.201** | 0.058 | 0.016 | –0.137 | 0.014 | 2160 |
| | (0.094) [0.033] | (0.067) [0.391] | (0.140) [0.908] | (0.126) [0.277] | (0.088) [0.875] | |
| Progressed one class (=1) | 0.138*** | 0.097** | –0.051 | –0.081 | –0.073 | 2249 |
| | (0.047) [0.003] | (0.042) [0.021] | (0.085) [0.547] | (0.066) [0.220] | (0.059) [0.212] | |
| Enrolled (=1) | –0.024 | –0.027 | –0.076 | –0.005 | 0.004 | 2249 |
| | (0.032) [0.454] | (0.029) [0.358] | (0.069) [0.271] | (0.043) [0.910] | (0.042) [0.930] | |
| Absent in last 3 months (=1) | –0.020 | –0.033 | 0.006 | 0.029 | –0.011 | 1818 |
| | (0.046) [0.669] | (0.042) [0.434] | (0.093) [0.950] | (0.065) [0.657] | (0.057) [0.846] | |
| General health (Z) | –0.029 | 0.020 | 0.347** | 0.011 | 0.011 | 2249 |
| | (0.072) [0.692] | (0.064) [0.748] | (0.154) [0.024] | (0.110) [0.920] | (0.085) [0.900] | |
| UTI health (Z) | –0.095 | –0.019 | 0.234 | 0.027 | 0.051 | 2249 |
| | (0.077) [0.216] | (0.058) [0.739] | (0.179) [0.191] | (0.115) [0.812] | (0.094) [0.585] | |
| Mental health (Z) | 0.037 | 0.092 | 0.017 | –0.060 | –0.087 | 2249 |
| | (0.088) [0.677] | (0.079) [0.248] | (0.175) [0.924] | (0.123) [0.624] | (0.119) [0.462] | |
| Self-esteem (Z) | 0.054 | 0.114 | 0.349* | –0.070 | –0.204* | 2249 |
| | (0.104) [0.599] | (0.075) [0.127] | (0.192) [0.069] | (0.142) [0.625] | (0.119) [0.085] | |
| Heart rate (Z) | –0.063 | –0.047 | 0.238 | –0.109 | –0.142 | 1897 |
| | (0.079) [0.431] | (0.069) [0.493] | (0.168) [0.156] | (0.114) [0.337] | (0.104) [0.173] | |
| Hygiene knowl. (Z) | 0.333*** | 0.491*** | 0.236 | –0.027 | –0.148 | 2160 |
| | (0.105) [0.001] | (0.076) [0.000] | (0.175) [0.178] | (0.137) [0.847] | (0.110) [0.180] | |
| Menstr. knowl. (Z) | 0.400*** | 0.471*** | 0.149 | –0.211** | –0.289*** | 2160 |
| | (0.086) [0.000] | (0.070) [0.000] | (0.119) [0.208] | (0.104) [0.043] | (0.087) [0.001] | |
| Hygiene behavior (Z) | 0.049 | 0.306*** | –0.221 | 0.148 | 0.040 | 2160 |
| | (0.113) [0.666] | (0.095) [0.001] | (0.172) [0.198] | (0.154) [0.335] | (0.117) [0.731] | |
| Menstr. behavior (Z) | 0.263*** | 0.441*** | 0.403*** | 0.226* | 0.274*** | 2249 |
| | (0.081) [0.001] | (0.066) [0.000] | (0.138) [0.003] | (0.119) [0.058] | (0.091) [0.003] | |
| Observed hygiene: Girls (Z) | –0.043 | 0.115* | 0.184 | –0.061 | –0.149 | 2160 |
| | (0.076) [0.569] | (0.068) [0.089] | (0.177) [0.297] | (0.138) [0.657] | (0.116) [0.196] | |
| Combined stigma index (Z) | 0.458*** | 0.821*** | 0.193 | –0.165 | –0.217** | 2249 |
| | (0.085) [0.000] | (0.061) [0.000] | (0.145) [0.182] | (0.113) [0.144] | (0.091) [0.018] | |
| Received info. (Z) | 0.105 | 0.583*** | 0.032 | 0.177 | –0.118 | 2249 |
| | (0.092) [0.251] | (0.079) [0.000] | (0.237) [0.891] | (0.142) [0.214] | (0.111) [0.291] | |
| Willingness to speak (Z) | 0.283*** | 0.648*** | 0.197 | –0.108 | 0.149 | 2243 |
| | (0.084) [0.001] | (0.069) [0.000] | (0.179) [0.271] | (0.131) [0.410] | (0.116) [0.200] | |
| Attitudes (Z) | 0.425*** | 0.679*** | 0.277* | –0.172 | –0.288*** | 2247 |
| | (0.083) [0.000] | (0.066) [0.000] | (0.167) [0.098] | (0.109) [0.113] | (0.096) [0.003] | |
| Norms (Z) | 0.297*** | 0.565*** | 0.036 | –0.067 | –0.228** | 2160 |
| | (0.082) [0.000] | (0.052) [0.000] | (0.169) [0.830] | (0.110) [0.540] | (0.089) [0.011] | |
| Shame response to vignettes (rev.) (Z) | 0.387*** | 0.533*** | 0.020 | –0.210** | –0.223*** | 2158 |
| | (0.082) [0.000] | (0.070) [0.000] | (0.138) [0.885] | (0.101) [0.037] | (0.082) [0.007] | |
| Observed shame (rev.) (Z) | 0.402*** | 0.306*** | 0.168 | –0.268*** | –0.200** | 2249 |
| | (0.081) [0.000] | (0.069) [0.000] | (0.155) [0.280] | (0.099) [0.007] | (0.085) [0.019] | |
| Bullying: teasing (Z) | 0.091 | 0.063 | –0.153 | 0.049 | –0.108 | 2160 |
| | (0.095) [0.341] | (0.084) [0.450] | (0.163) [0.349] | (0.133) [0.716] | (0.125) [0.388] | |
| Bullying: intimidation/harassment (Z) | –0.085 | –0.050 | –0.143 | 0.002 | –0.048 | 2160 |
| | (0.063) [0.175] | (0.057) [0.382] | (0.119) [0.229] | (0.095) [0.984] | (0.083) [0.563] | |
| Network (Z) | –0.063 | 0.164** | –0.185 | 0.485*** | 0.196 | 2160 |
| | (0.087) [0.473] | (0.080) [0.040] | (0.211) [0.382] | (0.145) [0.001] | (0.123) [0.110] | |

*Notes:* Each row represents a regression of a different outcome. Each column represents the coefficient on each term. *Menstr. at base.* is 1 when a girl was menstruating at baseline. All regressions include stratum fixed effects and the baseline controls selected by double LASSO, and cluster standard errors at the school level.

**Figure B.1.5:** *School-level relative frequency of number of pregnant girls in the sample*




*Notes*: Observation is at the school level. Shows the histogram of the number of pregnant girls in a school at endline, divided by the total number of schools in each treatment group in order to maintain comparability between *Base+YGL* (70 schools), *Base only* (35 schools) and *Control* schools (35 schools). Pregnancy = 1 if mothers responded that their daughter was currently pregnant or had been pregnant in the last year.

**Table B.1.22:** *Heterogeneity by whether girls started menstruating between baseline and endline*

| | Base Only | Base + YGL | Newly menstr. | Base × Newly menstr. | Base+YGL × Newly menstr. | N |
|---|---|---|---|---|---|---|
| Learning test (Z) | 0.131* | 0.054 | 0.009 | 0.080 | 0.080 | 2162 |
| | (0.071) [0.066] | (0.055) [0.327] | (0.095) [0.927] | (0.148) [0.587] | (0.101) [0.426] | |
| Progressed one class (=1) | 0.114*** | 0.083** | 0.078 | –0.023 | –0.062 | 2251 |
| | (0.044) [0.010] | (0.040) [0.039] | (0.068) [0.249] | (0.076) [0.763] | (0.069) [0.371] | |
| Enrolled (=1) | –0.021 | –0.021 | –0.058 | 0.010 | –0.002 | 2251 |
| | (0.024) [0.394] | (0.024) [0.388] | (0.039) [0.135] | (0.052) [0.850] | (0.047) [0.964] | |
| Absent in last 3 months (=1) | 0.002 | –0.018 | 0.006 | –0.040 | –0.063 | 1820 |
| | (0.037) [0.953] | (0.033) [0.591] | (0.071) [0.931] | (0.088) [0.652] | (0.073) [0.385] | |
| General health (Z) | –0.021 | –0.005 | –0.120 | –0.001 | 0.074 | 2251 |
| | (0.067) [0.755] | (0.051) [0.915] | (0.125) [0.338] | (0.160) [0.996] | (0.131) [0.575] | |
| UTI health (Z) | –0.123* | –0.037 | –0.227** | 0.152 | 0.155 | 2251 |
| | (0.064) [0.053] | (0.054) [0.498] | (0.113) [0.044] | (0.168) [0.366] | (0.126) [0.218] | |
| Mental health (Z) | –0.018 | 0.011 | –0.132 | 0.095 | 0.100 | 2251 |
| | (0.066) [0.784] | (0.065) [0.869] | (0.140) [0.344] | (0.151) [0.531] | (0.139) [0.470] | |
| Self-esteem (Z) | 0.002 | –0.033 | –0.140 | 0.009 | 0.150 | 2251 |
| | (0.073) [0.978] | (0.063) [0.606] | (0.136) [0.305] | (0.161) [0.955] | (0.148) [0.311] | |
| Heart rate (Z) | –0.104* | –0.110** | –0.067 | –0.031 | –0.011 | 1899 |
| | (0.060) [0.084] | (0.055) [0.046] | (0.108) [0.536] | (0.130) [0.814] | (0.116) [0.922] | |
| Hygiene knowl. (Z) | 0.316*** | 0.395*** | 0.048 | –0.015 | 0.058 | 2162 |
| | (0.081) [0.000] | (0.067) [0.000] | (0.120) [0.690] | (0.143) [0.919] | (0.122) [0.634] | |
| Menstr. knowl. (Z) | 0.295*** | 0.285*** | 0.443*** | –0.072 | 0.027 | 2162 |
| | (0.060) [0.000] | (0.053) [0.000] | (0.073) [0.000] | (0.090) [0.423] | (0.075) [0.723] | |
| Hygiene behavior (Z) | 0.095 | 0.292*** | –0.014 | –0.016 | 0.100 | 2162 |
| | (0.087) [0.275] | (0.071) [0.000] | (0.096) [0.881] | (0.138) [0.906] | (0.112) [0.370] | |
| Menstr. behavior (Z) | 0.271*** | 0.402*** | 1.671*** | 0.281** | 0.325*** | 2251 |
| | (0.062) [0.000] | (0.053) [0.000] | (0.086) [0.000] | (0.114) [0.013] | (0.090) [0.000] | |
| Observed hygiene: Girls (Z) | –0.169** | –0.029 | 0.132 | 0.248* | 0.158 | 2162 |
| | (0.069) [0.014] | (0.060) [0.626] | (0.105) [0.208] | (0.136) [0.068] | (0.117) [0.178] | |
| Combined stigma index (Z) | 0.335*** | 0.656*** | 0.101 | 0.034 | 0.089 | 2251 |
| | (0.070) [0.000] | (0.058) [0.000] | (0.089) [0.258] | (0.112) [0.762] | (0.101) [0.375] | |
| Received info. (Z) | 0.221** | 0.494*** | 0.136 | –0.191 | 0.004 | 2251 |
| | (0.086) [0.010] | (0.069) [0.000] | (0.105) [0.196] | (0.127) [0.132] | (0.122) [0.977] | |
| Willingness to speak (Z) | 0.171** | 0.657*** | 0.002 | 0.074 | 0.099 | 2245 |
| | (0.073) [0.019] | (0.060) [0.000] | (0.088) [0.979] | (0.119) [0.537] | (0.115) [0.387] | |
| Attitudes (Z) | 0.277*** | 0.473*** | 0.036 | 0.169 | 0.159 | 2249 |
| | (0.066) [0.000] | (0.064) [0.000] | (0.108) [0.741] | (0.127) [0.183] | (0.117) [0.176] | |
| Norms (Z) | 0.230*** | 0.413*** | 0.098 | 0.021 | 0.063 | 2162 |
| | (0.070) [0.001] | (0.060) [0.000] | (0.107) [0.361] | (0.127) [0.870] | (0.114) [0.581] | |
| Shame response to vignettes (rev.) (Z) | 0.282*** | 0.424*** | 0.203* | –0.086 | –0.106 | 2160 |
| | (0.063) [0.000] | (0.056) [0.000] | (0.104) [0.052] | (0.124) [0.489] | (0.107) [0.320] | |
| Observed shame (rev.) (Z) | 0.267*** | 0.194*** | 0.056 | –0.008 | 0.021 | 2251 |
| | (0.066) [0.000] | (0.056) [0.000] | (0.117) [0.631] | (0.136) [0.955] | (0.120) [0.861] | |
| Bullying: teasing (Z) | 0.107 | 0.036 | 0.091 | –0.018 | –0.116 | 2162 |
| | (0.074) [0.148] | (0.065) [0.583] | (0.133) [0.496] | (0.165) [0.914] | (0.149) [0.435] | |
| Bullying: intimidation/harassment (Z) | –0.064 | –0.028 | –0.022 | 0.003 | –0.115 | 2162 |
| | (0.062) [0.302] | (0.055) [0.615] | (0.131) [0.868] | (0.156) [0.982] | (0.129) [0.371] | |
| Network (Z) | 0.192** | 0.235*** | 0.020 | –0.176 | 0.073 | 2162 |
| | (0.092) [0.037] | (0.079) [0.003] | (0.133) [0.882] | (0.172) [0.306] | (0.154) [0.635] | |

*Notes:* Each row represents a regression of a different outcome. Each column represents the coefficient on each term. *Newly menstr.* takes the value 1 when a girl was not menstruating at baseline, but was menstruating by endline, and 0 otherwise. All regressions include stratum fixed effects and the baseline controls selected by double LASSO, and cluster standard errors at the school level.

**Table B.1.23:** *Effects on teachers*

| VARIABLES | (1) Progressive Menstruation Attitudes (Z) | (2) Mens. Knowledge Index (Z) | (3) Progressive Gender Attitudes (Z) | (4) Absent Last Month (=1) |
|---|---|---|---|---|
| Base Only | 0.6709*** | 0.1554 | 0.1129 | -0.0762 |
| | (0.0864) | (0.1139) | (0.1222) | (0.0547) |
| Base + YGL | 0.6839*** | 0.1864** | 0.1820 | 0.0051 |
| | (0.0876) | (0.0929) | (0.1118) | (0.0479) |
| Observations | 743 | 743 | 779 | 743 |
| Data source | Teachers | Teachers | Teachers | Teachers |
| Control mean | 0 | 0 | 0 | .497 |
| Control SD | 1 | 1 | 1 | .501 |
| p: Treated = 0 | <0.001 | .0481 | .13 | .6227 |
| p: Base Only = Base + YGL | .8157 | .7599 | .505 | .1093 |

*Notes:* The table displays results from the teachers' survey in endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and a dummy for a primary schools. *Progressive menstruation attitudes* combines questions about teachers' attitudes towards menstruation, e.g., should girls be allowed to go to school when menstruating. *Mens. knowledge index*: index of menstruation knowledge using a series of true or false questions. *Progressive gender attitudes*: index of progressive gender attitudes, e.g., using agreement with questions like "Boys are better leaders than girls". *Absent last month*=1 when the teacher reported being absent at least once last month, and is 0 otherwise.

**Table B.1.24:** *Changes in academic performance of nominated YGLs*

| | Change in grade (/20) | Change in class rank |
|---|---|---|
| Intercept | −1.173*** | 6.429*** |
| | (0.205) | (0.820) |
| Num.Obs. | 191 | 191 |
| Grade/rank at baseline | 11.7 | 4.8 |

*Notes:* Sample includes all girls who were selected as YGLs (in January 2022) *and* interviewed at endline (in summer 2023) and who were in school at endline. 255 YGLs were selected, 46 were not found at endline, and 18 of those found were no longer in school and thus did not have school marks to elicit. During YGL nomination and at endline, girls were asked their school marks (out of a maximum of 20) and their rank in class. Column (1) shows the change in school marks from nomination to endline. Column (2) shows the change in class rank from nomination to endline (with a higher rank indicating relatively worse academic performance). Regressions control for stratum fixed effects, and standard errors are clustered at the school level. Standard errors cluster at the school level. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

**Table B.1.25:** *Heterogeneity as predicted by generic machine learning: Best Linear Predictor*

| | (1) Treated pooled | | (2) Base + YGL | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HTE ($\beta_2$) | ATE ($\beta_1$) | HTE ($\beta_2$) |
| Learning test (Z) | .037** | .803 | .118 | .241 |
| Achievement tests only (Z) | .018** | .353 | .086* | .661 |
| Cognitive tests only (Z) | .134 | .203 | .302 | .138 |
| Math fluency (Z) | .060* | .600 | .136 | .433 |
| Reading fluency (Z) | .093* | .944 | .239 | .105 |
| Listening (Z) | .518 | .617 | .695 | .679 |
| Comprehension (Z) | .501 | .422 | .678 | .569 |
| Digit span (Z) | .409 | .494 | .659 | .194 |
| Reverse digit span (Z) | .229 | .971 | .148 | .201 |
| Progressed one class (=1) | .070* | .462 | .117 | .418 |
| Enrolled (=1) | .633 | .361 | .977 | .006*** |
| Absent in last 3 months (=1) | .577 | .293 | .358 | .342 |
| Hygiene knowl. (Z) | <0.001*** | .488 | <0.001*** | .777 |
| Menstr. knowl. (Z) | <0.001*** | .033** | <0.001*** | .011** |
| Hygiene behavior (Z) | .003*** | .553 | <0.001*** | .935 |
| Menstr. behavior (Z) | <0.001*** | .083* | <0.001*** | .397 |
| Observed hygiene: Girls (Z) | .671 | .124 | .508 | .587 |
| Received info. (Z) | <0.001*** | .102 | <0.001*** | .030** |
| Willingness to speak (Z) | <0.001*** | .886 | <0.001*** | .253 |
| Attitudes (Z) | <0.001*** | .324 | <0.001*** | .434 |
| Norms (Z) | <0.001*** | .681 | <0.001*** | .407 |
| Shame response to vignettes (rev.) (Z) | <0.001*** | .326 | <0.001*** | .821 |
| Observed shame (Z) | <0.001*** | .206 | .001*** | .514 |
| Bullying: Self-report (Z) | .980 | .235 | .728 | .953 |
| Heart rate (Z) | .048** | .352 | .038** | .890 |
| General health (Z) | .949 | .375 | .962 | .145 |
| UTI health (Z) | .613 | .419 | .809 | .960 |
| Mental health (Z) | .595 | .901 | .631 | .289 |
| Self-esteem (Z) | .722 | .246 | .826 | .215 |
| Network (Z) | .008*** | .475 | .002*** | .359 |

*Notes:* The table presents the results of the Best Linear Predictor of the Conditional Average Treatment Effect (CATE) on the ML proxies following Welz et al. (2022) over 1000 splits ($\alpha = .05$). We report the p-value for the hypothesis that the parameter is equal to zero against the two-sided alternative of the coefficients $\beta_1$ and $\beta_2$, which correspond to the ATE and heterogeneity loading (HTE) parameters in the *Best Linear Predictor*. * p< 0.1, ** p < 0.05, *** p < 0.01. The first column shows the outcomes from the girls' endline survey, and each row represents one estimation. In (1) we test heterogeneity comparing all the treated girls against the control, and in (2) we test heterogeneity comparing the girls in Base+YGL schools against the control.

**Table B.1.26:** *Heterogeneous effects by YGL positive deviance*

| | Base Only | Base + YGL | Base + YGL*Pos. Deviance Index | N |
|---|---|---|---|---|
| Learning test (Z) | 0.203*** | 0.107** | 0.095** | 2162 |
| | (0.072) [0.004] | (0.052) [0.042] | (0.038) [0.012] | |
| Progressed one class (=1) | 0.106*** | 0.070** | –0.029 | 2251 |
| | (0.039) [0.007] | (0.035) [0.044] | (0.018) [0.106] | |
| Enrolled (=1) | –0.021 | –0.023 | –0.004 | 2251 |
| | (0.024) [0.375] | (0.022) [0.296] | (0.016) [0.796] | |
| Absent in last 3 months (=1) | –0.006 | –0.033 | –0.024 | 1820 |
| | (0.030) [0.850] | (0.029) [0.258] | (0.020) [0.233] | |
| General health (Z) | –0.033 | 0.014 | –0.006 | 2251 |
| | (0.060) [0.586] | (0.048) [0.763] | (0.033) [0.865] | |
| UTI health (Z) | –0.085 | –0.003 | 0.067** | 2251 |
| | (0.053) [0.109] | (0.043) [0.944] | (0.034) [0.050] | |
| Mental health (Z) | 0.017 | 0.051 | 0.081* | 2251 |
| | (0.062) [0.784] | (0.060) [0.394] | (0.046) [0.080] | |
| Self-esteem (Z) | 0.021 | 0.013 | 0.015 | 2251 |
| | (0.068) [0.758] | (0.052) [0.803] | (0.036) [0.680] | |
| Heart rate (Z) | –0.128** | –0.125** | –0.045 | 1899 |
| | (0.057) [0.024] | (0.051) [0.014] | (0.033) [0.177] | |
| Hygiene knowl. (Z) | 0.340*** | 0.436*** | –0.002 | 2162 |
| | (0.077) [0.000] | (0.063) [0.000] | (0.041) [0.969] | |
| Menstr. knowl. (Z) | 0.314*** | 0.345*** | –0.019 | 2162 |
| | (0.057) [0.000] | (0.049) [0.000] | (0.035) [0.584] | |
| Hygiene behavior (Z) | 0.102 | 0.314*** | 0.054 | 2162 |
| | (0.086) [0.236] | (0.073) [0.000] | (0.040) [0.177] | |
| Menstr. behavior (Z) | 0.358*** | 0.556*** | 0.029 | 2251 |
| | (0.069) [0.000] | (0.053) [0.000] | (0.039) [0.464] | |
| Observed hygiene: Girls (Z) | –0.063 | 0.056 | –0.062 | 2162 |
| | (0.062) [0.311] | (0.055) [0.308] | (0.038) [0.103] | |
| Combined stigma index (Z) | 0.396*** | 0.731*** | –0.015 | 2251 |
| | (0.065) [0.000] | (0.055) [0.000] | (0.041) [0.709] | |
| Received info. (Z) | 0.193** | 0.532*** | –0.072* | 2251 |
| | (0.077) [0.012] | (0.063) [0.000] | (0.039) [0.062] | |
| Willingness to speak (Z) | 0.243*** | 0.719*** | 0.024 | 2245 |
| | (0.076) [0.001] | (0.065) [0.000] | (0.048) [0.617] | |
| Attitudes (Z) | 0.360*** | 0.559*** | 0.006 | 2249 |
| | (0.065) [0.000] | (0.058) [0.000] | (0.045) [0.901] | |
| Norms (Z) | 0.274*** | 0.468*** | –0.011 | 2162 |
| | (0.060) [0.000] | (0.049) [0.000] | (0.039) [0.771] | |
| Shame response to vignettes (rev.) (Z) | 0.303*** | 0.442*** | –0.036 | 2160 |
| | (0.058) [0.000] | (0.054) [0.000] | (0.038) [0.355] | |
| Observed shame (rev.) (Z) | 0.288*** | 0.221*** | –0.016 | 2251 |
| | (0.060) [0.000] | (0.053) [0.000] | (0.043) [0.722] | |
| Bullying: teasing (Z) | 0.102 | 0.016 | –0.053 | 2162 |
| | (0.065) [0.118] | (0.058) [0.782] | (0.040) [0.188] | |
| Bullying: intimidation/harassment (Z) | –0.085* | –0.067* | –0.078*** | 2162 |
| | (0.045) [0.061] | (0.040) [0.092] | (0.027) [0.004] | |
| Network (Z) | 0.167* | 0.259*** | 0.004 | 2162 |
| | (0.086) [0.053] | (0.074) [0.000] | (0.038) [0.907] | |

*Notes:* Each row is one regression for a given outcome. *Positive deviance index (Z)* is a school-level index constructed based on how high the selected *YGLs* score on an index of positive deviance, composed of willingness to adopt new behaviors, tell others about new things, discuss sensitive topics, convince others of new behaviors, and ease at explaining how to use a sanitary pad. The first column shows the coefficient on *Base Only*, the second column shows the coefficient on *Base+YGL*, and the third column shows the coefficient on (*Base+YGL* × *Positive Deviance Index*). All regressions control for stratum fixed effects, include LASSO selected controls, and cluster standard errors at the school level. Standard errors are in parentheses, and $p$-values are in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.1.27:** *Heterogeneous effects by YGL positive deviance, relative to positive deviance of other candidates in the same school*

| | Base Only | Base + YGL | Base + YGL × Rel. Pos. Deviance | N |
|---|---|---|---|---|
| Learning test (Z) | 0.204*** | 0.111** | 0.025 | 2162 |
| | (0.072) [0.005] | (0.053) [0.037] | (0.062) [0.685] | |
| Progressed one class (=1) | 0.108*** | 0.076** | –0.080*** | 2251 |
| | (0.039) [0.006] | (0.035) [0.031] | (0.024) [0.001] | |
| Enrolled (=1) | –0.020 | –0.019 | –0.041* | 2251 |
| | (0.024) [0.401] | (0.022) [0.383] | (0.023) [0.070] | |
| Absent in last 3 months (=1) | –0.006 | –0.033 | –0.017 | 1820 |
| | (0.030) [0.857] | (0.029) [0.254] | (0.030) [0.572] | |
| General health (Z) | –0.030 | 0.024 | –0.106*** | 2251 |
| | (0.060) [0.615] | (0.047) [0.612] | (0.036) [0.003] | |
| UTI health (Z) | –0.085 | 0.002 | –0.005 | 2251 |
| | (0.054) [0.116] | (0.043) [0.963] | (0.058) [0.928] | |
| Mental health (Z) | 0.016 | 0.054 | 0.033 | 2251 |
| | (0.062) [0.792] | (0.061) [0.381] | (0.061) [0.585] | |
| Self-esteem (Z) | 0.021 | 0.013 | 0.011 | 2251 |
| | (0.067) [0.760] | (0.053) [0.805] | (0.049) [0.827] | |
| Heart rate (Z) | –0.132** | –0.136*** | 0.081* | 1899 |
| | (0.057) [0.021] | (0.050) [0.007] | (0.047) [0.085] | |
| Hygiene knowl. (Z) | 0.337*** | 0.427*** | 0.092** | 2162 |
| | (0.077) [0.000] | (0.062) [0.000] | (0.045) [0.041] | |
| Menstr. knowl. (Z) | 0.311*** | 0.334*** | 0.100*** | 2162 |
| | (0.057) [0.000] | (0.049) [0.000] | (0.038) [0.008] | |
| Hygiene behavior (Z) | 0.100 | 0.310*** | 0.081 | 2162 |
| | (0.086) [0.242] | (0.074) [0.000] | (0.078) [0.296] | |
| Menstr. behavior (Z) | 0.363*** | 0.574*** | –0.155*** | 2251 |
| | (0.069) [0.000] | (0.054) [0.000] | (0.050) [0.002] | |
| Observed hygiene: Girls (Z) | –0.066 | 0.044 | 0.081* | 2162 |
| | (0.063) [0.300] | (0.056) [0.427] | (0.045) [0.074] | |
| Combined stigma index (Z) | 0.395*** | 0.726*** | 0.039 | 2251 |
| | (0.065) [0.000] | (0.055) [0.000] | (0.072) [0.589] | |
| Received info. (Z) | 0.193** | 0.529*** | –0.017 | 2251 |
| | (0.078) [0.013] | (0.063) [0.000] | (0.056) [0.768] | |
| Willingness to speak (Z) | 0.241*** | 0.713*** | 0.077 | 2245 |
| | (0.076) [0.001] | (0.064) [0.000] | (0.083) [0.351] | |
| Attitudes (Z) | 0.360*** | 0.559*** | –0.002 | 2249 |
| | (0.065) [0.000] | (0.057) [0.000] | (0.077) [0.980] | |
| Norms (Z) | 0.275*** | 0.469*** | –0.020 | 2162 |
| | (0.060) [0.000] | (0.049) [0.000] | (0.056) [0.721] | |
| Shame response to vignettes (rev.) (Z) | 0.303*** | 0.440*** | –0.004 | 2160 |
| | (0.059) [0.000] | (0.057) [0.000] | (0.084) [0.959] | |
| Observed shame (rev.) (Z) | 0.287*** | 0.213*** | 0.079* | 2251 |
| | (0.061) [0.000] | (0.054) [0.000] | (0.045) [0.077] | |
| Bullying: teasing (Z) | 0.101 | 0.011 | 0.021 | 2162 |
| | (0.065) [0.123] | (0.058) [0.854] | (0.053) [0.688] | |
| Bullying: intimidation/harassment (Z) | –0.083* | –0.065 | –0.084** | 2162 |
| | (0.045) [0.064] | (0.040) [0.102] | (0.041) [0.039] | |
| Network (Z) | 0.169** | 0.267*** | –0.091 | 2162 |
| | (0.086) [0.050] | (0.075) [0.000] | (0.057) [0.110] | |

*Notes:* Each row is one regression for a given outcome. *Relative Positive Deviance Index (Z)* is a school-level index constructed based on how high the selected *YGLs* score on an index of positive deviance, *relative to* the full pool of candidate YGLs in that same school. The first column shows the coefficient on *Base Only*, the second column shows the coefficient on *Base+YGL*, and the third column shows the coefficient on (*Base+YGL* × *Positive deviance index*). All regressions control for stratum fixed effects, include LASSO selected controls, and cluster standard errors at the school level. Standard errors are in parentheses, and *p*-values are in brackets. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

**Table B.1.28:** *Effects on marriage and fertility*

| VARIABLES | (1) Has been married (=1) | (2) Pregnant now or last year (=1) | (3) Dropout because of baby (=1) |
|---|---|---|---|
| Base Only | 0.009 | 0.016** | 0.019*** |
| | (0.008) | (0.007) | (0.006) |
| Base + YGL | 0.003 | 0.009 | 0.000 |
| | (0.007) | (0.007) | (0.005) |
| Observations | 2,282 | 2,282 | 2,251 |
| Data source | Mothers | Mothers | Girls |
| Control mean | .024 | .02 | .013 |
| Control SD | .152 | .141 | .113 |
| p: Treated = 0 | .4659 | .0665 | .1861 |
| p: Base Only = Base + YGL | .4125 | .3165 | <0.001 |

*Notes:* Table shows effects on marriage, pregnancy, and dropout due to pregnancy at endline. Columns (1) and (2) come from mothers' reports of the marriage and pregnancy status of their daughters. Column (3) comes from girls' own reports. Column (1) outcome is 1 if the girl has been married by endline, 0 otherwise. Column (2) is 1 if the girl has been pregnant in the last year or is currently pregnant, 0 otherwise. Column (3) is 1 if a girl reported dropping out of school because they had to look after a baby or because they were pregnant, 0 otherwise. All specifications include stratum fixed effects, baseline controls selected by double LASSO, and cluster standard errors at the school level.

## B.2 Data

### B.2.1 Construction of outcomes

This section documents the questions used to construct the main outcome indexes used in the paper.

Knowledge and behavior indexes are created by summing the number of correct answers. All other indexes are constructed by conducting an exploratory factor analysis of all questions in a module, and combining using factor loadings based on a single factor.

**General health**. Girls are asked how much the following things have been difficult for them over the last month, with possible answers as (i) never; (ii) almost never; (iii) sometimes; (iv) often; or (v) almost always:

- It's hard for me to walk more than a block.
- It's hard for me to run.
- It is difficult for me to play sports or exercise.
- It's hard for me to lift something heavy.
- It's difficult for me to take a bath or shower by myself.
- It is difficult for me to do household chores.
- I have pain or aches.
- I have little energy.

The index is constructed so that a higher value indicates *better* health (i.e., fewer difficulties).

**UTI health**. Girls are asked: "Please indicate if you have had the following symptoms for a few days in the last month." (Answer options: (i) No, (ii) Yes, light symptoms, (iii) Yes, moderate symptoms,

(iv) Yes, serious symptoms.):

- Frequent urination of small volumes of urine (going to the toilet very often)
- Urgent urination (strong, uncontrollable urge to urinate)
- Painful or burning sensation when urinating
- Sensation of incomplete emptying of the bladder after urination
- Feeling of pain or discomfort in the lower abdomen (suprapubic region)
- Presence of blood in the urine

They are also asked "Did these symptoms occur at the same time (on the same days)?". The index is constructed so that a higher value indicates *better* health, i.e., fewer symptoms of urinary tract infections.

**Mental health**. Girls are asked how many times they felt the following emotions over the last 7 days, based on an adapted version of the CESD index (Radloff, 1977):

- How many days have you been bothered by things that don't usually bother you?
- How many days have you not wanted to eat as usual?
- How many days did manage to shake sadness off, even with the help of family and friends?
- How many days were you happy? (rev.)
- How many days you could concentrate on the things you were doing (rev.)
- How many days did you feel discouraged?
- How many days did you feel that everything you did was difficult?
- How many days have you been full of hope for the future? (rev.)
- How many days have you felt so moody that you didn't want to talk to anyone?
- How many days were you afraid?
- How many days was your sleep restless?
- How many days have you been happy? (rev.)
- How many days have you spoken less than usual?
- How many days have you felt alone?
- How many days did others look out for you to chat? (rev.)
- How many days did you want to do your daily activities? (rev.)
- How many days have you cried so hard you couldn't stop?
- How many days have you felt sad?
- How many days have you told yourself that other girls don't like you?
- How many days have you lacked the desire to do things?

The index is constructed so that a higher value indicates *better* mental health (i.e., fewer symptoms of depression).

**Self esteem index**. Girls are asked "For each description, indicate how much it is true for you." Possible answers: Strongly disagree; disagree; agree; strongly agree. Questions are:

- Do you think you have a certain number of good qualities?
- All things considered, you are inclined to consider yourself a failure (rev.)
- you have a positive attitude towards yourself
- Overall, you are satisfied with yourself
- You wish you had more respect for yourself

- Sometimes you feel really useless (rev.)
- Sometimes you think you're good for nothing (rev.)

The index is constructed so that a higher value indicates *better* self esteem, i.e., more belief in oneself.

**Menstruation-related knowledge**. Index is constructed as the sum of correct answers to:

- How often does a girl usually get her period? (Ans: once per month / 28 days)
- How many days does the period last on average? (Ans: 3-7 days)
- At what age do women stop having their periods? (Ans: 40-60 years)
- What are the signs before or during your period? (Ans: at least 2 correct)
- What are healthy behaviors to adopt during your period? (Ans: at least 3 correct)
- T/F: When girls reach menarche, they can become pregnant from sexual intercourse?
- T/F: when blood flows during menstruation, it usually means that a woman is pregnant?
- T/F: is it common to experience pain and discomfort during menstruation?
- T/F: Menstrual blood is the destruction of tissue surrounding the uterus that occurs when fertilization does not occur.
- T/F: Is it the first sexual intercourse that causes a period for young girls?
- T/F: You should not shower during your period
- T/F: periods are symptoms of an illness
- List the different types of menstrual products that girls can use. (Ans: names at least 3)
- In your opinion, what are the most hygienic products to use during your period? (Ans: at least 2 correct)
- How many times a day should girls change/wash cloths or other menstrual products when they are on their period? (Ans: at least 3)
- How many times a day should girls change reusable pads when they are on their period? (Ans: at least 2)
- How often should girls wash when they are on their period? (Ans: at least 3x per day)
- What should you wash with during your period? (Ans: with soap)
- How much does one reusable sanitary napkin (made by a seamstress) cost? (Ans: between 5000 and 10,000 Ary)

**Hygiene knowledge**. Index is constructed as the sum of correct answers to:

- Name 3 good habits you can adopt that can help ensure you don't get sick.
- Name all the things you need to do to ensure you don't contaminate the water in a water container (Ans: at least 2 correct)
- Name all the things you can do to improve the hygiene of latrine use.
- To your knowledge, how many times a day should we wash our hands? (Ans: at least 3)
- To your knowledge, when are the most important times to wash your hands? (Ans: at least 2 correct)
- Can you name two methods for making water drinkable?

**Menstruation-related behaviour**. Index is constructed as the sum of "correct" answers to:

- During your last period, were you able to wash and change in private at home?
- During your last period, were you able to wash and change in private at school?
- Have you used reusable sanitary pads (made by a seamstress) in the last 3 months?

- In general, how often do you change your sanitary pads or other sanitary products when you have your period? (Ans: at least 3 times per day)
- The last time you used sanitary pads, where you wash them? (Ans: not in the river)
- The last time you used sanitary pads, where did you dry them? (Ans: outside in the sun)
- Where do you wash your underwear? (Ans: not in the river)
- Where do you dry your underwear? (Ans: outside in the sunshine)
- The last time you had your period, how many times did you wash during a typical day at home? (Ans: at least 3)
- The last time you had your period, how many times did you wash with soap on a typical day at home? (Ans: at least 3)
- The last time you had your period during term time, how many times did you wash during a typical day at school? (Ans: at least 3)
- The last time you had your period during school term, how many times did you wash with soap on a typical day at school? (Ans: at least 3)
- Generally, where do you dry your sanitary products? (Ans: outside in the sun)

**Hygiene behaviour**. Index is constructed as the sum of "correct" answers to:

- Over the past week, where did the water you used to wash your hands at home come from? (Ans: from clean source; from tap; from cleaned source; from village water system)
- During the last week, what did you wash your hands with at home? (Ans: with soap)
- During the last week, when did you wash your hands at home? (Ans: at least 3 moments)
- Think about the last time you went to urinate or defecate when you were at home. Where did you go? (Ans: in a toilet)
- How many times during the day did you wash your hands the last time you spent a day at home? (Ans: at least 3 times)
- During the last week you were at school, where did the water you used to wash your hands come from? (Ans: from clean source; from tap; from cleaned source; from village water system)
- During the last week you were at school, what did you wash your hands with? (Ans: soap)
- During the last week you were at school, when did you wash your hands? (Ans: at least 3 moments)
- Think about the last time you went to urinate or defecate at school. Where did you go? (Ans: in a toilet)
- How many times during the day did you wash your hands the last time you spent a day at school? (Ans: at least 3)
- During the last week, how many times have you gone swimming? Think about the last time you went swimming. Where did you go? (Ans: not in the river)

**Observed hygiene (girls)**. For the first 5 questions, the enumerator observes whether the girl is (i) very clean; (ii) quite clean; (iii) slightly dirty; or (iv) very dirty for the following: (i) hands, (ii) face, (iii) neck, (iv) clothes, (v) hair. The 6th question is whether the girl is wearing shoes. The index is constructed so that a higher value indicates *better* hygiene.

**Observed hygiene (home)**. Enumerator observes the following about the girl's homestead, with (+) indicating a positive hygienic observation, (-) indicating a negative reverse coded observation:

- Observe the presence of water at the place of hand washing. Check that the tap/pump, basin/bucket, water tank, or similar objects have/contain water. (+)
- Is there soap, detergent or ash at the hand-washing area? (+)
- Human and animal feces in and around the house (-)
- Waste accumulated around the house (-)
- Stagnant water around the house (-)
- Stables, pigs or kennels below or to the side of the house (-)
- Lack of ventilation inside the house (-)
- The terrace and the garden are clean (+)
- In the house there are separate spaces for cooking, sleeping and socializing (+)
- The house is neat (+)
- The kitchen area is neat (+)
- Animals have access to areas where food is prepared (-)
- There is water served in the kitchen (+)
- The dishes in the kitchen are clean (+)

The index is constructed so that a higher value indicates *better* hygiene.

**Observed hygiene (school)**. Enumerators observe the following about the school environment, all of which indicate more hygienic environment:

- Is there a cover for the holes in the toilet?
- Is there soap available to use at the water point?
- Are there garbage holes in the ground?
- Is the garbage separated?
- Are there trash collections in classrooms?
- Is there a water point for hand washing with soap?

The index is constructed so that a higher value indicates *better* hygiene.

**Received information about menstrual hygiene**. Girls are asked the following questions:

- In the past 3 months, how many times have you heard a classmate, or another student at school, talk about menstrual hygiene?
- In the past 3 months, how many times have you heard someone in your family talk about menstrual hygiene?
- In the past 3 months, how many times have you heard a teacher talk about menstrual hygiene?
- In the past 3 months, have you asked anyone a question about menstrual hygiene?
- In the past 3 months, has anyone given you information about menstrual hygiene, whether you asked for it or not?
- In the last 3 months, has [MOTHER] given you information about menstrual hygiene?

**Willingness to speak about menstruation**. Answer options: (i) strongly disagree; (ii) disagree; (iii) neither agree nor disagree; (iv) agree; (v) strongly agree. Girls are asked whether they would be willing to discuss...

- the importance of menstrual hygiene in a casual chat with friends, if only girls were present.

- the importance of menstrual hygiene in a casual chat with friends, if both boys and girls were present.
- the importance of menstrual hygiene in front of my class in a classroom, if only girls were present.
- the importance of menstrual hygiene in front of my class, in a classroom, if boys and girls were present.

**Attitudes towards menstruation**. Each question is a Likert scale, with possible answers: (i) strongly disagree; (ii) disagree; (iii) neither agree nor disagree; (iv) agree; (v) strongly agree.

- "Girls should be allowed to go to school when they are on their period"?
- "Girls should be required to hide the fact that they have their period while at school." ? (rev)
- "People should discuss menstruation openly with others"?
- "Mothers should broach the subject of menstruation with daughters only after their first period" (rev)
- "It is prohibited to dry washable towels in a public place" (rev)
- "Girls should be punished in case their brothers discover its intimate effects during the period of menstruation" (rev)
- "Girls should feel embarrassed or ashamed when they have their period in public." (rev)
- "Boys should know about menstruation."
- "Boys should be allowed to laugh at topics related to menstruation" (rev)
- "Periods are natural phenomena."

The index is constructed so that a higher value indicates *more progressive* attitudes towards menstruation (e.g., being more likely to say that girls should be allowed to go to school during menstruation).

**Norms**. Girls were asked how many other girls and then other boys in their school would agree with the following three statements:

- Girls should be obliged to hide the fact that they are on their periods when they are at school.
- People should be able to openly discuss menstruation with others.
- Boys should be allowed to laugh about things related to menstruation.

To aid with the questions, we asked how many girls (or boys) out of 10 would agree with the statement, using 10 small stones and asking girls to move stones to the right to represent people agreeing and to the left to represent people disagreeing.

Girls were then asked about their teachers' opinions of the same statements (answer options: all teachers; most teachers; half of teachers; only a few teachers; no teachers), and then their mother's opinion for the same statements (answer options: strongly disagree; disagree; neither disagree nor agree; agree; strongly agree). For teachers and mothers, we also asked perceived agreement with a fourth statement: "Girls should be allowed to come to school when they have their periods."

The index is constructed so that a higher value indicates that girls believe that others have *more progressive* attitudes towards menstruation (e.g., being more likely to say that girls should be allowed to go to school during menstruation).

**Shame in response to vignettes**. Enumerators describe 2 vignettes to girls:

1. Imagine that you went to school and you were speaking about periods with a friend. You turn around and realise that someone you don't know heard the whole conversation.

2. Imagine that you go to school and that someone who you don't know realises that you have started your periods, because he has seen a reusable sanitary pad in your bag.

After each vignette, girls are asked how much they feel the following emotions in response, with answer options: (i) I feel it strongly, (ii) I feel it a bit, (iii) I don't feel it at all

- Ashamed
- Indifferent (rev.)
- Humiliated
- Guilty
- That it's funny (rev.)
- Ridiculous
- That it's not a problem (rev.)

When used in the paper, the score is reversed so that a higher value indicates *less* shame in response to vignettes.

**Observed shame**. Enumerator is asked (after asking questions about menstruation) about the girl's behavior over the last 5 minutes of the survey:

- Was the girl looking at the floor when she spoke?
- Did the girl speak: (Quietly or very quietly)
- Did the girl laugh while answering or listening to the questions?
- How did the girl answer the questions? (Hesitated a bit or lots)
- Did the girl seem: (Uneasy or slightly uneasy)
- Did the girl seem: (More uneasy when menstruation was mentioned than other subjects)

When used in the paper, the index is constructed so that a higher value indicates *fewer* observable markers of shame.

**Network**. Girls are asked the following questions:

- Name your friends (how many girls; how many boys)
- Compared to the last school year, this year did you have [more friends vs less friends; girls and boys separately]
- Who are the girls or boys you play with at school? (how many girls; how many boys)
- Who are the girls or boys with whom you share the way to school? (how many girls; how many boys)
- Who are the girls or boys you talk to during recess? (how many girls; how many boys)
- Are there any girls or boys with whom you organize to clean or do the garden at school? (If yes: how many girls; how many boys)
- Do you sometimes share things (snacks, bikes, money to buy a snack, etc.) with your friends? (If yes: how many girls; how many boys)
- How often do you share with these people? (Every day, several times a week, once a week, once every 2 weeks, once a month, less than once a month)
- Do you sometimes do homework or school projects together with your classmates?
- Do your classmates help each other with school work in other ways?
- How often? (Every day, several times a week, once a week, once every 2 weeks, once a month, less than once a month)

- Who are the girls or boys you talk to about menstruation? (how many girls; how many boys)

**Bullying**. Girls are asked about a random subset of 2-3 other girls in the same grade-level as them in their school. If the respondent doesn't know the other girl being asked about, we ask about additional girls (up to a maximum of 5). Only girls whom the respondent knows are included in the results. Girls are also asked about bullying towards themselves. Teachers are asked about a random subset of 3 girls in their class.

The indexes are constructed so that higher values indicate *more* or more severe bullying.

Light teasing questions include:

- How often has at least one other classmate laughed at [NAME] in the past month? (Never, rarely, some days, most days, almost every day)
- How often has at least one other classmate teased [NAME] in the past month? (Never, rarely, some days, most days, almost every day)
- Please indicate how many other classmates tease [NAME]?
- Please indicate the severity of the teasing she receives from class members? (No teasing, light teasing, moderate teasing, serious teasing)

Severe intimidation/harassment questions include:

- How often has at least one other classmate intimidated/harassed [NAME] in the past month? (Never, rarely, some days, most days, almost every day)
- Please indicate how many other classmates intimidate/harass [NAME]? (Never, rarely, some days, most days, almost every day)
- Please indicate the severity of the intimidation/harassment inflicted on her by class members? (No intimidation, light teasing, light bullying, moderate intimidation, serious intimidation such as violence)

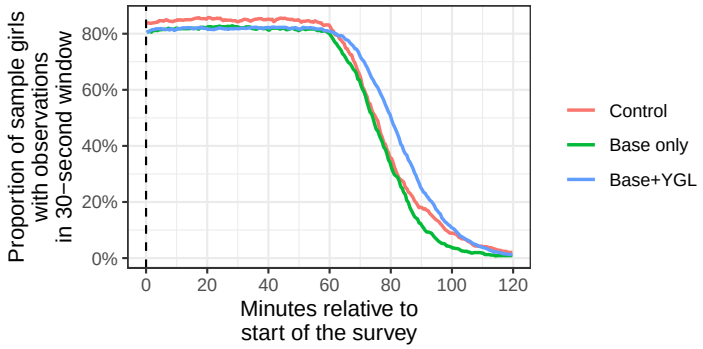
### B.2.2 Heart rate data

We use the *Scosche Rhythm 24* heart rate wristband monitor to measure girls' heart rate during the endline survey. This has been shown to have comparable accuracy to chest-strap monitors (Reece et al., 2021), but has the advantage of being significantly less invasive. Qualitative reports from enumerators indicate that girls frequently forgot that they were wearing the wristband by the end of the survey, suggesting that girls were not overly distracted or stressed by wearing the wristband itself.

*Sample*. In some cases, heart rate monitors were not fixed properly to girls' wrists, or the battery of monitors ran out on the field. Because of this, heart rate data is available for only 1904 girls. However, the proportion of data available is equal across treatment arms (Table B.1.5, column 6).

*Time-window restrictions*. For the main results in Figure 2.2, panel (a), we include all observations less than 120 minutes after the start of the survey. This includes almost all observations across all arms (see Figure B.2.6). We do not cut off at an earlier window in the main specification, because the average length of the survey is different across different treatments (e.g., because girls in *Base+YGL* schools are asked about the YGLs in their schools in a separate module), so any earlier cutoff could generate composition effects in the heart rate data. However, for tests of the *dynamics* over time (Figure 2.2 panel (b), and Table B.1.15), we use a 60 minute cutoff window. This avoids highly noisy time periods with few observations driving the results, since the number of girls with data available drops rapidly after 60 minutes because many girls finish the survey (see Figure B.2.6).

**Figure B.2.6:** *Number of girls with heart rate observations available in each 30-second window*




*Notes*: Shows the proportion of sample girls in each treatment arm who have heart rate data available in each 30-second window of the survey.

### B.2.3 Baseline variable definitions for LASSO

- *Girl's age (years):* Age of the girl in years.
- *Girl has a brother (=1):* Indicates whether the girl has a brother.
- *Distance to school (km):* Self-reported distance from the girl's home to her school in kilometers.
- *Time to school (min):* Self-reported time it takes for the girl to travel from home to school in minutes.
- *Was absent last 3 months (=1):* Indicates whether the girl was absent from school in the last three months (self-reported).
- *Days absent last 3 months:* The total number of days the girl was absent from school in the last three months (self-reported).
- *Years of school delay:* The number of years the girl is behind in her schooling compared to the typical age for her grade-level.
- *Father in household (=1):* Indicates whether the girl's father lives in the household.
- *Mother in household (=1):* Indicates whether the girl's mother lives in the household.
- *Knows how to read and write (=1) (guardian):* Indicates whether the girl's guardian can read and write.
- *No education (guardian) (=1):* Indicates whether the girl's guardian has no formal education.
- *Secondary or higher (guardian) (=1):* Indicates whether the girl's guardian has at least a secondary education.
- *Occupation: Agriculture (guardian) (=1):* Indicates whether the girl's guardian is employed in agriculture.
- *Household size:* The total number of people living in the girl's household.
- *Menstruation inhibiting school activity index (Z):* The sum of positive responses to whether the girl reported (over the last 3 months): (i) having missed school because of her period; (ii) having left school during the day because of her period; (iii) having not fully participated in school because of her period; (iv) having missed a social activity because of her period.
- *Menstrual health problems index (Z):* The sum of positive responses to whether the girl reported having had (over the last 3 months) (i) menstrual cramps, (ii) a urinal infection, and (iii) other health problems linked to menstruation.
- *Wealth index (Z):* is constructed from household module on asset ownership during the mother's interview.

Other indexes are constructed analogously to the endline outcomes documented in Section B.2.1.

### B.2.4 Other variable definitions

**Social desirability bias**. At baseline, we elicited a shortened version of the Crowne & Marlowe (1960) social desirability scale for the sample of girls, translated into the local language, Malagasy. The questions ask whether the respondent has a number of "too good to be true" traits (or the reverse), so that answering in the socially desirable way is likely to indicate a desire to please the interviewer. The questions used were:

1. It is sometimes difficult for me to continue my work if I am not encouraged. (-)
2. I sometimes feel resentful when I don't get what I want. (-)
3. On a few occasions, I have given up on doing something because I thought too little of my abilities. (-)

4. There were times when I wanted to rebel against the authorities, even though I knew they were right. (-)
5. No matter who I talk to, I always know how to listen.
6. There have been occasions when I have taken advantage of someone. (-)
7. I'm always willing to admit when I make a mistake.
8. I sometimes try to take revenge rather than forgive and forget. (-)
9. I am always polite, even to unpleasant people.
10. I have never been irritated when people expressed ideas very different from mine.
11. There were times when I was quite jealous of other people's good fortune. (-)
12. I sometimes get irritated by people who ask me for favors. (-)
13. I have deliberately said something that hurt someone. (-)

We construct an index of each girl's social desirability score using factor loadings, and then construct *High SDB*, which indicates whether a girl has an above-median social desirability score. This is the variable we use as an interaction in Table B.1.7.

**YGL positive deviance index**. During the YGL selection process, all girls nominated by teachers as potential candidates were asked a series of survey questions to evaluate their suitability for the role. The positive deviance index is constructed using a factor-analysis combination of the following questions:

1. How often do you tell your friends about new things you have learned and new experiences?
2. How much do you agree with the following statement: "If I learn new knowledge that I believe is good for me, I am ready to adopt it even if it is very different from my habits and those around me (for example, your friends, your sister, your classmates)."
3. If someone talks to you about a sensitive subject (like menstruation or sexual health), do you feel comfortable and open?
4. During the previous year, have you discussed and expressed your opinions in front of those around you about a sensitive subject such as menstruation or sexual health?
5. During these moments of discussion, overall, did you feel comfortable discussing the sensitive subject?
6. Imagine that you are discussing and expressing your opinions to those around you on a sensitive subject such as menstruation or sexual health. If the people around you feel uncomfortable or show embarrassment, frustration or anger, what action(s) would you take? (Options: I would insist / try to convince them more; I would ask them to talk to someone else; I would suggest another time to speak; I would give up)
7. Would you feel comfortable explaining how to use a sanitary pad in front of a group of *all the girls in your class*?
    - (If no) Would you feel comfortable explaining how to use a sanitary pad in front of a smaller group of girls?
    - (If yes) Would you feel comfortable explaining how to use a sanitary pad in front of a larger group of girls that also includes girls in other classes?

To construct a school-level indicator of YGL's positive deviance, we take the average positive deviance of all YGLs selected in the school in the first year of the program. To construct the *relative* measure

of YGL's positive deviance, we take this measure and subtract the average positive deviance of all the candidates who were in the potential pool of YGLs (whether they were selected or not).

### B.2.5 Tracking and attrition weights

*Tracking strategy*. Regular data collection occurred between May and July 2023. This phase included repeated visits to absent girls, mothers, and teachers. An intensive tracking phase was then organized in October 2023 following a multi-pronged approach to minimize attrition.

We revisited all schools where there was still a relatively large share of girls to track after the regular data collection for any of the following reasons: (i) schools where many girls had been absent during regular tracking because of seasonal work in mines, and where girls were expected to return to their homestead in October; (ii) schools where there were many refusals because the regular survey visit was too close to exams, and girls asked the team to return later; (iii) schools for which the survey team did not have enough time for the girls' survey during the regular phase; and (iv) schools where fewer than two teachers were surveyed in the regular tracking phase.

In addition to this set of schools, we randomly selected 19 of the remaining 62 schools where at least one girl was missing, and included these 19 schools for intensive tracking. This resulted in a total sample of 83 schools to be revisited.

In the revisited schools, all remaining girl, mother, and teacher surveys were targeted for completion. Some teams targeted the original school locality, while in parallel, other teams looked for girls who had moved in their migration destination. This latter step involved in-person interviews in common migration destinations (e.g., the capital city Antananarivo, and regions nearby the study region of Amoron'i Mania), along with phone surveys (for 3.7% of the sample) in the remaining destinations.

*Attrition weights*. Given the random selection of a subset of schools for intensive tracking, we can calculate *effective tracking rates* (Orr et al., 2003). Intensive tracking was successful in finding a very large share of targeted girls, leading to an effective tracking rate of 99%.

We calculate attrition weights that account for the random selection. Attrition weights are 1 for all girls found during regular tracking, or in schools that were selected for revisits during intensive tracking with probability 1 (as described above). Attrition weights are 62/19 for the girls in the additional schools randomly selected for intensive tracking.

Among the schools randomly selected for intensive tracking, 9 had a missing teacher, among the total of 38 remaining schools with missing teachers, so the attrition weight for teachers is set at 38/9 for teachers from these schools, and 1 for other teachers.

The results adjusting for attrition weights are shown in Table B.1.10 and Table B.1.11.

## B.3 Spillovers

Our sample of 140 schools was selected to avoid cases where schools were very close to each other. Nevertheless, since these schools were selected from only 3 districts in the region of Amoron'i Mania, schools and households are in some cases close enough that there is a risk of spillover effects from treatment schools to control schools.

To evaluate whether such spillovers threaten our main results, we estimate whether control girls who live within a certain radius of other sampled treatment girls have different outcomes. We use data on the GPS coordinates of the girl's survey at either baseline or endline.[1] We then use the following specification for girl $i$ in school $j$:

$$Y_{ij} = \beta_0 + \beta_1 BaseOnly_j + \beta_2 BaseYGL_j + \delta(Control_j \times AboveMedTreatedWithin2km_{ij}) + \mathbf{X}'_{ij}\Gamma + \varepsilon_{ij}$$

where $Control_j = 1$ for girls in control schools, and 0 otherwise; $AboveMedTreatedWithin2km_{ij}$=1 if girl $i$ has more than the median number of treated girls in our sample who were interviewed within 2km of $i$'s own interview, and 0 otherwise.[2] The coefficient $\delta$ tests whether outcomes are different for control girls with more than 3 treated girls in the sample within a 2km radius, where 3 is the median number of treated girls within 2km. $AboveMedTreatedWithin2km_{ij}$=1 for 273/540 (51%) of control girls.

We do not randomly vary the proportion of treated schools in a given locality, so this specification does not generate *causal* estimates of treatment spillovers. For example, $AboveMedTreatedWithin2km_{ij}$ is likely to be positively correlated with population density in an area, so our estimates of $\delta$ may partially pick up the effects of such density.

We find no consistent evidence in favour of spillovers that would undermine our main results (Table B.3.1). There are significant negative coefficients on grade progression and on observed hygiene for girls, but not on other related proxies of these measures (learning, hygiene knowledge, or hygiene behavior).

[1] Due to errors with the survey tablets, only 56% girls had GPS coordinates available at baseline. We use these where available, and use endline coordinates if they are not available.

[2] All other variables are defined analogously to Equation 2.1. We use the *number* rather than the *proportion* of treated girls, because there are many girls who have 0 girls from other schools within 2km, meaning that the denominator would be 0.

**Table B.3.1:** *Spillovers: heterogeneity by whether control group have above-median treated individuals living within 2km*

| | Control & above med. treated in 2km | N |
|---|---|---|
| Learning test (Z) | –0.001 | 2162 |
| | (0.093) [0.993] | |
| Progressed one class (=1) | –0.100* | 2251 |
| | (0.060) [0.097] | |
| Enrolled (=1) | 0.023 | 2251 |
| | (0.035) [0.508] | |
| Absent in last 3 months (=1) | –0.061 | 1820 |
| | (0.055) [0.269] | |
| Hygiene knowl. (Z) | 0.110 | 2162 |
| | (0.098) [0.263] | |
| Menstr. knowl. (Z) | 0.006 | 2162 |
| | (0.103) [0.954] | |
| Hygiene behavior (Z) | –0.072 | 2162 |
| | (0.128) [0.575] | |
| Menstr. behavior (Z) | 0.113 | 2251 |
| | (0.083) [0.176] | |
| Observed hygiene: Girls (Z) | –0.302*** | 2162 |
| | (0.090) [0.001] | |
| Received info. (Z) | –0.107 | 2251 |
| | (0.111) [0.335] | |
| Willingness to speak (Z) | –0.122 | 2245 |
| | (0.100) [0.226] | |
| Attitudes (Z) | 0.043 | 2249 |
| | (0.093) [0.649] | |
| Norms (Z) | 0.043 | 2162 |
| | (0.083) [0.604] | |
| Shame response to vignettes (rev.) (Z) | 0.089 | 2160 |
| | (0.094) [0.340] | |
| Observed shame (rev.) (Z) | –0.103 | 2251 |
| | (0.092) [0.260] | |
| Bullying: Self-report (Z) | 0.107 | 2162 |
| | (0.072) [0.137] | |
| Bullying: Self-report Teasing only (Z) | 0.127 | 2162 |
| | (0.112) [0.256] | |
| Bullying: Self-report Intimidation/ harassment only (Z) | 0.090 | 2162 |
| | (0.080) [0.263] | |
| Bullying: About menstr. only (Z) | –0.112 | 2162 |
| | (0.092) [0.223] | |
| Heart rate (Z) | 0.008 | 1899 |
| | (0.099) [0.933] | |
| General health (Z) | –0.139 | 2251 |
| | (0.087) [0.111] | |
| UTI health (Z) | –0.088 | 2251 |
| | (0.069) [0.201] | |
| Mental health (Z) | –0.187 | 2251 |
| | (0.132) [0.157] | |
| Self-esteem (Z) | –0.149 | 2251 |
| | (0.098) [0.128] | |
| Network (Z) | –0.111 | 2162 |
| | (0.140) [0.425] | |

*Notes:* Shows indications of spillovers within 2km of girls' homes. Each row represents a regression of the outcome on *Base Only*, *Base+YGL*, and *Control & above med. treated in 2km*. Coefficients on *Base Only* and *Base+YGL* are not shown. *Control & above med. treated in 2km* $= 1$ when a girl is in the control group, and the number of girls from a treated school interviewed within 2km was above median. The median is 3. *Control & above med. treated in 2km* $= 1$ for 273/540 (51%) control girls.

# Appendix C

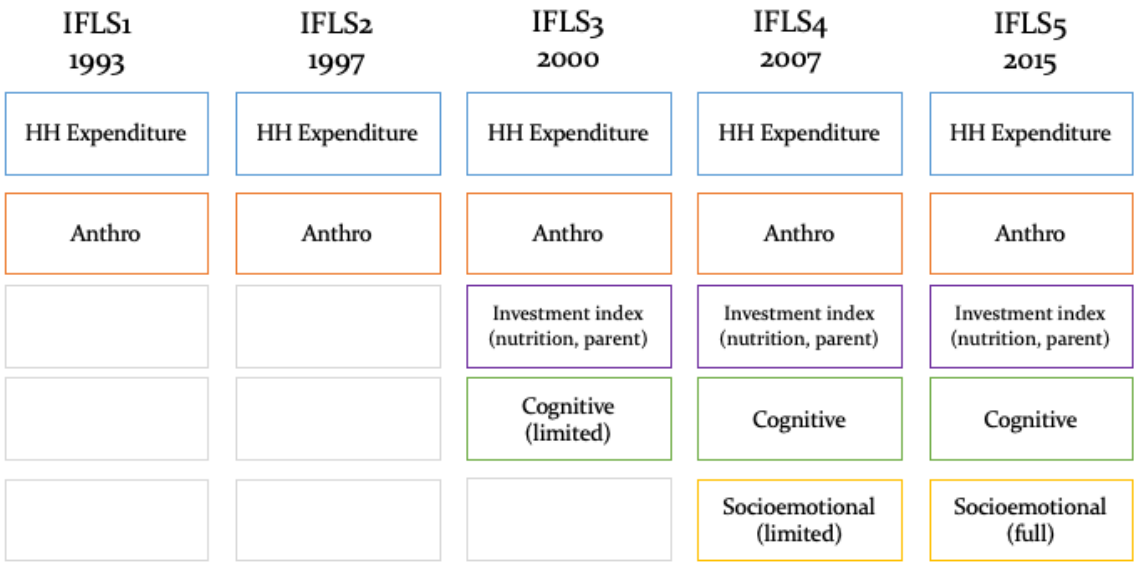
# Appendix for Chapter 3

## C.1 Data Appendix

### C.1.1 IFLS data

A summary of which outcome variables were available in each of the waves of IFLS (and the years of each survey) is found in Figure C.1.1. Figure C.1.2 shows the birth location of my main sample, which is concentrated in the 13 provinces covered by the IFLS data. The largest concentration of my sample is located in Western, Central, and Eastern Java.

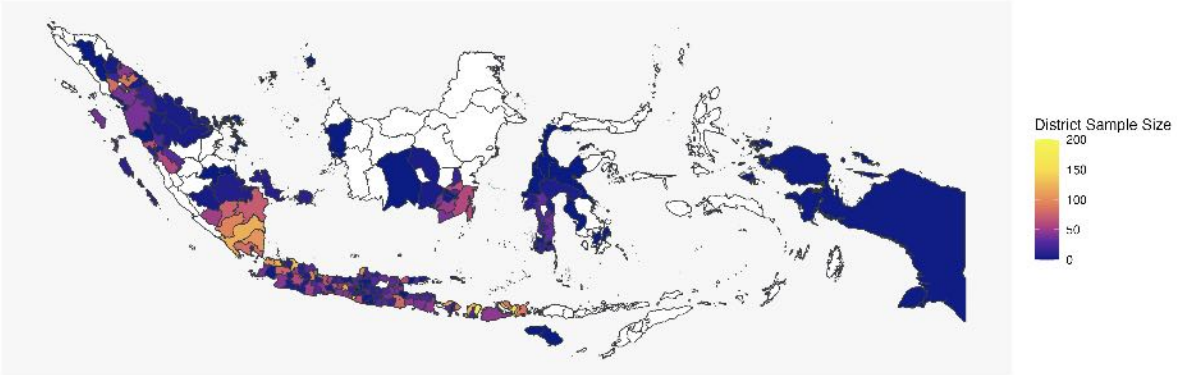
**Figure C.1.1:** *Summary of data availability for main outcome variables*




**Cognitive measures**. All IFLS respondents above the age of 15 are asked a number of questions that are designed to test cognitive ability:

1. *Raven's progressive matrices* (Raven, 2000) - participants have to fill in the gaps of a visual

**Figure C.1.2:** *Birth location of individuals in main sample*




matrix that follows a certain pattern. This is a common metric of "fluid intelligence".

2. *Maths - written* - participants have to answer 5 simple written maths questions.
3. *Word recall* - respondents were given a list of 10 words, and then asked to recall as many of the words as possible. They were asked to do this twice, once immediately after hearing the list, and another time 12-15 minutes later. (Score out of 10)
4. *Maths - oral* (IFLS5 only) - respondents were asked to serially subtract 7s from 100 (Score out of 10)
5. *Number series* (IFLS5 only) - participants are given a series of 3 numbers and are asked to fill in the fourth number in the series, e.g. "*2 4 6 ?*". The IFLS data then uses a Rasch scoring model to create a Woodcock-Johnson score based on a participant's ability to answer questions of varying difficulty.

The easier questions in the Raven's test and the written maths test are also given to children in IFLS households between the ages of 7-14. I use these to generate the results seen in Section 3.4.3.

**Socioemotional measures** - the full set of socioemotional measures is found in Table C.1.1. Depression measures are based on the CES-D scale. Personality measures are based on the Big 5 personality test.

**Household expenditure**. IFLS data contains detailed household expenditure data for each household. The categories asked are related to food, self-production, non-food, education, and rent (along with imputed rent). I impute the respondents' estimate of rent for home-owners. Total monthly expenditure is calculated by summing across all available categories. Total per capita monthly expenditure is calculated by dividing this by the reported household size. High quality price-adjustment data is not available for all waves of the IFLS data, so I use a variety of price-adjustments in the reported results:

1. *Nominal* - no price adjustment, available for all waves (IFLS1-5).
2. *"Real"* - the highest quality price adjustment available, from the IFLS data itself, that varies at the district (kabupaten) level, by rural/urban area, and by month-year. However this data is only available for IFLS2 and IFLS3.
3. *Inflation-adjusted ("Infl")* - a more comprehensive, but lower quality, price adjustment. It is

**Table C.1.1:** *Description of all socioemotional variables from IFLS*

| Module | Question Name | Pre question | Full Question | Reverse coded? | Included in IFLS4 | Answer Scale |
|---|---|---|---|---|---|---|
| Depression | Bothered | In the past week... | Bothered by things that usually don't bother me | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Concentration | In the past week... | I had trouble concentrating in what I was doing | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Depressed | In the past week... | I felt depressed | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Everything was an effort | In the past week... | I felt everything I did was an effort | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Hopeful | In the past week... | I felt hopeful about the future | - | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Fearful | In the past week... | I felt fearful | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Restless Sleep | In the past week... | My sleep was restless | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Happy | In the past week... | I was happy | - | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Lonely | In the past week... | I felt lonely | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Depression | Couldn't get going | In the past week... | I could not get going | Yes | Yes | 1 to 4 ("Rarely" to "Most of the time") |
| Personality | Agreeableness - Forgiving | I see myself as someone who ... | Has a forgiving nature | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Agreeableness - Considerate and Kind | I see myself as someone who ... | Is considerate and kind to almost everyone | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Agreeableness - Rude | I see myself as someone who ... | Is sometimes rude to others | Yes | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Conscientiousness - Thorough | I see myself as someone who ... | Does a thorough job | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Conscientiousness - Lazy | I see myself as someone who ... | Tends to be lazy | Yes | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Conscientiousness - Efficient | I see myself as someone who ... | Does things efficiently | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Extraversion - Talkative | I see myself as someone who ... | Is talkative | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Extraversion - Reserved | I see myself as someone who ... | Is reserved | Yes | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Extraversion - Outgoing | I see myself as someone who ... | Outgoing, sociable | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Neuroticism - Relaxed | I see myself as someone who ... | Is relaxed, handles stress well | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Neuroticism - Worries | I see myself as someone who ... | Worries a lot | Yes | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Neuroticism - Gets Nervous | I see myself as someone who ... | Gets nervous easily | Yes | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Openness - Original | I see myself as someone who ... | Is original, comes up with new ideas | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Openness - Active Imagination | I see myself as someone who ... | Has an active imagination | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Personality | Openness - Artistic and Aesthetic | I see myself as someone who ... | Values artistic, aesthetic experiences | - | - | 1 to 5 ("Disagree strongly" to "agree strongly") |
| Positive / Negative Affect | Frustrated | Yesterday, did you feel ... | Frustrated | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Sad | Yesterday, did you feel ... | Sad | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Enthusiastic | Yesterday, did you feel ... | Enthusiastic | - | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Lonely | Yesterday, did you feel ... | Lonely | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Content | Yesterday, did you feel ... | Content | - | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Worried | Yesterday, did you feel ... | Worried | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Bored | Yesterday, did you feel ... | Bored | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Happy | Yesterday, did you feel ... | Happy | - | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Angry | Yesterday, did you feel ... | Angry | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Tired | Yesterday, did you feel ... | Tired | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Stressed | Yesterday, did you feel ... | Stressed | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Positive / Negative Affect | Pain | Yesterday, did you feel ... | Pain | Yes | - | 1 to 5 ("Not at all" to "Very") |
| Subjective Wellbeing | Life Satisfaction (Cantrill Ladder) | How satisfied are you with your life as a whole | | - | - | 1 to 5 ("Not at all" to "Completely satisfied") |
| Subjective Wellbeing | Assessment of current situation | All things together how are things these days | | - | Yes | 1 to 4 ("Very unhappy" to "Very happy") |
| Risk/Time Preferences | Gamble Tradeoff 1 | Sure gain of 800k/month, or series of 50/50 gambles | | Yes | Yes | 1 to 5 (Least Risk Averse to Most Risk Averse) |
| Risk/Time Preferences | Gamble Tradeoff 2 | Sure gain of 4m/month, or series of 50/50 gambles | | Yes | Yes | 1 to 5 (Least Risk Averse to Most Risk Averse) |
| Risk/Time Preferences | 1 year time tradeoff | Receive 1m today or higher amount in 1 year | | - | Yes | 1 to 5 (Least Patient to Most Patient) |
| Risk/Time Preferences | 5 year time tradeoff | Receive 1m today or higher amount in 1 year | | - | Yes | 1 to 5 (Least Patient to Most Patient) |

less disaggregated than the "real" adjustment, and varies at the province level, by rural/urban area, and by year. However it covers IFLS1-4, so the sample size is around double for that of "real". I use this quantity in the main results.

**Anthropometrics**. IFLS respondents (including children) had their height and weight measured. To calculate height-for-age and weight-for-age z-scores for children, I combine these measurements with WHO's child growth standards data (WHO & UNICEF, 2009). Following WHO recommendations, I remove all outliers with z-scores outside of the range $[-5, 5]$.

**Rice production**. Detailed rice production data is available for IFLS4 and 5. For all farming households, detailed data is available on the most recent crop farmed in the last 12 months, including the main crop farmed, the area harvested, the amount harvested, and the value of the harvest. The results shown in Appendix Table C.4.1 include all harvests where the main crop was rice. I construct a measure of yield by dividing the amount harvested by the area harvested. To avoid dropping 0s, I use the inverse hyperbolic sine (IHS) transformation of the yield variables.

**Early-life nutrition**. I make use of the following variables to construct the nutrition index seen in Section 3.5.1:

1. Food frequency. Respondents are asked how often children eat (3+ times a day, 2 times a day, once a day, 5-6 times / week, 3-4 times / week, <3 times / week, or breastfeeding). When the response is breastfeeding, I assume this is equivalent to 3+ times / day. I transform this variable into an ordinal scale so that it can be age-standardised and converted to a Z-score according to the procedure described in Appendix Section C.2).
2. Vitamin-A frequency and iron frequency. The IFLS contains questions on the consumption of specific food groups, and I use these to construct measures of how often each child consumes (i) vitamin-A-rich foods (sweet potato, eggs, dairy, green leafy vegetables, papaya, carrot, and mango) and (ii) iron-rich foods (eggs, fish, meat, dairy). These categorisations are based on the FAO guidelines for constructing measures of dietary diversity Kennedy et al. (2011). This yields a measure from 0-7 indicating how many times in the week the child ate the relevant types of foods, which are also age-standardised and converted to a Z-score according to the procedure described in Appendix Section C.2).

**Parental time-use**. I make use of the following variables to construct the parental time-use index seen in Section 3.5.1: (i) primary activity is earning (mother), (ii) primary activity is earning (father), (iii) total weekly hours worked (mother), (iv) total weekly hours worked (father). All of these variables are also standardised according to the age of the child and converted to z-scores.

**Birth location and migration history**. IFLS contains data on all adults' birth district and migration history after the age of 12. I combine the data on district of birth, the district when surveyed, and migration history to construct a year-by-year history of the district each respondent has lived in since birth. This is subsequently used to match individuals to the SPEI values in each year of their life. The main specification in Section 3.3 also makes use of district location in the two years before birth. For these two years, individuals are matched to the location of their birth.

**Birth date**. I use the "best-guess" birth date from the IFLS data for each individual that incorporates information across multiple waves in cases of inconsistencies.

**Sibling fixed effects**. I combine information collected from both individuals and their parents in the IFLS survey to match individuals to their parents. This information is used to construct sibling

fixed effects. These are dummies that denote individuals in the same "family group", defined here as sharing at least one parent. 96% of family groups share exactly one mother and one father, indicating that this definition is unlikely to be capturing patterns due to remarriage and non-nuclear families. I then use the best-guess birth date of each individual in a family group to construct the birth-order fixed effects.

**Control variables**. As controls I make use of data on the gender and religion of each individual. For individuals I construct a parental education variable, either by matching individuals back to their parents if they are also IFLS respondents, or by using the self-reported data on parental education. The variable used in the paper is the highest level of education reached by either of the parents. The possible levels are: no education, elementary school, junior high school, senior high school, or university. I also have data on the identity of the household head which I use to construct the household head education variable in a similar way.

### C.1.2 District, Weather, and Crop Data

**District data**. District boundary data is taken from IPUMS International (Minnesota Population Center., 2018). This contains the shapefiles for the first-level regional boundaries (*province*) and the second-level district boundaries (*kabupaten*) used in the IFLS. There were multiple district boundary changes and splits (see e.g. Bazzi & Gudgeon, 2016) between the first and last waves of the IFLS. Since the IPUMS data is harmonized so that the geographical unit of analysis stays consistent over time, some split districts are kept together or combined into a larger unit in order to maintain consistency. The district (the main unit of analysis) used throughout the paper is thus the IPUMS 'harmonised' district. I also use the subdistrict (*kecamatan*) level for estimating the predicted-harvest shock, which was derived from data from the Indonesian government's statistics service (Statistik & SIAK, 2019).

**Climate data - SPEI**. The primary weather variable I use is the Standardised Precipitation Evapotranspiration Index (SPEI), a drought index developed recently by Vicente-Serrano et al. (2010). Although most of the economic literature on weather shocks focuses solely on precipitation and temperature, this neglects other features of the climate that can affect the growing cycle of a plant. In particular, the impact of rainfall on crop cycles depends on *potential evapotranspiration* (the ability of soil to retain water), which is in turn affected by multiple features of the environment, including temperature, latitude, windspeed, the number of sunlight hours. SPEI is a rich index that combines all of these features. It has been shown to be a better predictor of crop yields than other climate indexes Beguería et al. (2014). The methodology behind the construction of the SPEI index is described in more detail in Vicente-Serrano et al. (2010) and Beguería et al. (2014).

The Global SPEI database (Beguería & Vicente Serrano, 2017) uses precipitation and potential evaporation data to calculate monthly SPEI estimates in grids of 0.5 latitude by 0.5 longitude (corresponding to grids of approximately 50km by 50km at the equator). Monthly SPEI data is available for all years from 1900 to 2015. The index is standardised to have a mean of 0 and a standard deviation of 1 within each grid cell over the entire historical period (1900 to 2015), so that positive SPEI indexes generally represent good climactic conditions for crop-growing relative to the historical average, while negative SPEI indexes represent conditions more likely to lead to drought relative to the historical average. The data is fitted to a log-logistic distribution and can be normalised to a number of different time scales (e.g. 1, 2, 4, 6, 12 months, etc.). Since I am interested in the effect of weather conditions primarily through its effects on agriculture, I choose a short time frame (1-month) the SPEI calculation, while also showing robustness of my main results to the 4-month SPEI

seen in Harari & Ferrara (2018). This means the SPEI will reflect short- and medium-run changes in moisture conditions that will be relevant for the seasonal changes that are important for crop yields.

In order to calculate the monthly district-level SPEI, I match each district to all the grid-squares with which it overlaps. I then use a weighted average of SPEI for all the matched grid-squares, where the weights are proportional to the area of the square that overlaps with the district. For example, if exactly half of district $A$ sits inside grid-square $s_1$, and the other half sits inside grid-square $s_2$, then the estimated rainfall in district $A$ would be the (unweighted) arithmetic mean of the rainfall in squares $s_1$ and $s_2$.

My approach for calculating the district-level weather patterns is likely to be preferable to that seen in previous estimates from Indonesia (Maccini & Yang, 2009; Bharati & Chin, 2016). Maccini & Yang (2009) use data on the latitude and longitude of the *centroid* of each district to match each district to the nearest weather station, and then use monthly rainfall data from this station directly.[1] This measurement is likely to induce non-negligible measurement error; while the median distance between the district centroid and weather station in their sample period is only 14km, the 95th percentile is 70km and the maximum is approximately 230km. The main advantage of the approach that I take is that I make use of district shapefiles rather than just data on the district centroid. This means that I accurately account for the different shapes and sizes of each district when calculating district-level weather.

**Climate data - rainfall and temperature**. When developing the predicted harvest shock measure, I make use of precipitation and temperature data from SA-OBS dataset (van den Besselaar et al., 2017). This data is highly disaggregated at both temporal and geographical levels, with daily observations at the 0.25 latitude by 0.25 longitude grid level. I aggregate the gridded observations to the district (*kabupaten*) level in the same way as the SPEI data, as described above.

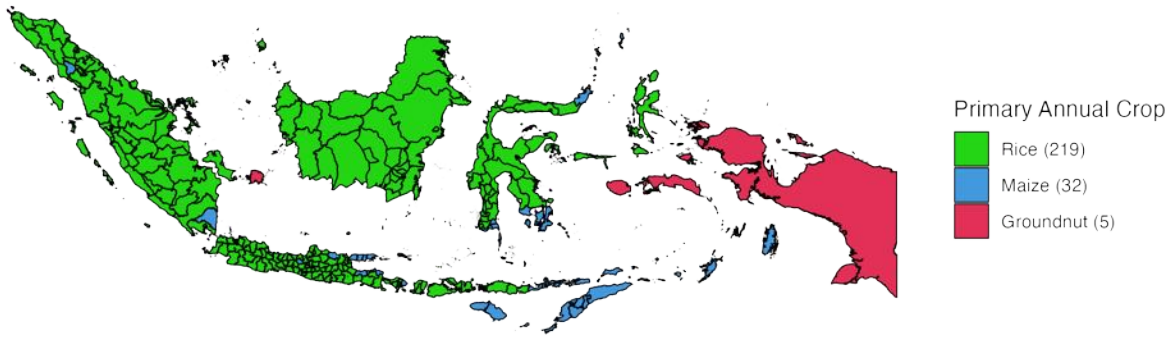
**Aggregate crop harvest data.** Crop harvested-area data comes from Monfreda et al. (2008), who combine national, state, and county level census data to create a global data set that describes the harvested area for 175 crops on a 5 minute by 5 minute (approximately 10km by 10km) grid, based on the closest available census data to the year 2000. Using this data, I calculate the primary annual crop in each district according to the total harvested area.[2] Figure C.1.3 shows the primary annual crop in each district in the data. Rice is the primary annual crop for 219 out of 256 districts, with production particularly concentrated in many areas of Java. The primary annual crop for 32 districts is maize, and for 5 it is groundnut. The districts in which groundnut is the most harvested annual crop tend to have a lower proportion of harvested area overall, and since the IFLS sample is taken from the 13 provinces indicated above in the western areas of Indonesia, these areas are unlikely to be driving the results I show in the main body of the paper. For example, only 9 individuals in the IFLS sample were born in the large eastern island of Papua highlighted in green on the map.

**Aggregate crop calendar data.** In order to identify the aggregate-level primary growing season in each district, I make use of crop calendar data from the University of Wisconsin-Madison's Nelson Institute (Sacks et al., 2010). This provides grid-cell level (0.5 x 0.5 degree) data for planting and harvesting dates of 19 different crops, including rice, maize, and groundnuts. I aggregate this to the district level by choosing the most common planting and harvesting dates over all the cells in the district.

[1] Their station-level rainfall data comes from Global Historical Climatology Network (GHCN) Precipitation and Temperature Data (Version 2) and the Badan Meterologi Dan Geofisika (BMG) agency in Indonesia.

[2] I exclude perennial crops such as oil palm from this calculation, since seasonal weather shocks are unlikely to affect production or yields.

**Figure C.1.3:** *Primary Annual Crops (by harvested area) in each Indonesian district*




### C.1.3 SPEI growing season definition

Although growing season timings and crops vary across districts, some patterns are common across many districts. The main crop for the majority of districts in my final sample is rice (Figure C.1.3). The majority of rice production occurs during the wet season, which typically takes place from around October to March in Indonesia (Food and Agriculture Organisation, 2005; Naylor et al., 2007) A secondary growing season for rice also occurs during the early dry season, but yields for this season are typically lower. I focus on the main growing season during the wet season for the rest of the paper.

**Shock definition**. In order to match individuals to shocks, I take the following steps, loosely following Maccini & Yang (2009). For individual $i$ in year $y$:

1. Identify the district $r$ that $i$ was in during year $y$ using the full migration history described above in Appendix Section C.1.1.
2. Identify the plant and harvest dates of the main crop in district $r$.
3. If $i$ was born before the harvest date of $y$, assign $i$ to the growing season in $y-1$. If $i$ was born after the harvest date of $y$, assign $i$ to the growing season in $y$.
4. Generate $SPEI_{iy}$ as the average SPEI over the assigned growing season in region $r$.

This shock definition is based on the assumption that weather patterns during the growing season affect agricultural income at harvest time, meaning that the effects of the shock will be felt only after harvest. Individuals are therefore matched to the growing season of the most recent harvest.

## C.2 Construction of indexes

As discussed in Section 3.2.3, I use a latent factor model in the style of Cunha et al. (2010) to combine data from multiple measures of skills to a lower-dimensional set of traits. The latent factor model explicitly allows for measurement error and combines data from multiple measures in order to mitigate the problem of measurement error on psychological variables.

I carried out an exploratory factor analysis (EFA) exercise on the cognitive measures and the socioemotional measures used as outcomes. The EFA process used was similar across all of these measures. The process is described in detail throughout the rest of this section.

For the outcomes and mediation measures of interest, I took the following five steps to carry out the exploratory factor analysis:

1. Correct survey responses for acquiescence bias (for socioemotional measures only).
2. Standardise measures by age (where required).
3. Determine the number of latent factors underlying the set of measured variables.
4. Estimate the factor loadings for each measure.
5. Use the factor loadings to generate factor scores for each individual in the survey

I describe each of these steps below. Note that for the cognitive and socioemotional outcome measures (only available in IFLS4 and IFLS5), I carried out separate factor analyses for each wave of the survey, so that there are two different factor structures, one for the questions contained in IFLS4 and one for the questions contained in IFLS5. For the mediation measures, I pooled data available from all waves in the analysis.

### C.2.1 Correcting for acquiescence bias (socioemotional only)

Acquiescence bias (also known as "yay-saying") is the tendency of a respondent to agree (or disagree) with a statement from the survey enumerator, even if doing so results in contradictory responses that are intended to capture the same trait. For example, a biased respondent may be likely to agree with both the statement "I see myself as someone who is talkative" *and* with the statement "I see myself as someone who is reserved", even though the two responses contradict each other. The acquiescence bias may be especially problematic given the lower-middle income context in Indonesia; cross-country evidence using the Big 5 personality tests tend to show higher acquiescence bias in lower income settings (Rammstedt & Farmer, 2013), and Laajaj & Macours (2019) find evidence of systematic measurement error related to acquiescence bias among farmers in western Kenya, which is more severe among the less educated. Using exploratory factor analysis without correcting for such bias can lead to misleading results, including the emergence of a factor that is an artefact of the response bias rather than underlying socioemotional variation. Laajaj & Macours (2019) further show that correcting for acquiescence bias substantially improves the reliability of socioemotional constructs, so I follow them and Laajaj et al. (2019) in carrying out a correction that is common in the psychometric literature (Soto et al., 2008; Rammstedt & Farmer, 2013). The correction makes use of the fact that for some socioemotional traits, there are some questions that are positively-coded, and some that are reverse-coded, and so we can detect the acquiescence pattern for a given individual. I take the following steps to correct for acquiescence on all socioemotional

outcomes that use a Likert scale:[3]

1. Reverse the reverse-coded items. For example, for a scale that goes from 1 to 5, 1 will be recoded as 5, 2 as 4, etc.
2. For every personality trait that has at least one reverse-coded item and one positively-coded item:
   (a) Take the average of the positively-coded items for each individual $i$.
   (b) Subtract from this the average of the reverse-coded items for the same individual $i$.
   (c) Divide this by two.
3. Calculate the overall acquiescence score $AS_i$ for individual $i$ by averaging over the differences for all personality traits calculated in step 2.
4. Correct individual $i$'s raw scores for acquiescence bias by adding $AS_i$ to every reverse-coded item, and subtracting $AS_i$ from every positively-coded item.

These corrected socioemotional scores are then used in all the steps below.

### C.2.2 Age standardisation

IFLS respondents are measured at different ages in adulthood, and both cognitive and socioemotional measures may vary by age. Therefore in order to get comparable measures of cognitive and socioemotional skills across different age groups I need to carry out some form of age standardisation. To do this, I make an internal standardisation following Rubio-Codina et al. (2015). To do this, I first compute the age-conditional mean of the cognitive / socioemotional measure using the fitted values of the regression:

$$Y_i = \alpha + \mathbf{X}_i'\beta + \varepsilon_i \tag{C.1}$$

where $Y_i$ is the raw cognitive / socioemotional measure, and $\mathbf{X}_i$ is a polynomial of order 6 in age. To calculate the conditional standard deviation, I take the estimated residuals from Equation C.1, denoted $\hat{\varepsilon}_i$, and regress them on the second order polynomial of age ($A_i$):

$$\hat{\varepsilon}_i = \pi_0 + \pi_1 A_i + \pi_2 A_i^2 + \nu_i$$

The estimated conditional standard deviation is then:

$$\widehat{SD}_i = \sqrt{\hat{\pi}_0 + \hat{\pi}_1 A_i + \hat{\pi}_2 A_i^2}$$

The age-specific z-score is then the result of taking the raw score, subtracting the age-conditional mean, and dividing by the age-conditional SD, as in:

$$ZScore_i = \frac{Y_i - (\hat{\alpha} + \mathbf{X}_i'\hat{\beta})}{\widehat{SD}_i}$$

[3]The 'Personality', 'Positive / Negative Affect', and 'Depression' modules used a Likert scale and had both positive and reverse-coded items. Because the phrasing of the Likert scale on each of these modules was different (see Table C.1.1), it led to a very different pattern of response bias, and so I generated separate acquiescence scores for each module and made one acquiescence correction for each module. In other words, I carried out steps 1-4 separately for each module, treating each Big 5 personality trait as a different trait in step 2, positive/negative affect as a single personality trait, and depression as a single personality trait.

As noted in Rubio-Codina et al. (2015), this form of standardisation gives results that are qualitatively similar to internal standardisation using the typical method (which involves calculating Z-scores within age-specific categories). The advantage here is that the approach is less sensitive to outliers and to having few observations within an age category than the standard method.

### C.2.3 Determining the number of factors

In order to determine the number of factors that should be extracted from the data, I make use of 4 commonly-used methods that suggest how many factors to extract. The methods are as follows:

1. *Kaiser's criterion* (Kaiser, 1958). This rule-of-thumb suggests retaining all factors with an eigenvalue of above 1. The intuition that underlies this is that factors should only be kept if they explain as much variance as the equivalent of one original variable. This tends to suggest retaining a low number of factors; newer literature often suggests using a threshold value of 0.7 instead. I show results for both.

2. *Scree plot* (Cattell, 1966). This test is based on visual inspection of a plot of eigenvalues associated with the data. The proposed number of factors is equal to the number of factors before the 'elbow' of the scree plot; that is, the number before the which the smooth decrease of eigenvalues appears to level off towards the right of the plot.

3. *Minimum Average Partial (MAP)* (Velicer, 1976). This rule chooses the number of factors that minimises the unexplained partial correlation. The procedure iterates as follows. For $r = 1$, we run a full factor analysis with one factor, partial out this factor from the correlation matrix of the variables of interest, and calculate the average squared coefficient in the off-diagonals of this partial correlation matrix. Then, for $r = 2$, we run a similar process in which we partial out two factors, and calculate the average squared off-diagonal from the resulting matrix. This process is repeated up until $r = K - 1$ where $K$ is the number of measurements. The recommended number of factors $r^*$ is the value of $r$ that minimises the average squared partial correlation. The intuition behind this process is that components are retained as long as the variance in the correlation matrix represents systematic variance. They are dropped when there is proportionately more unsystematic variance than systematic variance (see Attanasio, Cattan, et al., 2020, Appendix p. 15).

4. *Parallel Analysis* (Horn, 1965). Horn's test involves a Monte-Carlo procedure in which we simulate uncorrelated normal random variables of the same dimension as the actual data of interest. The eigenvalues derived from the actual data are then compared to the eigenvalues from the randomly generated data, and factors are kept if the eigenvalue is greater than the 95th percentile from the simulated data.

**Cognitive and socioemotional**. I carry out each of these methods first for all cognitive measures, and then for all socioemotional measures. All methods recommended one factor for cognitive measures in both IFLS4 and IFLS5. Table C.2.1 shows the recommendation of the number of factors using each method for socioemotional measures. Each different method tends to give different recommendations, so some judgement is required in deciding how many factors to retain. Ledesma et al. (n.d.) suggest that Velicer's MAP criterion and Horn's Parallel Analysis tend to be more reliable, so the number of factors I retain for analysis is the number most commonly recommended from all the methods, with more weight placed on the parallel analysis and MAP criteria. Since there were fewer socioemotional questions asked in IFLS4 than in IFLS5 (see Table C.1.1 for details), I expect there to be fewer factors for IFLS4; the fact that all methods suggest a smaller number pf factors in IFLS4 reflects this. For IFLS4, the scree plot (see Fig. C.2.1a), Kaiser criterion (1.0) and MAP all

suggest that there is 1 factor. For IFLS5, the same set of methods suggest a value of 3 (see also Fig. C.2.1b). In both cases, parallel analysis appears to overstate the number of latent factors. Based on this process, I retain 1 socioemotional factor for analysis in IFLS4, and 3 socioemotional factors in IFLS5.

I also carry out all of the above methods for each group of mediator variables that has three or more available measures (this therefore excludes parental education, which only has one measure). Based on the recommendations from each of the methods, I retain 2 factors for child health, 1 factor for early investment, 1 factor for parental health, and 2 factors for parental time use.

**Table C.2.1:** *Recommendation of how many factors to retain from each method*

| Method | IFLS4 | IFLS5 |
|---|---|---|
| Kaiser Criterion (1.0) | 1 | 3 |
| Kaiser Criterion (0.7) | 2 | 4 |
| Scree Plot | 1 | 3 |
| Parallel Analysis | 8 | 17 |
| Minimum Average Partial (MAP) | 1 | 3 |
| Retained for Analysis | 1 | 3 |

*Notes*: Scree plots for each IFLS wave can be seen in Figures C.2.1a and C.2.1b. Each method recommends fewer factors for IFLS4 than IFLS5 because IFLS4 contains fewer socioemotional questions than IFLS 5 (see Table C.1.1 for details).

### C.2.4 Estimating the factor analysis model

After identifying how many factors should be used in the factor analysis, I run an exploratory factor analysis (EFA) on measures (which have been corrected for acquiescence bias where relevant and age-standardised). The EFA estimates factor loadings by minimising the squared residuals[4], and then carries out a factor rotation that ensures that each measure primarily loads onto a single factor. I use an *oblimin* rotation to rotate factors; this has the advantage of being an *oblique* rotation rather than *orthogonal*, meaning that I do not have to make the strong assumption that factors are uncorrelated[5]. The aim of the rotation is to identify measures that primarily load onto a single factor, which can then be used as a 'dedicated' measure for that factor. Measures that load onto more than one factor, or onto no factors, will not be used in the measurement system.

Table C.2.3 shows the factor loadings for the socioemotional measures available in IFLS5, and Table C.2.4 shows the factor loadings for the socioemotional measures available in IFLS4. Following standard practice in the psychometric literature (Rammstedt & Farmer, 2013; Attanasio, Cattan, et al., 2020), I deem a measure to 'load' on a factor if the (absolute value of the) factor loading is greater than 0.3, and have highlighted cells in Tables C.2.3 and C.2.4 where the value meets this condition.

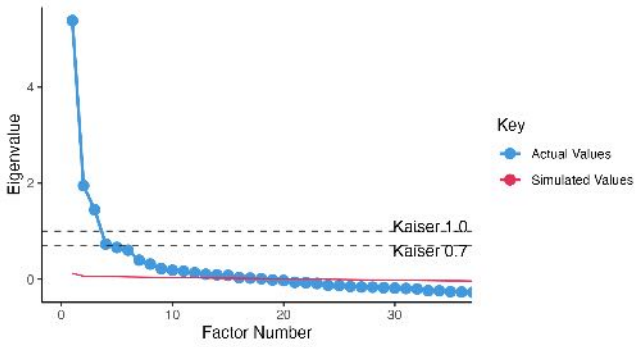
The factor loadings seen in Table C.2.3 shows that the underlying latent variables coincide relatively closely with the survey modules in which they are used, and thus permit a clear interpretation for each factor (see Table C.2.2) Factor 1 is loaded on strongly by all the questions in the Positive/Negative Affect (PNA) module, along with the two Subjective Well-being questions and the question from the Depression module relating to whether the individual felt happy in the last week.

[4]This has been shown to give unbiased estimates of factor loadings in the context of ordinal variables such as those used here, see Lee et al., 2012.

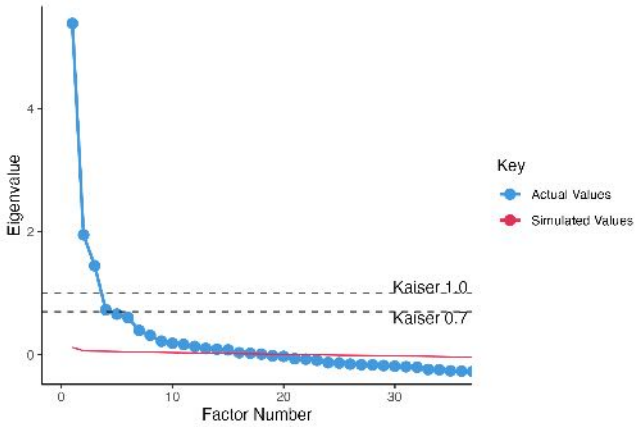
[5]This follows recent literature that also use oblique rotations, see Laajaj & Macours, 2019, Attanasio, Cattan, et al., 2020. Multiple methods for carrying out an oblique factor rotation are available, therefore as a robustness check I carried out the same exercise with both *geomin* and *quartimin* rotations and obtained similar results.

**Figure C.2.1:** *Scree Plots for Exploratory Factor Analysis*

**(a)** *Socioemotional variables in IFLS4*




**(b)** *Socioemotional variables in IFLS5*




*Notes:* 'Actual values' denotes the eigenvalues from the true data. 'Simulated values' (in red) denotes the values from the 95th percentile of the randomly generated uncorrelated normal variables. The black horizontal dashed lines indicate the threshold points for the Kaiser criteria (1.0 and 0.7), and the solid black marks indicate possible positions for the 'elbow' of the scree plot in Cattell (1966).

This factor can therefore be broadly interpreted as a measure of affect and well-being. Factor 3 has a similarly straightforward interpretation, being composed of all depression measurements apart from 'Everything was an effort', which does not load strongly (less than 0.3) on any factor. The factor loadings for the single factor in IFLS4 (Table C.2.4) closely accord with the factor loadings of Factor 3 in IFLS5, with all depression measures loading except 'Everything was an effort'. This implies that the factor structure here is congruent across the two time periods and populations.

**Table C.2.2:** *Dominant interpretation of each factor*

| Factor | Wave | Interpretation |
|---|---|---|
| Factor 1 | IFLS4 | "Depression" |
| Factor 1 | IFLS5 | "Affect / Wellbeing" |
| Factor 2 | IFLS5 | "Personality" |
| Factor 3 | IFLS5 | "Depression" |

*Notes*: Primary interpretations for each factor, based on the factor loadings seen in Tables C.2.3 and C.2.4

The factor loadings on Factor 2 in the personality module are less easily interpretable, as they do not conform with the 5-factor factor structure aligned with the "Big 5" personality traits. This result is in keeping with Laajaj & Macours (2019) and Laajaj et al. (2019), who show that factor structures tend to deviate from the Big 5 in low income contexts. The personality measures load onto a single factor, with the most consistent loadings coming from the *agreeableness*, *conscientiousness*, and *openness* traits, all of which have loadings around or above 0.3. Measures of *extroversion* and do not consistently load on Factor 2, despite a particularly strong factor loading for 'Extraversion - Outgoing' (0.543).

Notably, measures of risk and time preferences do not load onto any factor in either survey wave. This implies that such measures are subject to severe measurement error and would lead to misleading results if used as an outcome variable. This is in keeping with evidence that suggests self-reported responses for such measures that involve no real money at stake (unlike experimental measures of risk aversion or patience) tend to be highly unreliable (Binswanger, 1980; Y. Chuang & Schechter, 2015).

There is little evidence of acquiescence bias in the socioemotional factor loadings: there is no discernible pattern where only reverse-coded or positively-coded items load onto a factor. Given that such patterns were observable when I ran the factor analysis without acquiescence correction (not shown), this implies that the acquiescence correction described above is effective. Notably when looking at the interpretations of Factors 1 and 2, very similar questions appear to load on different factors depending on the module they are in. For example, both the Positive-Negative Affect (PNA) and the Depression modules ask if the individual was lonely, but the measurements do not load on the same factor. There may be two main reasons for this. First, we may be worried that the responses to similar questions in different modules are biased in different ways due to changes in response pattern, possibly due to the different scales used, or the fact that the Depression module comes significantly later in the survey than the PNA module. This worry is mitigated somewhat by the acquiescence correction, which accounts for the proportion of bias that is due to "yay-saying". Alternatively, the difference could be due to the change in temporal scope in each module. In other words, since the PNA module asks about feelings yesterday, while the Depression module asks about feelings over the past week (see Table C.1.1), respondents could give justifiably different answers. This consideration is confused somewhat by the fact that the subjective well-being questions (that

target longer-term cognitive evaluation of happiness) tend to load alongside the short-term affect questions rather than the medium-term depression questions. One possible explanation may be that the long-term subjective well-being questions are interpreted by respondents as an evaluation of short-term happiness; this type of problem is well known in the literature on measuring well-being (see e.g. White, 2018; Benjamin et al., 2020).

Whichever explanation is true, the considerations above point towards the importance of running a factor analysis exercise on socioemotional measures, rather than naively interpreting each question to be measuring the trait it is nominally intended to measure. If the answer to a question about loneliness is so affected by seemingly extraneous factors such as the phrasing of the question or the position within the model, then the answer to such a question should not be taken 'at face value'. Instead, it is better to assume that it is a proxy for some underlying latent factor, and that this factor can in turn be recovered from multiple proxy measures. Nevertheless, we must be equally wary in interpreting the latent factors themselves: even though they are designed to pick up true variation (instead of measurement error), this true variation may still be at least partly due to seemingly extraneous factors like survey design and response biases, and so should be interpreted with caution.

Table C.2.5 shows the cognitive factor loadings for IFLS4, and Table C.2.6 shows the cognitive factor loadings for IFLS5. In both waves, all the cognitive measures load onto the same single factor, implying that this factor can be clearly interpreted as measuring cognitive ability. The written tests (Raven, Written Maths, and Number Series), have higher loadings in both waves than the oral tests (Word Recall and Oral Maths), suggesting that these capture more of the variation in the latent cognitive factor. Overall, the loadings for cognitive a clear interpretation of a single cognitive factor that is in keeping with previous literature Laajaj & Macours (2019).

### C.2.5 Generating factor scores

I use the factor loadings from Tables C.2.3–C.2.6 to generate factor scores for each individual $i$. These factor scores represent the 'value' of that factor for each individual $i$. In order to generate the factor scores for factor $k$ I carry out the following steps:

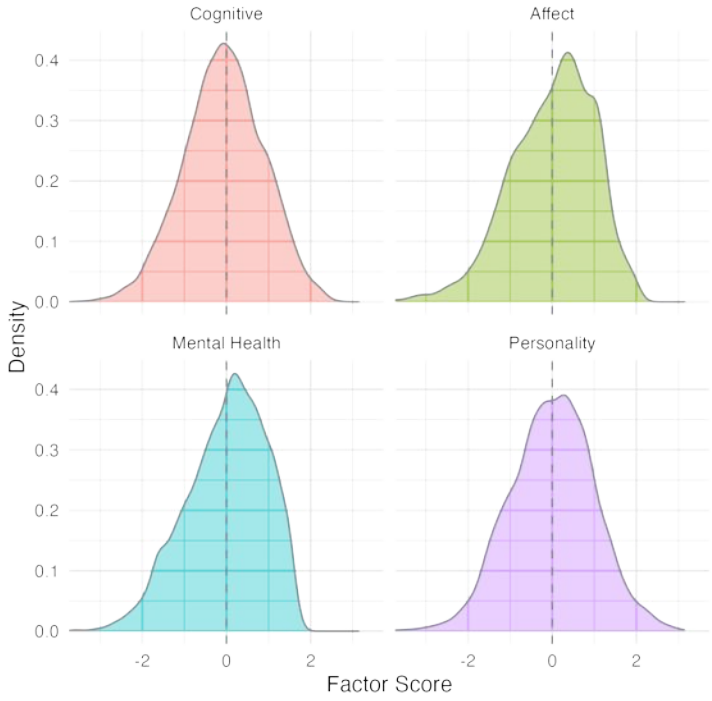
1. Assign each question to 'load' on factor $k$ if and only if the (absolute value of) the factor loading is greater than or equal to 0.3 (following standard practice in the psychometric literature, see Rammstedt & Farmer, 2013; Attanasio, Cattan, et al., 2020). All other questions have their factor loadings set to 0.
2. For each individual $i$, take a weighted mean of the questions that load onto factor $k$, where the weights correspond to the loadings obtained in Tables C.2.3, C.2.4, C.2.5 and C.2.6.
3. Standardise the factor scores so that each factor has a mean of 0 and a standard deviation of 1.

Using this method to obtain factor scores ensures that questions are only kept if they clearly load onto a particular factor, while questions with particularly weak loadings (such as the questions on risk and time preferences) are not included in the factor scores. Step 3 ensures that the factor scores are clearly interpretable. For example, if individual $i$ has a factor score of 1 on Factor 2 ("Personality") in IFLS5, this means that her non-cognitive personality skills are 1 standard deviation above the sample mean.

The process described here generates a separate factor scores in each wave for the two outcome

factors that are measured in both IFLS4 and IFLS5 (depression and cognitive).[6] The factor structures for these factors is strikingly similar across waves. I therefore construct the final outcome variables used in the main part of the paper by combining measurements across waves, in particular by taking a simple arithmetic mean of the factor score in IFLS4 and the factor score in IFLS5 for any individuals who were measured in both waves. The distribution of the finalised factor scores among my main sample are shown in Figure C.2.2. By construction, all the distributions are centred around 0, and they all appear to be approximately symmetrical.

**Figure C.2.2:** *Kernel density of each of the factor scores*




### C.2.6 Measures of noise

Table C.2.7 estimates two proxies for the amount of noise in each of the outcome variables, to act as a proxy for the likely degree of measurement error. First, I test for the correlation between the outcome for a given individual in 2008 and for the same individual in 2015. I can only calculate this measure for cognitive and depression scores, since these were the only outcomes available in 2008. The correlation is 0.52 for cognitive, suggesting strong correlation over the 7 year period. It is much lower for depression (0.15), which could be because this variable is measured with more error or simply because it varies more over time.

Second, I calculate the average inter-item correlation by taking the pairwise correlation between every measure used in an index, and then taking the mean for every pair. This indicates how inter-related the items in an index are, and therefore how likely they are to reliably measure the construct that they are intending to measure. The correlation takes a value of between 0.24 and 0.28 for cognitive, affect/wellbeing, and depression. This is between the recommended range of 0.15 and 0.5 (Clark & Watson, 1995). Personality has a lower inter-item correlation of 0.13, suggesting higher measurement error and possibly lower construct validity.

[6]Mediator variables, on the other hand, are already combined across waves as described in Step 2.

**Table C.2.3:** *Socio emotional factor Loadings for IFLS5*

| Module | Question | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|
| Positive / Negative Affect | Frustrated (-) | 0.607 | -0.004 | 0.005 |
| Positive / Negative Affect | Sad (-) | 0.694 | -0.021 | -0.011 |
| Positive / Negative Affect | Enthusiastic | 0.332 | 0.041 | 0.008 |
| Positive / Negative Affect | Lonely (-) | 0.519 | -0.017 | 0.046 |
| Positive / Negative Affect | Content | 0.632 | -0.028 | -0.034 |
| Positive / Negative Affect | Worried (-) | 0.467 | 0.005 | 0.039 |
| Positive / Negative Affect | Bored (-) | 0.594 | 0.087 | -0.052 |
| Positive / Negative Affect | Happy | 0.557 | 0.020 | 0.025 |
| Positive / Negative Affect | Angry (-) | 0.581 | 0.030 | 0.007 |
| Positive / Negative Affect | Tired (-) | 0.363 | 0.028 | -0.040 |
| Positive / Negative Affect | Stressed (-) | 0.673 | -0.007 | -0.027 |
| Positive / Negative Affect | Pain (-) | 0.301 | 0.000 | 0.026 |
| Subjective Wellbeing | Assessment of current situation | 0.365 | 0.008 | 0.029 |
| Subjective Wellbeing | Life satisfaction (Cantrill ladder) | 0.346 | 0.021 | -0.019 |
| Mental Health | Bothered (-) | 0.089 | -0.023 | 0.422 |
| Mental Health | Concentration (-) | 0.091 | 0.043 | 0.456 |
| Mental Health | Depressed (-) | 0.206 | -0.038 | 0.511 |
| Mental Health | Everything was an effort (-) | 0.182 | -0.011 | 0.198 |
| Mental Health | Hopeful | -0.191 | 0.011 | 0.726 |
| Mental Health | Fearful (-) | 0.108 | 0.067 | 0.446 |
| Mental Health | Restless sleep (-) | 0.159 | 0.037 | 0.363 |
| Mental Health | Happy | 0.444 | 0.013 | 0.321 |
| Mental Health | Lonely (-) | 0.071 | -0.024 | 0.495 |
| Mental Health | Couldn't get going (-) | 0.063 | 0.061 | 0.440 |
| Personality | Agreeableness - forgiving | -0.001 | 0.534 | -0.017 |
| Personality | Agreeableness - considerate and kind | -0.032 | 0.608 | -0.031 |
| Personality | Agreeableness - rude (-) | 0.020 | 0.386 | -0.024 |
| Personality | Conscientiousness - thorough | 0.026 | 0.480 | 0.038 |
| Personality | Conscientiousness - lazy (-) | -0.004 | 0.429 | -0.039 |
| Personality | Conscientiousness - efficient | 0.050 | 0.290 | 0.072 |
| Personality | Extroversion - talkative | -0.008 | 0.097 | 0.003 |
| Personality | Extroversion - reserved (-) | -0.030 | 0.163 | 0.053 |
| Personality | Extroversion - outgoing | 0.014 | 0.543 | 0.005 |
| Personality | Neuroticism - relaxed | 0.071 | 0.295 | 0.098 |
| Personality | Neuroticism - worries (-) | 0.130 | 0.103 | 0.170 |
| Personality | Neuroticism - gets nervous (-) | 0.063 | 0.288 | 0.049 |
| Personality | Openness - original | 0.007 | 0.405 | 0.043 |
| Personality | Openness - active imagination | -0.040 | 0.280 | 0.043 |
| Personality | Openness - artistic and aesthetic | -0.031 | 0.392 | 0.016 |
| Risk/Time Preferences | 1 year time tradeoff | -0.025 | 0.008 | 0.033 |
| Risk/Time Preferences | 5 year time tradeoff | -0.004 | 0.009 | 0.009 |
| N | | | | 5267 |

*Notes*: Cells highlighted in blue have an absolute value of greater than 0.3. '(-)' denotes a reverse-coded item.

**Table C.2.4:** *Socio emotional factor Loadings for IFLS4*

| Module | Question | Factor 1 |
|---|---|---|
| Mental Health | Bothered (-) | 0.635 |
| Mental Health | Concentration (-) | 0.533 |
| Mental Health | Depressed (-) | 0.613 |
| Mental Health | Everything was an effort (-) | 0.284 |
| Mental Health | Hopeful | 0.463 |
| Mental Health | Fearful (-) | 0.494 |
| Mental Health | Restless sleep (-) | 0.457 |
| Mental Health | Happy | 0.543 |
| Mental Health | Lonely (-) | 0.618 |
| Mental Health | Couldn't get going (-) | 0.634 |
| Subjective Wellbeing | Assessment of current situation | 0.243 |
| Risk/Time Preferences | 1 year time tradeoff | -0.035 |
| Risk/Time Preferences | 5 year time tradeoff | -0.042 |
| N | | 2242 |

*Notes*: Cells highlighted in blue have an absolute value of greater than 0.3. '(-)' denotes a reverse-coded item.

**Table C.2.5:** *Cognitive Factor Loadings (IFLS4)*

| Module | Question | Factor 1 |
|---|---|---|
| Cognitive | Raven (15-24) | 0.573 |
| Cognitive | Maths - written (15-24) | 0.533 |
| Cognitive | Word recall | 0.402 |
| N | | 3083 |

*Notes*: Cells highlighted in blue have an absolute value of greater than 0.3.

**Table C.2.6:** *Cognitive Factor Loadings (IFLS5)*

| Module | Question | Factor 1 |
|---|---|---|
| Cognitive | Raven (15-24) | 0.536 |
| Cognitive | Maths - written (15-24) | 0.545 |
| Cognitive | Word recall | 0.394 |
| Cognitive | Maths - oral | 0.380 |
| Cognitive | Number series | 0.603 |
| N | | 8003 |

*Notes*: Cells highlighted in blue have an absolute value of greater than 0.3.

**Table C.2.7:** *Indicators of noise*

| Measure | Cognitive | Affect/Wellbeing | Personality | Depression |
|---|---|---|---|---|
| $corr(Y_{2008}, Y_{2015})$ | 0.52 | - | - | 0.15 |
| N | 2424 | | | 1593 |
| Average inter-item correlation | 0.24 | 0.26 | 0.13 | 0.28 |
| N | 7109 | 4964 | 4956 | 7090 |

*Notes*: $corr(Y_{2008}, Y_{2015})$ denotes the correlation between the measure for a given individual in 2008 and the same measure in 2015. This correlation is only available for cognitive and mental health scores. Average inter-item correlation denotes the mean of every pairwise correlation in the set of measures used in that index.

### C.2.7 Investment indexes

The variables used in the investment indexes are constructed in a similar way to the cognitive and socioemotional scores. Age standardisation is performed as in Appendix Section C.2.2. Due to the small number of measures available, I assume that the nutrition variables constitute one index and the parental time measures constitute another and use Anderson (2008)'s invariance-covariance-weights to construct indexes, rather than performing an exploratory factor analysis and generating indexes based on the factor loadings. Table C.2.8 shows the weights on each measure when constructing the indexes.

**Table C.2.8:** *Inverse Covariance Weights for Investment Indexes*

| Index | Measure | Inverse covariance weight |
|---|---|---|
| Nutrition index | Food frequency | 0.353 |
| | Vitamin A frequency | 0.381 |
| | Iron frequency | 0.266 |
| Parental time index | Primary activity is earning (mother) | 0.247 |
| | Total weekly hours spent working (mother) | 0.195 |
| | Primary activity is earning (father) | 0.266 |
| | Total weekly hours spent working (father) | 0.292 |

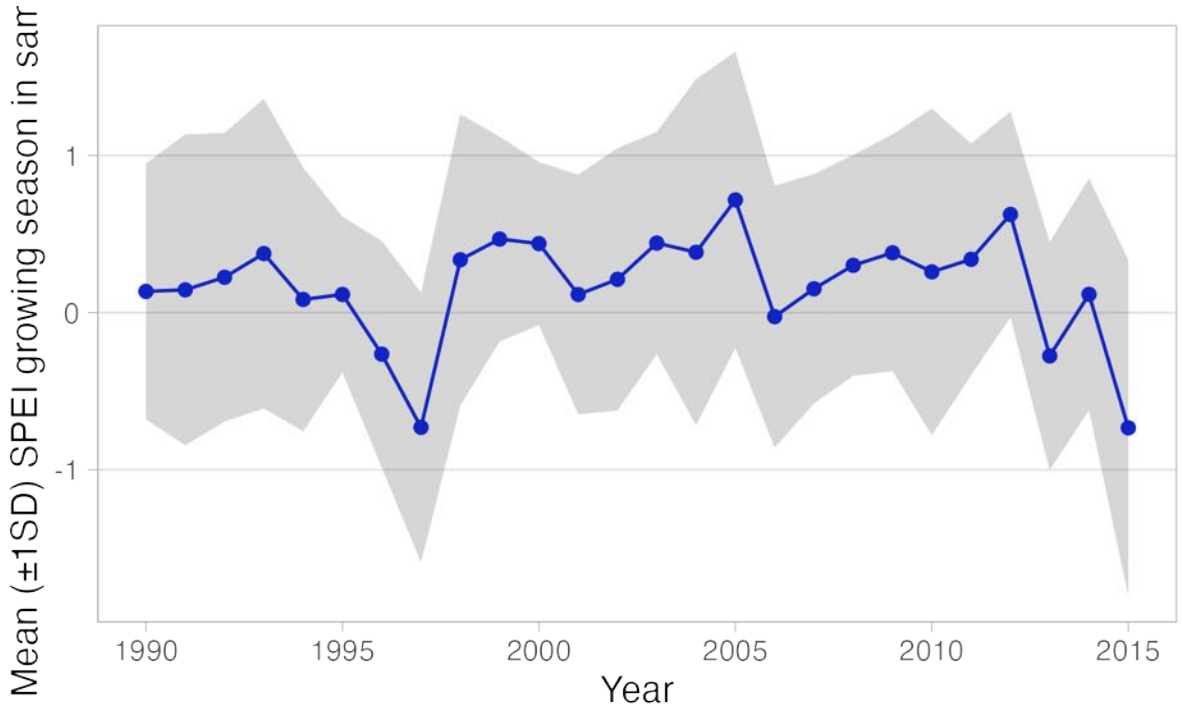
## C.3 Descriptive statistics

**Table C.3.1:** *Summary Statistics - Main Sample*

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|---|---|---|---|---|---|---|
| Factor score - cognitive | 5572 | 0.00 | 1.00 | −3.75 | 0.005 | 2.90 |
| Factor score - affect/wellbeing | 4956 | −0.00 | 1.00 | −4.35 | 0.13 | 2.05 |
| Factor score - personality | 4956 | −0.00 | 1.00 | −3.96 | 0.03 | 3.14 |
| Factor score - depression | 5565 | 0.00 | 1.00 | −4.35 | 0.11 | 1.79 |
| Born during the growing season (=1) | 5572 | 0.48 | 0.50 | 0 | 0 | 1 |
| SPEI growing season in t - 2 | 5572 | −0.06 | 0.88 | −3.33 | −0.002 | 2.95 |
| SPEI growing season in t - 1 | 5572 | 0.02 | 0.90 | −3.37 | 0.09 | 3.01 |
| SPEI growing season in t | 5572 | 0.05 | 0.88 | −3.20 | 0.12 | 3.01 |
| SPEI growing season in t + 1 | 5572 | 0.06 | 0.87 | −3.33 | 0.15 | 2.57 |
| SPEI growing season in t + 2 | 5572 | 0.09 | 0.85 | −3.20 | 0.17 | 2.57 |
| SPEI growing season in t + 3 | 5572 | 0.16 | 0.87 | −3.20 | 0.25 | 2.57 |
| SPEI growing season in t + 4 | 5572 | 0.13 | 0.86 | −3.03 | 0.21 | 2.57 |
| SPEI growing season in t + 5 | 5572 | 0.16 | 0.92 | −3.03 | 0.23 | 3.05 |
| SPEI growing season in t + 6 | 5572 | 0.17 | 0.91 | −2.90 | 0.27 | 3.05 |
| SPEI growing season in t + 7 | 5572 | 0.17 | 0.89 | −3.03 | 0.23 | 3.05 |
| SPEI growing season in t + 8 | 5572 | 0.18 | 0.91 | −2.90 | 0.22 | 3.05 |
| SPEI growing season in t + 9 | 5572 | 0.23 | 0.89 | −2.96 | 0.26 | 3.05 |
| SPEI growing season in t + 10 | 5572 | 0.33 | 0.81 | −2.87 | 0.32 | 3.66 |
| SPEI growing season in t + 11 | 5572 | 0.29 | 0.82 | −2.90 | 0.26 | 3.65 |
| SPEI growing season in t + 12 | 5572 | 0.32 | 0.85 | −2.96 | 0.31 | 3.66 |
| SPEI growing season in t + 13 | 5572 | 0.32 | 0.85 | −2.96 | 0.33 | 3.66 |
| SPEI growing season in t + 14 | 5572 | 0.27 | 0.87 | −2.90 | 0.28 | 3.66 |
| SPEI growing season in t + 15 | 5572 | 0.28 | 0.86 | −3.10 | 0.30 | 3.66 |
| Female (=1) | 5572 | 0.53 | 0.50 | 0 | 1 | 1 |
| Year of birth | 5572 | 1993.17 | 3.57 | 1988 | 1993 | 2000 |
| Highest level of parent eduction - elementary (=1) | 5572 | 0.52 | 0.50 | 0 | 1 | 1 |
| Highest level of parent eduction - junior High (=1) | 5572 | 0.20 | 0.40 | 0 | 0 | 1 |
| Highest level of parent eduction - senior High (=1) | 5572 | 0.19 | 0.40 | 0 | 0 | 1 |
| Highest level of parent eduction - university (=1) | 5572 | 0.08 | 0.28 | 0 | 0 | 1 |
| Religion - muslim (=1) | 5572 | 0.90 | 0.30 | 0 | 1 | 1 |
| Religion - protestant (=1) | 5572 | 0.04 | 0.19 | 0 | 0 | 1 |
| Religion - hindu (=1) | 5572 | 0.05 | 0.22 | 0 | 0 | 1 |
| Religion - other (=1) | 5572 | 0.01 | 0.11 | 0 | 0 | 1 |

*Notes*: Descriptive statistics for the "Main Sample" (row 1 in Table 3.2). This sample is at the individual level. Individuals who were born after 1988 were only tracked after moving household if they were part of the original sample from the first IFLS wave (1993), but will be missing if they were added in a later wave and moved households since IFLS4. Since Affect/Wellbeing and Personality were only measured in the 5th wave of IFLS (in 2015), any members interviewed only in IFLS4 (in 2007) have missing values for these variables. Therefore N is only 4956 for these factors.

**Figure C.3.1:** *SPEI growing season over each year*



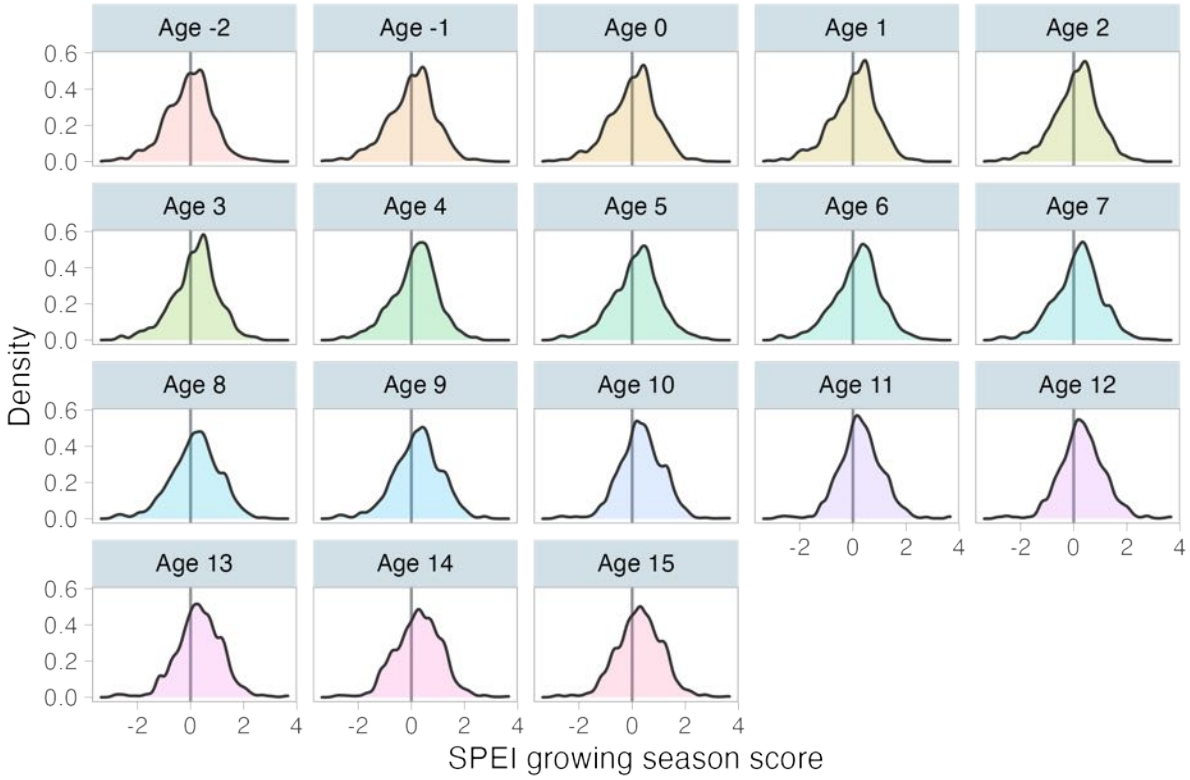

*Notes*: This shows the yearly mean and ±1 SD shaded intervals for the value of SPEI growing season experienced in each year for the individuals in the main sample. A negative spike is seen in 1997, which was when severe drought struck much of Indonesia.

**Table C.3.2:** *Correlation between SPEI growing season at different ages in the main sample*

| | Age -2 | Age -1 | Age 0 | Age 1 | Age 2 | Age 3 | Age 4 | Age 5 | Age 6 | Age 7 | Age 8 | Age 9 | Age 10 | Age 11 | Age 12 | Age 13 | Age 14 | Age 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age -2 | 1.00 | 0.11 | -0.02 | 0.06 | 0.02 | 0.03 | 0.05 | 0.07 | -0.06 | -0.03 | 0.14 | 0.09 | 0.14 | -0.03 | 0.05 | 0.08 | -0.01 | 0.04 |
| Age -1 | 0.11 | 1.00 | 0.13 | 0.02 | 0.07 | 0.02 | 0.04 | 0.04 | 0.08 | -0.07 | -0.06 | 0.14 | 0.11 | 0.16 | -0.01 | 0.09 | -0.01 | 0.01 |
| Age 0 | -0.02 | 0.13 | 1.00 | 0.20 | 0.02 | 0.07 | -0.02 | 0.04 | 0.04 | 0.07 | -0.07 | -0.03 | 0.10 | 0.13 | 0.18 | -0.01 | 0.06 | -0.08 |
| Age 1 | 0.06 | 0.02 | 0.20 | 1.00 | 0.19 | 0.10 | 0.10 | 0.03 | 0.01 | 0.04 | 0.07 | -0.06 | 0.04 | 0.05 | 0.17 | 0.19 | 0.00 | 0.07 |
| Age 2 | 0.02 | 0.07 | 0.02 | 0.19 | 1.00 | 0.18 | 0.10 | 0.08 | 0.03 | 0.00 | 0.02 | 0.04 | -0.11 | 0.05 | 0.05 | 0.19 | 0.15 | 0.01 |
| Age 3 | 0.03 | 0.02 | 0.07 | 0.10 | 0.18 | 1.00 | 0.18 | 0.15 | 0.07 | 0.00 | -0.01 | 0.00 | 0.02 | -0.14 | 0.07 | 0.05 | 0.16 | 0.16 |
| Age 4 | 0.05 | 0.04 | -0.02 | 0.10 | 0.10 | 0.18 | 1.00 | 0.16 | 0.10 | 0.01 | 0.00 | 0.03 | -0.08 | 0.00 | -0.15 | 0.12 | 0.06 | 0.14 |
| Age 5 | 0.07 | 0.04 | 0.04 | 0.03 | 0.08 | 0.15 | 0.16 | 1.00 | 0.17 | 0.04 | -0.04 | -0.02 | 0.09 | -0.12 | 0.06 | -0.10 | 0.04 | 0.10 |
| Age 6 | -0.06 | 0.08 | 0.04 | 0.01 | 0.03 | 0.07 | 0.10 | 0.17 | 1.00 | 0.09 | -0.00 | -0.01 | -0.04 | 0.03 | -0.06 | 0.11 | -0.20 | 0.01 |
| Age 7 | -0.03 | -0.07 | 0.07 | 0.04 | 0.00 | 0.00 | 0.01 | 0.04 | 0.09 | 1.00 | 0.11 | -0.04 | -0.04 | -0.07 | 0.03 | -0.03 | 0.12 | -0.23 |
| Age 8 | 0.14 | -0.06 | -0.07 | 0.07 | 0.02 | -0.01 | 0.00 | -0.04 | -0.00 | 0.11 | 1.00 | 0.06 | -0.03 | -0.01 | -0.07 | 0.02 | -0.05 | 0.10 |
| Age 9 | 0.09 | 0.14 | -0.03 | -0.06 | 0.04 | 0.00 | 0.03 | -0.02 | -0.01 | -0.04 | 0.06 | 1.00 | -0.02 | 0.01 | -0.01 | -0.07 | -0.03 | -0.05 |
| Age 10 | 0.14 | 0.11 | 0.10 | 0.04 | -0.11 | 0.02 | -0.08 | 0.09 | -0.04 | -0.04 | -0.03 | -0.02 | 1.00 | -0.03 | 0.09 | 0.04 | -0.09 | -0.07 |
| Age 11 | -0.03 | 0.16 | 0.13 | 0.05 | 0.05 | -0.14 | 0.00 | -0.12 | 0.03 | -0.07 | -0.01 | 0.01 | -0.03 | 1.00 | 0.01 | 0.06 | 0.06 | -0.05 |
| Age 12 | 0.05 | -0.01 | 0.18 | 0.17 | 0.05 | 0.07 | -0.15 | 0.06 | -0.06 | 0.03 | -0.07 | -0.01 | 0.09 | 0.01 | 1.00 | -0.00 | 0.03 | 0.03 |
| Age 13 | 0.08 | 0.09 | -0.01 | 0.19 | 0.19 | 0.05 | 0.12 | -0.10 | 0.11 | -0.03 | 0.02 | -0.07 | 0.04 | 0.06 | -0.00 | 1.00 | -0.02 | 0.02 |
| Age 14 | -0.01 | -0.01 | 0.06 | 0.00 | 0.15 | 0.16 | 0.06 | 0.04 | -0.20 | 0.12 | -0.05 | -0.03 | -0.09 | 0.06 | 0.03 | -0.02 | 1.00 | 0.01 |
| Age 15 | 0.04 | 0.01 | -0.08 | 0.07 | 0.01 | 0.16 | 0.14 | 0.10 | 0.01 | -0.23 | 0.10 | -0.05 | -0.07 | -0.05 | 0.03 | 0.02 | 0.01 | 1.00 |

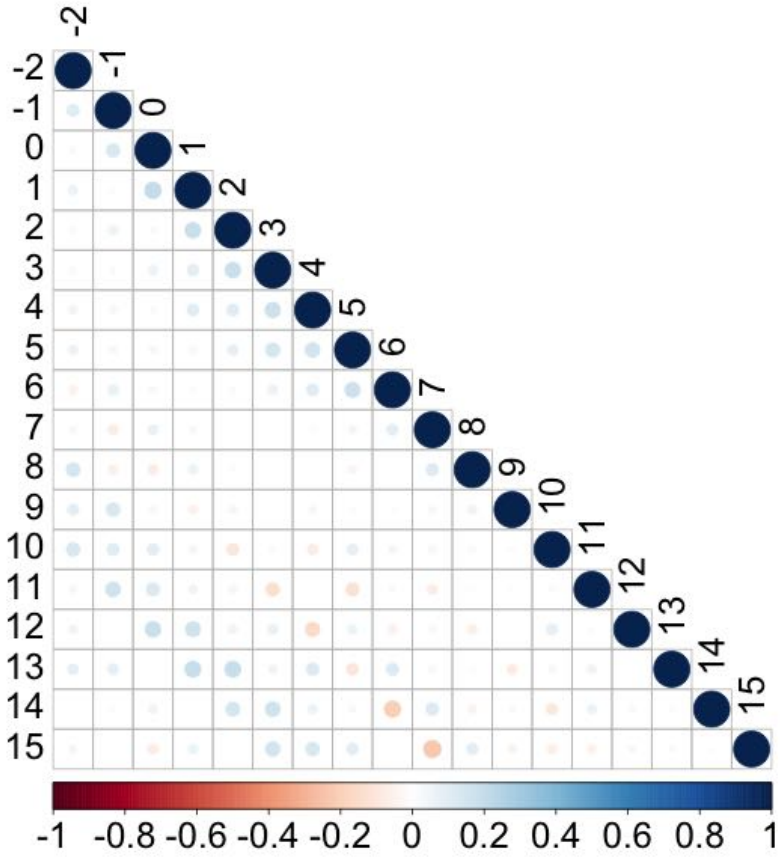
*Notes*: This is the correlation matrix for the values of SPEI growing season at different ages for individuals in the main sample. The $x$- and $y$-axes describe the age of the shock. The absolute value of the correlation between shocks in different periods is low, with no such correlation exceeding 0.2.

**Figure C.3.2:** *The distribution of SPEI growing season is similar at each age*




*Notes*: This shows the distribution of shocks experienced at each age by individuals in the main sample. Notably, the distribution of shocks is very similar across ages, despite some "spikes" in the yearly variation seen in Figure C.3.1. This is because individuals in the sample are born in different years.

**Figure C.3.3:** *Correlation between SPEI growing season at different ages in the main sample*




*Notes*: This is a visual representation of Table C.3.2, describing the correlation matrix for the values of SPEI growing season at different ages for individuals in the main sample. The $x$- and $y$-axes describe the age of the shock. The absolute value of the correlation between shocks in different periods is very low, with no such correlation exceeding 0.2.

## C.4 Construction of predicted harvest measure

To construct the predicted harvest index, I first take a number of climactic measures from the climatology literature that are likely to explain rice yields, and use them in as explanatory variables in a regression with rice yields from the IFLS data as an outcome.

### C.4.1 Summary of process

To construct the predicted harvest index, I first take a number of climactic measures from the climatology literature that are likely to explain rice yields, and use them in as explanatory variables in a regression with rice yields from the IFLS data as an outcome. The explanatory variables include the SPEI index at two different periods (0-2 months before planting, and planting to harvest), along with growing degree days (a measure of heat exposure widely used to predict crop yield (Schlenker et al., 2006)), measures of rainfall (total rainfall and number of rainy days (May, 2004)), the length of the longest dry spell over the harvest season (Tebaldi et al., 2006), and the delay in the arrival of the wet season (Naylor et al., 2007). For a rice crop $j$ farmed by household $h$ in district $r$ and in wave $t$, the estimating equation is:

$$IHS(RiceYield)_{jhrt} = \alpha + \sum_k \gamma_k WeatherRice_{jhrt,k} + \mathbf{X}'_{ht}\delta + \mathbf{Z}'_{jt}\lambda + \pi_r + \mu_t + \varepsilon_{jhrt} \tag{C.2}$$

$Weather_{jhrt,k}$ are the different weather-measure z-scores described above. $IHS(RiceYield)_{jhrt}$ is the inverse hyperbolic sine of the harvest yield, i.e. the harvest output in kg divided by the number of hectares of land harvested. The IHS specification allows me to include failed harvests, i.e. harvests with a yield of 0. $\mathbf{X}_{ht}$ is a vector of household controls that includes whether the household head is female and the religion of the household head. $\pi_r$ are district fixed effects, and $\mu_t$ are wave fixed effects.

The results in Table C.4.1 indicate that the main predictors of rice yields are (i) SPEI 0-2 months before average rice planting date, (ii) growing degree days over planting period, and (iii) wet season delay. Some models also show the longest dry spell to be significant. I use column (3) to construct a "predicted harvest index" for every year and district combination by taking the predicted value from this regression and normalising it to have mean 0 and standard deviation 1. The index can thus be thought of as a summary index for the weather features that predict high rice yields. show that using I use this predicted harvest index as an explanatory variable to complement the previous weather shock measure. I also use the models from columns (2) and (4)-(7) to construct alternative indexes that act as robustness checks for my results with the predicted harvest index. More details on the individual weather measures and the construction are given below.

### C.4.2 Weather definitions

In order to generate the richer set of weather measures, I use the following operational definitions:

- **Analysis year**. The analysis year starts on the 1st August, and finishes on 31st July the next year, following the approach seen in Naylor et al. (2007). The weather data available enable me to calculate measures for all the analysis years between 1981 and 2015.
- **Definition of start of wet season (yearly)**. The start of the wet season for a given district/subdistrict in a given analysis year is defined as the first day after August 1st that satisfies the following two conditions: (i) the cumulative rainfall for that analysis year is greater than 20cm (as in Naylor et al. (2007)); and (ii) the day is followed by at least 4 more consecutive rainy days (i.e. days with at least 0.1mm of rainfall), similar to the strategy seen in Macours et al. (2022).

**Table C.4.1:** *Impact of different weather features on rice yields*

| | IHS(harvest output / hectare) | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Growing degree days (Z) | 0.070 | 0.050** | 0.040* | 0.043* | 0.036 | | 0.040 |
| | (0.048) | (0.025) | (0.024) | (0.025) | (0.024) | | (0.024) |
| Wet season delay (Z) | −0.052** | −0.054** | −0.059** | −0.050** | −0.054** | −0.058** | |
| | (0.024) | (0.023) | (0.023) | (0.023) | (0.023) | (0.023) | |
| Longest dry spell (Z) | −0.062* | −0.041* | | −0.026 | | | |
| | (0.034) | (0.021) | | (0.021) | | | |
| SPEI (0-2m before rice planting) | 0.109*** | 0.116*** | 0.103** | | | 0.099** | 0.097** |
| | (0.040) | (0.039) | (0.040) | | | (0.040) | (0.040) |
| SPEI (rice planting to rice harvest) | −0.057 | | | | | | |
| | (0.061) | | | | | | |
| Num. rainy days (Z) | −0.036 | | | | | | |
| | (0.092) | | | | | | |
| Rain total (Z) | 0.004 | | | | | | |
| | (0.068) | | | | | | |
| Wave FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample mean | 9.079 | 9.079 | 9.079 | 9.079 | 9.079 | 9.079 | 9.079 |
| F statistic | 3.486 | 4.869 | 4.795 | 3.525 | 4.218 | 5.617 | 4.209 |
| p-value for F-test | 0.002 | 0.001 | 0.003 | 0.017 | 0.017 | 0.004 | 0.017 |
| $R^2$ | 0.189 | 0.189 | 0.188 | 0.186 | 0.186 | 0.187 | 0.187 |
| N | 4653 | 4653 | 4653 | 4653 | 4653 | 4653 | 4653 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by district. The construction of each of the weather variables used is found in Appendix Section C.4. The dependent variable in all columns is the inverse hyperbolic sine transformation of rice harvest output (in kg) divided by the area harvested (in hectares). Rice production data is only available in the IFLS data in waves 4 and 5, corresponding to 2007 and 2015. All explanatory variables are standardised to have a mean of 0 and a standard deviation of 1 within each district over the period 1981-2015. Model (3) is the one that is used to construct the main predicted harvest index used in the paper. Other models are later used as robustness checks, with corresponding results presented in the Appendix. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

- **Definition of end of wet season (yearly)**. The wet season is the first date for which at least 90% of cumulative rainfall in that analysis year has fallen, as in Naylor et al. (2007).
- **Average harvest/plant date**. For each district/subdistrict, I use the data on rice farming from IFLS4 and IFLS5 to calculate the average day on which rice is planted and harvested. To do this, I assign a given rice crop to be in the wet season if it was harvested earlier than the day that lies 2 months after the average end of the wet season (said average being taken over all years in that district/subdistrict). I then calculate the average harvest date for all crops in a given district/subdistrict that were assigned to the wet season.

### C.4.3 Raw weather measures

Here I describe the weather measures used to predict rice yield, and then as inputs to the predicted harvest measure.

The following measures are calculated at two different levels. For the rice crops, I have more detailed location data, and so can calculate weather measures at the *kecamatan* (subdistrict) level. I generate a number of measures denoted as $WeatherRice_{jhrt,k}$, where $j$ is a rice crop farmed by household $h$ in subdistrict $r$ and in survey-wave $t$, and $k = \{1, 2, ..., K\}$ denote the different weather measures. For individuals' historical shocks, I am only able to calculate weather measures at the *kabupaten* (district) level. To do this I calculate measures denoted as $WeatherDistrict_{ry,k}$, where $r$ is the district, $y$ is the analysis year, and $k$ denotes the different measures. I use these to then calculate a series of shocks for each individual over their life-course based on the monthly-migration history constructed from the IFLS data.

- **Wet season delay**. The average wet season start is calculated by taking the number of days after 1st August for the start of the wet season in every year in the data, and then taking the mean of this value. "Wet season delay" is then calculated as the difference between the wet season start for that year relative to the district/subdistrict average over all years. Since rice planting decisions are typically based on reaching a set threshold of accumulated rainfall, a delay in the monsoon rains can delay rice planting and reduce yields (Maccini & Yang, 2009). Naylor et al. (2007) demonstrate that delays in the monsoon rains of 30 days or more in Java and Bali can have significant negative impacts on rice production during the main rice harvest season.[7]

  Rice crops are assigned to a given analysis year by assigning them to the most recent year for which the rice crop was planted *later than* the day lying exactly 2 months before the average start of the rainy season (in order to allow for variation in rice planting due to yearly variation in the start of the season.)
- **SPEI (0-2m before planting)**. This is the average SPEI index calculated over $\{0, 1, 2\}$ months before planting. This time window is likely to be important because there needs to be sufficient moisture before and around planting time in order to prepare the land for cultivation and to facilitate the early rooting of transplanted seedlings. However, excessive water at the vegetative growth stage soon after planting can hamper rooting and reduce yields (De Datta, 1981; Naylor et al., 2007).
- **SPEI (during growing)**. This is the mean SPEI index between the month after planting and the month of harvest. For rice crops this measure is calculated relative to the actual plant date

[7] For example, they show that a 30-day delay causes rice production between January and April to fall by an average of 11% in East Java / Bali, and by 6.5% in West/Central Java.

of the crop measured, while for the individual/historical measure it is calculated relative to the average plant date of that district/subdistrict.

For all the following measures, they are defined in slightly different ways when applied to rice crops and when applied to individuals, given the differing data available for each observation type. When applying to rice crops, the measures are calculated using daily data from between the planting date *of that crop*, and the harvest date *of that crop*. Such a calculation is not possible when looking at the weather shocks experienced by individuals throughout their life course; instead, the measures are calculated using daily data from between the start and the end of the wet season for each year.

- **Growing degree days**. Following the climate change literature, growing degree days are constructed using the number of days with temperature within certain bounds, i.e. according to the following function:

$$g(x) = \begin{cases} 0 & \text{if } x < DD_{min} \\ x - DD_{min} & \text{if } DD_{min} < x < DD_{max} \\ DD_{max} - DD_{min} & \text{if } x > DD_{max} \end{cases}$$

  where $x$ is the temperature on a given day, $DD_{min} = 8°C$ and $DD_{max} = 32°C$. The growing degree days measure is then calculated as the sum of the $g(x)$ measure for the relevant date ranges outlined above. Such a measure is commonly used in the climate change and agronomy literature as a measure of heat exposure that predicts crop yields in a wide variety of contexts (Schlenker & Roberts, 2006; Schlenker et al., 2006; Ritchie & Nesmith, 2015; Fishman, 2018).

- **Rain total**. Rain total is the sum of daily rainfall in mm, calculated over the relevant date ranges outlined above.

- **Number of rainy days**. The number of rainy days is the total number of days for which rain was greater than 0.1mm within the relevant date ranges outlined above. This measure is commonly used in the climate change literature, and, when included alongside the total rainfall, can be interpreted as a measure of the variability of daily rainfall *within* a rainy season (Fishman, 2016, 2018; May, 2004).

- **Longest dry spell**. The length in days of the longest dry spell (calculated as consecutive days with no rain (<0.1mm)), calculated over the relevant date ranges outlined above. This measure is also used in the climate change literature and captures a type of intraseasonal variability in rainfall that is correlated with droughts and crop stress (Fishman, 2016, 2018; Tebaldi et al., 2006).

All the above weather measures are then standardised at the district/subdistrict level by taking the deviation from the average and dividing by the standard deviation. These standardised measures are taken over the entire period for which data is available (1981 to 2015), and exclude the year for which the z-score is being calculated.

### C.4.4 Construction of predicted harvest measure

I use the results in Table C.4.1 to construct a predicted harvest index $PredHarvest_{ry}$ for each district $r$ and analysis year $y$ in the following way:

$$PredHarvest_{ry} := W'_{ry}\hat{\gamma} \tag{C.3}$$

where $W_{ry}$ denotes the vector of weather variables defined at the district × year level seen in Table C.4.1, and $\hat{\gamma}$ is the vector of predicted coefficients on the weather variables seen in Table C.4.1. In

the main specification (based on column (3)), $W_{ry} \in \mathbb{R}^3$ includes Growing Degree Days (Z), Wet Season Delay (Z), and SPEI measured 0-2 months before rice planting.

## C.5 Empirical specification for timing

### C.5.1 Contemporaneous timing tests

In Sections 3.5.1 and 3.5.2 I test for the timing of the impact of weather variation on contemporaneous investments and expenditure using the specification described in this section. I define two 6-month periods during the year.[8] Let $\mathbb{1}\{AfterHarvest\}_{iry}$ be an indicator such that:

$$\mathbb{1}\{AfterHarvest\}_{iry} := \begin{cases} 0 & \text{when investment in } i \text{ is measured between 3 months before} \\ & \quad \text{harvest and 2 months after (inclusive)} \\ 1 & \text{when investment in } i \text{ is measured between 3 and 8 months} \\ & \quad \text{after harvest (inclusive)} \end{cases}$$

Using this variable I can run the following empirical specification that allows for a more granular test of timing at the 6-monthly level:

$$\begin{aligned} I_{iry} = \alpha + \beta_0 Weather_{ry} + \beta_1 Weather_{r,y-1} + \delta \cdot \mathbb{1}\{AfterHarvest\}_{iry} \; + \\ (\gamma_0 \cdot \mathbb{1}\{AfterHarvest\}_{iry} \cdot Weather_{ry}) \; + \; (\gamma_1 \cdot \mathbb{1}\{AfterHarvest\}_{iry} \cdot Weather_{r,y-1}) \\ + \; \pi_r + \mu_y + \varepsilon_{iry} \end{aligned} \tag{C.4}$$

Using the appropriate sums of the coefficients $\beta_0, \beta_1, \gamma_0,$ and $\gamma_1$ I can calculate the effect of yearly weather variation when measured in 4 six-month windows relative to the harvest, namely -3 to 2, 3 to 8, 9 to 14, and 15 to 20 months after harvest.[9]

### C.5.2 Effects on cognitive: 6-month intervals

Based on the results seen in the expenditure section, weather variation "bites" and affects household expenditure and nutritional investment between 3 and 8 months after harvest for that year. I use this result when analysing the timing of the effect of early life weather shocks on adult cognitive score.

For these results, I match individuals to weather variation by first using the migration history described in Appendix Section C.1.1 to calculate the region of residence for each sample individual $i$ in any given month $m$, call this $r(i,m)$. I then match individuals to weather variation at two levels, annual and 6-monthly. Each temporal level uses a slightly different approach:

**Annual shock measures**. I first define a *monthly* weather variation measure that accounts for the fact that the shock bites between 3 and 8 months after harvest. So for month $m$:

$$WeatherMonth_{rm} = \begin{cases} WeatherYear_{ry} & \text{if } m \text{ is between 3 and 8 months} \\ & \quad \text{after the avg. harvest month in year } y \\ 0 & \text{otherwise} \end{cases}$$

I then amalgamate these monthly shock measures to the annual level by taking the average shock value for each age. Consider an individual $i$. Let $\mathscr{A}(i,a)$ denote the set of months in which $i$ is age $a$.

[8] In Appendix Figure C.15.1 I examine the timing of weather effects at a more granular monthly level. The predicted harvest index appears to lead increases in nutritional investment and household expenditure for the period approximately between 3 and 8 months after harvest, thus motivating the 6-month periods chosen here.

[9] The effects of interest that are reported in Table 3.7 are $\beta_0$ for -3 to 2 months; $\beta_0 + \gamma_0$ for 3-8 months; $\beta_1$ for 9-14 months; $\beta_1 + \gamma_1$ for 15-20 months.

For example, if $i$ was someone born on 1st Jan 2000, $A(i,1) = \{$Jan 2001, Feb 2001, ..., Dec 2001$\}$. For individual $i$ at age $a$ residing in district $r$:

$$WeatherIndAnnual_{iar} = \frac{1}{12} \sum_{m \in \mathscr{A}(i,a)} WeatherMonth_{r(i,m),m}$$

Intuitively, I take the average value of the monthly shock variable over all the 12 months in which $i$ has been age $a$.

**6 month shock measures**. The monthly weather shock takes the same value for 6 months at a time. This means that if I try to calculate a 6-monthly shock in the same way as the annual shock, there will be very high pseudo-serial correlation between shocks (i.e., high correlation between the value of $i$'s shock for age $a$ and $a-1$). This can generate artefacts in the results, including misleading alternating patterns in the data where a strong positive coefficient is followed by a strong negative coefficient.[10]

In order to avoid this problem, I instead assign individuals in a given period to a 6-month "bucket", in the following sense. Starting with an individual $i$'s birth month, I split up their life into 6-month life-periods (I denote the period 1-6 months after birth as age 0, 7-12 months after birth as age 0.5, 13-18 months after birth as age 1, etc.) In a similar way, I split up the weather in a given district into 6-month buckets, where a month will be in the first bucket for that year if $m$ is between 3 months before and 2 months after the average harvest month, and will be in the second bucket for that year if $m$ is between 3 and 8 months after the average harvest month. The value of the predicted harvest measure will naturally take the value of 0 for the first bucket, and to $WeatherYear_{ry}$ when in the second bucket.

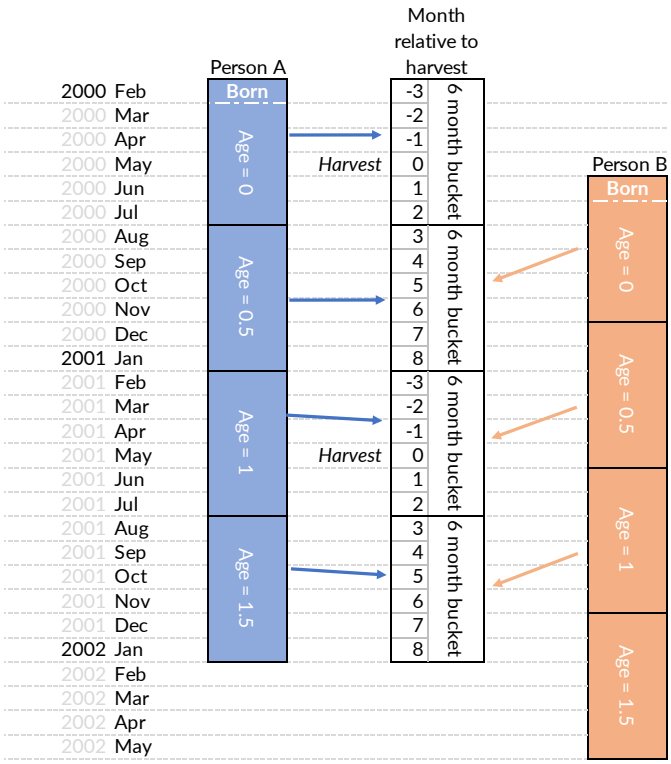
Then, for every 6-month life-period for every individual, I match it to the *single* 6-month weather bucket for that district, where the matching weather bucket will be the one that overlaps the most with the 6-month life period. Figure C.5.1 gives an example of this approach. In the district of analysis, the average harvest month is May. The 6-month weather buckets are thus defined by the white cells in the centre column. Person A was born in February 2000. Her life periods align perfectly with the weather buckets, so her "age = 0" period is assigned to the first bucket, "age = 0.5" is assigned to the second bucket, etc. Person B was born in June 2000. Because Person B's first life period ("age = 0") overlaps more with the second weather bucket (4 month overlap) than the first (2 month overlap), her first life period is assigned to the second weather bucket. Following similar logic, her "age = 0.5" life period is assigned to the third weather bucket, and "age = 1" is assigned to the fourth weather bucket, and so on.

One concern with this strategy is that this generates variation in shock-exposure even within a given bucket (and a given district). For example, in Figure C.5.1, if we think that the shock "bites" in months 3-8 after harvest, then person A will get 6 months of exposure to the shock at age 0.5, while person B will only get 4-months of exposure to the shock at age 0, even though they are assigned the same value of the shock variable. I deal with this issue in two ways. First, I control for district-of-birth × birth-month-of-year fixed effects in all specifications with bucket-assigned shocks. This ensures that the treatment effects are not estimated using variation from individuals born in different months of the year within the same district. The cumulative effect of being born in a month with 4-months of exposure (for example), compared to being born in a month with 6-months of exposure,

[10] Simulations confirm that such a pattern is likely to result when there is sufficient noise in the model, and when the shock variables are sufficiently serially correlated, even when such an alternating pattern does not exist in the data-generating process.

will be captured by the fixed effects and so won't confound the treatment effect estimates. Second, I run a robustness check (Appendix Table C.12.13) in which I multiply individuals' shock variable by their level of "exposure" to the shock. For example, person A's shock would remain unchanged, while person B's shock would be multiplied by a factor of $\frac{4}{6}$ to account for the weaker level of exposure. These do not qualitatively change the results.

**Figure C.5.1:** *6-month bucket assignment process*




### C.5.3 Bootstrapped standard errors for predicted harvest measure

For all results that make use of the predicted harvest measure, I bootstrap standard errors to take into account variation caused by the construction of the regressors. For example, to calculate the standard errors on the estimates of predicted harvest index on adult cognitive outcomes in Figure

3.3, I follow the process below:

1. I perform wild cluster resampling of the rice production sample. In other words, I resample at the district level, keeping all observations in a given district. I then calculate the effects of each weather measure $\hat{\gamma}$ using equation C.2.
2. I construct $PredHarvest_{ry}$ using the calculated coefficients $\hat{\gamma}$.
3. I construct 6-month shock measures based on the values of $PredHarvest_{ry}$ using the steps outlined in the previous section.
4. I match individuals in the main sample to these constructed shock measures.
5. I perform wild cluster resampling of the main cognitive sample, clustering at the district level.
6. I calculate the effect of these 6-month shock measures on adult cognitive scores in the bootstrapped sample.

I repeat this process 1000 times, and the standard errors reported in the tables are the standard deviations of the resampled treatment effects from the last step. The bootstrapped standard errors for the other tables using predicted harvest index are constructed in an analogous fashion.

## C.6 Conceptual framework

Here, I describe the theoretical framework on which my empirical results are founded. This framework allows me to give a precise definition of the concept of critical periods. It is based on the human capital production function model seen in Cunha & Heckman (2007) and Cunha et al. (2010) in which the timing of investment into human capital during childhood matters. Let $\boldsymbol{\theta}_t$ denote the vector of a child's stock of skills (including, for example, cognitive and socioemotional skills). In this model, skills evolve over the course of childhood in a way that depends on parental characteristics and human capital ($\boldsymbol{h}^P$), the child's existing stock of skills ($\boldsymbol{\theta}_t$), and parental investments in each period ($I_t$). Given my empirical setting, I add a variable $\mu_t$ that describes the exogenous weather shock to investment in period $t$.[11] $I_t$ can be thought of as denoting the component of investment resulting from a variety of exogenous and endogenous factors, whereas $\mu_t$ captures the specific exogenous component of investment that is driven by weather shocks. The way skills evolve from one age-period to the next is described using:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{f}_t(\boldsymbol{h}^P, \boldsymbol{\theta}_t, (I_t + \mu_t)) \tag{C.5}$$

All characters in boldface denote vectors. I assume for simplicity that investment and shocks are unidimensional scalars. In my empirical analysis, I allow for 17 distinct time periods to affect adult skills, ranging from 2 years before birth up to age 15.[12] I therefore assume that adult skills settle by age 16, and will not be affected by any changes thereafter. Whereas Equation C.5 is written in recursive form, we can substitute in each equation for $\boldsymbol{\theta}_{15}, \boldsymbol{\theta}_{14}, \boldsymbol{\theta}_{13}$, etc. to get an expression for the adult stock of skills ($\boldsymbol{h}$) as a function of initial endowments (from before birth, e.g. genetics) and

[11] This is similar to the adaptation seen in Currie & Almond (2011); Almond et al. (2018).

[12] The lower bound is based on evidence cited in the Introduction that suggests that cognitive traits are affected by shocks in utero. The upper bound is based on the fact that adult cognitive ability is measured for all adults above aged 15 in the IFLS data, and also based loosely on the discussion in Cunha & Heckman (2009, p. 331), which suggests that socioemotional skills in particular are "malleable until later ages".

all past investments and shocks:

$$\boldsymbol{h} := \boldsymbol{\theta}_{16} = \boldsymbol{m}\Big(\boldsymbol{h}^P, \boldsymbol{\theta}_{-2}, (I_{-2}+\mu_{-2}), (I_{-1}+\mu_{-1}), ..., (I_{15}+\mu_{15})\Big)$$

Or for a specific skill $k$ (e.g. cognitive ability), adult skills $h_k$ are

$$h_k = m_k\Big(\boldsymbol{h}^P, \boldsymbol{\theta}_{-2}, (I_{-2}+\mu_{-2}), (I_{-1}+\mu_{-1}), ..., (I_{15}+\mu_{15})\Big)$$

There are two quantities of interest when examining the effects of weather shocks on adult human capital. The first is $\partial m_k(.)/\partial \mu_t$, which I will call the "direct effect". This is the direct impact of a change in investment on adult human capital, ignoring any subsequent investment responses by parents. The second is $dm_k(.)/d\mu_t$, which I will call the "reduced-form effect". This describes the overall impact on adult human capital $h_k$ *while accounting for all parental responses*. In general, these two quantities will not be the same, because parents may react to early shocks by changing investments in later periods. If these later investments are reinforcing, the reduced-form effect will be larger than the direct effect; if they are compensating, the reduced-form effect will be smaller than the direct effect.

Corresponding to these two definitions of effects, we can define two notions of critical periods (building on the definitions seen in Cunha & Heckman (2007)). In particular, define $C$ as a time interval lasting from $t_{min}$ to $t_{max}$:

$$C = [t_{min}, t_{max}]$$

Then $C$ is a *direct effect critical period* for skill $k$ if and only if:

$$\frac{\partial m_k(.)}{\partial \mu_t} > 0 \quad \forall t \in C \qquad \text{and} \qquad \frac{\partial m_k(.)}{\partial \mu_t} = 0 \quad \forall t \notin C$$

And $C$ is a *reduced-form critical period* for skill $k$ if and only if:

$$\frac{dm_k(.)}{d\mu_t} > 0 \quad \forall t \in C \qquad \text{and} \qquad \frac{dm_k(.)}{d\mu_t} = 0 \quad \forall t \notin C$$

Intuitively, $C$ is a critical period when *only* shocks within the time interval have a (direct or reduced-form) effect on adult skills $h_k$, while shocks outside of the time interval have no effect on adult skills $h_k$.[13]

### C.6.1 Simple model with welfare

The simplified model in this section shows that we can only write the welfare impacts of an early life shock in terms of the "direct effect" in the specific case where parents are optimising agents and have full knowledge of the human capital production function (which allows us to use the envelope conditions to ignore investment responses).

Assume a simpler version of the model in Section C.6, in which human capital is a scalar, there is only 1 shock variable $\mu_1$ and we ignore $h^P$ and $\theta_1$ for simplicity. Parents can choose $I_2$ to react to the shock $\mu_1$, and $I_1$ is fixed at $\bar{I}$.

$$h(I_2; \mu_1) = m\Big(\mu_1, \bar{I}_1, I_2(\mu_1)\Big)$$

[13] The biomedical literature typically uses a less restrictive definition, where effects in critical periods are stronger but effects outside critical periods need not be 0. This corresponds to the definition of "sensitive periods" in Cunha & Heckman (2007).

Parental utility is determined by their lifetime consumption $C$, along with their child's human capital $h$. $\alpha$ determines their relative preference between these two factors:

$$U = (1-\alpha)\ln C + \alpha \ln h$$

The parental budget constraint is

$$\begin{aligned} Y &= C + p\bar{I}_1 + \frac{pI_2}{(1+r)} \\ \implies C(I_2) &= Y - p\left[\bar{I}_1 + \frac{I_2}{(1+r)}\right] \end{aligned}$$

where $p$ is the relative price of investment, and $r$ is the real exchange rate, and $Y$ is (exogenous) income. We can substitute the budget constraint and technology into the utility function to get it solely in terms of $I_2$ and $\mu_1$:

$$U(I_2;\mu_1) = (1-\alpha)\log\{C(I_2)\} + \alpha\log\{h(I_2;\mu_1)\}$$

If the prerequisites for the envelope conditions do not hold (e.g. because the parent does not optimise or doesn't have full knowledge of the human capital production function), then the parent chooses $\widetilde{I_2}$ and the welfare impact will be:

$$\frac{dU}{d\mu_1} = (1-\alpha)\frac{1}{C(\widetilde{I_2})}\frac{dC(\widetilde{I_2})}{d\mu_1} + \alpha\frac{1}{h(\widetilde{I_2};\mu_1)}\frac{dh(\widetilde{I_2};\mu_1)}{d\mu_1}$$

i.e. the welfare impact will be related to the *reduced form* impact of the shock $\mu_1$ on $h$.

We can derive expression for the welfare impact of $\mu_1$ in terms of the direct effect, but only in the specific case where we can assume that the envelope conditions hold, i.e. when the parent is maximising and has full knowledge of the human capital production function, and $\mu_1$ is marginal. In this case, the agent will maximise to choose some optimal $I_2^*(\mu_1)$ and welfare (or indirect utility) will be:

$$W(\mu_1) = \max_{I_2}\{U(I_2;\mu_1)\}$$

Then the envelope conditions for $I_2$ allow us to simplify the expression for $dW/d\mu_1$:

$$\begin{aligned} \frac{dW}{d\mu_1} &= \frac{dU(I_2^*(\mu_1),\mu_1)}{d\mu_1} \\ &= \underbrace{\frac{\partial U(I_2^*(\mu_1),\mu_1)}{\partial I_2}}_{=0}\cdot\frac{dI_2^*}{d\mu_1} + \frac{\partial U}{\partial \mu_1} = \frac{\partial U}{\partial \mu_1} = \frac{\partial U}{\partial h}\frac{\partial h}{\partial \mu_1} \\ \implies \frac{dW}{d\mu_1} &= \alpha\cdot\frac{1}{h}\cdot\frac{\partial h}{\partial \mu_1} \end{aligned}$$

So in this specific case, to capture the full welfare effects of a change in $\mu_1$ we only need to know the "direct effect" i.e. $\partial h/\partial t$, along with the parent's relative preference for human capital $\alpha$.

If the policy maker has another objective altogether, such as maximising government revenue, the reduced form effect will also be the determinant of welfare. For example, suppose a representative agent model in which individuals earn $hw$ and are taxed at an ad valorem rate of $\tau$. Then government revenue will be:

$$G = hw(1-\tau)$$

And the impact of the shock on government revenue will be:

$$\frac{dG}{d\mu_1} = \frac{dh}{d\mu_1} \cdot w(1-\tau)$$

But if the prerequisites for the envelope conditions do not hold (e.g. because the parent does not optimise), and the parent chooses $\widetilde{I_2}$ instead of $I_2^*$, then the welfare impacts will be:

$$\frac{dW}{d\mu_1} = (1-\alpha)\frac{1}{C}\frac{dC}{d\mu_1} + \alpha\frac{1}{h}\frac{dh}{d\mu_1}$$

## C.7 Cognitive placebo tests

The effect of *SPEI growing season* on cognitive skills is isolated to a single spike at age 2. This raises concerns that the age 2 coefficient is the unexpected result of some artefact in the data or some aspect of the empirical strategy. For example, we might worry that shocks are correlated across- and within-individuals in just such a way that, by chance, random noise in the data "aligns" across individuals to create an apparently strong effect at a single age. Alternatively, the patterns of correlation between shocks across and within individuals may imply that inference using F-tests is not sufficiently conservative. To alleviate these concerns, I construct two simulated placebo tests of my empirical specification.

**Before-conception shocks**. I randomly select (without replacement) 18 years from the set of years between 27 and 3 years before birth. I use the 18 corresponding shock variables as explanatory variables in the main estimating equation 3.1. This acts as a check of the F-test, ensuring that random variation from *before* an individual's life cannot generate seemingly significant joint effects on adult cognitive skills.

$SPEI_{i(t+j)}$ denotes the weather experienced by individual $i$ in the $j$th year after his/her birth in year $t$. When $j$ is negative, let $SPEI_{i(t+j)}$ be the weather in $i$'s district of birth $|j|$ years before he/she was born. For example, $SPEI_{i(t-10)}$ would be the SPEI index in $i$'s district of birth from 10 years before $i$ was born. The before-conception shocks placebo test is constructed as follows.

1. I construct a set $\mathscr{J}_{\text{hist}}$ of possible "historical" time periods from well before birth that should not affect adult outcomes. In this case, I choose $\mathscr{J}_{\text{hist}} = \{-27, -26, ..., -3\}$.[14]
2. For iteration $k = 1$:
   (a) Randomly select (without replacement) a subset $\mathscr{Z}_k$ of length 18 from the set $\mathscr{J}_{\text{hist}}$.[15]
   (b) Run the main specification but using the time periods from $\mathscr{Z}_k$, i.e. run
   $$Y_{irt} = \alpha + \sum_{z \in \mathscr{Z}_k} \beta_z SPEI_{i(t+z)} + \pi G_{ir} + \boldsymbol{X}_i'\gamma \ + \delta_r + \mu_t + \varepsilon_{irt}$$
3. Repeat step 2. for all iterations $k \in \{2, 3, ..., 1000\}$.

**Within-lifetime iterations**. I randomly "reshuffle" the within-lifetime shocks. Separately for each individual, the shock values for ages $(-2, -1, ..., 15)$ is reassigned to a different year in the set $\{-2, -1, ..., 15\}$. So for example, the shock value for age 0 might be reassigned to age 13, and the shock value for age 5 might be reassigned to age 1. The reassigned shock variables are then used as explanatory variables in equation 3.1. This can be seen as a verification of the q-value hypothesis tests, since it checks whether the chance alignment of noisy within-lifetime variation across individuals can generate patterns that look like strong effects in a single period (or critical periods). Let $\mathscr{J} = (-2, -1, ..., 14, 15)$ be the (ordered) list of time periods used in the main specification. Let $j_p$ be the $p$th element of $\mathscr{J}$.

[14] $-27$ is the minimum value because the oldest individual in my sample was born in 1988, and my SPEI data ranges back to 1961. This means that the furthest back before birth I can construct shocks without generating missing values is 27 years before birth. $-3$ is the maximum value because it is the closest time period to the year of birth for which there is very unlikely to be any effects of weather on adult outcomes ($-2$ and $-1$ may have effects due to sensitivity to shocks *in utero*).

[15] 18 values are chosen because there are 18 coefficients in my main specification.

1. For iteration $k = 1$:
   (a) For individual $i = 1$:
      i. Create a new set $\mathscr{Q}_i^{\text{reordered}}$ by taking $\mathscr{J}$ and reordering the items randomly (such that none of the items are in the position they were originally at in $\mathscr{J}$). Let $q_p$ denote the $p$th element of $\mathscr{Q}_i^{\text{reordered}}$. For example, if 5 is the 1st element of $\mathscr{Q}_i^{\text{reordered}}$, then $q_1 = 5$.
      ii. Define new placebo weather shock variables for individual $i$ according to the elements in $\mathscr{Q}_i^{\text{reordered}}$ in the following way: $Placebo_{ip} = SPEI_{i(t+q_p)}$. Suppose, for example, that for the first placebo shock that $q_1 = 5$ as above, then $Placebo_{i,1} = SPEI_{i(t+5)}$. This will yield 18 different $Placebo_{i,p}$ variables for each individual.
   (b) Repeat step (a) for all individuals $i \in \{2, ..., N\}$, such that each of their placebo shock variables are iterated differently.
   (c) Run a regression, analogous to the main specification but using the placebo variables as the main treatment variables. In other words, run:

$$Y_{irt} = \alpha + \sum_{\{1,...,18\}} \beta_p Placebo_{i,p} + \pi G_{ir} + \boldsymbol{X}_i'\gamma + \delta_r + \mu_t + \varepsilon_{irt} \tag{C.6}$$

2. Repeat step 2. for all iterations $k \in \{2, 3, ..., 1000\}$.

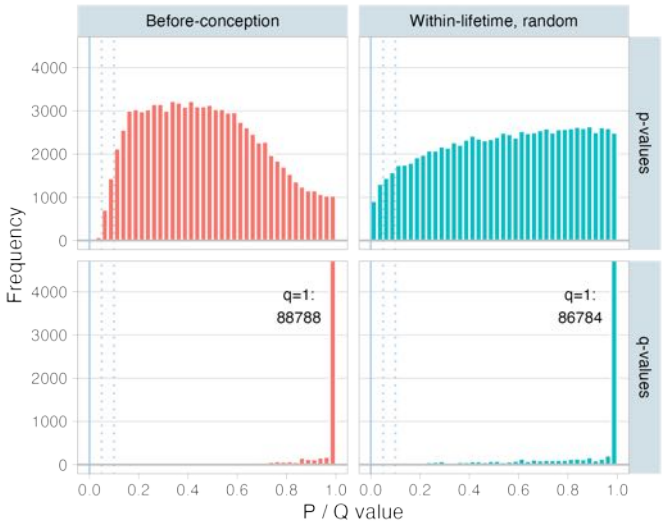
**Results**: The results of both placebo tests are found in Table C.7.1 and Figure C.7.2. For both types of placebos, I run three types of hypothesis tests: (i) whether the p-value is less than $\alpha = 0.05$ for each shock *coefficient* in the model, (ii) whether the q-value is less than $\alpha = 0.05$ for each shock *coefficient* in the model, and (iii) whether the F-test is significant at the $\alpha = 0.05$ level (one test for each *model*). For the p- and q-value tests, I also report at the model level whether at least one p- or q-value is significant in the model. Each placebo test is re-run for 5000 models.

The results here confirm that my results from Section 3.4.1 are unlikely to be driven by artefacts in the data. The before-conception placebo generates some significant p-values, but no significant q-values. This suggests that the strongly significant results on adult cognitive ability are highly likely to be driven by true variation from within an individual's lifetime, rather than just chance alignment of random noise. The F-test is never significant, which indicates that the inference on the F-statistic is not insufficiently conservative. The within-lifetime placebo also generates results that support the methodology. A high proportion of simulations (35.3%) lead to at least one significant p-value, but only 0.1% of simulations include at least one significant q-value. This indicates that the FDR-controlled multiple hypothesis adjustment is effective at reducing the likelihood of false positives.[16] The low proportion of simulations that result in significant q-values, even at the 5% level, means that we can be highly confident that the strongly significant coefficient at age 2 in my main specification (with a q-value of 0.002) is capturing a true phenomenon rather than random noise.

[16] Since the model for my main results (Section 3.4) leads to a rejection of the F-test, and the within-lifetime placebo test always uses the same variation but randomly "reshuffled", we also expect the F-test to be rejected in almost all cases, as indeed it is.

**Figure C.7.2:** *Placebo test results*

**(a)** *Distribution of p- and q-values for individual coefficients after 1000 simulations of each placebo method*




**(b)** *Distribution of p-values for F-statistics on all coefficients after 1000 simulations of each placebo method*

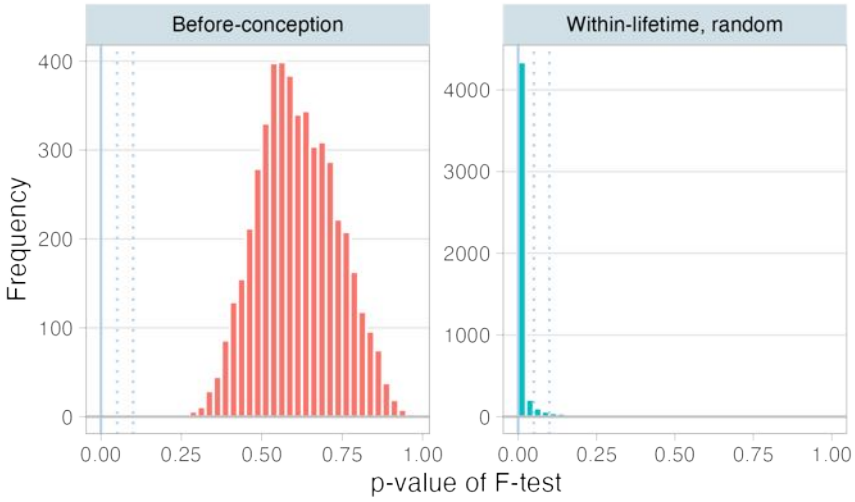


**Table C.7.1:** *Results of placebo tests*

| Placebo Type | Test Type | Total Tests | # Significant | Prop Significant |
|---|---|---|---|---|
| Before-conception | p - by coefficient | 90000 | 79 | 0.001 |
| Before-conception | p - at least one | 5000 | 74 | 0.015 |
| Before-conception | q - by coefficient | 90000 | 0 | 0.000 |
| Before-conception | q - at least one | 5000 | 0 | 0.000 |
| Before-conception | F | 5000 | 0 | 0.000 |
| Within-lifetime, random | p - by coefficient | 90000 | 2225 | 0.025 |
| Within-lifetime, random | p - at least one | 5000 | 1812 | 0.362 |
| Within-lifetime, random | q - by coefficient | 90000 | 50 | 0.001 |
| Within-lifetime, random | q - at least one | 5000 | 50 | 0.010 |
| Within-lifetime, random | F | 5000 | 4551 | 0.910 |

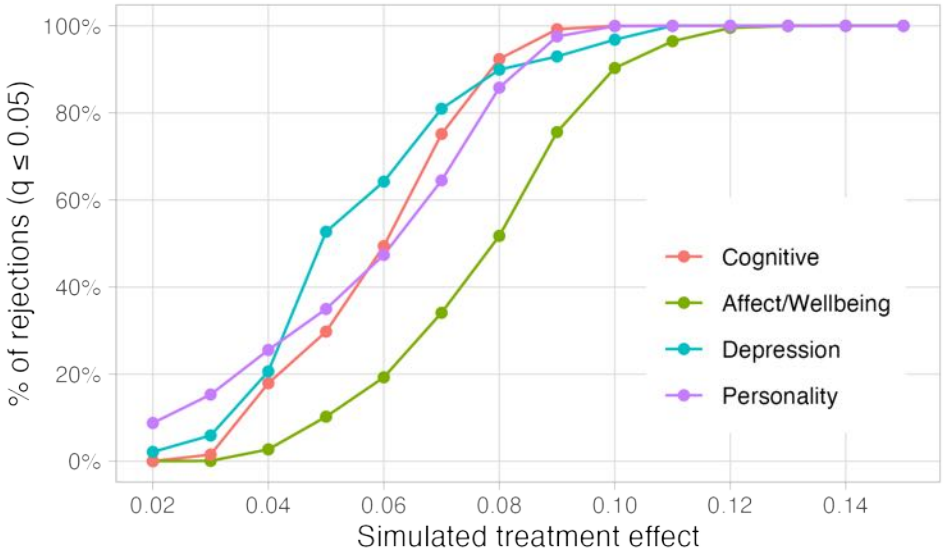
*Notes*: The "before-conception shocks" placebo describes the results of randomly selecting 18 shocks from significantly before an individual was born (27-3 years before birth) as the explanatory variables in Equation 3.1. The "within-lifetime, random" placebo describes the results of randomly reordering the actual shocks experienced (separately for each individual in the sample) and using these synthetic placebo shocks as the explanatory variables in Equation 3.1. Each model is estimated 5000 times. All tests are tested at the significance level $\alpha = 0.05$. "p, by coefficient" describes how many *coefficients* have a p-value below 0.05 (yielding $18 \times 5000 = 90,000$ tests), while "p, at least one" describes how many models include at least one p-value below 0.05. "q, by coefficient" and "q, at least one" are defined similarly but use the q-values calculated using the sharpened FDR control procedure from Anderson (2008). The "F" rows describe how many models result in a rejected F-test at the 5% level.

## C.8 Power simulations

I run simulations to estimate the power of my test for critical periods. I take before-conception shocks for each outcome as a null distribution, and then simulate treatment effects of size $\kappa$. More specifically, for each outcome variable:

1. I use the methodology from Section C.7 to generate simulated data that takes before-conception shocks as the main explanatory variables. I use the specification seen in equation C.6. The distribution of the estimated treatment effects of the placebo shocks ($\hat{\beta}_p$) is a proxy for the distribution of treatment effects in the main specification (equation 3.1) under the null that $\beta_j = 0$ for all $j$.
2. For a randomly selected placebo shock, I simulate a linear treatment effect of that shock of $\kappa$.
3. I repeat this process 5000 times and count the proportion of simulations in which the randomly selected shock has a q-value of less than 0.05 (i.e., is rejected after adjusting for multiple hypothesis testing). This is the estimated power for the given level of $\kappa$.

**Figure C.8.3:** *Simulated power: proportion of rejected q-values as a function of simulated treatment effect*




*Notes*: This shows the simulated power of each outcome at different levels of treatment effect $\kappa$, as described in the methodology in this section. The simulated data uses before-conception shocks as placebo shocks that proxy the null distribution, and then simulates an additional linear treatment effect on all individuals in the data of size $\kappa$ for one randomly selected placebo shock. This simulation is repeated 5000 times for each point on the graph. Each point denotes the proportion of q-values below 0.05 for the shock with a simulated treatment effect.

## C.9 Heterogeneous effects

A recent body of literature (e.g. Goodman-Bacon, 2018; de Chaisemartin & d'Haultfoeuille, 2020) has pointed out that in the presence of heterogeneous treatment effects there may be difficulties with the interpretation of the type of fixed-effect regressions that I use in my main specification, namely those with multiple groups $g$ and multiple time periods $t$. For example, (de Chaisemartin & d'Haultfoeuille, 2020) show that in a simple fixed-effects model with a common trends assumption, the coefficient on a fixed effect regression ($\beta_{fe}$) will estimate a *weighted* average of the $(g,t)$-cell-specific ATTs (average treatment effect on the treated). In general, the weights on each cell will not be equal, and some can be negative, implying that with heterogeneous treatment effects $\beta_{fe}$ can even be of a different sign to all the cell-specific ATTs.

This framework implies that the coefficient estimates in my main results will only have a causal interpretation if one of the following two conditions hold:

- *Homogenous effects*. The ATT of weather variation at a given age is constant across all (birth-district, birth-year) cells.
- *Uncorrelated heterogeneous effects*. The ATT of weather variation at a given age may vary across (birth-district, birth-year) cells, but the weight attached to each (birth-district, birth-year) in the fixed effect regression is uncorrelated with the ATT in that cell.

The first condition is implausible in my context. There are many reasons why the ATT of weather variation may be higher in some districts and time periods than others, because for example they have a more educated population (which could change how parental investment reacts to shocks), because some districts are more suitable for growing crops that are more sensitive to weather vari-

ation, or because there is a greater rural population that is dependent on farming, thus altering changing the general equilibrium effects of weather shocks on the local economy.

On the other hand, the uncorrelated heterogeneous effects assumption is plausible. Note first that the weights attached to each cell depend on the treatment status of that cell (see e.g. Proposition 1 in (de Chaisemartin & d'Haultfoeuille, 2020)). This means that any correlation between the treatment status of a cell (i.e. the weather variation experienced) and the ATT of that cell will threaten the validity of my fixed-effect coefficient estimates. However, the main treatment variables I am examining are year-on-year weather variations (away from the historical average), which are plausibly as-good-as random, and thus exogenous to determinants of the ATT. For example, whether a district experiences a positive weather shock in a given year is determined by climatological factors and will thus be independent of factors like the proportion of adults who have completed a primary school in that district, the level of urbanisation in that district, the suitability of different crops, etc. that may influence the ATT. If this is true, then the fixed-effects estimates used in my main results can be interpreted causally. Note that the assumption required here is similar to the main identification assumption described in the empirical strategy, although here I require conditional exogeneity with respect to the ATT instead of to the outcome variable.

I am also able to show empirical evidence to suggest that the uncorrelated heterogeneous effects assumption is likely to hold. To do this, I first replicate my main empirical specification (Equation 3.1) with adult cognitive scores as an outcome, and use the framework in (de Chaisemartin & d'Haultfoeuille, 2020) to estimate the weights attached to each district-year cell. Table C.9.1 then shows the results of regressing these weights on a number of observable characteristics that vary at the district-year level. None of the coefficients indicate a significant correlation between the estimated weights on SPEI growing season (at ages 0, 1, or 2) and any of the characteristics, which include the proportion of district area dedicated to rice farming, the population density, a measure of education (adult primary school completion), and the proportion of households in rural areas. The fact that these observable district characteristics are not correlated with the weights of the fixed-effect regression lends credence to the claim that the ATTs of each cell are also not correlated with the weights. This would imply that the fixed-effects regressions used in the main text successfully identify a meaningful causal effect.

**Table C.9.1:** *No significant correlation between the weights given to a (year × district) cell and observable characteristics of the cell*

| Variable | Treatment at age | Coefficient estimate | Std. error | p value | Number of cells |
|---|---|---|---|---|---|
| Area (proportion) dedicated to rice farming | 0 | 0.05 | 0.20 | 0.80 | 1463 |
| | 1 | 0.18 | 0.13 | 0.17 | 1463 |
| | 2 | -0.03 | 0.08 | 0.67 | 1463 |
| Population density (persons per km$^2$) | 0 | 465.25 | 988.62 | 0.64 | 1463 |
| | 1 | 58.71 | 670.22 | 0.93 | 1463 |
| | 2 | -264.32 | 434.17 | 0.54 | 1463 |
| Proportion of adults with completed primary school | 0 | -0.09 | 0.09 | 0.31 | 1463 |
| | 1 | -0.07 | 0.08 | 0.36 | 1463 |
| | 2 | -0.01 | 0.05 | 0.78 | 1463 |
| Proportion rural | 0 | 0.17 | 0.19 | 0.36 | 1463 |
| | 1 | 0.10 | 0.16 | 0.53 | 1463 |
| | 2 | 0.11 | 0.08 | 0.18 | 1463 |

*Notes*: The results in this table are calculated as follows. First, I run a fixed-effects regression (Equation 3.1) with adult cognitive score as the outcome variable. For the coefficients on SPEI growing season at ages 0, 1, and 2, I use the results in de Chaisemartin & d'Haultfoeuille, 2020 to calculate the weights attached to each (birth-district, birth-year) "cell" in this regression. Denote as $w^a_{r,t}$ the weight attached the cell $(r,t)$ for coefficient on age $a \in \{0,1,2\}$, where $r$ is a birth-district, and $t$ is a birth-year. I then regress these weights $w^a_{r,t}$ on the variable in the first column. Apart from the area dedicated to rice farming, the measures in the first column are calculated using Indonesian census data (Minnesota Population Center, 2020), and are allowed to vary at the district-year level by matching the year of the treatment variable (i.e. birth-year + 0, birth-year + 1, birth-year + 2) to the closest census date and using the data from that census year. Proportion rural denotes the proportion of households that are in rural areas. Area dedicated to rice farming is calculated from the Monfreda et al. (2008) data, and is time invariant (based on data from 2000). The "p-value" column denotes the p-value of the t-test that the coefficient is significantly different from 0.

## C.10 Multiple testing adjustment

The false discovery rate (FDR) is the expected proportion of false rejections from testing a set of null hypotheses $H_1, ..., H_n$ i.e. $E(H_i \text{ is true} | \text{Test rejects } H_i)$. For each of the main results tables in the paper, I calculate the Benjamini & Hochberg (1995) FDR-adjusted p-values using the sharpened procedure below based on the one seen in Anderson (2008). The basic Benjamini-Hochberg procedure is as follows:

1. For each of the coefficients $\beta_j, j \in \{-2,-1,...,15\}$, get the corresponding p-values from the primary specification (denoted $p_{-2}, p_{-1}, ..., p_{15}$)
2. Sort the p-values in ascending order so we have $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(17)}$
3. To test whether $\beta_j$ is rejected at the significance level $\alpha$:
4. (a) Find the largest $k$ such that $p_{(k)} \leq k \times \alpha/17$
   (b) Reject all hypotheses with a lower p-value than $p_{(k)}$ i.e. the coefficients associated with the p-values $p_{(1)}, ..., p_{(k)}$
5. Repeat step 3 for all values $\alpha$ (starting from 1.000, then 0.999, then 0.998 etc.). The q-value $q_j$ for $\beta_j$ is the smallest value of $\alpha$ at which step 3 rejects the hypothesis for that coefficient.

The above procedure is actually over-conservative if the true (but unknown) number of false null hypotheses is greater than 0. Given that, it is possible to "sharpen" the procedure and increase power using the following steps:

1. Apply the above procedure at a significance level $q' = q/(1+q)$. Let $c$ be the number of hypotheses rejected. If $c = 0$ then stop; otherwise go to step 2.

2. Let $\hat{m}_0 = M - c$.

3. Apply the above procedure at a level $q^* = q'M/\hat{m}_0$.

The degree to which the Benjamini & Hochberg (1995) procedure is conservative varies based on the dependency structure of the tested p-values $p_{-2}, p_{-1}, ..., p_{15}$. Suppose we set a significance threshold $\alpha_0 = 0.05$ such that we reject the null for $\beta_j$ if $\beta_j \leq \alpha_0$. If the p-values are independent or positively dependent, then the procedure above will ensure that $FDR \leq \alpha_0$, implying that we are controlling for the false discovery rate at the desired level of 5%. If, on the other hand, any of the p-values are negatively dependent, the procedure is overly conservative as it ensures that $FDR \lessapprox \alpha_0 / \ln(n) = \alpha_0 / \ln(17)$. Since the right-hand side of this equation is smaller than $\alpha_0$, this means we are controlling for a smaller $FDR$ than intended, and being overly conservative in inferring significant effects. Given my empirical setting, the p-values are likely to be positively dependent (for example, the effect of weather shocks at age 1 is likely to be positively correlated with the effect of weather shocks at age 2). But if they are negatively dependent, I am still successfully controlling for Type I errors.

## C.11 Sample selection

A relevant worry in my empirical setting is whether the main sample I am using is subject to a selection effect that could bias causal estimates of the impact of early-life weather shocks on adult outcomes. Ideally, my main sample would be a random sample of all individuals born between 1988 and 2000 in rural Indonesia (in the provinces surveyed by the IFLS). However, we may be worried about two primary sources of selection that could lead to a non-random sample. First, there may be selective mortality. For example, it may be the case that extreme weather shocks such as drought lead to infant death, and that deaths are concentrated on infants with poor health. If poor health in infancy is negatively correlated with adult cognitive ability, this would lead to a positively biased estimate in my main specification. Second, children who were born in IFLS households are only tracked when they move households if they were part of one of the original households from the very first wave, IFLS1. If children who were only incorporated into the IFLS as a result of later waves non-randomly migrate out of their households, the sample of individuals in my main sample will be similarly non-random.

In order to test for such selection effects, I make use of the detailed pregnancy history taken of each women above aged 15 in the IFLS survey. Using this data, I have a record of every pregnancy and birth for IFLS respondents, which amounts to 9226 records.[17] 54.5% of these records (N = 5028) can be matched to individuals in the main sample used in the analysis. But the rest are missing from the main sample: for example, 16.2% (1493) die before the age of 5 or result in still birth, and so do not have their adult outcomes measured. If this attrition is non-random and correlated with early life weather shocks, this could lead to biased estimates in my main analysis. The test proposed will not detect all sources of attrition bias: in particular, it does not account for any individuals who are not recorded in the pregnancy history data, which could result from a mothers' recall bias or refusal to respond. On the other hand, the data available give me a unique opportunity to test for selective mortality directly. Table C.11.1 shows the results of regressing measures of early life SPEI growing season on indicators for attrition, controlling as for the main analysis for birth-year and birth-district fixed effects.

In columns (1) and (2), I test whether early life SPEI affects the probability of dying before the age of 5 (or resulting in still birth). It appears that in utero $SPEI$ has a positive effect on the probability of dying before age 5.[18] The coefficient of 0.018 on $SPEI_{t-1}$ suggests that an increase in SPEI growing season of 1 standard deviation leads to an increase in the probability of dying before aged 5 by 1.8 percentage points, and the coefficient is significant at the 1% level. The direction of this effect is surprising given the main results in the paper: since SPEI is positively associated with household economic wellbeing and child anthropometrics, we would expect higher SPEI to have a negative impact on infant mortality. On the other hand, there may be non-economic explanations for such a result. For example, the impact shown here could possibly be explained by a change in the disease environment in utero. For example, wetter environments (higher SPEI) may lead to higher incidence of malaria and an increased risk of infant death (Kudamatsu et al., 2012). If in utero SPEI does indeed have a positive effect on the probability of infant death, this would likely lead to a positive bias on the results found in my main analysis: I would be missing weaker children (who are likely to have lower cognitive ability) that are subject to high SPEI from my analysis. However, this bias is likely to be isolated to the *in utero* period, rather than at age 2 where we see positive impacts of SPEI growing season on cognitive ability in the main results. The coefficient on $SPEI_{t-1}$

[17] I match pregnancies to districts using data on the mother's migration history.
[18] This appears to be driven mostly by an increase in the probability of a still birth.

is also relatively small. Thus while I cannot rule out selection bias due to mortality, it is unlikely to explain the large effects seen on adult cognitive skills at age 2 in Section 3.4.1.

In columns (3) and (4), I run a similar test to see whether early life SPEI affects the overall probability to be included in my main sample. No coefficients on early life SPEI are significant at the 10% level, suggesting no evidence for any selective inclusion in the final sample. Despite the potential selective mortality seen in columns (1) and (2), I cannot reject that there is no selection effect of $SPEI_{t-1}$ overall, implying that selection bias in the sample is unlikely to drive the main results in the paper.

**Table C.11.1:** *Testing for sample selection*

| | Dep Var: Died before age 5 | | | Dep Var: Included in the main sample | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | q value (2) | (3) | (4) | q value (4) |
| (Mean) SPEI growing season, ages -1 to 4 | 0.001 | | | 0.012 | | |
| | (0.023) | | | (0.033) | | |
| SPEI growing season, age = -1 | | 0.018*** | 0.009 | | −0.015 | 1 |
| | | (0.005) | | | (0.012) | |
| SPEI growing season, age = 0 | | −0.004 | 1 | | 0.011 | 1 |
| | | (0.006) | | | (0.007) | |
| SPEI growing season, age = 1 | | 0.003 | 1 | | 0.004 | 1 |
| | | (0.007) | | | (0.010) | |
| SPEI growing season, age = 2 | | −0.003 | 1 | | 0.015 | 1 |
| | | (0.006) | | | (0.009) | |
| SPEI growing season, age = 3 | | −0.005 | 1 | | 0.003 | 1 |
| | | (0.006) | | | (0.010) | |
| SPEI growing season, age = 4 | | 0.005 | 1 | | −0.009 | 1 |
| | | (0.006) | | | (0.008) | |
| SPEI growing season, age = 5 | | −0.005 | 1 | | −0.003 | 1 |
| | | (0.005) | | | (0.009) | |
| Birth year fixed effects | Yes | Yes | | Yes | Yes | |
| Birth district fixed effects | Yes | Yes | | Yes | Yes | |
| Sample mean | 0.162 | 0.162 | | 0.545 | 0.545 | |
| F statistic | 0.001 | 2.926 | | 0.141 | 1.391 | |
| p-value for F-test | 0.975 | 0.006 | | 0.714 | 0.294 | |
| Adj. $R^2$ | 0.024 | 0.025 | | 0.140 | 0.140 | |
| N | 9226 | 9226 | | 9226 | 9226 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. Sample includes all children born between 1988 and 2000 whose mother was an IFLS respondent and gave a pregnancy history in at least one IFLS wave. "Died before age 5" is an dummy variable that takes the value 1 if pregnancy ended in still birth or miscarriage, or if child was born and died before 5th birthday. "Included in Main Sample" is an indicator variable for whether child is in the main sample used in the paper (see Table 3.2 and surrounding discussion). Mean SPEI growing season ages -1 to 4 takes the mean of the SPEI growing season an individual experiences across ages -1 to 4 inclusive. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

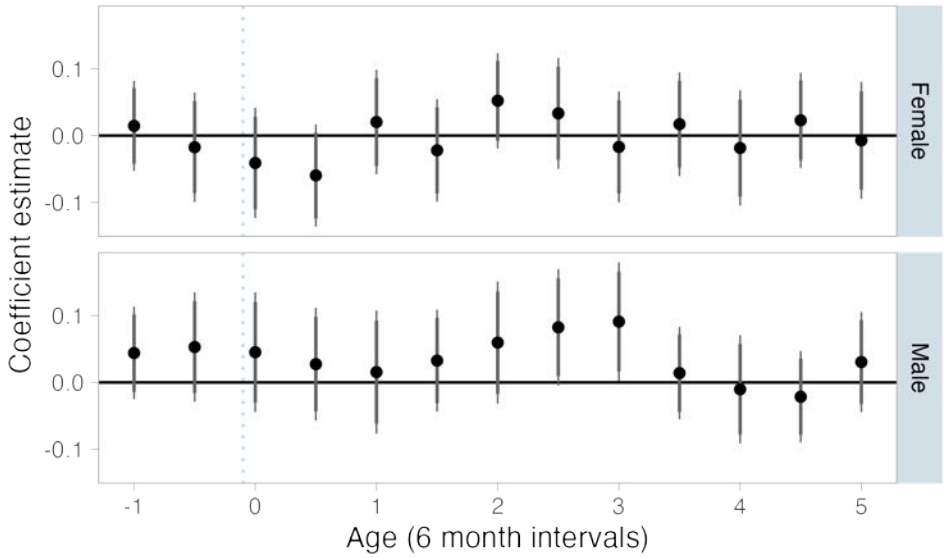
**Table C.11.2:** *Evidence against selective migration*

| | Dep Var: Individual Migrates in Year t | | Dep Var: Probability that migration destination at t is district d | |
|---|---|---|---|---|
| | SPEI in starting district (1) | q value (1) | SPEI in destination district (2) | q value (2) |
| SPEI growing season in t - 5 | 0.0002 (0.0017) | [1] | 0.0015 (0.0011) | [1] |
| SPEI growing season in t - 4 | −0.0050 (0.0037) | [1] | −0.0005 (0.0009) | [1] |
| SPEI growing season in t - 3 | 0.0041 (0.0034) | [1] | 0.0009 (0.0009) | [1] |
| SPEI growing season in t - 2 | −0.0008 (0.0034) | [1] | 0.0007 (0.0005) | [1] |
| SPEI growing season in t - 1 | −0.0046 (0.0069) | [1] | −0.0002 (0.0009) | [1] |
| SPEI growing season in t | 0.0034 (0.0043) | [1] | 0.0003 (0.0004) | [1] |
| SPEI growing season in t + 1 | 0.0002 (0.0021) | [1] | 0.0010 (0.0008) | [1] |
| SPEI growing season in t + 2 | 0.0032 (0.0016)* | [1] | 0.0013 (0.0008) | [1] |
| SPEI growing season in t + 3 | −0.0100 (0.0071) | [1] | −0.0007 (0.0005) | [1] |
| SPEI growing season in t + 4 | 0.0010 (0.0047) | [1] | 0.0001 (0.0006) | [1] |
| SPEI growing season in t + 5 | 0.0007 (0.0032) | [1] | −0.0001 (0.0005) | [1] |
| Individual FEs | Yes | | No | |
| Year FEs | Yes | | Yes | |
| Sample mean | 0.0282 | | 0.0054 | |
| F statistic | 1.5401 | | 0.9695 | |
| p-value for F-test | 0.1869 | | 0.5064 | |
| $R^2$ | 0.1334 | | 0.0940 | |
| N | 106375 | | 3355 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by individual. Includes all individuals in the main sample. Column (1) uses a panel dataset with observations at the (Individual × year) level. The dependent variable is a dummy variable indicating whether the individual migrated in that year. The explanatory variables are the SPEI growing season in the starting district (pre-migration). Column (2) uses a dataset at the (District × year) level. The outcome variable is the probability that an individual in the sample chooses district $d$ as their destination when migrating, calculated by taking the number of individuals who migrated to $d$ in year $t$ and dividing by the total number of individuals who migrated in year $t$.

## C.12 Additional tables: cognitive outcomes

**Figure C.12.1:** *Effect of SPEI growing season on cognitive Z-score using 6-month intervals, by gender*




*Notes*: The dependent variable in all models is the adult cognitive factor score (measured after age 15). The cognitive factor score is standardised to have 0 mean and standard deviation of 1. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. Individuals are matched to values of predicted harvest index using the process described in Section C.5.2. All coefficients control for birth-year fixed effects and individual-level controls (dummies for highest level of parental education, sex, and religion). 6-monthly specifications control for (district × month-of-birth fixed effects), to account for variation due to the timing of birth relative to the 6-month "bucket" individuals are assigned to (see Appendix Section C.5.2 for details.) All models use the individual-level attrition-corrected weights provided in the IFLS data.

**Table C.12.1:** *Cognitive robustness checks*

| | Sample 1985-2000 (1) | q value (1) | Sample 1990-2000 (2) | q value (2) | Sibling FEs (3) | q value (3) | SPEI 4 month (4) | q value (4) | Anderson Index (5) | q value (5) | Mean Index (6) | q value (6) | Urban (7) | q value (7) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SPEI growing season, age = -2 | −0.009 (0.014) | [0.826] | −0.005 (0.019) | [1] | 0.056 (0.030)* | [0.341] | −0.012 (0.019) | [1] | −0.003 (0.016) | [1] | −0.001 (0.016) | [1] | −0.013 (0.020) | [0.676] |
| SPEI growing season, age = -1 | 0.024 (0.013)* | [0.24] | 0.035 (0.016)** | [0.372] | 0.072 (0.028)** | [0.096] | 0.022 (0.014) | [1] | 0.027 (0.015)* | [1] | 0.027 (0.015)* | [1] | −0.030 (0.020) | [0.594] |
| SPEI growing season, age = 0 | −0.006 (0.013) | [0.882] | 0.003 (0.018) | [1] | 0.034 (0.030) | [0.532] | −0.006 (0.017) | [1] | −0.002 (0.015) | [1] | −0.004 (0.015) | [1] | 0.020 (0.028) | [0.676] |
| SPEI growing season, age = 1 | 0.020 (0.014) | [0.458] | 0.021 (0.021) | [1] | 0.044 (0.030) | [0.401] | 0.017 (0.018) | [1] | 0.020 (0.018) | [1] | 0.019 (0.018) | [1] | 0.025 (0.024) | [0.594] |
| SPEI growing season, age = 2 | 0.042 (0.013)*** | [0.027] | 0.057 (0.017)*** | [0.017] | 0.104 (0.032)*** | [0.022] | 0.057 (0.017)*** | [0.016] | 0.045 (0.014)*** | [0.039] | 0.048 (0.015)*** | [0.023] | 0.043 (0.024)* | [0.594] |
| SPEI growing season, age = 3 | 0.029 (0.014)** | [0.155] | 0.016 (0.021) | [1] | 0.061 (0.031)** | [0.337] | 0.031 (0.017)* | [1] | 0.018 (0.014) | [1] | 0.019 (0.014) | [1] | 0.021 (0.021) | [0.594] |
| SPEI growing season, age = 4 | 0.020 (0.016) | [0.501] | −0.019 (0.023) | [1] | −0.025 (0.031) | [0.713] | −0.006 (0.016) | [1] | −0.005 (0.016) | [1] | −0.005 (0.016) | [1] | −0.029 (0.022) | [0.594] |
| SPEI growing season, age = 5 | 0.034 (0.013)** | [0.106] | 0.005 (0.022) | [1] | 0.011 (0.030) | [0.913] | 0.014 (0.017) | [1] | 0.014 (0.015) | [1] | 0.016 (0.015) | [1] | 0.011 (0.031) | [0.676] |
| SPEI growing season, age = 6 | −0.006 (0.014) | [0.882] | −0.005 (0.021) | [1] | 0.045 (0.032) | [0.401] | −0.006 (0.017) | [1] | −0.007 (0.015) | [1] | −0.006 (0.015) | [1] | −0.048 (0.028)* | [0.594] |
| SPEI growing season, age = 7 | −0.001 (0.015) | [1] | −0.020 (0.025) | [1] | 0.055 (0.031)* | [0.341] | −0.024 (0.018) | [1] | −0.025 (0.017) | [1] | −0.024 (0.017) | [1] | −0.017 (0.027) | [0.676] |
| SPEI growing season, age = 8 | 0.021 (0.013) | [0.358] | −0.000 (0.021) | [1] | −0.013 (0.030) | [0.913] | −0.008 (0.014) | [1] | 0.010 (0.015) | [1] | 0.010 (0.015) | [1] | 0.023 (0.023) | [0.594] |
| SPEI growing season, age = 9 | 0.013 (0.015) | [0.769] | −0.012 (0.026) | [1] | 0.054 (0.034) | [0.396] | −0.005 (0.018) | [1] | −0.013 (0.017) | [1] | −0.009 (0.018) | [1] | −0.013 (0.026) | [0.676] |
| SPEI growing season, age = 10 | −0.017 (0.016) | [0.638] | −0.030 (0.026) | [1] | 0.002 (0.030) | [1] | −0.007 (0.019) | [1] | −0.018 (0.017) | [1] | −0.018 (0.017) | [1] | 0.039 (0.024) | [0.594] |
| SPEI growing season, age = 11 | −0.004 (0.015) | [0.928] | −0.039 (0.025) | [1] | 0.006 (0.031) | [1] | −0.021 (0.018) | [1] | −0.020 (0.018) | [1] | −0.020 (0.018) | [1] | 0.012 (0.029) | [0.676] |
| SPEI growing season, age = 12 | 0.031 (0.014)** | [0.155] | 0.011 (0.025) | [1] | 0.017 (0.033) | [0.868] | 0.027 (0.018) | [1] | 0.026 (0.020) | [1] | 0.026 (0.020) | [1] | −0.030 (0.027) | [0.594] |
| SPEI growing season, age = 13 | 0.012 (0.014) | [0.769] | −0.013 (0.027) | [1] | −0.004 (0.030) | [1] | 0.018 (0.019) | [1] | 0.004 (0.019) | [1] | 0.003 (0.019) | [1] | −0.046 (0.028)* | [0.594] |
| SPEI growing season, age = 14 | −0.002 (0.013) | [1] | −0.008 (0.020) | [1] | 0.017 (0.032) | [0.868] | 0.003 (0.015) | [1] | 0.008 (0.015) | [1] | 0.009 (0.015) | [1] | −0.045 (0.025)* | [0.594] |
| SPEI growing season, age = 15 | −0.010 (0.014) | [0.769] | −0.004 (0.022) | [1] | 0.016 (0.031) | [0.868] | 0.004 (0.017) | [1] | −0.010 (0.016) | [1] | −0.010 (0.016) | [1] | −0.030 (0.022) | [0.594] |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | | No | | Yes | | Yes | | Yes | | Yes | |
| Siblings FEs | No | | No | | Yes | | No | | No | | No | | No | |
| Birth Order FEs | No | | No | | Yes | | No | | No | | No | | No | |
| F statistic | 3.362 | | 2.036 | | 1.740 | | 1.979 | | 2.122 | | 2.180 | | 1.749 | |
| p-value for F-test | 0.000 | | 0.010 | | 0.028 | | 0.013 | | 0.007 | | 0.005 | | 0.035 | |
| $R^2$ | 0.195 | | 0.216 | | 0.879 | | 0.212 | | 0.207 | | 0.210 | | 0.262 | |
| N | 7309 | | 4483 | | 5572 | | 5572 | | 5572 | | 5572 | | 3416 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. Column (1) extends the sample by including individuals born between 1985 and 2000 (instead of just 1988 and 2000). Column (2) uses individuals born between 1990 and 2000. Column (3) uses siblings fixed effects and birth order fixed effects instead of birth district fixed effects. Column (4) uses a definition of SPEI based on a 4-month rolling average, rather than the 1-month average used in the main specification. Column (5) constructs the cognitive z-score using inverse covariance weighted indices from Anderson (2008). Column (6) uses cognitive indexes based on taking an unweighted sum of the z-scores for each sub-measure (Kling et al., 2007). Column (7) uses a different sample of individuals who are equivalent to the main sample, but were born in urban areas instead of in rural areas. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.2:** *Cognitive results are robust to the exclusion of later shocks*

| | Not demeaned | | | | Demeaned SPEI | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Up to age 2 (1) | q value (1) | Up to age 7 (2) | q value (2) | All ages (3) | q value (3) | Up to age 2 (4) | q value (4) | Up to age 7 (5) | q value (5) |
| SPEI growing season, age = -2 | −0.004 (0.016) | [0.794] | −0.001 (0.017) | [1] | −0.003 (0.018) | [1] | −0.004 (0.016) | [0.794] | −0.001 (0.017) | [1] |
| SPEI growing season, age = -1 | 0.019 (0.015) | [0.548] | 0.023 (0.016) | [0.438] | 0.026 (0.015)* | [0.884] | 0.019 (0.015) | [0.548] | 0.023 (0.016) | [0.438] |
| SPEI growing season, age = 0 | −0.003 (0.014) | [0.794] | 0.001 (0.015) | [1] | −0.000 (0.015) | [1] | −0.003 (0.014) | [0.794] | 0.001 (0.015) | [1] |
| SPEI growing season, age = 1 | 0.019 (0.017) | [0.548] | 0.019 (0.017) | [0.615] | 0.018 (0.019) | [1] | 0.019 (0.017) | [0.548] | 0.019 (0.017) | [0.615] |
| SPEI growing season, age = 2 | 0.063 (0.016)*** | [0.001] | 0.063 (0.016)*** | [0.001] | 0.063 (0.016)*** | [0.002] | 0.063 (0.016)*** | [0.001] | 0.063 (0.016)*** | [0.001] |
| SPEI growing season, age = 3 | | | 0.032 (0.014)** | [0.121] | 0.030 (0.015)* | [0.796] | | | 0.032 (0.014)** | [0.121] |
| SPEI growing season, age = 4 | | | −0.005 (0.018) | [1] | −0.003 (0.017) | [1] | | | −0.005 (0.018) | [1] |
| SPEI growing season, age = 5 | | | 0.019 (0.013) | [0.438] | 0.017 (0.016) | [1] | | | 0.019 (0.013) | [0.438] |
| SPEI growing season, age = 6 | | | 0.000 (0.017) | [1] | 0.000 (0.017) | [1] | | | 0.000 (0.017) | [1] |
| SPEI growing season, age = 7 | | | −0.016 (0.017) | [0.615] | −0.020 (0.020) | [1] | | | −0.016 (0.017) | [0.615] |
| SPEI growing season, age = 8 | | | | | 0.012 (0.016) | [1] | | | | |
| SPEI growing season, age = 9 | | | | | 0.000 (0.020) | [1] | | | | |
| SPEI growing season, age = 10 | | | | | −0.015 (0.019) | [1] | | | | |
| SPEI growing season, age = 11 | | | | | −0.021 (0.019) | [1] | | | | |
| SPEI growing season, age = 12 | | | | | 0.022 (0.021) | [1] | | | | |
| SPEI growing season, age = 13 | | | | | −0.003 (0.020) | [1] | | | | |
| SPEI growing season, age = 14 | | | | | −0.001 (0.016) | [1] | | | | |
| SPEI growing season, age = 15 | | | | | −0.009 (0.018) | [1] | | | | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | | Yes | | Yes | | Yes | |
| F statistic | 4.686 | | 3.708 | | 2.490 | | 4.686 | | 3.708 | |
| p-value for F-test | 0.000 | | 0.000 | | 0.001 | | 0.000 | | 0.000 | |
| $R^2$ | 0.210 | | 0.211 | | 0.212 | | 0.210 | | 0.211 | |
| N | 5572 | | 5572 | | 5572 | | 5572 | | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. Columns (1) and (2) show that early coefficients are largely unchanged when removing later SPEI growing season variables. Columns (3) to (5) run a similar analysis but using a dataset in which all SPEI variables are demeaned so that the mean of each SPEI growing season variable is 0. By enforcing assumption (ii) in Footnote 12, this specification avoids the bias that can arise if there are omitted interaction effects on shock variables. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.3:** *Cognitive "missing middle" by gender*

| | Dep var: Cog at 7-14 | | Dep var: Cog at 15+ | | | |
|---|---|---|---|---|---|---|
| | Female (1) | Male (2) | Female (3) | Female (4) | Male (5) | Male (6) |
| SPEI growing season, age = -1 | 0.014 | 0.023 | 0.025 | 0.022 | 0.045* | 0.040* |
| | (0.026) | (0.024) | (0.023) | (0.023) | (0.024) | (0.024) |
| SPEI growing season, age = 0 | −0.023 | 0.030 | 0.004 | 0.010 | 0.007 | −0.001 |
| | (0.027) | (0.027) | (0.023) | (0.022) | (0.024) | (0.022) |
| SPEI growing season, age = 1 | −0.001 | 0.021 | 0.031 | 0.031 | 0.006 | 0.000 |
| | (0.025) | (0.026) | (0.022) | (0.021) | (0.028) | (0.026) |
| SPEI growing season, age = 2 | 0.024 | 0.033 | 0.062** | 0.056** | 0.074** | 0.066** |
| | (0.025) | (0.032) | (0.024) | (0.024) | (0.029) | (0.027) |
| SPEI growing season, age = 3 | 0.026 | 0.030 | 0.018 | 0.011 | 0.053** | 0.045** |
| | (0.027) | (0.026) | (0.023) | (0.023) | (0.023) | (0.020) |
| SPEI growing season, age = 4 | −0.020 | −0.019 | −0.007 | −0.001 | −0.012 | −0.007 |
| | (0.026) | (0.027) | (0.031) | (0.029) | (0.024) | (0.022) |
| SPEI growing season, age = 5 | 0.024 | −0.029 | 0.008 | 0.001 | −0.005 | 0.003 |
| | (0.024) | (0.024) | (0.024) | (0.023) | (0.022) | (0.021) |
| SPEI growing season, age = 6 | 0.006 | 0.008 | 0.021 | 0.019 | −0.009 | −0.011 |
| | (0.027) | (0.024) | (0.025) | (0.024) | (0.024) | (0.023) |
| Cog score (aged 7-14) | | | | 0.256*** | | 0.251*** |
| | | | | (0.026) | | (0.024) |
| Birth year FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth district FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Siblings FEs | No | No | No | No | No | No |
| Birth Order FEs | No | No | No | No | No | No |
| F statistic | 0.694 | 0.791 | 1.783 | 1.361 | 2.171 | 1.961 |
| p-value for F-test | 0.696 | 0.612 | 0.085 | 0.218 | 0.033 | 0.055 |
| $R^2$ | 0.218 | 0.203 | 0.286 | 0.326 | 0.231 | 0.275 |
| N | 2181 | 2189 | 2181 | 2181 | 2189 | 2189 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with only SPEI growing seasons from age -1 to age 6 as explanatory variables. The first two columns use the cognitive score measured between ages 7 and 14 as the dependent variable. Columns (3) to (6) use cognitive score measured in adulthood (above aged 15) as the dependent variable. The column titles indicate whether only females or males are included in the sample. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.4:** *Cognitive outcomes are robust to inclusion of age fixed effects*

| | Dep Var: Adult Cognitive Score | | | |
|---|---|---|---|---|
| | Main spec | q value | Age FEs | q value |
| | (1) | (1) | (2) | (2) |
| SPEI growing season, age = -2 | 0.007 (0.017) | [1] | 0.007 (0.017) | [1] |
| SPEI growing season, age = -1 | 0.028 (0.016)* | [1] | 0.030 (0.016)* | [0.941] |
| SPEI growing season, age = 0 | −0.002 (0.015) | [1] | −0.001 (0.016) | [1] |
| SPEI growing season, age = 1 | 0.012 (0.019) | [1] | 0.011 (0.018) | [1] |
| SPEI growing season, age = 2 | 0.065 (0.017)*** | [0.003] | 0.066 (0.017)*** | [0.002] |
| SPEI growing season, age = 3 | 0.023 (0.015) | [1] | 0.025 (0.015)* | [0.941] |
| SPEI growing season, age = 4 | −0.005 (0.017) | [1] | −0.005 (0.017) | [1] |
| SPEI growing season, age = 5 | 0.012 (0.015) | [1] | 0.010 (0.015) | [1] |
| SPEI growing season, age = 6 | 0.005 (0.017) | [1] | 0.004 (0.016) | [1] |
| SPEI growing season, age = 7 | −0.017 (0.020) | [1] | −0.015 (0.020) | [1] |
| SPEI growing season, age = 8 | 0.006 (0.015) | [1] | 0.006 (0.015) | [1] |
| SPEI growing season, age = 9 | −0.004 (0.019) | [1] | −0.005 (0.019) | [1] |
| SPEI growing season, age = 10 | −0.018 (0.018) | [1] | −0.016 (0.018) | [1] |
| SPEI growing season, age = 11 | −0.024 (0.018) | [1] | −0.025 (0.017) | [1] |
| SPEI growing season, age = 12 | 0.020 (0.021) | [1] | 0.020 (0.021) | [1] |
| SPEI growing season, age = 13 | −0.002 (0.019) | [1] | −0.001 (0.019) | [1] |
| SPEI growing season, age = 14 | −0.004 (0.017) | [1] | −0.004 (0.018) | [1] |
| SPEI growing season, age = 15 | −0.017 (0.017) | [1] | −0.019 (0.017) | [1] |
| Birth year FEs | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | |
| Age (Year) FEs | No | | Yes | |
| Sample mean | −0.003 | | −0.003 | |
| F statistic | 2.859 | | 2.934 | |
| p-value for F-test | 0.000 | | 0.000 | |
| $R^2$ | 0.190 | | 0.191 | |
| N | 7109 | | 7109 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. Instead of taking the average adult cognitive score when it is measured in multiple waves, I here leave the observations at the (individual × wave) level to allow the inclusion of age fixed effects. Therefore, while the set of individuals is exactly the same as in the main results (e.g. Table 3.3), the number of observations is greater. Age fixed effects are dummies for each year of age. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.5:** *Shocks in early life have weakly cumulative effects*

| | Dep var: Adult Cognitive Score | | |
|---|---|---|---|
| | Shock: SPEI≤-0.5 (1) | Shock: SPEI≤0 (2) | Shock: SPEI≤0.5 (3) |
| # negative shocks = 1 | −0.043 | −0.025 | 0.034 |
| | (0.030) | (0.030) | (0.064) |
| # negative shocks = 2 | −0.069 | −0.121*** | −0.028 |
| | (0.047) | (0.045) | (0.063) |
| # negative shocks ≥ 3 | −0.190** | −0.143** | −0.085 |
| | (0.079) | (0.062) | (0.059) |
| Birth year FEs | Yes | Yes | Yes |
| Birth district FEs | Yes | Yes | Yes |
| $R^2$ | 0.208 | 0.209 | 0.209 |
| N | 5572 | 5572 | 5572 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The outcome variable is cognitive factor score measured in adulthood (above aged 15). Cognitive factor score is internally standardised, so the sample mean is 0 by construction. The table examines the effect of *how many* negative shocks are experienced in early life. *# negative shocks* is defined as the number of years between ages 1 and 3 that SPEI was below a certain threshold. Column (1) defines a negative shock as one where SPEI growing season is below -0.5, column (2) defines it as below 0, and column (3) as below 0.5. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data.

**Table C.12.6:** *Dynamic complementarity in adolescence is mostly driven by males*

| | Dep Var: Adult Cog Z-score | | | | | |
|---|---|---|---|---|---|---|
| | Female (1) | Male (2) | Female (3) | Male (4) | Female (5) | Male (6) |
| SPEI growing season, ages 0-3 | 0.073* | 0.180*** | 0.053 | 0.101* | 0.050 | 0.120* |
| | (0.043) | (0.058) | (0.043) | (0.059) | (0.047) | (0.062) |
| SPEI growing season, ages 10-12 | −0.019 | −0.006 | | | | |
| | (0.041) | (0.049) | | | | |
| SPEI growing season, (ages 0-3 x ages 10-12) | −0.000 | −0.113 | | | | |
| | (0.060) | (0.082) | | | | |
| SPEI growing season, ages 13-15 | | | −0.021 | 0.038 | | |
| | | | (0.046) | (0.046) | | |
| SPEI growing season, (ages 0-3 x ages 13-15) | | | 0.078 | 0.101** | | |
| | | | (0.055) | (0.048) | | |
| SPEI growing season, ages 10-15 | | | | | −0.054 | 0.045 |
| | | | | | (0.087) | (0.073) |
| SPEI growing season, (ages 0-3 x ages 10-15) | | | | | 0.086 | 0.049 |
| | | | | | (0.074) | (0.078) |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth district fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.251 | 0.227 | 0.252 | 0.227 | 0.252 | 0.226 |
| N | 2971 | 2601 | 2971 | 2601 | 2971 | 2601 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running versions of Equation 3.1. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. All columns use cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Columns (1), (3) and (5) only include females, and other columns only include males. *SPEI growing season, ages 0-3* denotes the mean value of SPEI growing season for an individual across ages 0, 1, 2, and 3. The age ranges 13-15 and 10-15 are defined analogously. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.7:** *Dynamic complementarity: effects on cognitive are not driven by changes in education*

**Panel A:** No interaction terms

| | Dep Var: Years of schooling | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Both (1) | Female (2) | Male (3) | Both (4) | Female (5) | Male (6) | Both (7) | Female (8) | Male (9) |
| SPEI growing season, ages 0-3 | −0.092 | −0.287* | 0.097 | | | | | | |
| | (0.126) | (0.153) | (0.177) | | | | | | |
| SPEI growing season, ages 10-12 | | | | −0.065 | −0.049 | −0.053 | | | |
| | | | | (0.106) | (0.140) | (0.132) | | | |
| SPEI growing season, ages 13-15 | | | | | | | −0.169* | −0.225* | −0.185 |
| | | | | | | | (0.096) | (0.120) | (0.140) |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth district fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Age fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample mean | 10.129 | 10.248 | 10.005 | 10.129 | 10.248 | 10.005 | 10.129 | 10.248 | 10.005 |
| $R^2$ | 0.340 | 0.384 | 0.361 | 0.340 | 0.383 | 0.361 | 0.341 | 0.384 | 0.362 |
| N | 5010 | 2732 | 2278 | 5010 | 2732 | 2278 | 5010 | 2732 | 2278 |

**Panel B:** Interaction terms

| | Dep Var: Years of schooling | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Both (1) | Female (2) | Male (3) | Both (4) | Female (5) | Male (6) | Both (7) | Female (8) | Male (9) |
| SPEI growing season, ages 0-3 | −0.144 | −0.325* | 0.041 | −0.049 | −0.316* | 0.187 | −0.071 | −0.358* | 0.173 |
| | (0.138) | (0.190) | (0.179) | (0.142) | (0.160) | (0.213) | (0.163) | (0.200) | (0.223) |
| SPEI growing season, ages 10-12 | −0.040 | −0.020 | −0.036 | | | | | | |
| | (0.109) | (0.143) | (0.135) | | | | | | |
| SPEI growing season, (ages 0-3 x ages 10-12) | 0.153 | 0.105 | 0.169 | | | | | | |
| | (0.181) | (0.219) | (0.283) | | | | | | |
| SPEI growing season, ages 13-15 | | | | −0.164 | −0.132 | −0.239 | | | |
| | | | | (0.100) | (0.125) | (0.154) | | | |
| SPEI growing season, (ages 0-3 x ages 13-15) | | | | −0.022 | 0.184 | −0.112 | | | |
| | | | | (0.125) | (0.149) | (0.205) | | | |
| SPEI growing season, ages 10-15 | | | | | | | −0.282 | −0.199 | −0.379 |
| | | | | | | | (0.179) | (0.252) | (0.231) |
| SPEI growing season, (ages 0-3 x ages 10-15) | | | | | | | 0.046 | 0.284 | −0.075 |
| | | | | | | | (0.219) | (0.282) | (0.299) |
| Birth year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth district fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Age fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample mean | 10.129 | 10.248 | 10.005 | 10.129 | 10.248 | 10.005 | 10.129 | 10.248 | 10.005 |
| $R^2$ | 0.341 | 0.384 | 0.361 | 0.341 | 0.385 | 0.362 | 0.341 | 0.385 | 0.362 |
| N | 5010 | 2732 | 2278 | 5010 | 2732 | 2278 | 5010 | 2732 | 2278 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running versions of Equation 3.1. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. All columns use *years of schooling* measured in above age 15 as the dependent variable. To control for age, age fixed effects are included. *SPEI growing season, ages 0-3* denotes the mean value of SPEI growing season for an individual across ages 0, 1, 2, and 3. The age ranges 13-15 and 10-15 are defined analogously. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.8:** *Effect of SPEI growing season on cognitive: 2 year intervals*

| | (1) | q value (1) |
|---|---|---|
| SPEI growing season, ages -2 to -1 | 0.022 | [1] |
| | (0.024) | |
| SPEI growing season, ages 0 to 1 | 0.005 | [1] |
| | (0.015) | |
| SPEI growing season, ages 2 to 3 | 0.064*** | [0.001] |
| | (0.014) | |
| SPEI growing season, ages 4 to 5 | 0.006 | [1] |
| | (0.014) | |
| SPEI growing season, ages 6 to 7 | −0.004 | [1] |
| | (0.015) | |
| SPEI growing season, ages 8 to 9 | 0.010 | [1] |
| | (0.015) | |
| SPEI growing season, ages 10 to 11 | −0.016 | [1] |
| | (0.017) | |
| SPEI growing season, ages 12 to 13 | 0.023 | [1] |
| | (0.019) | |
| SPEI growing season, ages 14 to 15 | −0.001 | [1] |
| | (0.015) | |
| F statistic | 3.779 | |
| p-value for F-test | 0.000 | |
| $R^2$ | 0.211 | |
| N | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are in parentheses and are clustered by birth district. The table uses a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. *SPEI growing season, ages 2 to 3* is the mean SPEI growing season for the individual, taken across ages 2 and 3. Other age ranges are defined similarly. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.9:** *Effect of weather on cognitive development, 6 month intervals and extended age range*

| | Dep Var: Cognitive Z Score | | | |
|---|---|---|---|---|
| | SPEI growing season (1) | q value (1) | Predicted harvest index (2) | q value (2) |
| Age = -2 | 0.033 (0.029) | [1] | −0.003 (0.031) | [1] |
| Age = -1.5 | −0.004 (0.025) | [1] | −0.015 (0.030) | [1] |
| Age = -1 | 0.028 (0.025) | [1] | −0.028 (0.028) | [1] |
| Age = -0.5 | 0.013 (0.028) | [1] | 0.015 (0.030) | [1] |
| Age = 0 | 0.024 (0.029) | [1] | 0.018 (0.031) | [1] |
| Age = 0.5 | −0.021 (0.028) | [1] | −0.013 (0.029) | [1] |
| Age = 1 | 0.025 (0.028) | [1] | −0.027 (0.029) | [1] |
| Age = 1.5 | −0.006 (0.027) | [1] | −0.006 (0.027) | [1] |
| Age = 2 | 0.057 (0.028)** | [1] | −0.002 (0.027) | [1] |
| Age = 2.5 | 0.079 (0.027)*** | [0.149] | 0.076 (0.029)*** | [0.5] |
| Age = 3 | 0.036 (0.030) | [1] | 0.017 (0.027) | [1] |
| Age = 3.5 | 0.016 (0.026) | [1] | −0.014 (0.030) | [1] |
| Age = 4 | −0.004 (0.030) | [1] | 0.015 (0.028) | [1] |
| Age = 4.5 | −0.014 (0.024) | [1] | −0.029 (0.029) | [1] |
| Age = 5 | 0.029 (0.029) | [1] | −0.028 (0.028) | [1] |
| Age = 5.5 | 0.018 (0.026) | [1] | −0.013 (0.030) | [1] |
| Age = 6 | −0.022 (0.032) | [1] | −0.032 (0.028) | [1] |
| Age = 6.5 | 0.010 (0.025) | [1] | −0.044 (0.032) | [1] |
| Age = 7 | −0.005 (0.031) | [1] | −0.027 (0.031) | [1] |
| Age = 7.5 | −0.017 (0.027) | [1] | −0.010 (0.029) | [1] |
| Age = 8 | 0.024 (0.030) | [1] | −0.015 (0.028) | [1] |
| Age = 8.5 | −0.001 (0.026) | [1] | −0.038 (0.030) | [1] |
| Age = 9 | 0.045 (0.030) | [1] | −0.017 (0.031) | [1] |
| Age = 9.5 | −0.039 (0.027) | [1] | −0.044 (0.032) | [1] |
| Age = 10 | −0.055 (0.031)* | [1] | −0.011 (0.030) | [1] |
| Age = 10.5 | −0.015 (0.029) | [1] | −0.038 (0.032) | [1] |
| Age = 11 | −0.011 (0.033) | [1] | −0.004 (0.037) | [1] |
| Age = 11.5 | −0.024 (0.031) | [1] | −0.038 (0.034) | [1] |
| Age = 12 | 0.008 (0.032) | [1] | −0.007 (0.033) | [1] |
| Age = 12.5 | 0.021 (0.034) | [1] | −0.035 (0.033) | [1] |
| Age = 13 | −0.007 (0.033) | [1] | 0.027 (0.033) | [1] |
| Age = 13.5 | 0.013 (0.033) | [1] | −0.036 (0.032) | [1] |
| Age = 14 | 0.022 (0.030) | [1] | 0.040 (0.031) | [1] |
| Age = 14.5 | −0.009 (0.029) | [1] | −0.003 (0.033) | [1] |
| Age = 15 | −0.012 (0.031) | [1] | −0.005 (0.029) | [1] |
| Birth year FEs | Yes | | Yes | |
| District x month-of-birth FEs | Yes | | Yes | |
| $R^2$ | 0.424 | | 0.424 | |
| N | 5572 | | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is the adult cognitive factor score (measured after age 15). The cognitive factor score is standardised to have 0 mean and standard deviation of 1. Column (1) uses the SPEI growing season as the main explanatory variable, and column (2) uses the predicted harvest index. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. Individuals are matched to values of predicted harvest index using the process described in Section C.5.2. 6-monthly specifications control for (district × month-of-birth fixed effects), to account for variation due to the timing of birth relative to the 6-month "bucket" individuals are assigned to (see Appendix Section C.5.2 for details.) All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.10:** *Cognitive results using predicted harvest while including later ages*

| | Dep Var: Adult Cognitive Score | | | | | |
|---|---|---|---|---|---|---|
| | Both (1) | q value (1) | Females Only (2) | q value (2) | Males Only (3) | q value (3) |
| Predicted Harvest Index, Age = -2 | −0.012 (0.021) | [1] | −0.026 (0.037) | [1] | 0.003 (0.033) | [1] |
| Predicted Harvest Index, Age = -1 | −0.001 (0.020) | [1] | 0.017 (0.030) | [1] | 0.000 (0.029) | [1] |
| Predicted Harvest Index, Age = 0 | 0.015 (0.021) | [1] | 0.018 (0.029) | [1] | 0.016 (0.032) | [1] |
| Predicted Harvest Index, Age = 1 | −0.014 (0.019) | [1] | −0.023 (0.029) | [1] | 0.004 (0.025) | [1] |
| Predicted Harvest Index, Age = 2 | 0.038 (0.021)* | [0.956] | −0.001 (0.029) | [1] | 0.082 (0.028)*** | [0.067] |
| Predicted Harvest Index, Age = 3 | 0.020 (0.019) | [1] | 0.001 (0.028) | [1] | 0.051 (0.029)* | [0.751] |
| Predicted Harvest Index, Age = 4 | −0.005 (0.019) | [1] | −0.043 (0.028) | [1] | 0.032 (0.030) | [0.947] |
| Predicted Harvest Index, Age = 5 | −0.023 (0.021) | [1] | −0.016 (0.034) | [1] | −0.020 (0.028) | [1] |
| Predicted Harvest Index, Age = 6 | −0.040 (0.021)* | [0.956] | −0.027 (0.035) | [1] | −0.047 (0.026)* | [0.751] |
| Predicted Harvest Index, Age = 7 | −0.008 (0.021) | [1] | 0.001 (0.037) | [1] | −0.017 (0.028) | [1] |
| Predicted Harvest Index, Age = 8 | −0.005 (0.020) | [1] | 0.030 (0.035) | [1] | −0.037 (0.026) | [0.841] |
| Predicted Harvest Index, Age = 9 | −0.015 (0.024) | [1] | −0.010 (0.033) | [1] | −0.036 (0.034) | [0.947] |
| Predicted Harvest Index, Age = 10 | −0.023 (0.023) | [1] | −0.019 (0.031) | [1] | −0.025 (0.038) | [1] |
| Predicted Harvest Index, Age = 11 | −0.035 (0.022) | [0.956] | −0.032 (0.035) | [1] | −0.044 (0.037) | [0.947] |
| Predicted Harvest Index, Age = 12 | 0.003 (0.019) | [1] | −0.023 (0.026) | [1] | 0.015 (0.030) | [1] |
| Predicted Harvest Index, Age = 13 | 0.014 (0.021) | [1] | −0.017 (0.036) | [1] | 0.040 (0.029) | [0.841] |
| Predicted Harvest Index, Age = 14 | 0.038 (0.023)* | [0.956] | 0.038 (0.032) | [1] | 0.043 (0.032) | [0.841] |
| Predicted Harvest Index, Age = 15 | −0.017 (0.022) | [1] | 0.010 (0.034) | [1] | −0.026 (0.034) | [1] |
| Birth year FEs | Yes | | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | | Yes | |
| District x month-of-birth FEs | No | | No | | No | |
| F statistic | 1.435 | | 0.712 | | 2.189 | |
| p-value for F-test | 0.120 | | 0.796 | | 0.005 | |
| $R^2$ | 0.212 | | 0.252 | | 0.237 | |
| N | 5572 | | 2971 | | 2601 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is the adult cognitive factor score (measured after age 15). The cognitive factor score is standardised to have 0 mean and standard deviation of 1. Column (1) includes both genders, column (2) only includes females, and column (3) only includes males. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. Individuals are matched to values of predicted harvest index using the process described in C.5.2. Yearly specifications control for birth district fixed effects. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.11:** *Cognitive results are robust to other weather indexes (yearly intervals)*

| | GDD + Wet | | | | | | GDD + SPEI | | | | | | Wet + SPEI | | | | | | GDD + Wet + SPEI + Dry Spell | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Both Sexes (1) | q value (1) | Male Only (2) | q value (2) | Female Only (3) | q value (3) | Both Sexes (4) | q value (4) | Male Only (5) | q value (5) | Female Only (6) | q value (6) | Both Sexes (7) | q value (7) | Male Only (8) | q value (8) | Female Only (9) | q value (9) | Both Sexes (10) | q value (10) | Male Only (11) | q value (11) | Female Only (12) | q value (12) |
| Age = -1 | 0.004 | [1] | 0.011 | [1] | 0.012 | [1] | −0.002 | [1] | −0.010 | [0.805] | 0.023 | [1] | 0.000 | [1] | −0.005 | [1] | 0.022 | [1] | −0.005 | [1] | −0.015 | [1] | 0.019 | [1] |
| | (0.020) | | (0.031) | | (0.028) | | (0.018) | | (0.024) | | (0.028) | | (0.018) | | (0.025) | | (0.028) | | (0.019) | | (0.028) | | (0.027) | |
| Age = 0 | 0.020 | [1] | 0.024 | [0.948] | 0.019 | [1] | 0.018 | [1] | 0.026 | [0.627] | 0.018 | [1] | 0.020 | [1] | 0.026 | [0.623] | 0.021 | [1] | 0.013 | [1] | 0.013 | [1] | 0.022 | [1] |
| | (0.019) | | (0.029) | | (0.030) | | (0.018) | | (0.025) | | (0.028) | | (0.018) | | (0.026) | | (0.028) | | (0.018) | | (0.026) | | (0.028) | |
| Age = 1 | 0.002 | [1] | 0.037 | [0.948] | −0.034 | [1] | −0.005 | [1] | 0.010 | [0.805] | −0.010 | [1] | −0.004 | [1] | 0.016 | [0.659] | −0.013 | [1] | −0.013 | [1] | 0.016 | [1] | −0.034 | [1] |
| | (0.019) | | (0.026) | | (0.031) | | (0.017) | | (0.023) | | (0.025) | | (0.018) | | (0.024) | | (0.026) | | (0.018) | | (0.025) | | (0.029) | |
| Age = 2 | 0.045** | [0.201] | 0.067** | [0.1] | 0.028 | [1] | 0.041** | [0.244] | 0.085*** | [0.016] | 0.004 | [1] | 0.047** | [0.141] | 0.093*** | [0.008] | 0.007 | [1] | 0.039* | [0.598] | 0.081*** | [0.032] | 0.001 | [1] |
| | (0.020) | | (0.027) | | (0.032) | | (0.019) | | (0.027) | | (0.023) | | (0.020) | | (0.028) | | (0.025) | | (0.020) | | (0.028) | | (0.026) | |
| Age = 3 | −0.015 | [1] | 0.027 | [0.948] | −0.053* | [0.838] | 0.027 | [0.5] | 0.038 | [0.452] | 0.031 | [1] | 0.025 | [0.877] | 0.042* | [0.319] | 0.023 | [1] | 0.012 | [1] | 0.041* | [0.371] | −0.003 | [1] |
| | (0.019) | | (0.026) | | (0.029) | | (0.017) | | (0.023) | | (0.025) | | (0.018) | | (0.024) | | (0.026) | | (0.018) | | (0.024) | | (0.025) | |
| Age = 4 | −0.004 | [1] | 0.023 | [0.948] | −0.039 | [1] | −0.001 | [1] | 0.034 | [0.512] | −0.036 | [1] | −0.001 | [1] | 0.037 | [0.392] | −0.039 | [1] | −0.013 | [1] | 0.027 | [1] | −0.062*** | [0.054] |
| | (0.022) | | (0.033) | | (0.029) | | (0.016) | | (0.025) | | (0.024) | | (0.017) | | (0.026) | | (0.024) | | (0.017) | | (0.028) | | (0.023) | |
| Age = 5 | −0.022 | [1] | −0.025 | [0.948] | −0.008 | [1] | 0.002 | [1] | 0.007 | [0.805] | 0.003 | [1] | −0.000 | [1] | 0.004 | [1] | 0.003 | [1] | −0.005 | [1] | 0.002 | [1] | −0.008 | [1] |
| | (0.019) | | (0.029) | | (0.032) | | (0.020) | | (0.026) | | (0.030) | | (0.020) | | (0.028) | | (0.032) | | (0.020) | | (0.028) | | (0.032) | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| District x month-of-birth FEs | No | | No | | No | | No | | No | | No | | No | | No | | No | | No | | No | | No | |
| F statistic | 1.144 | | 1.970 | | 1.123 | | 1.063 | | 1.874 | | 1.006 | | 1.094 | | 2.108 | | 0.885 | | 0.835 | | 1.934 | | 1.302 | |
| p-value for F-test | 0.338 | | 0.062 | | 0.351 | | 0.389 | | 0.077 | | 0.429 | | 0.369 | | 0.046 | | 0.520 | | 0.559 | | 0.068 | | 0.252 | |
| $R^2$ | 0.209 | | 0.231 | | 0.251 | | 0.209 | | 0.231 | | 0.250 | | 0.209 | | 0.232 | | 0.250 | | 0.209 | | 0.231 | | 0.251 | |
| N | 5569 | | 2600 | | 2969 | | 5572 | | 2601 | | 2971 | | 5572 | | 2601 | | 2971 | | 5572 | | 2601 | | 2971 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is the adult cognitive factor score (measured after age 15). The cognitive factor score is standardised to have 0 mean and standard deviation of 1. The first column for each index includes both genders, the second column only includes females, and the third column only includes males. The first three columns use an index constructed in an analogous way to the predicted harvest index but based on just using Wet Season Delay and the SPEI to construct the index (column (2) of Table C.4.1). Columns (4)-(6) use an index constructed from column (5) of Table C.4.1, columns (7)-(9) use an index constructed from column (7) of Table C.4.1, and columns (10)-(12) use an index constructed from column (6) of Table C.4.1. The harvest date is defined as the average harvest date of rice in that district based on IFLS harvest data (see Section C.4.2). Individuals are matched to values of predicted harvest index using the process described in C.5.2. The yearly specifications control for birth district fixed effects. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.12.12:** *Cognitive results are robust to other weather indexes (6 month intervals)*

| | GDD + Wet | | | | | | GDD + SPEI | | | | | | Wet + SPEI | | | | | | GDD + Wet + SPEI + Dry Spell | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Both Sexes (1) | q value (1) | Male Only (2) | q value (2) | Female Only (3) | q value (3) | Both Sexes (4) | q value (4) | Male Only (5) | q value (5) | Female Only (6) | q value (6) | Both Sexes (7) | q value (7) | Male Only (8) | q value (8) | Female Only (9) | q value (9) | Both Sexes (10) | q value (10) | Male Only (11) | q value (11) | Female Only (12) | q value (12) |
| Age = -12 to -7 months | −0.000 | [1] | −0.015 | [1] | −0.000 | [1] | −0.032 | [1] | −0.037 | [1] | −0.007 | [1] | −0.030 | [1] | −0.033 | [1] | −0.010 | [1] | −0.025 | [1] | −0.039 | [1] | −0.006 | [1] |
| | (0.024) | | (0.042) | | (0.039) | | (0.024) | | (0.040) | | (0.046) | | (0.025) | | (0.040) | | (0.047) | | (0.025) | | (0.043) | | (0.044) | |
| Age = -6 to -1 months | 0.018 | [1] | −0.048 | [0.689] | 0.042 | [1] | 0.019 | [1] | −0.028 | [1] | 0.044 | [1] | 0.021 | [1] | −0.035 | [1] | 0.048 | [1] | 0.017 | [1] | −0.042 | [1] | 0.050 | [0.807] |
| | (0.029) | | (0.044) | | (0.043) | | (0.025) | | (0.039) | | (0.038) | | (0.025) | | (0.041) | | (0.038) | | (0.026) | | (0.045) | | (0.038) | |
| Age = 0 to 5 months | 0.034 | [1] | 0.013 | [1] | 0.013 | [1] | 0.012 | [1] | 0.027 | [1] | 0.007 | [1] | 0.011 | [1] | 0.026 | [1] | 0.002 | [1] | 0.015 | [1] | 0.009 | [1] | 0.020 | [1] |
| | (0.029) | | (0.054) | | (0.044) | | (0.026) | | (0.038) | | (0.036) | | (0.028) | | (0.040) | | (0.038) | | (0.026) | | (0.040) | | (0.036) | |
| Age = 6 to 11 months | −0.042 | [0.924] | −0.062 | [0.604] | −0.051 | [1] | 0.001 | [1] | 0.003 | [1] | −0.019 | [1] | 0.001 | [1] | −0.006 | [1] | −0.015 | [1] | −0.015 | [1] | −0.027 | [1] | −0.015 | [1] |
| | (0.028) | | (0.043) | | (0.049) | | (0.026) | | (0.038) | | (0.040) | | (0.026) | | (0.038) | | (0.041) | | (0.026) | | (0.038) | | (0.040) | |
| Age = 12 to 17 months | 0.014 | [1] | −0.001 | [1] | 0.011 | [1] | −0.016 | [1] | −0.014 | [1] | −0.002 | [1] | −0.014 | [1] | −0.012 | [1] | −0.006 | [1] | −0.025 | [1] | −0.019 | [1] | −0.026 | [1] |
| | (0.025) | | (0.035) | | (0.039) | | (0.025) | | (0.035) | | (0.036) | | (0.026) | | (0.035) | | (0.037) | | (0.025) | | (0.033) | | (0.036) | |
| Age = 18 to 23 months | −0.020 | [1] | 0.008 | [1] | −0.096** | [0.303] | −0.000 | [1] | 0.003 | [1] | −0.028 | [1] | 0.001 | [1] | 0.007 | [1] | −0.036 | [1] | 0.000 | [1] | 0.022 | [1] | −0.051 | [0.807] |
| | (0.027) | | (0.048) | | (0.046) | | (0.025) | | (0.046) | | (0.037) | | (0.026) | | (0.047) | | (0.039) | | (0.026) | | (0.045) | | (0.040) | |
| Age = 24 to 29 months | 0.016 | [1] | 0.056 | [0.604] | −0.026 | [1] | 0.014 | [1] | 0.064* | [0.373] | −0.036 | [1] | 0.016 | [1] | 0.074** | [0.208] | −0.043 | [1] | 0.005 | [1] | 0.068* | [0.429] | −0.063* | [0.701] |
| | (0.027) | | (0.043) | | (0.041) | | (0.024) | | (0.036) | | (0.031) | | (0.024) | | (0.037) | | (0.032) | | (0.025) | | (0.039) | | (0.034) | |
| Age = 30 to 35 months | 0.089*** | [0.022] | 0.160*** | [0.002] | 0.030 | [1] | 0.071*** | [0.115] | 0.130*** | [0.024] | −0.008 | [1] | 0.078*** | [0.07] | 0.146*** | [0.008] | −0.005 | [1] | 0.076*** | [0.098] | 0.149*** | [0.004] | −0.001 | [1] |
| | (0.028) | | (0.041) | | (0.043) | | (0.027) | | (0.041) | | (0.034) | | (0.028) | | (0.041) | | (0.036) | | (0.028) | | (0.040) | | (0.039) | |
| Age = 36 to 41 months | −0.025 | [1] | 0.012 | [1] | −0.039 | [1] | 0.026 | [1] | 0.039 | [1] | 0.009 | [1] | 0.024 | [1] | 0.044 | [1] | 0.002 | [1] | −0.001 | [1] | 0.038 | [1] | −0.036 | [0.807] |
| | (0.025) | | (0.043) | | (0.036) | | (0.022) | | (0.037) | | (0.034) | | (0.023) | | (0.038) | | (0.034) | | (0.023) | | (0.038) | | (0.027) | |
| Age = 42 to 47 months | −0.001 | [1] | 0.064 | [0.604] | −0.026 | [1] | −0.005 | [1] | 0.023 | [1] | −0.004 | [1] | −0.005 | [1] | 0.033 | [1] | −0.010 | [1] | −0.011 | [1] | 0.038 | [1] | −0.027 | [1] |
| | (0.031) | | (0.047) | | (0.044) | | (0.025) | | (0.034) | | (0.040) | | (0.026) | | (0.035) | | (0.041) | | (0.026) | | (0.034) | | (0.039) | |
| Age = 48 to 53 months | −0.002 | [1] | 0.079 | [0.604] | −0.088** | [0.303] | 0.026 | [1] | 0.084** | [0.154] | −0.033 | [1] | 0.024 | [1] | 0.089** | [0.171] | −0.042 | [1] | 0.011 | [1] | 0.096** | [0.159] | −0.067** | [0.472] |
| | (0.026) | | (0.052) | | (0.037) | | (0.023) | | (0.036) | | (0.032) | | (0.024) | | (0.039) | | (0.033) | | (0.023) | | (0.042) | | (0.030) | |
| Age = 54 to 59 months | −0.008 | [1] | 0.012 | [1] | −0.007 | [1] | −0.016 | [1] | 0.009 | [1] | −0.039 | [1] | −0.018 | [1] | 0.012 | [1] | −0.040 | [1] | −0.021 | [1] | −0.007 | [1] | −0.038 | [1] |
| | (0.029) | | (0.043) | | (0.054) | | (0.024) | | (0.037) | | (0.040) | | (0.024) | | (0.038) | | (0.042) | | (0.025) | | (0.037) | | (0.041) | |
| Age = 60 to 65 months | −0.048* | [0.525] | −0.066* | [0.604] | −0.040 | [1] | −0.004 | [1] | −0.016 | [1] | 0.006 | [1] | −0.009 | [1] | −0.021 | [1] | 0.001 | [1] | −0.014 | [1] | −0.014 | [1] | −0.008 | [1] |
| | (0.025) | | (0.039) | | (0.035) | | (0.021) | | (0.031) | | (0.036) | | (0.021) | | (0.031) | | (0.036) | | (0.022) | | (0.031) | | (0.033) | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Birth district FEs | No | | No | | No | | No | | No | | No | | No | | No | | No | | No | | No | | No | |
| District x month-of-birth FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| F statistic | 2.029 | | 2.887 | | 1.627 | | 1.306 | | 1.787 | | 0.687 | | 1.389 | | 1.963 | | 0.819 | | 1.332 | | 1.943 | | 1.336 | |
| p-value for F-test | 0.016 | | 0.001 | | 0.082 | | 0.202 | | 0.049 | | 0.774 | | 0.157 | | 0.027 | | 0.639 | | 0.187 | | 0.029 | | 0.196 | |
| $R^2$ | 0.422 | | 0.570 | | 0.552 | | 0.421 | | 0.568 | | 0.550 | | 0.421 | | 0.569 | | 0.550 | | 0.421 | | 0.570 | | 0.551 | |
| N | 5569 | | 2600 | | 2969 | | 5572 | | 2601 | | 2971 | | 5572 | | 2601 | | 2971 | | 5572 | | 2601 | | 2971 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is the adult cognitive factor score (measured after age 15). The cognitive factor score is standardised to have 0 mean and standard deviation of 1. The first column for each index includes both genders, the second column only includes females, and the third column only includes males. The first three columns use an index constructed in an analogous way to the predicted harvest index but based on just using Wet Season Delay and the SPEI to construct the index (column (2) of Table C.4.1). Columns (4)-(6) use an index constructed from column (5) of Table C.4.1, columns (7)-(9) use an index constructed from column (7) of Table C.4.1, and columns (10)-(12) use an index constructed from column (6) of Table C.4.1. Individuals are matched to values of predicted harvest index using the process described in C.5.2. The 6-monthly specifications control for (District × month of birth fixed effects), to account for variation due to the timing of birth relative to the 6-month "bucket" individuals are assigned to (see Appendix Section C.5.2 for details.) All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

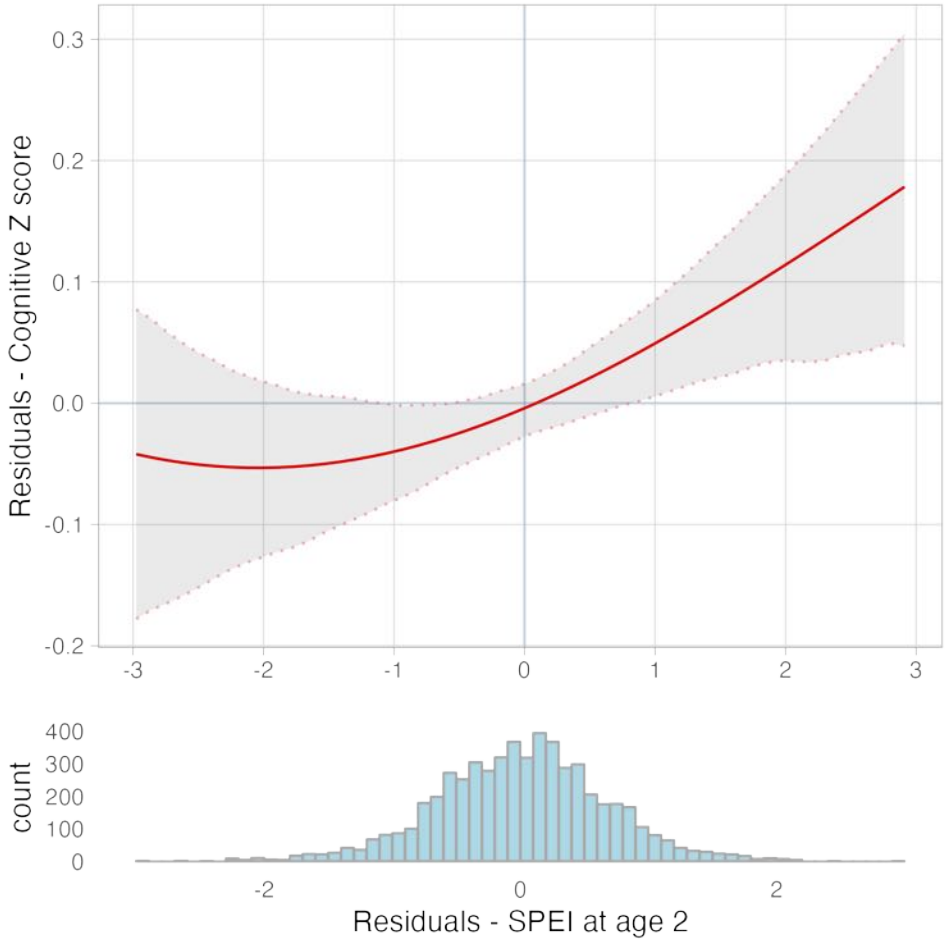
**Table C.12.13:** *Cognitive results are robust to using adjusted shock measure (weighted to account for differential exposure to shocks)*

| | Dep Var: Cognitive Z Score | | | |
|---|---|---|---|---|
| | SPEI growing season (1) | q value (1) | Predicted Harvest Index (2) | q value (2) |
| Age = -1 | 0.028 (0.031) | [1] | −0.023 (0.033) | [1] |
| Age = -0.5 | 0.029 (0.034) | [1] | 0.030 (0.033) | [1] |
| Age = 0 | 0.007 (0.037) | [1] | 0.033 (0.037) | [1] |
| Age = 0.5 | −0.024 (0.035) | [1] | −0.006 (0.035) | [1] |
| Age = 1 | 0.024 (0.035) | [1] | −0.014 (0.034) | [1] |
| Age = 1.5 | 0.013 (0.033) | [1] | 0.003 (0.033) | [1] |
| Age = 2 | 0.074 (0.037)** | [0.411] | 0.016 (0.033) | [1] |
| Age = 2.5 | 0.108 (0.034)*** | [0.021] | 0.110 (0.036)*** | [0.03] |
| Age = 3 | 0.027 (0.036) | [1] | 0.024 (0.031) | [1] |
| Age = 3.5 | 0.039 (0.032) | [1] | 0.002 (0.034) | [1] |
| Age = 4 | −0.017 (0.037) | [1] | 0.042 (0.032) | [1] |
| Age = 4.5 | −0.012 (0.029) | [1] | −0.010 (0.032) | [1] |
| Age = 5 | 0.014 (0.034) | [1] | −0.020 (0.029) | [1] |
| Birth year FEs | Yes | | Yes | |
| District x month-of-birth FEs | Yes | | Yes | |
| F statistic | 1.578 | | 1.605 | |
| p-value for F-test | 0.085 | | 0.077 | |
| $R^2$ | 0.421 | | 0.422 | |
| N | 5572 | | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is the adult cognitive factor score (measured after age 15). The cognitive factor score is standardised to have 0 mean and standard deviation of 1. Column (1) includes both genders, column (2) only includes females, and column (3) only includes males. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. Individuals are matched to values of predicted harvest index using the process described in C.5.2. All models in this table use an adjusted shock measure that accounts for differential exposure to shocks in each 6-month interval. For example, person *a* is born at the start of the shock-exposure period (3 months after harvest), so their first 6 months are exposed fully to the shock and the shock variable is not adjusted. However, person *b* is born in the middle of the shock-exposure period (say 5 months after harvest), and so only 4 out of the 6 first months of their life are exposed to the shock. I therefore adjust person *b*'s shock measure to be 2/3 as large as person *a*'s. The yearly specifications control for birth district fixed effects, while the 6-monthly specifications control for (District × month of birth fixed effects), to account for variation due to the timing of birth relative to the 6-month "bucket" individuals are assigned to (see Appendix Section C.5.2 for details.) All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

## C.13 Additional tables: alternative non-linear models

**Figure C.13.1:** *Non-parametric kernel regression of $SPEI_{i(t+2)}$ on the residuals of cognitive factor score*




*Notes*: The red line denotes the estimates from a non-parametric kernel regression of $SPEI_{i(t+2)}$ on the residuals of cognitive factor score after conditioning for birth-year fixed effects, birth-district fixed effects, SPEI growing season from ages -2, -1, 0, 1, and 3-15, and individual controls (parental education, gender, and religion). The specification is a local-linear regression that uses a Gaussian kernel and chooses a bandwidth of 0.95 using a process of least-squares cross-validation. The dashed lines show the bootstrapped 95% confidence intervals. The light-blue histogram at the bottom of the chart shows the distribution of residuals in this sample.

**Table C.13.1:** *Distinguishing between positive and negative shocks - linear model with different coefficients for positive and negative shocks*

| | Dep Var: Adult Cog (1) |
|---|---|
| Negative: | |
| SPEI growing season $< 0$, age = -2 | 0.003 (0.030) |
| SPEI growing season $< 0$, age = -1 | 0.049 (0.028)* |
| SPEI growing season $< 0$, age = 0 | 0.008 (0.029) |
| SPEI growing season $< 0$, age = 1 | 0.016 (0.033) |
| SPEI growing season $< 0$, age = 2 | 0.049 (0.033) |
| SPEI growing season $< 0$, age = 3 | 0.014 (0.031) |
| SPEI growing season $< 0$, age = 4 | 0.009 (0.033) |
| SPEI growing season $< 0$, age = 5 | 0.007 (0.037) |
| SPEI growing season $< 0$, age = 6 | 0.030 (0.036) |
| SPEI growing season $< 0$, age = 7 | −0.017 (0.040) |
| SPEI growing season $< 0$, age = 8 | 0.009 (0.039) |
| SPEI growing season $< 0$, age = 9 | −0.014 (0.039) |
| SPEI growing season $< 0$, age = 10 | 0.052 (0.049) |
| SPEI growing season $< 0$, age = 11 | −0.095 (0.053)* |
| SPEI growing season $< 0$, age = 12 | 0.064 (0.056) |
| SPEI growing season $< 0$, age = 13 | −0.108 (0.051)** |
| SPEI growing season $< 0$, age = 14 | −0.044 (0.049) |
| SPEI growing season $< 0$, age = 15 | −0.021 (0.038) |
| Positive: | |
| SPEI growing season $\geq 0$, age = -2 | −0.001 (0.033) |
| SPEI growing season $\geq 0$, age = -1 | 0.001 (0.031) |
| SPEI growing season $\geq 0$, age = 0 | 0.001 (0.033) |
| SPEI growing season $\geq 0$, age = 1 | 0.035 (0.038) |
| SPEI growing season $\geq 0$, age = 2 | 0.080 (0.031)** |
| SPEI growing season $\geq 0$, age = 3 | 0.052 (0.028)* |
| SPEI growing season $\geq 0$, age = 4 | 0.000 (0.029) |
| SPEI growing season $\geq 0$, age = 5 | 0.025 (0.027) |
| SPEI growing season $\geq 0$, age = 6 | −0.004 (0.029) |
| SPEI growing season $\geq 0$, age = 7 | −0.036 (0.033) |
| SPEI growing season $\geq 0$, age = 8 | 0.039 (0.033) |
| SPEI growing season $\geq 0$, age = 9 | 0.015 (0.037) |
| SPEI growing season $\geq 0$, age = 10 | −0.032 (0.032) |
| SPEI growing season $\geq 0$, age = 11 | 0.005 (0.035) |
| SPEI growing season $\geq 0$, age = 12 | 0.006 (0.037) |
| SPEI growing season $\geq 0$, age = 13 | 0.042 (0.030) |
| SPEI growing season $\geq 0$, age = 14 | 0.049 (0.034) |
| SPEI growing season $\geq 0$, age = 15 | 0.001 (0.029) |
| Birth year fixed effects | Yes |
| Birth district fixed effects | Yes |
| F statistic | 2.486 |
| p-value for F-test | 0.000 |
| $R^2$ | 0.214 |
| N | 5572 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. The first set of coefficients (rows 1 to 18) describe the effect of a variable that takes the value of $SPEI_{i(t+j)}$ when $SPEI_{i(t+j)} < 0$, and 0 otherwise, so it describes the (linear) effect of $SPEI_{i(t+j)}$ when it is negative. Similarly, the second set of coefficients (rows 19 to 36) describe the effect of a variable that takes the value of $SPEI_{i(t+j)}$ when $SPEI_{i(t+j)} \geq 0$, and 0 otherwise, i.e. the (linear) effect of $SPEI_{i(t+j)}$ when it is positive. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.13.2:** *Distinguishing between positive and negative shocks - binary shock variables*

| | Large positive shocks (>2 SD) (1) | q value (1) | Large negative shocks (<-2 SD) (2) | q value (2) | Binary shock (>0) (3) | q value (3) |
|---|---|---|---|---|---|---|
| SPEI growing season, age = -2 | −0.210 (0.098)** | [0.255] | 0.046 (0.100) | [1] | −0.004 (0.029) | [1] |
| SPEI growing season, age = -1 | −0.055 (0.105) | [1] | −0.147 (0.094) | [0.43] | 0.038 (0.027) | [1] |
| SPEI growing season, age = 0 | 0.115 (0.093) | [0.674] | −0.205 (0.091)** | [0.173] | −0.019 (0.029) | [1] |
| SPEI growing season, age = 1 | −0.000 (0.072) | [1] | −0.175 (0.095)* | [0.345] | 0.025 (0.034) | [1] |
| SPEI growing season, age = 2 | 0.088 (0.126) | [1] | −0.235 (0.090)** | [0.173] | 0.103 (0.031)*** | [0.018] |
| SPEI growing season, age = 3 | 0.155 (0.068)** | [0.255] | −0.207 (0.086)** | [0.173] | 0.013 (0.031) | [1] |
| SPEI growing season, age = 4 | −0.166 (0.090)* | [0.339] | −0.107 (0.151) | [1] | −0.021 (0.028) | [1] |
| SPEI growing season, age = 5 | 0.040 (0.092) | [1] | −0.092 (0.097) | [1] | −0.016 (0.030) | [1] |
| SPEI growing season, age = 6 | −0.077 (0.116) | [1] | −0.097 (0.114) | [1] | −0.008 (0.031) | [1] |
| SPEI growing season, age = 7 | −0.028 (0.086) | [1] | −0.016 (0.132) | [1] | −0.036 (0.036) | [1] |
| SPEI growing season, age = 8 | −0.144 (0.114) | [0.674] | −0.032 (0.148) | [1] | −0.003 (0.030) | [1] |
| SPEI growing season, age = 9 | 0.053 (0.091) | [1] | −0.212 (0.123)* | [0.354] | −0.016 (0.034) | [1] |
| SPEI growing season, age = 10 | −0.282 (0.113)** | [0.255] | −0.187 (0.201) | [1] | −0.018 (0.033) | [1] |
| SPEI growing season, age = 11 | 0.061 (0.086) | [1] | −0.059 (0.176) | [1] | −0.046 (0.028)* | [1] |
| SPEI growing season, age = 12 | −0.017 (0.092) | [1] | −0.154 (0.231) | [1] | 0.044 (0.036) | [1] |
| SPEI growing season, age = 13 | 0.140 (0.084)* | [0.411] | 0.021 (0.175) | [1] | −0.023 (0.033) | [1] |
| SPEI growing season, age = 14 | 0.019 (0.087) | [1] | −0.010 (0.195) | [1] | −0.021 (0.030) | [1] |
| SPEI growing season, age = 15 | −0.006 (0.102) | [1] | 0.001 (0.117) | [1] | 0.008 (0.026) | [1] |
| Birth year fixed effects | Yes | | Yes | | Yes | |
| Birth district fixed effects | Yes | | Yes | | Yes | |
| F statistic | 4.540 | | 2.451 | | 1.644 | |
| p-value for F-test | 0.000 | | 0.001 | | 0.054 | |
| $R^2$ | 0.212 | | 0.209 | | 0.212 | |
| N | 5572 | | 5572 | | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Column (1) shows the effect of large positive shocks (i.e. $SPEI_{i(t+j)} > 2$) , Column (2) shows the effect of large negative shocks (i.e. $SPEI_{i(t+j)} < -2$), and Column (3) shows the effect of a binary variable with the threshold at 0, (i.e. $SPEI_{i(t+j)} > 0$). Around 2% of the sample experience large positive shocks in a given year, and around 2% of the sample experience large negative shocks in a given year. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.13.3:** *Different thresholds for extreme negative shocks*

| | <-2 SD (1) | q value (1) | <-1.5 SD (2) | q value (2) | <-1 SD (3) | q value (3) | <-0.5 SD (4) | q value (4) |
|---|---|---|---|---|---|---|---|---|
| SPEI growing season, age = -2 | 0.046 (0.100) | [1] | −0.046 (0.063) | [0.834] | 0.051 (0.042) | [0.874] | −0.016 (0.036) | [1] |
| SPEI growing season, age = -1 | −0.147 (0.094) | [0.43] | −0.083 (0.073) | [0.801] | −0.081 (0.041)** | [0.682] | −0.048 (0.029) | [0.684] |
| SPEI growing season, age = 0 | −0.205 (0.091)** | [0.173] | −0.092 (0.074) | [0.801] | −0.060 (0.042) | [0.874] | 0.014 (0.031) | [1] |
| SPEI growing season, age = 1 | −0.175 (0.095)* | [0.345] | −0.130 (0.070)* | [0.801] | 0.006 (0.043) | [1] | −0.004 (0.037) | [1] |
| SPEI growing season, age = 2 | −0.235 (0.090)** | [0.173] | −0.131 (0.073)* | [0.801] | −0.062 (0.047) | [0.874] | −0.105 (0.033)*** | [0.036] |
| SPEI growing season, age = 3 | −0.207 (0.086)** | [0.173] | −0.100 (0.065) | [0.801] | −0.087 (0.049)* | [0.761] | −0.024 (0.034) | [1] |
| SPEI growing season, age = 4 | −0.107 (0.151) | [1] | −0.035 (0.078) | [0.867] | −0.022 (0.047) | [1] | −0.002 (0.033) | [1] |
| SPEI growing season, age = 5 | −0.092 (0.097) | [1] | −0.113 (0.078) | [0.801] | −0.052 (0.044) | [0.874] | −0.019 (0.039) | [1] |
| SPEI growing season, age = 6 | −0.097 (0.114) | [1] | −0.156 (0.069)** | [0.801] | −0.004 (0.044) | [1] | 0.006 (0.033) | [1] |
| SPEI growing season, age = 7 | −0.016 (0.132) | [1] | 0.011 (0.087) | [0.867] | −0.026 (0.063) | [1] | 0.066 (0.037)* | [0.684] |
| SPEI growing season, age = 8 | −0.032 (0.148) | [1] | −0.042 (0.076) | [0.867] | −0.038 (0.052) | [1] | −0.014 (0.034) | [1] |
| SPEI growing season, age = 9 | −0.212 (0.123)* | [0.354] | −0.058 (0.069) | [0.801] | 0.012 (0.051) | [1] | 0.041 (0.039) | [1] |
| SPEI growing season, age = 10 | −0.187 (0.201) | [1] | −0.037 (0.121) | [0.867] | −0.147 (0.062)** | [0.49] | 0.007 (0.046) | [1] |
| SPEI growing season, age = 11 | −0.059 (0.176) | [1] | 0.032 (0.108) | [0.867] | 0.014 (0.072) | [1] | 0.054 (0.048) | [1] |
| SPEI growing season, age = 12 | −0.154 (0.231) | [1] | −0.097 (0.184) | [0.867] | −0.145 (0.094) | [0.874] | −0.071 (0.045) | [0.684] |
| SPEI growing season, age = 13 | 0.021 (0.175) | [1] | 0.168 (0.143) | [0.801] | 0.046 (0.062) | [1] | 0.045 (0.047) | [1] |
| SPEI growing season, age = 14 | −0.010 (0.195) | [1] | −0.149 (0.091) | [0.801] | −0.010 (0.061) | [1] | 0.056 (0.036) | [0.684] |
| SPEI growing season, age = 15 | 0.001 (0.117) | [1] | −0.105 (0.077) | [0.801] | −0.003 (0.057) | [1] | 0.004 (0.040) | [1] |
| Birth year fixed effects | Yes | | Yes | | Yes | | Yes | |
| Birth district fixed effects | Yes | | Yes | | Yes | | Yes | |
| F statistic | 2.451 | | 1.205 | | 1.912 | | 2.347 | |
| p-value for F-test | 0.001 | | 0.261 | | 0.017 | | 0.002 | |
| $R^2$ | 0.209 | | 0.210 | | 0.211 | | 0.212 | |
| N | 5572 | | 5572 | | 5572 | | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. The column title indicates the threshold use to define the negative shock variable, e.g. column (1) shows the effect of negative shocks defined as $\mathbb{1}\left\{SPEI_{i(t+j)} < -2\right\}$ where $\mathbb{1}\{.\}$ is the indicator function. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.13.4:** *Different thresholds for extreme positive shocks*

| | >2 SD (1) | q value (1) | >1.5 SD (2) | q value (2) | >1 SD (3) | q value (3) | >0.5 SD (4) | q value (4) |
|---|---|---|---|---|---|---|---|---|
| SPEI growing season, age = -2 | −0.210 (0.098)** | [0.255] | −0.077 (0.056) | [1] | −0.021 (0.043) | [0.564] | 0.021 (0.030) | [1] |
| SPEI growing season, age = -1 | −0.055 (0.105) | [1] | 0.026 (0.075) | [1] | 0.047 (0.037) | [0.39] | −0.011 (0.029) | [1] |
| SPEI growing season, age = 0 | 0.115 (0.093) | [0.674] | −0.027 (0.056) | [1] | −0.053 (0.046) | [0.426] | 0.001 (0.029) | [1] |
| SPEI growing season, age = 1 | −0.000 (0.072) | [1] | 0.134 (0.079)* | [1] | 0.063 (0.045) | [0.39] | 0.032 (0.034) | [1] |
| SPEI growing season, age = 2 | 0.088 (0.126) | [1] | 0.090 (0.069) | [1] | 0.089 (0.038)** | [0.176] | 0.047 (0.028)* | [1] |
| SPEI growing season, age = 3 | 0.155 (0.068)** | [0.255] | 0.136 (0.048)*** | [0.103] | 0.088 (0.040)** | [0.176] | 0.036 (0.028) | [1] |
| SPEI growing season, age = 4 | −0.166 (0.090)* | [0.339] | 0.001 (0.075) | [1] | 0.012 (0.043) | [0.615] | −0.003 (0.030) | [1] |
| SPEI growing season, age = 5 | 0.040 (0.092) | [1] | 0.074 (0.063) | [1] | 0.048 (0.037) | [0.39] | 0.000 (0.029) | [1] |
| SPEI growing season, age = 6 | −0.077 (0.116) | [1] | −0.036 (0.066) | [1] | −0.001 (0.038) | [0.705] | −0.002 (0.032) | [1] |
| SPEI growing season, age = 7 | −0.028 (0.086) | [1] | 0.010 (0.069) | [1] | −0.064 (0.044) | [0.39] | −0.045 (0.035) | [1] |
| SPEI growing season, age = 8 | −0.144 (0.114) | [0.674] | 0.027 (0.062) | [1] | 0.080 (0.034)** | [0.176] | 0.007 (0.032) | [1] |
| SPEI growing season, age = 9 | 0.053 (0.091) | [1] | 0.088 (0.067) | [1] | 0.030 (0.048) | [0.564] | −0.012 (0.036) | [1] |
| SPEI growing season, age = 10 | −0.282 (0.113)** | [0.255] | −0.028 (0.068) | [1] | −0.019 (0.032) | [0.564] | −0.035 (0.033) | [1] |
| SPEI growing season, age = 11 | 0.061 (0.086) | [1] | 0.054 (0.058) | [1] | 0.011 (0.045) | [0.615] | −0.060 (0.035)* | [1] |
| SPEI growing season, age = 12 | −0.017 (0.092) | [1] | −0.007 (0.060) | [1] | 0.068 (0.038)* | [0.282] | 0.022 (0.032) | [1] |
| SPEI growing season, age = 13 | 0.140 (0.084)* | [0.411] | 0.015 (0.058) | [1] | 0.027 (0.033) | [0.564] | 0.048 (0.031) | [1] |
| SPEI growing season, age = 14 | 0.019 (0.087) | [1] | 0.074 (0.057) | [1] | 0.063 (0.030)** | [0.176] | 0.003 (0.031) | [1] |
| SPEI growing season, age = 15 | −0.006 (0.102) | [1] | −0.022 (0.052) | [1] | −0.043 (0.033) | [0.39] | −0.025 (0.031) | [1] |
| Birth year fixed effects | Yes | | Yes | | Yes | | Yes | |
| Birth district fixed effects | Yes | | Yes | | Yes | | Yes | |
| F statistic | 4.540 | | 1.551 | | 2.313 | | 2.011 | |
| p-value for F-test | 0.000 | | 0.077 | | 0.003 | | 0.011 | |
| $R^2$ | 0.212 | | 0.210 | | 0.213 | | 0.211 | |
| N | 5572 | | 5572 | | 5572 | | 5572 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. The column title indicates the threshold use to define the negative shock variable, e.g. column (1) shows the effect of negative shocks defined as $\mathbb{1}\left\{SPEI_{i(t+j)} > 2\right\}$ where $\mathbb{1}\{.\}$ is the indicator function. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.13.5:** *Effect of SPEI growing season on adult cognitive, with squared term specification*

| | Model with squared terms (1) |
|---|---|
| SPEI growing season, age = -2 | 0.003 (0.018) |
| SPEI growing season, age = -1 | 0.026 (0.017) |
| SPEI growing season, age = 0 | 0.003 (0.016) |
| SPEI growing season, age = 1 | 0.018 (0.019) |
| SPEI growing season, age = 2 | 0.066 (0.017)*** |
| SPEI growing season, age = 3 | 0.029 (0.017)* |
| SPEI growing season, age = 4 | 0.001 (0.019) |
| SPEI growing season, age = 5 | 0.015 (0.016) |
| SPEI growing season, age = 6 | 0.012 (0.017) |
| SPEI growing season, age = 7 | −0.029 (0.020) |
| SPEI growing season, age = 8 | 0.025 (0.018) |
| SPEI growing season, age = 9 | −0.003 (0.020) |
| SPEI growing season, age = 10 | 0.009 (0.022) |
| SPEI growing season, age = 11 | −0.046 (0.022)** |
| SPEI growing season, age = 12 | 0.042 (0.026)* |
| SPEI growing season, age = 13 | −0.027 (0.024) |
| SPEI growing season, age = 14 | 0.006 (0.019) |
| SPEI growing season, age = 15 | −0.008 (0.019) |
| (SPEI growing season)$^2$, age = -2 | −0.000 (0.012) |
| (SPEI growing season)$^2$, age = -1 | −0.014 (0.011) |
| (SPEI growing season)$^2$, age = 0 | −0.009 (0.012) |
| (SPEI growing season)$^2$, age = 1 | −0.010 (0.013) |
| (SPEI growing season)$^2$, age = 2 | 0.011 (0.013) |
| (SPEI growing season)$^2$, age = 3 | −0.000 (0.010) |
| (SPEI growing season)$^2$, age = 4 | −0.003 (0.011) |
| (SPEI growing season)$^2$, age = 5 | 0.004 (0.012) |
| (SPEI growing season)$^2$, age = 6 | −0.014 (0.012) |
| (SPEI growing season)$^2$, age = 7 | −0.005 (0.014) |
| (SPEI growing season)$^2$, age = 8 | −0.002 (0.013) |
| (SPEI growing season)$^2$, age = 9 | 0.003 (0.014) |
| (SPEI growing season)$^2$, age = 10 | −0.025 (0.014)* |
| (SPEI growing season)$^2$, age = 11 | 0.022 (0.017) |
| (SPEI growing season)$^2$, age = 12 | −0.017 (0.018) |
| (SPEI growing season)$^2$, age = 13 | 0.024 (0.015) |
| (SPEI growing season)$^2$, age = 14 | 0.009 (0.017) |
| (SPEI growing season)$^2$, age = 15 | 0.001 (0.012) |
| Birth year fixed effects | Yes |
| Birth district fixed effects | Yes |
| F statistic | 2.520 |
| p-value for F-test | 0.000 |
| $R^2$ | 0.214 |
| N | 5572 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

## C.14 Additional tables: socioemotional outcomes

**Table C.14.1:** *Effect of SPEI growing season on socioemotional factor scores by gender*

| | Affect/Wellbeing | | | | Personality | | | | Depression | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Female (1) | q value (1) | Male (2) | q value (2) | Female (3) | q value (3) | Male (4) | q value (4) | Female (5) | q value (5) | Male (6) | q value (6) |
| SPEI growing season, age = -2 | 0.050* | [0.416] | −0.042 | [1] | −0.045 | [0.527] | −0.003 | [1] | 0.015 | [1] | 0.016 | [1] |
| | (0.029) | | (0.036) | | (0.030) | | (0.031) | | (0.023) | | (0.031) | |
| SPEI growing season, age = -1 | 0.065** | [0.416] | 0.017 | [1] | 0.013 | [1] | 0.002 | [1] | 0.044* | [0.664] | 0.041 | [1] |
| | (0.032) | | (0.031) | | (0.028) | | (0.031) | | (0.024) | | (0.025) | |
| SPEI growing season, age = 0 | 0.030 | [0.463] | −0.007 | [1] | −0.024 | [0.853] | −0.004 | [1] | −0.000 | [1] | −0.018 | [1] |
| | (0.030) | | (0.039) | | (0.027) | | (0.037) | | (0.023) | | (0.030) | |
| SPEI growing season, age = 1 | −0.050 | [0.427] | 0.028 | [1] | 0.011 | [1] | 0.015 | [1] | −0.005 | [1] | 0.001 | [1] |
| | (0.032) | | (0.037) | | (0.033) | | (0.035) | | (0.024) | | (0.028) | |
| SPEI growing season, age = 2 | 0.053* | [0.416] | −0.004 | [1] | −0.030 | [0.711] | −0.012 | [1] | −0.032 | [0.989] | −0.004 | [1] |
| | (0.029) | | (0.039) | | (0.024) | | (0.040) | | (0.023) | | (0.029) | |
| SPEI growing season, age = 3 | 0.010 | [0.68] | −0.036 | [1] | −0.010 | [1] | −0.027 | [1] | −0.008 | [1] | 0.026 | [1] |
| | (0.025) | | (0.035) | | (0.027) | | (0.034) | | (0.025) | | (0.028) | |
| SPEI growing season, age = 4 | 0.049 | [0.427] | 0.006 | [1] | 0.028 | [0.864] | 0.028 | [1] | 0.006 | [1] | 0.018 | [1] |
| | (0.033) | | (0.034) | | (0.038) | | (0.033) | | (0.031) | | (0.031) | |
| SPEI growing season, age = 5 | −0.049 | [0.427] | 0.087** | [0.256] | −0.074** | [0.4] | −0.045 | [1] | −0.056** | [0.664] | 0.021 | [1] |
| | (0.032) | | (0.034) | | (0.031) | | (0.030) | | (0.027) | | (0.028) | |
| SPEI growing season, age = 6 | −0.027 | [0.463] | 0.025 | [1] | −0.060* | [0.4] | 0.013 | [1] | −0.013 | [1] | 0.001 | [1] |
| | (0.028) | | (0.030) | | (0.031) | | (0.037) | | (0.025) | | (0.029) | |
| SPEI growing season, age = 7 | 0.011 | [0.68] | 0.048 | [1] | 0.006 | [1] | −0.074** | [0.445] | 0.034 | [0.989] | 0.023 | [1] |
| | (0.031) | | (0.038) | | (0.029) | | (0.035) | | (0.026) | | (0.031) | |
| SPEI growing season, age = 8 | −0.017 | [0.654] | 0.032 | [1] | 0.001 | [1] | −0.020 | [1] | −0.025 | [1] | 0.004 | [1] |
| | (0.033) | | (0.034) | | (0.025) | | (0.035) | | (0.026) | | (0.028) | |
| SPEI growing season, age = 9 | −0.019 | [0.654] | −0.007 | [1] | −0.030 | [0.853] | −0.005 | [1] | 0.001 | [1] | −0.001 | [1] |
| | (0.031) | | (0.045) | | (0.035) | | (0.039) | | (0.029) | | (0.035) | |
| SPEI growing season, age = 10 | 0.038 | [0.463] | 0.053* | [1] | −0.027 | [0.853] | 0.004 | [1] | 0.016 | [1] | −0.018 | [1] |
| | (0.036) | | (0.032) | | (0.029) | | (0.042) | | (0.026) | | (0.026) | |
| SPEI growing season, age = 11 | −0.055* | [0.416] | −0.002 | [1] | −0.044 | [0.527] | −0.019 | [1] | −0.052* | [0.664] | −0.046 | [1] |
| | (0.029) | | (0.030) | | (0.030) | | (0.042) | | (0.028) | | (0.029) | |
| SPEI growing season, age = 12 | 0.018 | [0.654] | 0.044 | [1] | −0.066* | [0.4] | 0.064* | [0.562] | 0.001 | [1] | 0.021 | [1] |
| | (0.033) | | (0.044) | | (0.035) | | (0.035) | | (0.030) | | (0.035) | |
| SPEI growing season, age = 13 | −0.017 | [0.654] | 0.054 | [1] | −0.012 | [1] | −0.028 | [1] | 0.008 | [1] | −0.010 | [1] |
| | (0.032) | | (0.036) | | (0.036) | | (0.038) | | (0.029) | | (0.031) | |
| SPEI growing season, age = 14 | −0.102*** | [0.199] | −0.032 | [1] | −0.057* | [0.4] | 0.022 | [1] | 0.016 | [1] | 0.019 | [1] |
| | (0.039) | | (0.040) | | (0.030) | | (0.035) | | (0.032) | | (0.033) | |
| SPEI growing season, age = 15 | 0.015 | [0.654] | 0.029 | [1] | 0.026 | [0.854] | −0.098*** | [0.144] | 0.034 | [0.989] | 0.034 | [1] |
| | (0.032) | | (0.040) | | (0.033) | | (0.036) | | (0.026) | | (0.026) | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Birth district FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| F statistic | 2.165 | | 1.801 | | 1.427 | | 1.764 | | 1.639 | | 1.050 | |
| p-value for F-test | 0.006 | | 0.029 | | 0.124 | | 0.034 | | 0.056 | | 0.408 | |
| $R^2$ | 0.107 | | 0.106 | | 0.099 | | 0.116 | | 0.129 | | 0.111 | |
| N | 2717 | | 2239 | | 2717 | | 2239 | | 2967 | | 2598 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running Equation 3.1 with the three socioemotional factors measured in adulthood (above aged 15) as the dependent variable for one gender at a time. Columns (1), (3), and (5) only include females in the sample, and columns (2), (4) and (6) only include males in the sample. All socioemotional scores are internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.14.2:** *Effect of SPEI growing season on socioemotional factor scores with siblings fixed effects*

| | Affect/Wellbeing (1) | q value (1) | Personality (2) | q value (2) | Depression (3) | q value (3) |
|---|---|---|---|---|---|---|
| SPEI growing season, age = -2 | 0.021 (0.044) | [1] | 0.059 (0.043) | [0.658] | 0.042 (0.035) | [0.672] |
| SPEI growing season, age = -1 | 0.029 (0.046) | [1] | −0.006 (0.045) | [1] | −0.009 (0.034) | [1] |
| SPEI growing season, age = 0 | 0.012 (0.046) | [1] | 0.009 (0.041) | [1] | −0.065 (0.036)* | [0.384] |
| SPEI growing season, age = 1 | 0.017 (0.049) | [1] | −0.034 (0.046) | [0.846] | −0.010 (0.037) | [1] |
| SPEI growing season, age = 2 | −0.012 (0.047) | [1] | −0.059 (0.044) | [0.658] | −0.100 (0.040)** | [0.205] |
| SPEI growing season, age = 3 | −0.049 (0.045) | [1] | −0.052 (0.043) | [0.807] | −0.039 (0.035) | [0.672] |
| SPEI growing season, age = 4 | 0.002 (0.047) | [1] | 0.009 (0.045) | [1] | −0.030 (0.042) | [1] |
| SPEI growing season, age = 5 | −0.014 (0.049) | [1] | −0.107 (0.047)** | [0.265] | −0.070 (0.037)* | [0.384] |
| SPEI growing season, age = 6 | −0.093 (0.047)* | [0.377] | −0.042 (0.048) | [0.807] | −0.045 (0.042) | [0.672] |
| SPEI growing season, age = 7 | −0.059 (0.053) | [1] | 0.004 (0.047) | [1] | 0.001 (0.041) | [1] |
| SPEI growing season, age = 8 | −0.030 (0.051) | [1] | −0.105 (0.043)** | [0.265] | −0.095 (0.040)** | [0.205] |
| SPEI growing season, age = 9 | −0.019 (0.047) | [1] | 0.039 (0.045) | [0.807] | −0.042 (0.040) | [0.672] |
| SPEI growing season, age = 10 | 0.057 (0.049) | [1] | −0.064 (0.046) | [0.658] | −0.011 (0.041) | [1] |
| SPEI growing season, age = 11 | −0.064 (0.045) | [1] | −0.033 (0.049) | [0.846] | −0.019 (0.040) | [1] |
| SPEI growing season, age = 12 | −0.125 (0.052)** | [0.306] | −0.072 (0.045) | [0.658] | −0.055 (0.041) | [0.672] |
| SPEI growing season, age = 13 | −0.019 (0.047) | [1] | −0.023 (0.041) | [0.975] | −0.007 (0.040) | [1] |
| SPEI growing season, age = 14 | −0.098 (0.044)** | [0.306] | −0.010 (0.046) | [1] | 0.017 (0.041) | [1] |
| SPEI growing season, age = 15 | −0.040 (0.044) | [1] | −0.065 (0.044) | [0.658] | 0.059 (0.038) | [0.517] |
| Birth year FEs | Yes | | Yes | | Yes | |
| Birth district FEs | No | | No | | No | |
| Sibling FEs | Yes | | Yes | | Yes | |
| Birth Order FEs | Yes | | Yes | | Yes | |
| Sample mean | 0.003 | | −0.006 | | 0.009 | |
| F statistic | 1.240 | | 1.702 | | 1.400 | |
| p-value for F-test | 0.221 | | 0.034 | | 0.122 | |
| $R^2$ | 0.835 | | 0.843 | | 0.820 | |
| N | 4956 | | 4956 | | 5565 | |

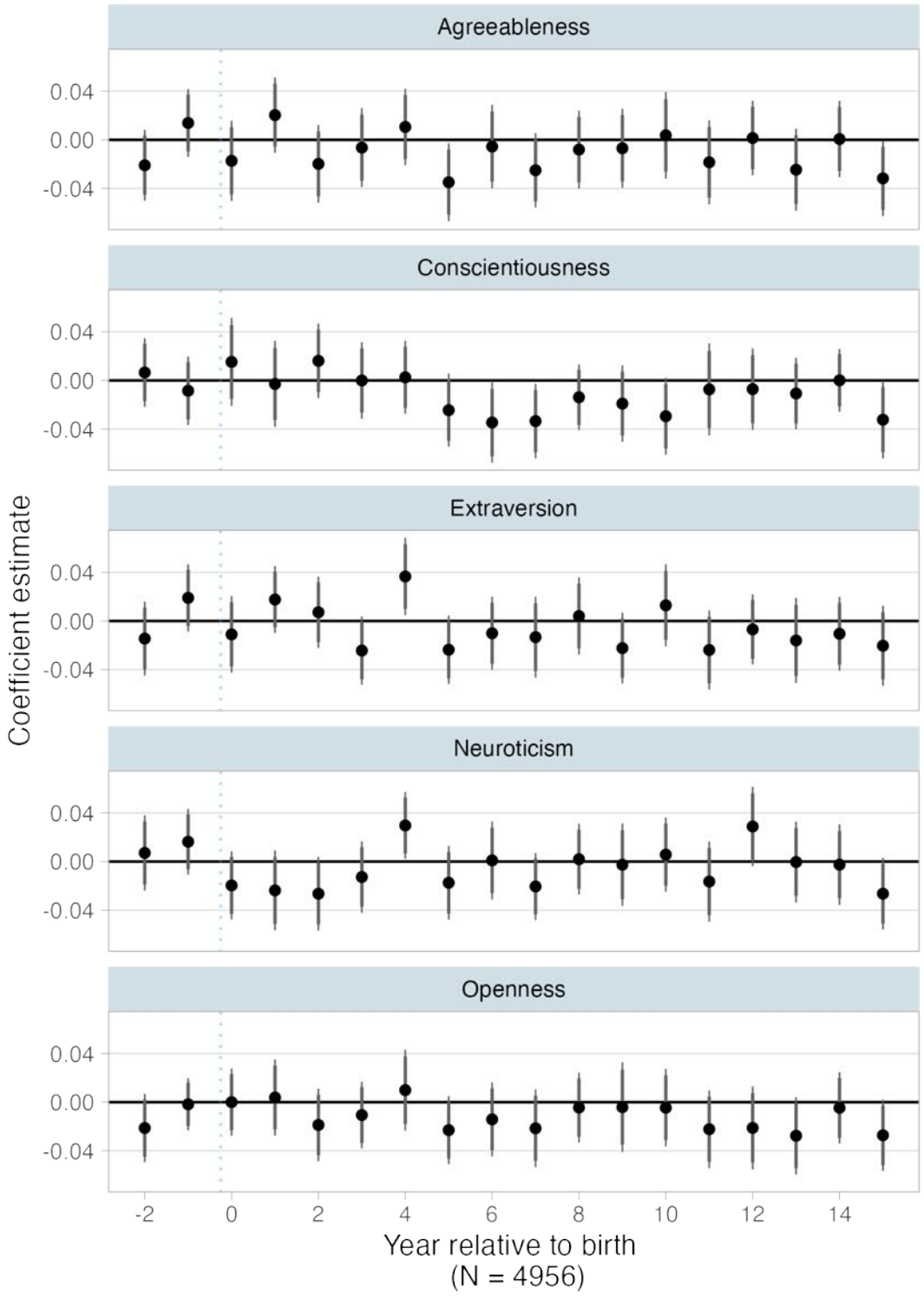
*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running Equation 3.1 with the three socioemotional factors measured in adulthood (above aged 15). All regressions include both sibling fixed effects and birth order fixed effects, as defined in Appendix Section C.1.1. All socioemotional scores are internally standardised, so the sample mean is 0 by construction. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.14.3:** *Effect of SPEI growing season on female's fertility*

| | Dep Var: Individual Has Given Birth (=1) | |
|---|---|---|
| | (1) | q value (1) |
| SPEI growing season, age = 10 | 0.009 | [0.72] |
| | (0.012) | |
| SPEI growing season, age = 11 | 0.016 | [0.72] |
| | (0.012) | |
| SPEI growing season, age = 12 | 0.020* | [0.72] |
| | (0.012) | |
| SPEI growing season, age = 13 | 0.015 | [0.72] |
| | (0.012) | |
| SPEI growing season, age = 14 | 0.008 | [0.72] |
| | (0.012) | |
| SPEI growing season, age = 15 | 0.009 | [0.72] |
| | (0.011) | |
| Birth year FEs | Yes | |
| Birth district FEs | Yes | |
| Sample mean | 0.415 | |
| F statistic | 0.717 | |
| p-value for F-test | 0.637 | |
| $R^2$ | 0.440 | |
| N | 2754 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The sample used is all females in the "main sample" for whom socioemotional variables are measured (as defined in Table 3.2). The dependent variable is an indicator for whether the individual has given birth by the survey in IFLS5. Since affect/wellbeing is only measured at one period in time (the 5th wave of IFLS in 2015), and I am controlling for birth-year fixed effects, I am effectively controlling for age fixed effects in this regression. All socioemotional scores are internally standardised, so the sample mean is 0 by construction. Additional controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Figure C.14.1:** *Effect of SPEI growing season on each Big 5 personality trait separately*




*Notes*: Each point represents the effect of an SPEI shock in the specific year relative to birth on the respective socioemotional Z-score in adulthood. Socioemotional measures are separated according to the "Big 5" personality traits, and combined using inverse-covariance-weighted indices (Anderson, 2008). All coefficients control for birth year fixed effects, birth district fixed effects, and individual-level controls (dummies for highest level of parental education, sex, and religion). Error bars are 90% and 95% confidence intervals based on standard errors clustered at the district level (not adjusted for multiple testing). Points are highlighted in red when the q-value is less than 0.05. . No points are highlighted because no coefficient has a q-value below 0.05.

## C.15 Additional tables: household expenditure

**Monthly expenditure empirical strategy**: Consider a household $h$ in district $r$ assigned to the harvest in year $y$ (as in Equation 3.2). Let $Month_k$ be an indicator that takes the value 1 when the household was interviewed $k$ months after the average harvest date for the district $r$, and 0 otherwise. Let $\mathcal{K} = \{-3, -2, ..., 8\}$ be the set of integers incorporating the values for months from -3 to 8 (that covers all the possible values for $k$). Then the monthly specification is:

$$\begin{aligned} Exp_{hry} = \alpha + \beta_0 Weather_{ry} + \beta_1 Weather_{r,y-1} + \sum_{k \in \mathcal{K}} \delta_k Month_k \\ + \sum_{k \in \mathcal{K}} \gamma_{0,k} Weather_{ry} + \sum_{k \in \mathcal{K}} \gamma_{1,k} Weather_{r,y-1} + \pi_r + \mu_y \end{aligned} \tag{C.7}$$

with all the other terms defined as in Equation 3.2. The coefficients seen in Figure C.15.1 are *sums* of coefficients from Equation C.7. For example, the reported coefficient for "Month 3" will be $\tau_3$ where $\tau_j$ is defined as below. There are two cases:

$$\tau_j = \begin{cases} \beta_0 + \gamma_{0,j} & \text{if } j \in \{-3, -2, ..., 8\} \\ \beta_1 + \gamma_{1,j-12} & \text{if } j \in \{9, 10, ..., 20\} \end{cases}$$

$\tau_j$ captures the overall effect of the the predicted harvest measure in the month $j$ relative to the harvest. For example, for $j = 3$ (3 months after harvest), I take the effect of the predicted harvest measure for the whole year around harvest ($\beta_0$), and add the coefficient specific to that month ($\gamma_{0,3}$). For $j = 15$, I take the coefficient for the effect of the predicted harvest measure from the previous year ($\beta_1$), and add it to the coefficient specific to that same month relative to harvest but from the previous year ($\gamma_{1,3}$). The standard errors of the $\tau_j$ coefficients are easily calculated using the variance covariance matrices of the $\beta$ and $\gamma$ coefficients, since:

$$Var(\tau_j) = \begin{cases} Var(\beta_0) + Var(\gamma_{0,j}) + 2Cov(\beta_0, \gamma_{0,j}) & \text{if } j \in \{-3, -2, ..., 8\} \\ Var(\beta_1) + Var(\gamma_{1,j-12}) + 2Cov(\beta_1, \gamma_{1,j-12}) & \text{if } j \in \{9, 10, ..., 20\} \end{cases}$$

The 6-month coefficients and standard errors in Table 3.8 are calculated in an analogous way.

**Table C.15.1:** *Predicted harvest index expenditure results are robust to different price adjustments and index constructions*

**Panel A:** Index Construction: Predicted Harvest Index

| | Nominal | | Infl [IFLS1-4] | | Real [IFLS2/3] | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Year 0 | 0.064 (0.022)*** | | 0.068 (0.028)** | | 0.116 (0.043)*** | |
| Year 1 | −0.000 (0.016) | | −0.001 (0.023) | | −0.014 (0.034) | |
| Months -3 to 2 | | 0.032 (0.028) | | −0.021 (0.039) | | 0.026 (0.066) |
| Months 3 to 8 | | 0.074 (0.026)*** | | 0.099 (0.031)*** | | 0.149 (0.048)*** |
| Months 9 to 14 | | 0.034 (0.020)* | | −0.009 (0.029) | | −0.064 (0.060) |
| Months 15 to 20 | | −0.028 (0.018) | | −0.002 (0.026) | | −0.027 (0.034) |
| $R^2$ | 0.802 | 0.802 | 0.118 | 0.120 | 0.123 | 0.126 |
| N | 15419 | 15415 | 11561 | 11557 | 5620 | 5619 |

**Panel B:** Index Construction: GDD + Wet Season Delay + SPEI + Dry Spell Index

| | Nominal | | Infl [IFLS1-4] | | Real [IFLS2/3] | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Year 0 | 0.062 (0.022)*** | | 0.071 (0.030)** | | 0.107 (0.040)*** | |
| Year 1 | 0.011 (0.014) | | 0.002 (0.022) | | −0.003 (0.029) | |
| Months -3 to 2 | | 0.024 (0.029) | | −0.009 (0.051) | | 0.001 (0.064) |
| Months 3 to 8 | | 0.075 (0.024)*** | | 0.095 (0.030)*** | | 0.140 (0.040)*** |
| Months 9 to 14 | | 0.039 (0.017)** | | −0.007 (0.028) | | −0.063 (0.056) |
| Months 15 to 20 | | −0.013 (0.016) | | 0.004 (0.026) | | −0.013 (0.029) |
| $R^2$ | 0.802 | 0.802 | 0.119 | 0.120 | 0.123 | 0.126 |
| N | 15419 | 15415 | 11561 | 11557 | 5620 | 5619 |

**Panel C:** Index Construction: GDD + Wet Season Delay Index

| | Nominal | | Infl [IFLS1-4] | | Real [IFLS2/3] | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Year 0 | 0.073 (0.019)*** | | 0.069 (0.027)** | | 0.209 (0.054)*** | |
| Year 1 | 0.009 (0.019) | | −0.005 (0.026) | | −0.024 (0.045) | |
| Months -3 to 2 | | 0.045 (0.028) | | 0.026 (0.055) | | 0.105 (0.093) |
| Months 3 to 8 | | 0.083 (0.024)*** | | 0.080 (0.026)*** | | 0.243 (0.053)*** |
| Months 9 to 14 | | 0.021 (0.019) | | −0.001 (0.026) | | −0.071 (0.087) |
| Months 15 to 20 | | −0.004 (0.028) | | −0.005 (0.035) | | −0.020 (0.041) |
| $R^2$ | 0.802 | 0.802 | 0.119 | 0.119 | 0.128 | 0.131 |
| N | 15419 | 15415 | 11561 | 11557 | 5620 | 5619 |

*Continued on next page...*

**Table C.15.1:** *Household expenditure results are robust to different price adjustments and weather indexes (**continued**)*

**Panel D:** Index Construction: GDD + SPEI Index

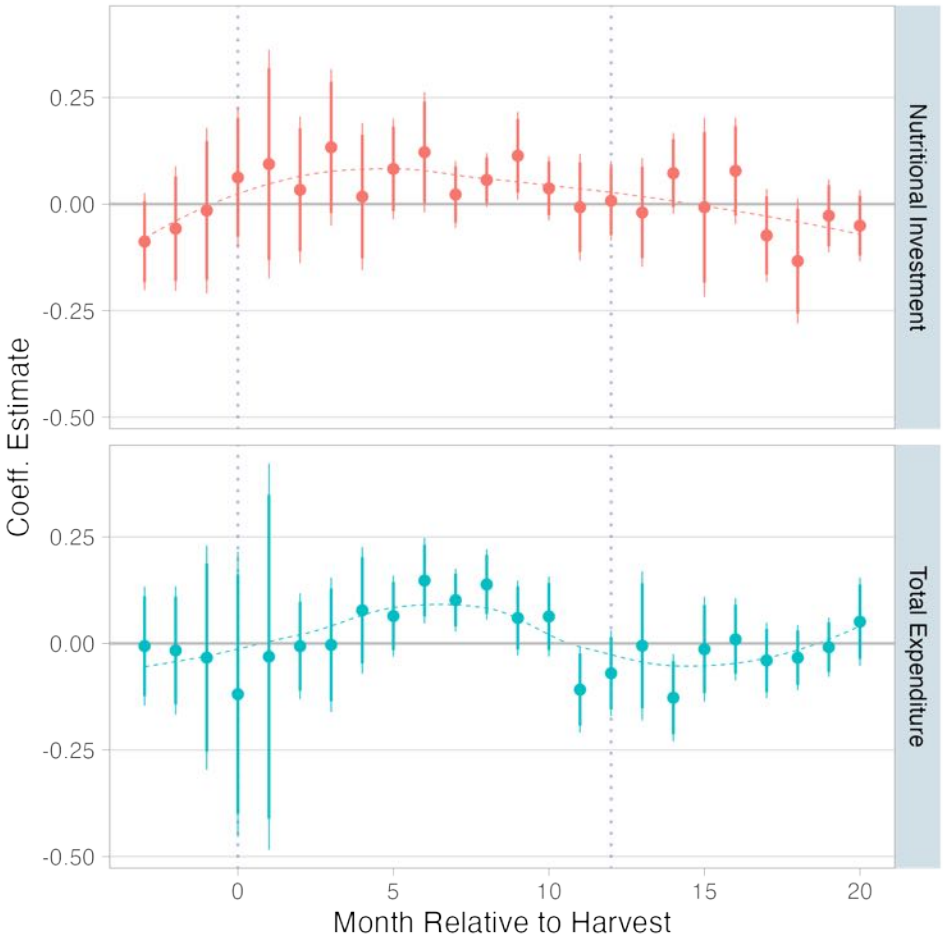
| | Nominal | | Infl [IFLS1-4] | | Real [IFLS2/3] | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Year 0 | 0.037 (0.018)** | | 0.047 (0.022)** | | 0.073 (0.033)** | |
| Year 1 | −0.003 (0.014) | | 0.001 (0.020) | | 0.003 (0.029) | |
| Months -3 to 2 | | 0.008 (0.025) | | −0.027 (0.028) | | 0.007 (0.055) |
| Months 3 to 8 | | 0.048 (0.023)** | | 0.076 (0.028)*** | | 0.097 (0.037)** |
| Months 9 to 14 | | 0.037 (0.021)* | | −0.010 (0.030) | | −0.032 (0.052) |
| Months 15 to 20 | | −0.030 (0.015)** | | −0.000 (0.023) | | −0.008 (0.032) |
| $R^2$ | 0.801 | 0.802 | 0.117 | 0.119 | 0.121 | 0.123 |
| N | 15419 | 15415 | 11561 | 11557 | 5620 | 5619 |

**Panel E:** Index Construction: Wet Season Delay + SPEI Index

| | Nominal | | Infl [IFLS1-4] | | Real [IFLS2/3] | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Year 0 | 0.050 (0.021)** | | 0.056 (0.025)** | | 0.091 (0.038)** | |
| Year 1 | −0.003 (0.015) | | 0.000 (0.022) | | −0.005 (0.031) | |
| Months -3 to 2 | | 0.021 (0.028) | | −0.029 (0.033) | | 0.015 (0.059) |
| Months 3 to 8 | | 0.061 (0.025)** | | 0.088 (0.030)*** | | 0.119 (0.042)*** |
| Months 9 to 14 | | 0.035 (0.020)* | | −0.011 (0.031) | | −0.060 (0.053) |
| Months 15 to 20 | | −0.030 (0.016)* | | −0.001 (0.025) | | −0.015 (0.033) |
| $R^2$ | 0.802 | 0.802 | 0.117 | 0.119 | 0.122 | 0.124 |
| N | 15419 | 15415 | 11561 | 11557 | 5620 | 5619 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by district. The unit of observation is (household × wave). The dependent variable is based on log total household expenditure. In columns (1) and (2) of all panels, the nominal expenditure is used (available for all IFLS waves 1-5). In columns (3) and (4), expenditure is adjusted for inflation at the district and year level (available for IFLS waves 1-4). In columns (5) and (6), expenditure is adjusted for inflation at the district and month-year level, but this is only available for IFLS2 and IFLS3 (see more on this in Appendix Section C.1.1). The explanatory variables used in Panel A are based on the predicted harvest index based on column (3) of Table C.4.1, as in the main specification. Panel B uses an index constructed in an analogous way but based on just using Wet Season Delay and the SPEI to construct the index (column (2) of Table C.4.1). Panel C uses an index constructed from column (5) of Table C.4.1, Panel D uses an index constructed from column (7) of Table C.4.1, and Panel E uses an index constructed from column (6) of Table C.4.1. The harvest date is defined as the average harvest date of rice in that district based on IFLS harvest data (see Section C.4.2). "Year 0" is the period between 3 months before and 8 months after harvest (inclusive). Analogously, "Year 1" is the period 9 months and 20 months after harvest (inclusive).

**Figure C.15.1:** *Effect of predicted harvest index on log household per capita expenditure and nutritional investment (monthly)*




*Notes*: Results of the estimation process from Equation C.7.

## C.16 Additional tables: alternative mechanisms

**Table C.16.1:** *SPEI at age 1 has strong effects on the small subsample who don't breastfeed*

| | Dep Var: Cognitive Z Score |
|---|---|
| | (1) |
| SPEI growing season, age = 0 | 0.011 |
| | (0.018) |
| SPEI growing season, age = 1 | 0.015 |
| | (0.020) |
| SPEI growing season, age = 2 | 0.056** |
| | (0.022) |
| SPEI growing season, age = 3 | 0.014 |
| | (0.017) |
| Never breastfed | −0.089 |
| | (0.094) |
| (SPEI growing season, age = 0) x (Never breastfed) | −0.141 |
| | (0.155) |
| (SPEI growing season, age = 1) x (Never breastfed) | 0.240** |
| | (0.114) |
| (SPEI growing season, age = 2) x (Never breastfed) | 0.038 |
| | (0.138) |
| (SPEI growing season, age = 3) x (Never breastfed) | −0.031 |
| | (0.142) |
| Birth year fixed effects | Yes |
| Birth district fixed effects | Yes |
| $R^2$ | 0.224 |
| N | 3690 |

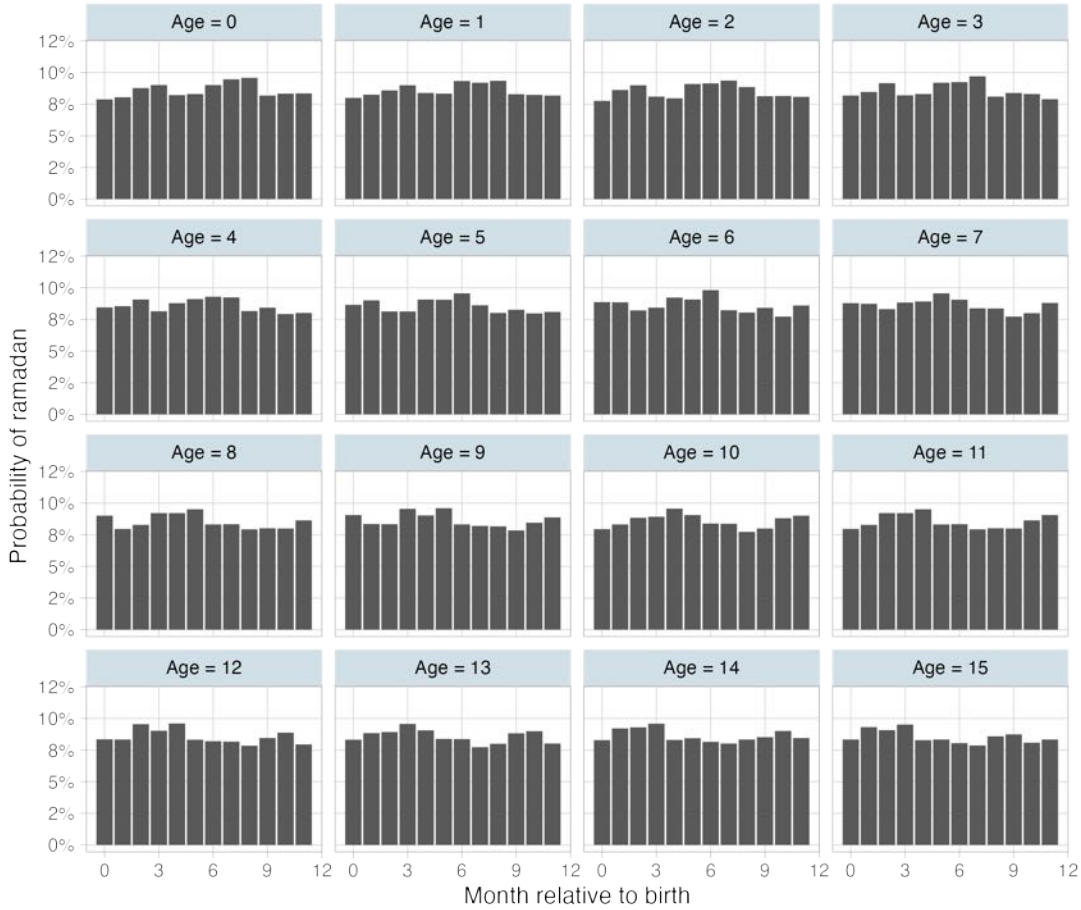
*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The table shows the result of running a version of Equation 3.1 with cognitive factor score measured in adulthood (above aged 15) as the dependent variable. Cognitive factor score is internally standardised, so the sample mean is 0 by construction. The sample used is all the main sample for whom the IFLS contains data on whether they were never breastfed. This sample includes all individuals whose mother was interviewed while the individual was a child. *Never breastfed* is an indicator variable that takes the value 1 if the caretaker never breastfed the child. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. Controls include dummies for the education level of the most-educated parent, dummies for religion, and a dummy for gender. All models use the individual-level attrition-corrected weights provided in the IFLS data.

**Table C.16.2:** *Weather may also have an effect on maternal nutrition*

| | | Dep Var: Mother Nutrition Index | | | |
|---|---|---|---|---|---|
| | Outcome measured in: | (1) | (2) | (3) | (4) |
| SPEI growing season | Year 0 | 0.019 | | | |
| | | (0.021) | | | |
| SPEI growing season | Year 1 | −0.020 | | | |
| | | (0.016) | | | |
| SPEI growing season | Months -3 to 2 after harvest | | 0.003 | | |
| | | | (0.025) | | |
| SPEI growing season | Months 3 to 8 after harvest | | 0.044 | | |
| | | | (0.034) | | |
| SPEI growing season | Months 9 to 14 after harvest | | −0.028 | | |
| | | | (0.022) | | |
| SPEI growing season | Months 15 to 20 after harvest | | −0.001 | | |
| | | | (0.022) | | |
| Predicted harvest index | Year 0 | | | 0.039 | |
| | | | | (0.027) | |
| Predicted harvest index | Year 1 | | | −0.012 | |
| | | | | (0.026) | |
| Predicted harvest index | Months -3 to 2 after harvest | | | | −0.007 |
| | | | | | (0.039) |
| Predicted harvest index | Months 3 to 8 after harvest | | | | 0.069* |
| | | | | | (0.036) |
| Predicted harvest index | Months 9 to 14 after harvest | | | | 0.004 |
| | | | | | (0.028) |
| Predicted harvest index | Months 15 to 20 after harvest | | | | −0.025 |
| | | | | | (0.033) |
| Year FEs | | Yes | Yes | Yes | Yes |
| District FEs | | Yes | Yes | Yes | Yes |
| Sample mean | | 0.065 | 0.065 | 0.065 | 0.065 |
| $R^2$ | | 0.167 | 0.168 | 0.172 | 0.173 |
| N | | 3479 | 3479 | 3479 | 3478 |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by district. The unit of observation is (individual × wave). The sample used is mothers of children under 5 in rural farming households. The dependent variable is a nutritional investment index for the mothers, composed of food frequency and dietary diversity measures in exactly the same way as the child nutrition index from the main results. The survey questions used are described in Appendix Section C.1.1. The construction of the indexes is described fully in Appendix Section C.2.7. *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. For the SPEI growing season, the harvest date is defined as the average harvest date of the main crop grown in that district (as according to aggregate data, see description in Section C.1.2). For the Predicted Harvest Index, the harvest date is defined as the average harvest date of rice in that district based on IFLS harvest data (see Section C.4.2). "Year 0" is the period between 3 months before and 8 months after harvest (inclusive). Analogously, "Year 1" is the period 9 months and 20 months after harvest (inclusive). For the 6-monthly effects, the coefficients reported come from a version of Equation C.4 (but with investment index as the dependent variable) and are $\beta_0$ for -3 to 2 months; $\beta_0 + \gamma_0$ for 3-8 months; $\beta_1$ for 9-14 months; $\beta_1 + \gamma_1$ for 15-20 months.

**Figure C.16.1:** *Timing of Ramadan is balanced across sample at different ages*




*Notes*: This graph shows the probability that Ramadan occurred at each month of age for the main sample. There are no notable differences between the timing of Ramadan at each age for the sample, suggesting that this cannot drive the impacts on cognitive development seen in the main results.

## C.17 Height-for-age

The evidence suggests that weather variation leads to increased nutritional investment. I here examine whether this change in nutritional investment translates into improved height-for-age in early life. Table C.17.1 displays the results of running a version of Equation 3.1 with child height-for-age Z-scores as the outcome variable. Here, I use data for all children measured in the IFLS between the ages of 2 and 6.[19] Panel A displays the results when using SPEI growing season as the main treatment variable. Although Sections 3.5.1 and 3.5.2 indicate that the SPEI growing season increases nutritional investment and household expenditure, there is no apparent effect on early life height-for-age Z-scores, with no statistically significant coefficients across any age range and gender at the 5% level even before adjusting for multiple hypothesis testing. Panel B shows some suggestive evidence that the predicted harvest index does have an effect on children's height-for-age. We see a strong positive coefficient of 0.088 for shocks at age 1 for females. The q-value for this coefficient is 0.072, indicating that it is weakly significant after adjusting for multiple hypothesis testing. No

[19] I divide this group into two overlapping age ranges (2-4 year olds and 4-6 year olds inclusive). This helps me to retain a large enough sample size to maintain statistical power, while also being able to examine the effect of later shocks (at age 2 and 3) with the older group.

other coefficients are significant at the 5% level.[20] Taken together, the results suggest that the nutritional investment effects of the predicted harvest index can translate into health benefits for young children that can be detected in changes in height-for-age, although this effect is not ubiquitous or consistent across age-groups and gender.

[20] In Appendix Table C.17.2, I use the methodology described in Appendix Section C.5.2 to analyse the effects of shocks at a more granular level, namely at 6-month age intervals. This analysis indicates that the yearly results may be disguising some sensitive periods, as it suggests that there may be positive effects of an increased predicted harvest index for girls across a wide range of infant ages (from 12 months to 36 months), and that there may be some positive effect of weather before birth (on both sexes).

## Table C.17.1: *Effect of early life weather on height-for-age Z scores*

**Panel A:** SPEI growing season

| | Both Sexes | | | | Male Only | | | | Female Only | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2-4 y.o. (1) | q value (1) | 4-6 y.o. (2) | q value (2) | 2-4 y.o. (3) | q value (3) | 4-6 y.o. (4) | q value (4) | 2-4 y.o. (5) | q value (5) | 4-6 y.o. (6) | q value (6) |
| SPEI Shock, Age = -1 | −0.007 | [1] | −0.019 | [1] | 0.014 | [0.879] | −0.049* | [0.444] | −0.028 | [1] | 0.001 | [1] |
| | (0.024) | | (0.018) | | (0.037) | | (0.026) | | (0.035) | | (0.030) | |
| SPEI Shock, Age = 0 | −0.003 | [1] | −0.001 | [1] | −0.017 | [0.879] | −0.041 | [0.444] | 0.017 | [1] | 0.028 | [1] |
| | (0.026) | | (0.020) | | (0.033) | | (0.028) | | (0.029) | | (0.029) | |
| SPEI Shock, Age = 1 | −0.035* | [0.406] | 0.001 | [1] | −0.054* | [0.21] | 0.017 | [0.604] | −0.014 | [1] | 0.003 | [1] |
| | (0.021) | | (0.018) | | (0.028) | | (0.024) | | (0.034) | | (0.027) | |
| SPEI Shock, Age = 2 | | | 0.015 | [1] | | | 0.040 | [0.449] | | | −0.011 | [1] |
| | | | (0.023) | | | | (0.033) | | | | (0.030) | |
| SPEI Shock, Age = 3 | | | 0.014 | [1] | | | 0.001 | [0.631] | | | 0.027 | [1] |
| | | | (0.017) | | | | (0.025) | | | | (0.025) | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| (Current) District FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Age (Months) FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Sample mean | −1.848 | | −1.716 | | −1.850 | | −1.770 | | −1.845 | | −1.661 | |
| F statistic | 0.956 | | 0.480 | | 1.358 | | 1.425 | | 0.328 | | 0.393 | |
| p-value for F-test | 0.415 | | 0.791 | | 0.258 | | 0.218 | | 0.805 | | 0.853 | |
| $R^2$ | 0.105 | | 0.134 | | 0.143 | | 0.154 | | 0.145 | | 0.182 | |
| N | 5608 | | 5781 | | 2896 | | 2889 | | 2712 | | 2892 | |

**Panel B:** Predicted Harvest Index

| | Both Sexes | | | | Male Only | | | | Female Only | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2-4 y.o. (1) | q value (1) | 4-6 y.o. (2) | q value (2) | 2-4 y.o. (3) | q value (3) | 4-6 y.o. (4) | q value (4) | 2-4 y.o. (5) | q value (5) | 4-6 y.o. (6) | q value (6) |
| Predicted Harvest Index, Age = -1 | 0.012 | [1] | 0.054 | [0.343] | −0.009 | [1] | 0.052 | [0.582] | 0.030 | [1] | 0.048 | [0.713] |
| | (0.033) | | (0.033) | | (0.038) | | (0.035) | | (0.049) | | (0.051) | |
| Predicted Harvest Index, Age = 0 | −0.004 | [1] | 0.039 | [0.343] | −0.041 | [1] | 0.005 | [1] | 0.046 | [1] | 0.060 | [0.499] |
| | (0.026) | | (0.028) | | (0.037) | | (0.036) | | (0.041) | | (0.043) | |
| Predicted Harvest Index, Age = 1 | −0.034 | [1] | 0.043* | [0.343] | −0.030 | [1] | 0.006 | [1] | −0.034 | [1] | 0.088** | [0.072] |
| | (0.031) | | (0.025) | | (0.035) | | (0.033) | | (0.041) | | (0.035) | |
| Predicted Harvest Index, Age = 2 | | | 0.022 | [0.372] | | | 0.022 | [1] | | | 0.030 | [0.713] |
| | | | (0.026) | | | | (0.032) | | | | (0.038) | |
| Predicted Harvest Index, Age = 3 | | | 0.014 | [0.372] | | | 0.055 | [0.582] | | | −0.016 | [0.713] |
| | | | (0.025) | | | | (0.038) | | | | (0.035) | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| (Current) District FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| District x month-of-birth FEs | No | | No | | No | | No | | No | | No | |
| Age (month) FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Sample mean | −1.848 | | −1.716 | | −1.850 | | −1.771 | | −1.845 | | −1.661 | |
| F statistic | 0.471 | | 1.870 | | 0.574 | | 1.146 | | 0.693 | | 1.597 | |
| p-value for F-test | 0.703 | | 0.102 | | 0.633 | | 0.339 | | 0.558 | | 0.164 | |
| $R^2$ | 0.104 | | 0.133 | | 0.136 | | 0.151 | | 0.146 | | 0.183 | |
| N | 5608 | | 5782 | | 2896 | | 2890 | | 2712 | | 2892 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is *height-for-age Z score*, calculated using the WHO's child growth standards data (WHO & UNICEF, 2009). *SPEI growing season* is constructed by taking the average value of the SPEI index over a district's growing season, standardised to have a standard deviation of 1 and a mean of 0 over the period 1961-2015. A full migration history for each individual is used to match individuals to the value of SPEI growing season that they experienced at every age. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

**Table C.17.2:** *Effect of predicted harvest index on HAZ in 6-monthly age intervals*

| | Both Sexes | | | | Male Only | | | | Female Only | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2-4 y.o. (1) | q value (1) | 4-6 y.o. (2) | q value (2) | 2-4 y.o. (3) | q value (3) | 4-6 y.o. (4) | q value (4) | 2-4 y.o. (5) | q value (5) | 4-6 y.o. (6) | q value (6) |
| Predicted Harvest Index, Age = -12 to -7 months | −0.051 | [1] | 0.008 | [1] | −0.099 | [1] | −0.008 | [1] | 0.011 | [1] | 0.019 | [1] |
| | (0.040) | | (0.034) | | (0.064) | | (0.054) | | (0.065) | | (0.051) | |
| Predicted Harvest Index, Age = -6 to -1 months | 0.025 | [1] | 0.071** | [0.372] | −0.022 | [1] | 0.093* | [0.681] | 0.048 | [1] | 0.009 | [1] |
| | (0.039) | | (0.036) | | (0.063) | | (0.054) | | (0.068) | | (0.059) | |
| Predicted Harvest Index, Age = 0 to 5 months | −0.015 | [1] | 0.007 | [1] | −0.042 | [1] | −0.033 | [0.947] | −0.020 | [1] | −0.010 | [1] |
| | (0.040) | | (0.037) | | (0.060) | | (0.058) | | (0.068) | | (0.059) | |
| Predicted Harvest Index, Age = 6 to 11 months | 0.005 | [1] | −0.012 | [1] | −0.023 | [1] | −0.064 | [0.681] | 0.051 | [1] | 0.006 | [1] |
| | (0.039) | | (0.032) | | (0.065) | | (0.050) | | (0.064) | | (0.052) | |
| Predicted Harvest Index, Age = 12 to 17 months | −0.012 | [1] | 0.055* | [0.372] | −0.020 | [1] | −0.005 | [1] | 0.016 | [1] | 0.114** | [0.12] |
| | (0.036) | | (0.032) | | (0.058) | | (0.049) | | (0.061) | | (0.054) | |
| Predicted Harvest Index, Age = 18 to 23 months | | | 0.038 | [0.513] | | | −0.040 | [0.947] | | | 0.112** | [0.12] |
| | | | (0.033) | | | | (0.052) | | | | (0.049) | |
| Predicted Harvest Index, Age = 24 to 29 months | | | 0.055* | [0.372] | | | 0.085* | [0.681] | | | 0.105** | [0.12] |
| | | | (0.031) | | | | (0.050) | | | | (0.047) | |
| Predicted Harvest Index, Age = 30 to 35 months | | | 0.015 | [1] | | | −0.015 | [1] | | | 0.094* | [0.124] |
| | | | (0.032) | | | | (0.051) | | | | (0.052) | |
| Predicted Harvest Index, Age = 36 to 41 months | | | 0.035 | [0.513] | | | 0.062 | [0.681] | | | 0.031 | [0.981] |
| | | | (0.032) | | | | (0.050) | | | | (0.045) | |
| Birth year FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| (Current) District FEs | No | | No | | No | | No | | No | | No | |
| District x month-of-birth FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Age (month) FEs | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | |
| Sample mean | −1.848 | | −1.716 | | −1.850 | | −1.771 | | −1.845 | | −1.661 | |
| F statistic | 0.543 | | 1.469 | | 0.517 | | 1.214 | | 0.260 | | 1.664 | |
| p-value for F-test | 0.744 | | 0.154 | | 0.763 | | 0.282 | | 0.935 | | 0.093 | |
| $R^2$ | 0.325 | | 0.340 | | 0.483 | | 0.473 | | 0.510 | | 0.519 | |
| N | 5608 | | 5782 | | 2896 | | 2890 | | 2712 | | 2892 | |

*Notes*: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Standard errors are reported in parentheses and are clustered by birth district. The dependent variable in all models is *height-for-age Z score*, calculated using the WHO's child growth standards data (WHO & UNICEF, 2009). The calculation of predicted harvest in 6-month age intervals is described in Appendix Section C.5.2. *Predicted harvest index* is a weather index constructed by taking the coefficients from model (2) in Table C.4.1 and calculating the predicted rice yields from this model for each year × district. It is standardised to so that within each district there is a mean of 0 and standard deviation of 1 over the period 1981-2015. All models use the individual-level attrition-corrected weights provided in the IFLS data. The q-value is the FDR-adjusted p-value, calculated according to the sharpened process seen in Anderson (2008) by pooling all coefficients in a given column. F-statistic and F-test are for the joint test of the null that all the shown coefficients in the model are equal to 0.

# Résumé de thèse

**Duncan Webb, Paris School of Economics**

Cette thèse explore deux thèmes majeurs en économie du développement : l'influence des interactions sociales sur les résultats économiques des groupes marginalisés, et la manière dont les inégalités dans le capital humain se propagent et peuvent être résolues dans les pays à faible et moyen revenu.

Le premier chapitre, intitulé "Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India", examine comment la communication entre les membres d'un groupe majoritaire peut réduire la discrimination envers une minorité marginalisée. À travers une expérience de terrain menée en Inde urbaine (N=3 397), l'étude mesure la discrimination à l'encontre des personnes transgenres et teste deux types de communication : la communication horizontale entre pairs et la communication verticale venant d'autorités. Les résultats montrent que les discussions de groupe entre voisins réduisent fortement la discrimination ultérieure, même lors de choix privés. Cet effet est 1,7 fois plus important que celui d'une communication top-down informant les participants sur les droits légaux des personnes transgenres. Les effets semblent provenir de la persuasion mutuelle entre participants, en particulier parce que les individus favorables aux personnes transgenres sont les plus vocaux dans les discussions.

Le deuxième chapitre, "Menstrual Stigma, Hygiene, and Human Capital: Experimental Evidence from Madagascar", co-écrit avec Karen Macours et Julieta Vera Rueda, étudie comment la stigmatisation menstruelle peut être combattue en s'appuyant sur des "déviants positifs" - des personnes prêtes à s'exprimer contre cette stigmatisation. À travers un essai contrôlé randomisé dans 140 écoles à Madagascar, l'étude évalue un programme ciblant à la fois l'hygiène et la stigmatisation menstruelle. Les résultats montrent qu'un ensemble d'interventions comprenant des infrastructures sanitaires, des produits menstruels et la sensibilisation des enseignants conduit à des améliorations substantielles (0,15 écart-type) des résultats d'apprentissage. L'étude évalue également l'effet additionnel de la nomination de "jeunes filles leaders" - des écolières prêtes à s'exprimer contre la stigmatisation menstruelle. Cette composante génère des améliorations significatives des connaissances et comportements en matière d'hygiène et de la stigmatisation menstruelle.

Le troisième chapitre, "Critical Periods in Cognitive and Socioemotional Development: Evidence from Weather Shocks in Indonesia", examine comment le moment des chocs pendant la petite enfance peut avoir des conséquences durables sur le développement du capital humain. En étudiant l'impact des chocs météorologiques sur les résultats cognitifs et socio-émotionnels à l'âge adulte d'individus nés en Indonésie rurale entre 1988 et 2000, l'étude met en évidence une période critique forte à l'âge de 2 ans pour le développement cognitif, mais ne trouve pas de période critique similaire pour le développement socio-émotionnel. Les effets semblent être dus à des changements dans le revenu agricole et l'investissement nutritionnel, et n'apparaissent qu'après l'âge de 15 ans.

Dans l'ensemble, cette thèse apporte des contributions importantes à notre compréhension des dynamiques sociales qui peuvent réduire la discrimination et la stigmatisation, ainsi que des périodes critiques du développement du capital humain. Elle fournit des implications concrètes pour la conception de politiques visant à améliorer les résultats des groupes marginalisés et le développement du capital humain dans les pays à faible et moyen revenu.

# Chapitre 1 : Communication entre pairs et réduction de la discrimination

Le premier chapitre examine comment la communication au sein d'un groupe majoritaire peut affecter la discrimination envers une minorité. Bien que la discrimination individuelle soit souvent attribuée à des préjugés profondément ancrés difficiles à changer, on observe parfois au niveau sociétal des réductions rapides des préférences discriminatoires. Ces observations suggèrent que les interactions sociales et la communication qu'elles impliquent pourraient être des moteurs de tels changements.

## Contexte et méthodologie

L'étude se concentre sur la communauté transgenre en Inde, appelée localement "thirunangai". Cette communauté, qui a un rôle culturel et religieux historique dans la société indienne, fait face à une discrimination économique importante. L'expérience de terrain a été menée à Chennai, en Inde urbaine, auprès de 3 397 participants.

La discrimination est mesurée à travers une série de choix de recrutement : les participants se voient offrir une livraison gratuite de courses et doivent choisir entre différents livreurs, certains étant transgenres. L'étude teste deux types de communication :

- La communication horizontale : des discussions de groupe entre trois voisins
- La communication verticale : des informations sur les droits légaux des personnes transgenres

Les participants sont randomisés entre plusieurs conditions :

1. Discussion à trois personnes : les participants discutent en groupe de leurs choix de recrutement
2. Discussion à deux personnes avec un auditeur : deux participants discutent pendant qu'un troisième écoute
3. Pas de discussion (public) : les choix individuels sont révélés au groupe sans discussion
4. Pas de discussion (privé) : condition de contrôle avec des choix individuels privés

## Résultats principaux

Les résultats montrent une discrimination substantielle dans le groupe de contrôle : les participants sont 19 points de pourcentage moins susceptibles de choisir un travailleur transgenre qu'un travailleur non-transgenre. Cette discrimination représente un sacrifice équivalent à 1,9 fois leurs dépenses alimentaires quotidiennes pour éviter d'interagir avec une personne transgenre pendant 15 minutes.

La discussion de groupe réduit fortement cette discrimination : les participants ayant pris part à une discussion sont 17 points de pourcentage plus susceptibles de sélectionner un travailleur transgenre que le groupe de contrôle. En moyenne, la discrimination disparaît complètement dans ce groupe. Ces effets persistent partiellement : un mois plus tard, ces participants restent 4 points de pourcentage plus susceptibles de choisir des travailleurs transgenres.

L'information sur les droits légaux réduit également la discrimination, mais dans une moindre mesure : l'effet est 59% plus faible que celui de la discussion de groupe. Les participants informés des droits légaux sont 10,3 points de pourcentage plus susceptibles de sélectionner un travailleur transgenre que le groupe de contrôle.

## Mécanismes

L'étude examine plusieurs mécanismes potentiels pour expliquer les effets de la discussion :

1. Correction d'une norme mal perçue : Les participants pourraient initialement surestimer le niveau de discrimination de leurs pairs. Cependant, ce mécanisme ne peut expliquer que 21% de l'effet du traitement.
2. Signal vertueux : Les participants pourraient vouloir paraître non-discriminatoires en groupe. Toutefois, le traitement "Pas de discussion (public)" n'a pas d'effet significatif, suggérant que ce mécanisme n'est pas suffisant.
3. Persuasion : Les résultats suggèrent que les participants se persuadent mutuellement pendant la discussion. Les "auditeurs" silencieux montrent une réduction similaire de la discrimination, indiquant que l'exposition aux arguments est suffisante. De plus, les participants favorables aux personnes transgenres sont plus vocaux dans les discussions.

## Implications

Cette recherche démontre que la communication horizontale entre pairs peut réduire rapidement la discrimination, même sans injection d'information externe. Cela contraste avec les théories économiques standard qui attribuent la discrimination à des préférences ou des croyances difficiles à modifier. Les résultats suggèrent que les environnements de décision devraient être conçus pour favoriser les discussions de groupe plutôt que les décisions individuelles dans les contextes où la discrimination peut se produire (recrutement, logement, admissions universitaires, etc.). L'étude souligne également l'importance du droit comme outil de réduction de la discrimination sociétale, tout en montrant que la communication entre pairs peut être encore plus efficace.

Ces résultats ont des implications importantes pour la conception de politiques visant à réduire la discrimination. Ils suggèrent qu'il peut être plus efficace de créer des situations où les minorités sont naturellement discutées plutôt que d'essayer de changer directement les attitudes individuelles. Cependant, l'étude souligne également que ces discussions ne sont bénéfiques que dans certaines conditions - notamment lorsque la

discrimination n'est pas trop forte et qu'il existe des individus prêts à défendre la minorité.

# Chapitre 2 : Stigmatisation menstruelle, hygiène et capital humain

Le deuxième chapitre, co-écrit avec Karen Macours et Julieta Vera Rueda, examine comment la stigmatisation menstruelle peut être combattue en s'appuyant sur des "déviants positifs" - des personnes prêtes à s'exprimer contre cette stigmatisation - et comment les programmes axés sur l'hygiène peuvent générer des améliorations du capital humain dans les écoles à faible revenu. Cette recherche s'inscrit dans un contexte où la stigmatisation menstruelle est très répandue, avec environ une femme sur trois qui risque la honte et le harcèlement en raison de la menstruation. Une telle stigmatisation peut avoir des impacts négatifs sur la santé, la fréquentation scolaire et l'apprentissage.

L'étude présente les résultats d'un essai contrôlé randomisé mené dans 140 écoles à Madagascar. Dans ce contexte, la menstruation est fortement stigmatisée et les comportements d'hygiène sont généralement médiocres. Au début de l'étude, seules 45% des filles avaient entendu une camarade de classe parler de menstruation, 39% pensaient qu'on ne devrait pas en discuter ouvertement, et seulement 55% avaient reçu des informations de leur mère sur le sujet. De plus, seules 11% des filles utilisaient des serviettes hygiéniques, les autres improvisant avec des tissus ou des vêtements usagés.

Le design expérimental évalue les effets de deux composantes du programme. Premièrement, il examine un ensemble d'interventions qui comprend simultanément l'engagement et la formation des enseignants pour promouvoir l'hygiène et réduire la stigmatisation menstruelle dans les écoles, ainsi que la construction d'infrastructures d'hygiène (lavabos et latrines) et la distribution gratuite de serviettes hygiéniques. Deuxièmement, il évalue l'effet additionnel d'une approche innovante consistant à identifier et à amplifier la voix des déviants positifs. Dans chaque école traitée, environ quatre "Jeunes Filles Leaders" sont sélectionnées sur la base de leur volonté de parler activement et ouvertement de la menstruation, malgré la nature taboue du sujet, et de leurs compétences générales en leadership.

Les résultats montrent que le programme de base seul conduit à des impacts importants sur les résultats d'apprentissage des filles. Le programme génère des améliorations substantielles (0,15 écart-type) des tests d'apprentissage et des notes scolaires officielles, sans affecter l'assiduité ou la santé. Le rythme cardiaque des filles pendant l'enquête finale est inférieur de 0,12 écart-type dans les écoles traitées, suggérant que les bénéfices d'apprentissage peuvent provenir d'une réduction du stress et d'une amélioration de l'environnement psychosocial à l'école.

Les améliorations de l'environnement psychosocial sont corroborées par d'autres résultats. L'étude constate une augmentation de 0,23 écart-type d'un indice de connexions en réseau, basé sur les liens des filles avec les autres à l'école, et une réduction de la gravité du harcèlement scolaire (-0,08 écart-type). Ces améliorations pourraient favoriser l'apprentissage soit en réduisant le stress, soit en augmentant la collaboration et l'apprentissage entre pairs, soit en améliorant la motivation des élèves.

L'intervention des Jeunes Filles Leaders, combinée au programme de base, s'avère significativement plus efficace que le programme de base seul pour améliorer plusieurs mesures le long de la théorie du changement. Alors que les deux versions du programme conduisent à des améliorations substantielles des connaissances en matière d'hygiène (0,3-0,4 écart-type), ces améliorations apparaissent plus rapidement lorsque la composante JFL est incluse. La composante JFL génère également des améliorations supplémentaires significatives du comportement en matière d'hygiène, suggérant que les JFL réussissent à réduire partiellement l'écart entre connaissances et comportements.

En ce qui concerne la stigmatisation menstruelle, les résultats montrent une réduction importante dans les écoles Base+JFL par rapport au contrôle. Un indice combiné de stigmatisation indique une amélioration de 0,7 écart-type dans les écoles Base+JFL (contre 0,4 écart-type dans les écoles Base seul). Par exemple, les filles déclarent être plus disposées à parler ouvertement de menstruation, avoir des attitudes plus progressistes concernant les comportements liés à la menstruation, et croire que les autres dans leur réseau ont des attitudes plus progressistes.

L'étude examine également si les écoles avec des JFL particulièrement "déviantes positives" obtiennent de meilleurs résultats. Les résultats montrent que ces écoles ont de meilleurs scores aux tests d'apprentissage, des taux plus faibles de harcèlement grave et une meilleure santé mentale déclarée. Cependant, cette hétérogénéité semble être fortement liée à l'environnement scolaire initial, suggérant que les JFL positivement déviantes ne peuvent être trouvées que dans des environnements scolaires déjà plus progressistes, ou ne peuvent être efficaces que dans ces environnements.

Enfin, l'étude documente des preuves suggestives que le programme de base peut générer des conséquences négatives imprévues, mais que les JFL peuvent compenser ces effets. Par exemple, il y a des preuves que les grossesses augmentent dans les écoles Base seul, et les mères rapportent un comportement psychosocial plus mauvais pour leurs filles dans ce même groupe. Ces effets négatifs ne se matérialisent pas dans les écoles Base+JFL. En revanche, il y a des indications que le programme JFL évince certains des effets d'apprentissage du programme de base, avec des estimations ponctuelles généralement plus faibles sur les tests d'apprentissage dans les écoles Base+JFL par rapport à Base seul.

Cette recherche apporte plusieurs contributions importantes. Elle montre qu'un programme axé sur la réduction de la stigmatisation menstruelle et l'amélioration des comportements d'hygiène peut générer des effets substantiels sur l'apprentissage des adolescentes. Elle démontre également que le travail avec des déviants positifs peut être une tactique efficace pour s'attaquer aux normes sociales néfastes. Ces résultats pourraient être appliqués dans d'autres contextes où il existe des déviants positifs dont la voix peut être amplifiée, comme la lutte contre la discrimination anti-minoritaire ou contre des pratiques néfastes comme les mutilations génitales féminines.

# Chapitre 3 : Périodes critiques dans le développement cognitif et socio-émotionnel

Le troisième chapitre examine comment le moment des chocs pendant la petite enfance peut avoir des conséquences durables sur le développement du capital humain d'un individu. Bien que les circonstances de la petite enfance soient des déterminants importants des résultats en termes de capital humain et de bien-être à long terme, nous avons peu de preuves causales sur l'importance relative du moment de ces circonstances. Les 1000 premiers jours de vie sont souvent cités comme une "période critique" pour le développement de l'enfant, mais cette notion a rarement été testée directement.

Cette recherche utilise un cadre où les enfants sont potentiellement soumis à des chocs à chaque année de leur enfance pour estimer l'impact des chocs météorologiques en début de vie sur les résultats cognitifs et socio-émotionnels à l'âge adulte. L'étude se concentre sur les individus nés dans l'Indonésie rurale entre 1988 et 2000, utilisant les données de l'Indonesia Family Life Survey (IFLS), une enquête longitudinale représentative de 83% de la population indonésienne.

La méthodologie exploite plusieurs avantages clés du contexte. Premièrement, l'Indonésie présente une grande diversité géographique, s'étendant sur 5 100 km d'est en ouest, ce qui génère une variation significative des conditions météorologiques entre les provinces et les districts. Deuxièmement, avec une importante population rurale et un grand secteur agricole, la population étudiée fait l'expérience des chocs météorologiques comme des chocs de revenu agricole qui affectent les enfants à travers des changements dans la nutrition et d'autres canaux d'investissement. Troisièmement, les données de l'IFLS permettent de reconstruire l'historique détaillé de la localisation des individus depuis leur naissance, grâce à des taux d'attrition très faibles et des données complètes sur les migrations.

La variable météorologique principale est l'indice standardisé de précipitations et d'évapotranspiration (SPEI), un indice de sécheresse qui combine des informations sur les précipitations, la température, la latitude, la vitesse du vent et le nombre d'heures d'ensoleillement. Cet indice est particulièrement pertinent car il prédit mieux les rendements des cultures que d'autres indices climatiques basés uniquement sur la température ou les précipitations.

Les résultats montrent l'existence d'une période critique forte à l'âge de 2 ans pour le développement cognitif. Une amélioration d'un écart-type de l'indice SPEI à l'âge de 2 ans conduit à une augmentation significative des capacités cognitives à l'âge adulte de 0,063 écart-type. Des variations météorologiques similaires n'ont pas d'effet statistiquement significatif à d'autres âges, y compris in utero. L'effet à l'âge de 2 ans est très robuste à différentes spécifications empiriques, et des tests de placebo démontrent qu'un résultat aussi fortement significatif est extrêmement improbable d'être le résultat d'artefacts dans les données.

En revanche, il n'y a pas de preuve positive d'une période critique pour le développement socio-émotionnel : les effets sont soit nuls, soit ne peuvent pas être isolés à une seule période critique. Cette distinction entre le développement cognitif et socio-émotionnel contribue aux preuves que le capital humain ne peut pas être considéré comme un "objet unidimensionnel".

L'étude montre des preuves suggestives que la variation météorologique affecte les capacités cognitives adultes en influençant les investissements nutritionnels pendant l'enfance. Les ménages des zones rurales font l'expérience de la variation météorologique comme un choc sur le bien-être global du ménage, affectant leurs dépenses par habitant et l'investissement dans la nutrition des jeunes enfants. Les effets peuvent être isolés à une période de 3 à 8 mois après la récolte, permettant de préciser le moment des impacts des chocs.

Un résultat surprenant est l'absence d'effet avant l'âge de 2 ans, contrairement à la littérature existante qui propose les 1000 premiers jours comme période critique pour le développement cognitif. L'étude suggère que cela pourrait être dû au fait que les enfants sont protégés des chocs nutritionnels lorsqu'ils sont allaités. Les données montrent que les mères adaptent leur comportement d'allaitement de manière à compenser les chocs négatifs : lorsque les récoltes sont mauvaises autour de l'âge d'un an, les mères prolongent l'allaitement.

L'étude examine également si les effets à long terme des chocs en début de vie sur les compétences cognitives peuvent être détectés plus tôt. Les résultats montrent que l'impact des chocs météorologiques précoces sur le score cognitif ne peut pas être détecté pendant l'enfance intermédiaire (entre 7 et 14 ans). Les effets sont initialement latents et n'apparaissent qu'après l'âge de 15 ans, ce qui fournit un avertissement contre l'utilisation des scores cognitifs pendant l'enfance intermédiaire comme statistique suffisante pour les impacts à long terme sur le développement cognitif.

Ces résultats ont des implications importantes pour la conception des politiques. Dans ce contexte, l'impact des chocs météorologiques sur le développement cognitif est isolé à l'année entre le deuxième et le troisième anniversaire. Si des périodes critiques similaires existent pour les effets d'autres politiques qui agissent par des canaux similaires, cibler ces âges pourrait considérablement augmenter leur rapport coût-efficacité. L'étude souligne également l'importance d'améliorer l'assurance contre les chocs pour les familles ayant des enfants récemment sevrés, étant donné le rôle protecteur apparent de l'allaitement.