Essays in Gender and Development Economics



Claire Cullen

St Antony's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Public Policy

Trinity 2020

Acknowledgements

To my parents, Karen Webb Cullen and Charlie Cullen- you set me on the path towards this DPhil a long time ago when you patiently volleyed my endless questions of 'why' and gave me the space and support to go in search of answers.

I am so very grateful to my supervisors, Julien Labonne and Stefan Dercon, who provided invaluable guidance every step of the way. I would like to thank Julien for his thoughtful advice, support, generosity, and humor. His academic mentorship has been a big part of why this DPhil has been one of the most rewarding and enjoyable experiences I've had. Stefan's guidance and perspective throughout this work have helped me to look farther ahead and reach higher. I know I won the supervisor lottery when I got you both.

I have also benefited greatly from comments from my examiners Anandi Mani, Cheryl Doss, and Simon Quinn. I am grateful to the many others who have generously shared their time, advice and experience with me throughout this project, including faculty, postdocs, and students at the Centre for the Study of African Economies (CSAE), the Blavatnik School of Government, the Economics Department, and the International Development Department. My thanks also go to Ted Miguel and UC Berkeley faculty for their feedback in developing this work.

I am so grateful to staff at the Government of Rwanda's Ministry of Gender and Family Promotion, and to colleagues at the World Bank, particularly my co-authors Arthur-Alik Lagrange, Müthoni Ngatia, and Julia Vaillant, as well as Patricia Fernandes, Harald Hinkel, Markus Goldstein, Jean-Paul Ngabonziza, Sreelakshmi Papineni, and Verena Phipps for their collaboration so fundamental to this research. And to the staff at Laterite and TNS Global who did such fantastic, careful work in Rwanda and Nigeria, and were so patient in the face of all the challenges of data collection. And to the most important people without whom this research could never have been done- the women and men in Rwanda and Nigeria who agreed to give their time and share their personal experiences so we could learn more about the causes of, and how to combat, gender inequality and intimate partner violence.

I could not have completed this DPhil without funding from the Blavatnik School of Government, CSAE, St Antony's College, the Fund for Women Graduates, and Graduate Women International.

Finally, last but by no means least, thanks to my stellar DPhil classmates and my friends and family in Oxford, Australia, the US, Rwanda, and around the world who read drafts, gave feedback, challenged my ideas, and nurtured this thesis (and me) in myriad ways. This DPhil took a whole global village. I'll always look back on this experience with such wonderful memories- I'm so grateful to all who made it possible.

Abstract

This dissertation analyzes how social and behavioral frictions in decision-making contribute to gender inequality in developing economies. It studies factors that have received relatively little attention in economics – namely social norms, role models, and concerns about preserving identity – and examines their impact on behavior, including intimate partner violence (IPV). I use field and survey experiments in Rwanda and Nigeria to provide empirical evidence on these relationships.

Economic progress has brought about significant improvements in global living conditions, yet it has not always translated into improvements for women and girls (Dhar, Jain, & Jayachandran 2018; Duflo 2012). Currently, women comprise only 39 percent of the world's labor force, they spend two to ten times more time on unpaid domestic and care work than men, and one in three women experiences IPV at some point in her lifetime (Ferrant, Pesando, & Nowacka, 2014; WHO, 2013). The persistence of IPV and other gender gaps in all economies and at all income levels suggests that there is no mechanical improvement in gender equality as countries become richer. Instead, sticky gender identities and social norms appear to play a role in maintaining gender inequality (Alesina, Brioschi, & La Ferrara 2020; Jayachandran 2015).

In my first paper, co-authors and I conducted a two-level randomized control trial to test the effectiveness of a government-run IPV prevention program in Rwanda, randomly assigning 98 villages and 2,042 couples into treatment and control groups. The intervention involved 22 weekly facilitated group discussions designed to reduce IPV by improving couples' communication skills, changing individuals' gender attitudes, and spreading a message of gender equality through the community. Randomizing at two levels– villages and couples– we assess the program's effect both on trained couples and on 'spillover couples'– control couples living in treated villages. Six months after the program ended, we find that it caused large, unintended increases in IPV. Women assigned to the couples training report experiencing more emotional, physical and sexual IPV compared to control couples living in control villages, with increases of 11%, 15% and 24% respectively. Spillover couples have even larger IPV increases when compared to control couples, with increases of 11%, 33% and 40% respectively. We present evidence suggesting that the IPV impacts represent a true increase in violence rather than an increase in women's likelihood of reporting. Evidence on mechanisms suggests that the program may have increased violence because men backlashed against the perceived threat to their identity

posed by program messages about women's empowerment, and against their wives' more progressive gender attitudes and greater aspirations.

In my second paper, I use a measurement experiment in Rwanda and Nigeria to analyze the magnitude and predictors of misreporting on intimate partner and sexual violence. I randomly assigned respondents to answer questions about IPV using one of three survey methods: an indirect method (list experiment) that gives respondents anonymity; a direct, self-administered method that increases privacy; and the standard, direct face-to-face method. In Rwanda, IPV rates are double, and in Nigeria, they increase by up to 39 percent when measured using the list method compared with the direct methods. Misreporting was associated with indicators commonly targeted in women's empowerment interventions, including gender norms. This suggests that the standard way of measuring IPV may result in significant prevalence underestimates and biased correlations and treatment effect estimates.

For my third paper, I ran a survey experiment in Rwanda to test whether an individual's perceptions about social norms influences their own attitudes and behaviors. I provide evidence that while the majority of men and women privately do not hold conservative gender attitudes, most over-estimate the share of others who do. I then randomly assigned women and men who over-estimated conservative gender norms to receive information updates with the 'true' norm (measured 12 months prior), also varying the type of role model about which people are given information. I measure the impact of this information update on an index of self-reported gender attitudes and behavior in an incentivized spouse dictator game. I find that men's gender attitudes improve and they give more to their spouse after having their misperceptions corrected; there is no average impact for women. On role models, I find that men give more to their spouse after being told about all role models, but there is suggestive evidence that they respond most to information about village leaders. For women, there is no significant difference between being updated about men or women role models, though results suggest that women may respond negatively to having their beliefs corrected about other women. My results indicate a potentially important factor in gender inequality, where misperceived social norms affect individuals' own gender attitudes and related behaviors.

Together, my papers contribute to our understanding of the measurement of IPV, the social and behavioral factors that contribute to gender inequality, and the effectiveness of interventions to reduce it. My research suggests that identity and norms likely represent important frictions in decision-making, and may therefore be important to address to overcome gender gaps and other forms of inequality. My research also provides evidence that interventions that challenge prevailing identities can provoke large backlashes, and that even programs intended to target individuals can cause negative spillovers to communities. This has important implications for social scientists studying gender, IPV, and other topics tied to identities, and for policymakers trying to change behaviors rooted in cultural norms.

Contents

1 Introduction **1**

2 The Unintended Impacts of an Intimate Partner Violence Prevention Program: Experimental Evidence from Rwanda **7**

3 Method Matters: Underreporting of Intimate Partner Violence in Nigeria and Rwanda **79**

4 Can role models and correcting misperceived norms change gender attitudes? Evidence from a survey experiment in Rwanda **127**

5 Conclusions and Policy Implications **173**

1 Introduction

This dissertation analyzes how social and behavioral frictions in decision-making contribute to gender inequality in developing economies. It studies factors that have received relatively little attention in economics – namely social norms, role models, and identity – and examines their impact on behavior including intimate partner violence. Field and survey experiments in Rwanda and Nigeria are used to provide empirical evidence on these relationships.

In Chapter 2, co-authors and I conducted a two-level randomized control trial to test the effectiveness of a government-run IPV prevention program in Rwanda, randomly assigning 98 villages and 2,042 couples into treatment and control groups. The intervention involved 22 weekly facilitated group discussions designed to reduce IPV by improving couples' communication skills, changing individuals' gender attitudes, and spreading a message of gender equality through the community. Randomizing at two levels– villages and couples– we assess the program's effect both on trained couples and on 'spillover couples'– control couples living in treated villages. Six months after the program ended, we find that it caused large, unintended increases in IPV. Women assigned to the couples training reported experiencing more emotional, physical, and sexual IPV compared to control couples living in control villages, with increases of 11%, 15%, and 24% respectively. Spillover couples had even larger IPV increases when compared to control couples, with increases of 11%, 33%, and 40% respectively. We present evidence suggesting that this represents a true increase in violence rather than an increase in women's likelihood of reporting. Evidence on mechanisms suggests that the program may have increased violence because men backlashed against the perceived threat to their identity posed by program messages about women's empowerment, and against improvements to their wives' gender attitudes and greater aspirations.

This chapter makes three key contributions to the literature. First, our study is one of the first large-scale randomized studies of an IPV prevention program in a low-income country. We contribute to the growing empirical literature on the impact of such programs-most of which is from smaller-scale and underpowered or non-causal studies, or studies where treatment and control groups live in the same community. These studies have typically found null or promising beneficial effects of such prevention programs (Hossain et al. 2014; Abramsky et al. 2014; Chatterji et al. 2020; Doyle et al. 2018). We demonstrate that these programs can have large unintended negative effects, including negative externalities in communities.

We also contribute to the theoretical and empirical economic literature on IPV, providing some of the first evidence that IPV can increase even in the absence of any material changes in relative economic welfare between spouses, and that a threat to identity potentially explains this. Our paper thus also relates to the nascent literature on the role of identity in IPV, and the broader literature demonstrating that social norms and individual identity explain behavioral and economic outcomes and can shape responses to public policy (Tur-Prats 2019; Lowes et al. 2017; Alesina, Giuliano, and Nunn 2013; Guiso et al. 2006; Bhalotra et al. 2020; Oh 2019; Akerlof and Kranton 2000; Karing 2018).

Third, this paper is one of the few studies of a gender or public health information intervention that measures spillovers to non-participants in the community. Of the existing studies, primarily of edutainment programs intended to change broader norms, few find spillovers beyond targeted individuals (Banerjee, La Ferrara, and Orozco 2019; Green, Wilke, and Cooper 2020). Our results therefore also have important implications for both health prevention and other social programs that seek to, or may unintentionally, have an impact beyond direct program participants. A related methodological contribution of our paper is to highlight the importance of study designs that avoid contamination of control groups. The large externalities we observe suggest that studies of similar programs designed to change culturally sensitive issues should avoid having treated and control groups living in the same community or social network. We thus contribute to the growing literature on spillovers, social networks, and research design in the face of potential interference (Haushofer, Reisinger, and Shapiro 2015; Egger et al. 2019; Baird et al. 2014; DiTraglia et al. 2018; Miguel and Kremer 2004).

Chapter 3 investigates misreporting bias in survey questions about intimate partner and sexual violence in Nigeria and Rwanda. I randomly assigned women and men to answer questions using one of three survey methods: an indirect method (list experiment) that gives respondents anonymity; a direct, self-administered method that increases privacy; and the standard, direct face-to-face method. In Rwanda, IPV rates were double, and in Nigeria, up to 39 percent greater when measured using the list method than direct

methods. Misreporting was systematic and correlated with widely theorized causes of IPV and variables often targeted in women's empowerment programs such as gender norms and education. My results suggest that in some contexts, standard survey methods are likely to result in significant underestimates of the prevalence and welfare costs of IPV. Further, the systematic nature of the misreporting bias highlights the risk that standard survey methods may generate biased correlations and treatment effect estimates, leading to inaccurate conclusions about IPV's causes and potential solutions.

This paper contributes to the growing social science literature studying alternative survey methods that may reduce misreporting, including on IPV measurement. Recent studies on IPV measurement have compared respondents' experience of violence across the face-to-face and list experiment methods. They found mixed results, with one study in Peru finding the same rates of IPV between face-to-face and list methods (Agüero and Frisancho 2017) and others in Vietnam, India, and Nigeria finding a difference (Bulte and Lensink 2019; Joseph et al. 2017; Banerjee, La Ferrara, and Orozco as cited in World Bank 2017). However, in several of these studies, respondents were asked about their IPV experience using two methods within the same survey about similar, but slightly different, violent acts. This potentially limits comparability across methods. To avoid this issue, we conducted one of the first IPV measurement experiments that randomly assigns respondents to answer questions using only one method rather than asking respondents two overlapping questions using different methods. Also, this paper aims to contribute to the limited evidence on who misreports on sensitive questions and why, the implications of misreporting for bias, and the trade-offs between different survey methods that seek to address misreporting.

This is the first paper to compare the prevalence of sexual and intimate partner violence using three survey methods that offer respondents differing levels of privacy and anonymity that compares across two countries, and for women and men. It is also one of the first papers to study IPV misreporting in Africa and to demonstrate that misreporting is substantial in some contexts, varies by the type of violence, and is correlated with characteristics commonly targeted in women's empowerment interventions.

In Chapter 4, I provide evidence from rural Rwanda that while the majority of men and women privately do not hold conservative attitudes about household gender roles and IPV, the majority over-estimate the share of others who do. I randomly assigned women and men who over-estimated conservative gender norms to receive information updates with the 'true' norm (measured 12 months prior), also varying the type of role model about which people are given information. I measure the impact of this information update on an index of self-reported gender attitudes and behavior in an incentivized spouse dictator game. I find that men's gender attitudes improve and they give more to their spouse after having their misperceptions corrected; there is no average impact

for women. On role models, I find that while men give more to their spouse after being told about all role models, results suggest they may respond most to information about village leaders. For women, while there is no significant difference between being updated about men or women role models, I find suggestive evidence that women may respond negatively to having their beliefs corrected about other women. My results indicate a potentially important factor in gender inequality, where misperceived social norms affect individuals' own gender attitudes and related behaviors.

There are three main contributions of this paper. First, this paper contributes to the growing literature demonstrating that individuals misperceive, and are influenced by their perceptions of, social norms. Our focus on norms about gender roles and IPV, and their impact on attitudes and behavior, builds on a large economic literature on gender norms and economic behavior in low-income countries (Banerjee, La Ferrara, and Orozco 2019; Green, Wilke, and Cooper 2020; Field et al. 2019; Jayachandran 2020; Beaman et al. 2012) and more broadly on the role of norms information on decision-making (Grigorieff, Roth, and Ubfal 2018; L Bursztyn, Egorov, and Fiorin 2019; Leonardo Bursztyn, González, and Yanagizawa-Drott 2018). Our results - that men change their attitudes and behavior in response to having their norms beliefs corrected - highlight a potentially important margin for promoting social change.

Second, this paper contributes to the economic literature on role models and aspirations (Del Carpio and Guadalupe 2018; LaFerrara 2019; Mani and Riley 2019; Ray 2006; Lybbert and Wydick 2018; Angrist 2020; LaFerrara, Chong, and Duryea 2012; G.A. Akerlof 1997). As yet, there is limited evidence on the types of role models and social referents that may be more or less effective at influencing gender norms. We contribute to this research, showing that people have different perceptions about the beliefs of different groups, often holding the most accurate beliefs about those they are most proximate to. Our findings also suggest that individuals respond differently to norms information about different role model 'types', potentially based on a shared identity (in our case, sex), or those in positions of authority (in our case, village leaders).

Finally, our results contribute to the literature on motivated beliefs, suggesting that some people may self-deceive and be resistant to, or backlash against, attempts to shift their attitudes with norms information (Bénabou and Tirole 2016; Zimmermann 2020; Kappes and Sharot 2019; Kahan 2013). Overall, this paper builds on the emerging evidence that social norms and other behavioral factors may be an important decision-making friction that contributes to gender inequality.

Together, my three papers contribute to our understanding of the measurement of IPV, the social and behavioral factors that contribute to gender inequality, and the effectiveness of interventions to reduce it. My research suggests that individuals' economic behavior

is influenced both by their desire to preserve their identity and to behave in accordance with social norms. Identity and norms thus likely represent important frictions in decision-making about behaviors related to gender inequality and may, therefore, be important to address in order to overcome it.

Notes for the reader: This dissertation comprises an introduction, three stand-alone papers, and a conclusion, followed by references for the introduction and conclusion. The total word count is approximately 48,000 words.
My second and third papers are solo-authored. My first paper is co-authored with three current and former economists at the World Bank Gender Innovation Lab: Arthur Alik-Lagrange, Müthoni Ngatia, and Julia Vaillant. I did the majority of the work on this paper. The initial idea and design were developed by Alik-Lagrange, Vaillant, and myself. I led on survey instrument development, field data collection, cleaning and analyzing the data, and writing the paper.

2

The Unintended Impacts of an Intimate Partner Violence Prevention Program: Experimental Evidence from Rwanda

The Unintended Impacts of an Intimate Partner Violence Prevention Program: Experimental Evidence from Rwanda

Claire Cullen, Arthur Alik-Lagrange, Mũthoni Ngatia, Julia Vaillant[†]

Abstract

One in three women worldwide experiences intimate partner violence (IPV) at some point in her lifetime, but little is known about how to prevent it. We conduct one of the first large-scale randomized control trials of an IPV prevention program in a low-income country. The program, delivered in rural Eastern Rwanda, involved 22 weekly facilitated group discussions designed to reduce IPV by improving couples' communication skills, changing individuals' gender attitudes, and spreading a message of gender equality in the community. Randomizing at two levels– villages and couples– we assess the program's effect both on trained couples and on 'spillover couples'– control couples living in treated villages. Six months after the program ended, we find that it caused large, unintended increases in IPV. Women assigned to the couples training report experiencing more emotional, physical and sexual IPV compared to control couples, with increases of 11%, 15% and 24% respectively. Spillover couples have even larger IPV increases, of 11%, 33% and 40%, respectively. We present evidence suggesting that this represents a true increase in violence rather than an increase in women's likelihood of reporting. Evidence on mechanisms suggests that the program may have increased violence because men backlashed against the perceived threat to their identity posed by program messages about women's empowerment, and against improvements to their wives' gender attitudes and greater aspirations.

JEL categories: C93, J12, J16, O12
Keywords: Intimate Partner Violence, Gender, Identity, Spillovers, Norms, Field Experiments

[†]The authors can be reached at Cullen (claire.cullen@bsg.ox.ac.uk), Alik-Lagrange (aaliklagrange@worldbank.org), Ngatia (ngatia@gmail.com), Vaillant (jvaillant@worldbank.org). Acknowledgments: We thank Julien Labonne, Stefan Dercon, Markus Goldstein, and Rachael Pierotti for support throughout the project. We thank the respondents for participating in the surveys, and the staff at Laterite Ltd for their outstanding data collection work. We also thank Jean Paul Ngabonziza, Victoria Isika, Harald Hinkel, Verena Phipps, Patricia Fernandes, Jessica Ott, and Alex Twahirwa and other staff at MIGEPROF, and CARE, the Rwandan Women's Network, and Rwandan Men's Resource Centre for support, programming, and insights into program administration. We are grateful to Lori Heise, Kristin Dunkle, and Erin Stern, and other members of the Indashyikirwa 1 design, implementation, and research teams for their roles in the original program and advice on revisions, and for fruitful discussions of the results. We are grateful for comments and suggestions from Diana Arango, Leonardo Burzstyn, Cheryl Doss, Karlijn Morsink, Anandi Mani, Ted Miguel, Jeremy Magruder, Rossa O'Keefe O'Donovan, Berk Özler, Simon Quinn, and seminar participants at the World Bank, Blavatnik School of Government, and CSAE. All errors and opinions are our own and do not represent the views of the World Bank or its member countries. The study protocols were approved by the Rwandan National Ethics Committee (722/RNEC/2018) and the Oxford University Central University Research Ethics Committee (R50347), and a research visa was granted by the National Institute of Statistics Rwanda (0676/2018/10/NISR).

1. Introduction

While economic progress has brought about significant improvements in global living conditions, it has not always translated into improvements for women and girls (Duflo 2012; Jayachandran 2015). One widespread form of gender inequality is intimate partner violence (IPV), which affects one in three women in their lifetime (García-Moreno et al. 2015) and in some countries, affects as many as 70 percent (WHO 2013).[1] The persistence of IPV and other gender gaps in all economies and at all income levels suggests that there is no inevitable improvement in gender equality as countries become richer. Instead, sticky gender identities and social norms appear to play a role in maintaining gender inequality and may be vital to tackle to overcome high rates of IPV in equilibrium. Although IPV prevalence varies widely across contexts, some of the highest rates are found in sub-Saharan Africa, the setting for this study (Devries et al. 2013; Cools and Kotsadam 2017).[2] However, little is known about how to prevent IPV.

In this paper, we present the results of one of the first large-scale randomized control trials (RCTs) designed to study the impact of an IPV prevention program. The intervention, called *Indashyikirwa 2*, meaning 'agents of change,' was a group-based couples training program led by the Rwandan Ministry of Gender and Family Promotion and implemented by local NGOs. The program, developed for the Rwandan context, consisted of 22 weekly facilitated group discussions with approximately 20 couples per group. It was designed to prevent IPV by improving couples' marital communication skills, shifting conservative gender attitudes, and promoting new progressive gender norms in the community. Trained local facilitators delivered information and skills through structured sessions, including group activities and discussion, deliberation, and couples' take-home exercises.[3] Anecdotal evidence also suggests that in treated villages, village leaders also asked participating couples to promote their experience of the program and its message of gender equality during Rwanda's monthly compulsory community-wide meetings.

Using a two-level randomized trial design, 98 villages and 2,000 couples were assigned to treatment or control groups, producing three study groups: treated couples (couples living

[1] IPV is estimated to cost the global economy between 1.8 and 5 percent of global GDP (World Bank, 2018; Fearon & Hoeffler, 2014). The costs include negative impacts on women's labor force participation, and women's and children's physical and mental health. These impacts can be passed on through generations, with the children of IPV victims having lower birth weight and IQ, worse school performance, and being more likely to enter abusive relationships as adults (Koenen et al. 2003; Aizer 2011; Carrell, Hoekstra, and Kuka 2018; WHO 2013; Roy et al. 2018; Hidrobo et al. 2014)

[2] Most perpetrators of violence against women are male intimate partners or former partners (WHO 2012). The standardized measurement of IPV typically includes emotional, physical, and sexual abuse by an intimate partner (WHO 2012). Male-on-female IPV is the most common form of IPV (Devries et al. 2013). This study will focus on understanding the causes of IPV perpetrated against women by their married or de facto partners.

[3] The training curriculum covered topics including the differences between gender and sex, gendered power dynamics, marital communication and conflict management, alcohol abuse, household budget management, parenting, Rwandan laws against domestic violence, and strategies to promote gender equality in the community. The intervention is discussed in more detail below; the 22 sessions are listed in Figure A2 in the appendix.

in treated villages and assigned to receive the couples training); spillover couples (control couples living in treated villages); and 'pure' control couples living in control villages.[4] This design allows us to identify both the direct and 'spillover effects' of the intervention by comparing the treated and the spillover groups with the pure control group, respectively. Six months post-program, we measured impacts on women's self-reported experience of emotional, physical, and sexual IPV, and collected rich data on other indicators of gender attitudes, wellbeing, and economic empowerment.

We find large, unintended, negative impacts of the IPV prevention program on both treated and spillover couples. Relative to pure control couples, women who attended the couples training report *more* emotional, physical, and sexual intimate partner violence (7, 5, and 10 percentage points more, respectively). Spillover couples have even larger increases, of 7, 11, and 17 percentage points, respectively. The negative IPV impacts are also reflected in broadly negative impacts across other wellbeing indicators, particularly for spillover couples. For example, depression, alcohol consumption, and IPV against men also increase.

Evidence on mechanisms suggests that the program may have threatened men's sense of identity by empowering women in the community, provoking a widespread IPV backlash. For example, the program increased treated and spillover women's progressive gender attitudes and aspirations for input in decision-making. It also widened the gap between a wife's progressive gender attitudes and her husband's, with this gap widening the most for spillover couples. For treated couples, the program appears to have partially mitigated the worst of the IPV backlash (relative to spillover couples) as treated couples show some improvements in relationship quality and cooperation, and treated men adopt some new progressive gender attitudes. However, for *all* couples (treated and spillover men and women), there is a worsening in beliefs about whether a wife has the right to refuse sex.

Anecdotal evidence supports a mechanism of backlash against threatened gender identities. Qualitative reports indicate that village leaders in treated villages asked trained couples to speak during Rwanda's monthly compulsory village-wide meetings about their experience of the program and its messages about gender equality. These reports suggest that some of these messages were met with vocal public resistance at community meetings, including the

[4] All study couples were existing members of Village Savings and Loan Associations, meaning that we measure treatment effects by comparing treated VSLA members to VSLA members living in pure control villages. Similarly, spillover effects are measured by comparing spillover couples (members of VSLAs where some other members were treated) with VSLA members living in control villages. A baseline comparison sample found VSLA members are very similar to average community members on key observables. While the VSLA could be the source of the dissemination of spillover effects, we believe that the spillover effects are more likely to reflect a general community 'spillover effect' because: VSLAs are not formed organically with groups of friends; VSLAs dissolve after one or two years; respondents told us they are no more likely to be friends with people in their VSLA than other non-VSLA community members; and Rwandan communities have many other compulsory or highly-attended community meetings where people interact with others and where this program was discussed (VSLAs meet about once a week, about as often as the other community meetings respondents are part of). It is, therefore, possible that VSLAs themselves are *not* the main channel for the negative spillovers, and that the spillovers may represent broader community spillovers.

notion of a woman's right to refuse and initiate sex with her husband.[5] Reports also suggest that trained men who adopted new behaviors such as sharing of household chores or who reduced drinking at bars were shamed and became the subject of gossip, sometimes reverting to old behaviors.

The primary outcomes of women's experience of emotional, physical, and sexual IPV are self-reported. We measure IPV using the most common, globally standardized WHO IPV survey instrument. We also sought to minimize reporting bias using survey methods that increase respondent privacy and have been shown, in our context, to increase women's willingness to report IPV (Cullen 2020).[6] It is possible that the IPV treatment effects reflect an increase in women's propensity to report rather than (or as well as) a true increase in IPV.[7] However, we consider evidence on three key reasons why women's propensity to report might increase because of the program: (1) strategic misreporting to get access to the program; (2) expanded understanding of IPV question terms; and (3) reduced stigma or empowerment to report. We find suggestive evidence that the impact likely represents a true increase, not just increased reporting.[8]

This study brings together two strands of the growing economic literature on the causes of IPV- the literature that has focused on women's relative bargaining power in the household, and literature on identity and persistent cultural factors such as norms that affect current-day economic behavior (Alesina, Brioschi, and La Ferrara 2020; Aizer 2010; Stevenson and

[5] This evidence on the community-wide dissemination of program messages suggests that the couples training program may have been complemented with an informal community-level 'intervention,' rather than being a targeted program that was then disseminated to others in the village. If the spillovers were primarily caused by the former community meetings pathway, this would imply 'spillover couples' are not 'spillover' in the standard sense of the term (Baird et al. 2014). We will use the term 'spillovers' in this paper given we are unable to definitively test with our design the dominant causal pathway through which the spillover effects occurred, and as they are 'spillover' according to the original study design and intervention intent.

[6] We use Audio Computer-Assisted Self Interview (ACASI) to interview women about IPV. During this study's baseline, Cullen (2020) experimentally tested reporting rates under different survey methods, the face-to-face method used in most DHSs, ACASI, and the list experiment. While she found the list experiment led to the highest reporting rates, and thus likely the lowest misreporting, we chose to use ACASI to measure outcomes in this study due to concerns about statistical power to detect treatment effects with the list experiment, given the large standard errors it generates.

[7] For example, women in the treated or spillover groups may be more likely to report IPV due to: strategically reporting to increase one's chance of program access in the future; an expanded understanding of what the terms in the IPV questions mean; or reduced IPV reporting stigma or otherwise feeling more empowered to report it.

[8] First, we present evidence that suggests we can rule out increased strategic reporting to maximize the likelihood of being 'selected' for future access to the program and its small travel stipend. This evidence includes, for example, all groups reporting an increase in men's alcohol consumption, no impact on self-reported financial need, and leaders lamenting the random selection rather than targeting of high IPV couples. We also find evidence that indicates we can rule out reporting bias from an expanded understanding of what the IPV terms mean. For example, the IPV increase is smaller for treated women who spent 22 weeks discussing gender than for spillover women; and we find similar negative impacts across IPV questions that use more subjective and objective terms. Finally, we rule out women increasing reporting because of a reduction in the stigma of reporting. For example, we find almost no treatment effect on women's refusal-to-answer rate on IPV questions, though it is marginally *higher* for spillover women on physical IPV questions; and spillover women who experience the largest increase in IPV are also no more likely to speak to others about the violence that they experienced. Finally, women's reports of experiencing violence are correlated with several non-self-reported indicators related to violence that have less risk of measurement bias, for example, questions answered by the enumerator about whether the woman had visible signs of injury, and was referred to and spoke to a counsellor after the survey. Further, we also find that spillover men report being more likely to perpetrate IPV in a set of hypothetical situations, corroborating their wives' reports of an IPV increase (though treated men report being less likely to perpetrate violence).

Wolfers 2006; Akerlof and Kranton 2000; Haushofer et al. 2019). Most theoretical models of the economic behavior underlying IPV (Farmer and Tiefenthaler 1997; Tauchen, Witte, and Long 1991; Bloch and Rao 2002; Eswaran and Malhotra 2011) frame it as the result of noncooperative intrahousehold bargaining between partners, where men's utility for violence is constrained by the female partner's outside option (Roy et al., 2018; Haushofer et al., 2019). IPV has also been presented as either instrumental – to extract pecuniary resources from the female partner – or expressive – where the husband receives non-pecuniary returns from violence, for example, if he enjoys asserting dominance (Haushofer et al. 2019). To date, most economic literature on IPV has focused on shifting bargaining power dynamics either from tangible material resources (such as cash transfers to wives) or changes in outside options that could affect material resources or stress levels (such as men's unemployment, eased divorce laws or women's relative wages). The empirical literature primarily finds that boosting women's bargaining power tends to reduce or have no effect on IPV (Buller et al. 2018; Baranov et al. 2020). Further, most economic models of IPV do not explain the underlying drivers of expressive violence, nor why expressive violence occurs if men do not have an implicit taste or preference for it.

This paper makes three key contributions to the literature. First, we contribute to the theoretical and empirical economic literature on IPV, providing some of the first evidence that IPV can increase even in the absence of any material changes in relative economic welfare between spouses, and that a threat to identity potentially explains this. Our paper thus also contributes to the nascent literature on the role of identity in IPV, and the literature demonstrating that social norms and individual identity explain behavioral and economic outcomes and can shape responses to public policy (Tur-Prats 2019; Lowes et al. 2017; Alesina, Giuliano, and Nunn 2013; Guiso et al. 2006; Bhalotra et al. 2020; Oh 2019; Akerlof and Kranton 2000; Karing 2018).

Second, our study is one of the first large-scale randomized studies of an IPV prevention program.[9] We contribute to the growing empirical literature on the impact of such programs in low-income countries- most of which is from smaller-scale and potentially underpowered or non-causal studies. These studies have typically found null or promising beneficial effects

[9] Most of the existing causal research on IPV prevention from low- and middle-income countries comes from studies conducted in a small number of communities, for example, between 4-16 communities, and risk being underpowered. Some of this research also risks generating biased treatment effects if controls live in the same communities and social networks as treated individuals. In Rwanda, Doyle et al. (2018) randomized couples within the same village to treatment and control groups, and found that a parenting and IPV prevention couples training program caused a significant reduction in IPV. Further, several cluster randomized trials of IPV prevention programs that were conducted in few communities do not cluster standard errors at the highest unit of randomization, citing statistical power constraints (Dunkle, Stern, and Heise 2020; Chatterji et al. 2020). While resulting in smaller standard errors, this analytical approach is unconventional in economics. Finally, during sample recruitment in some of these studies, couples were recruited differently in the treatment and control arms and are different on key pre-intervention observable characteristics, including IPV, potentially biasing treatment effect estimates (Dunkle, Stern, and Heise 2020; Chatterji et al. 2020). We could only identify one study of a training-based IPV prevention program, by Vaillant et al. (2020), that was well-powered, clustered standard errors at the unit of randomization, and where the control group did not live in the same village as the treatment group. The relevant literature is discussed in more detail below.

of such prevention programs (Dunkle, Stern, and Heise 2020; Chatterji et al. 2020; Doyle et al. 2018; Hossain et al. 2014; Abramsky et al. 2014). We identified only one training-based IPV prevention study, by Vaillant et al. (2020), that was well-powered, clustered standard errors at the unit of randomization, and where the control group lived in different communities to the treatment group, thus avoiding potential contamination. This study, of a 16-session group discussion program in Eastern DRC, worked with groups of men, rather than couples, to change gender attitudes, and found no impact on wives' reported experience of IPV. We demonstrate that programs targeted to couples can have large unintended negative effects, including large externalities in communities.

Third, this is one of the few studies of a targeted gender or public health information intervention that measures spillovers to non-participants in the community. Of the existing studies, primarily of edutainment programs intended to change broader norms, few find spillovers beyond participating individuals (Banerjee, La Ferrara, and Orozco 2019; Green, Wilke, and Cooper 2020). Therefore, our results also have important implications for other social programs that seek to, or may unintentionally, have an impact beyond direct program participants. Our paper's related methodological contribution highlights the importance of study designs that avoid contamination of control groups. The large externalities we observe suggest that studies of similar programs designed to change culturally sensitive issues should avoid having treated and control groups living in the same community or social network. Had our control group lived in the same community as our treatment group, we would have concluded the program caused a significant reduction in IPV. We thus contribute to the growing literature on spillovers, social networks, and research design in the face of potential interference (Haushofer, Reisinger, and Shapiro 2015; Egger et al. 2019; Baird et al. 2014; Miguel and Kremer 2004).

This study also ties into other strands of the literature. We contribute to the growing theoretical and empirical literature on the mechanisms driving gender gaps and efforts to reduce them (Beaman et al. 2012; Dhar, Jain, and Jayachandran 2018; La Ferrara, Chong, and Duryea 2012; Jensen and Oster 2009; Efferson et al. 2015), and to the literature showing the fragility of norm compliance and that anti-social behavior is more contagious than pro-social behavior (Dimant 2019; Bicchieri et al. 2020). Finally, we also contribute to broader literatures on aspirations and agency, households, persuasion, and preference formation (Bowles 1998; DellaVigna and Gentzkow 2010; Becker 2009, 1973; Benabou and Tirole 2011; Lybbert and Wydick 2018; Mani and Riley 2019).

This paper is structured as follows. The next section briefly discusses our conceptual framework and the literature, followed by the study context, the intervention, and study design. Section 3 discusses the data and empirical strategy. Section 4 presents the results on primary

and secondary outcomes. In Section 5, we discuss mechanisms and robustness checks. Section 6 discusses evidence on whether impacts reflect true impacts or differences in reporting, and the final section concludes.

2. Study Context and Design

2.1 Conceptual Framework

We briefly discuss the application of the Akerlof and Kranton (2000) identity model to the IPV prevention program in our setting. This model serves to structure our analysis, guide our consideration of the literature, and inform our understanding of the mechanisms through which a prevention intervention might affect IPV.

As discussed in the introduction, the primary focus of the economic literature on IPV has been on material transfers or changes in men's and women's relative income or bargaining power. In the absence of a material change in relative bargaining power (when men might use IPV instrumentally to extract resources), economic models of domestic violence imply that the 'expressive' use of IPV would only result because men have a preference or taste for using violence. We provide some of the first evidence that IPV can increase even in the absence of any material changes in relative economic welfare between spouses. We also suggest that the identity model, which predicts a backlash against threatened male identity, better explains IPV in our context than the removal of a constraint to men's taste for violence. This model provides a framework for understanding the gender identity factors causing IPV; the mechanisms through which this, and other, IPV prevention programs try to reduce IPV; and how such programs may inadvertently risk prompting a backlash.

The identity model was informed by an extensive cross-disciplinary literature on the importance of social categories for behavior (Akerlof and Kranton 2000). Akerlof and Kranton define identity as one's sense of self or one's sense of belonging to one or multiple social categories. In the model, one's identity directly enters the utility function. People's economic actions are therefore influenced by a desire to conform with one's sense of self. Importantly, one can experience disutility from observing *others* violating social categories. Akerlof and Kranton highlight three implications from their model and empirical tests that are relevant for our context: (1) people have identity-based payoffs derived from their own actions; (2) people have identity-based payoffs derived from others' actions; and (3) third parties can generate persistent changes in these payoffs.

Akerlof and Kranton draw four conclusions about identity's role in explaining economic behavior that are relevant in our context of an IPV prevention program designed to change gender attitudes and norms, and in the face of an apparent backlash. First, they suggest that identity can explain behavior that appears detrimental. "People behave in ways that

would be considered maladaptive or even self-destructive by those with other identities. The reason for this behavior may be to bolster a sense of self or to salve a diminished self-image."

Second, they note that identity presents a new type of externality. In this way, one person's behavior can have meaning for and provoke responses in others. For example, a man choosing to take on housework, care for his children, or spend leisure time with his wife instead of going to the bar with friends may threaten the identity of other men. This externality may then itself create further externalities if other men in the community, or his wife, react negatively (or positively) to his behavior changes, potentially evoking further reactions.

Third, identity implies a way for preferences to be changed, noting that conceptions of identity evolve and that governments or other agents have incentives to manipulate them. These efforts, for example a government or NGO introducing an IPV prevention program, may seek to change the identity payoffs by changing social categories and their associated prescriptions to ultimately change economic behaviors.

Finally, Akerlof and Kranton conclude that because identity is fundamental to behavior, one's choice of identity may be one of the most important economic decisions one makes and one of the most important determinants of economic wellbeing (Akerlof and Kranton 2000).

Given its relevance to our context, we present the most basic elements of the model below. Akerlof and Kranton propose the following utility function:

$$U_j = U_j(a_j, a_{-j}, I_j)$$

Such that utility depends on j's identity or self-image I_j, as well as on the other standard vectors of j's and other's actions, a_j and $, a_{-j}$ respectively. They propose the following representation of I_j:

$$I_j = I_j(a_j, a_{-j}, c_j, \in_j, P)$$

A person j's identity I_j depends on j's social categories c_j, and each social category is associated with a social status such that those in the higher social status category may enjoy higher self-image. Identity also depends on how much j's own given characteristics $\in_j$ match the ideal of j's category, captured by prescriptions P. Finally, Aklerof and Kranton conclude that identity depends on the extent to which j's own and others' actions match prescribed behavior indicated by P. They call increases or decreases in utility deriving from I_j gains or losses in identity.

Their model demonstrates that people have identity-related payoffs from others' actions, which they illustrate with an example "A woman working in a 'man's' job may make male

colleagues feel less like 'men.' To allay these feelings, they may act to affirm their masculinity and act against female co-workers." In such cases, an action by others, a_{-j} impacts I_j, which may be countered by an action a_j. Further, they note that because of j's identification with other people, it may affect j's identity when another person in j's social category violates prescriptions. Scorn or ostracism are common responses and serve to re-affirm one's own self-image and to distance oneself from the deviant (Akerlof and Kranton 2000). This model would explain large externalities when a program's arrival threatens prevailing gender identities. It would also predict gossip, and reversion to old behaviors for individuals exhibiting new behaviors that threaten others' identity.

2.2 Related Literature

There is significant demand from governments and NGOs for interventions that can effectively prevent IPV. Two of the most common types of prevention interventions, primarily in lower-income countries, are designed to change conservative gender attitudes and social norms that condone violence. They include small group workshops for men *or* women, and large-scale 'edutainment' or community campaigns. The former are designed to challenge existing beliefs, promote reflection and debate, and encourage collective action (Green et al. 2015; Gupta et al. 2013; Hossain et al. 2014; Kerr-Wilson et al. 2020; Vaillant et al. 2020; Doyle et al. 2018). The latter programs typically attempt to use compelling narratives or media messages to promote community 'change agents', train influential community leaders, or use public screenings of edutainment videos to change attitudes and norms (Banerjee, La Ferrara, and Orozco 2019; Green, Wilke, and Cooper 2018; Abramsky et al. 2014). Both types of strategies have sometimes resulted in modest changes to gender attitudes and, in rare cases, small reductions in women's self-reported experience of IPV.

Learning from the relatively disappointing primarily null impacts of men-only and women-only IPV prevention programs on women's experience of IPV, recent interest has grown in interventions that try to change dynamics *within* couples rather than men or women separately. Potentially, couples-based programs may be effective where programs only engaging with women may have been hampered by their inability to address the drivers of men's behavior, and where programs only engaging with men may have been unable to challenge them to truly reflect on power imbalances and to develop the necessary understanding and empathy necessary for attitude and behavior change.

Couples-based programs thus work with groups of existing couples and seek to critically engage men and confront gendered power dynamics that may constrain women's empowerment. These programs seek to build an understanding of power, communication and conflict resolution skills, and shift gender attitudes and behaviors within couples, reinforcing them through weeks of facilitated group activities, discussion, and deliberation. Couples are then encouraged to model and advocate for more gender equal, violence-free relationships in their

communities. Couples-based IPV prevention programs are now being tested or scaled up by governments and NGOs in Sub-Saharan Africa. Yet there is limited evidence on their effectiveness and the mechanisms through which they work; we found no evidence that they may cause a negative IPV backlash.[10] In this paper, we provide some of the first rigorous evidence on the effectiveness of a couples-based IPV prevention program.

2.3 Country context

Rwanda has undergone enormous social upheaval over the 25 years since the genocide, including a radical social shift focused on women's status and empowerment. The enshrining of gender equality in law over the past two decades rapidly overturned patriarchal institutions and signifies a divide between the country's more conservative historical gender norms and more progressive modern policy and law led by the centralized state.

Under the colonial period and with the influence of the church, power was centralized in institutions dominated by men, limiting women's political power and reinforcing patriarchal gender norms where women were subservient to men (Longman 2006 in Asiedu et al. 2018). Current attitudes and behaviors still broadly reflect this patriarchal history. Violence against women is widely tolerated, and women experience high rates of IPV. For example, a Rwandan proverb states that "a woman who is not yet battered is not a real woman" (Human Rights Watch 1996 in La Mattina 2017). In Rwanda's most recent nationally representative Demographic and Health Survey (DHS) conducted in 2014-15, 34% of women report having experienced physical and/or sexual IPV in her lifetime, and 21% of women reported experiencing it in the previous 12 months. This was down from rates reported in the 2010 DHS, when 56% reported lifetime IPV, and 44% had experienced it in the preceding 12 months.[11] Social norms that condone IPV are also widespread, with 41% of Rwandan women and 18% of men believing that violent behavior by a husband towards his wife is justified under specific circumstances (Mannell et al. 2018). These IPV rates and IPV-supportive attitudes are higher in rural areas.

[10] There are two studies of similar couples-based IPV prevention programs, both in Rwanda. A study of a parenting and IPV prevention intervention for new parents found that the 15 session couples discussion group program caused a significant reduction in IPV (Doyle et al. 2018). However, in this study the control group comprised individuals in the same village as treated couples, which may result in biased impact estimates if there were within-village spillovers. The second study, a Rwandan couples training and community activism IPV prevention program, called '*Indashyikirwa 1*', which was the basis for the couples curriculum evaluated in this paper, was recently evaluated using a randomized trial design with a synthetic control group (Chatterji et al. 2020). The researchers found that the multi-component program was associated with large and statistically significant reductions in IPV for couples who went through the training program. However, there are several significant programmatic, study design, and analytical differences between our study and that of *Indashyikirwa 1* that mean we cannot directly compare results. The differences are discussed in more detail in the intervention section below. Some key design and analytical differences include, for the *Indashyikirwa 1* study, a different non-blinded approach to recruiting the couples into the 14 treatment and 14 control sectors (with resulting differences in baseline IPV prevalence), potentially biasing results, and standard errors not clustered at the unit of randomization.

[11] DHS data on IPV is collected using face-to-face interviews with a data collector.

In parallel, over the past two decades, the Rwandan government has been at the forefront of global efforts to promote gender equality. After the 1994 genocide, when women comprised 70% of the population, women increasingly took on traditionally male roles and women's organizations promoted female representation in the interim government of 1994 (Hansén 2017). Since then, the Rwandan Government has expressed its commitment to gender equality and has introduced numerous legal and policy reforms to help realize it. In 1999, the government legislated that women could own and use land on equal status with men (Polavarapu 2011). Since 2003, when the country introduced a 30% gender quota for female civil servants and parliamentarians, the country has surpassed this objective and now has the world's highest rate of female political representation in the world. In 2009, the government passed an equal pay for equal work law and a comprehensive Gender Based Violence Bill which obliges all Rwandans to report violence: 'Any person must prevent gender-based violence, rescue and call for rescue the victims of this violence'. Penalties for perpetrating IPV, for example, 'harassing one's spouse' and 'conjugal rape' is imprisonment for 8 months to 2 years (Mannell et al. 2018). In more recent legislation, the government overturned the legal provision stating that men are the heads of household (Abbott and Malunda 2016). The Government has also funded one-stop centers to support IPV survivors with specialized shelter, medical, legal and police support services. And at compulsory monthly community meetings for all citizens, gender-based violence is often discussed. Women can now also inherit property, receive alimony on divorce, and get equal rights as if married even when they have not undergone an official civil marriage ceremony. These progressive legal reforms have been mirrored in health and education policies for women and girls, which have resulted in a significant decline in maternal mortality and the unmet need for contraception, and a narrowing of the gender gap in educational attainment (DHS 2016).

In spite of these many achievements in progressive gender reforms, there is an apparent gulf between national laws and policies, and the aforementioned beliefs, attitudes, and behaviors of many, particularly in rural areas. Gender norms about women's subservience to men appear to have remained intact. In our sample of rural couples, many people either are unaware of the significant changes to gender law, or they believe the new laws are out of step with society. For example, only 24% of men and 12% of women in the sample for this study know that husbands and wives are equal heads of household under Rwandan law, and 98% of women agreed with the statement "Rwandan laws make it too easy to bring a violence charge against a man."

2.4 Intervention

Rwanda's Ministry of Gender Equality and Family Promotion (MIGEPROF) worked with the NGOs CARE, the Rwandan Women's Network, and Rwandan Men's Resource Centre (RWAMREC), to implement a holistic IPV prevention intervention. The key component of

the intervention, a five-month couples training, was complemented by two other components (not evaluated in this study): a condensed week-long training of selected local leaders on gender equality, and the opening of a number of safe spaces to provide information or support for those wanting to report abuse or access services. The leaders training and safe spaces were available in both treated and control areas and are therefore not assessed in this study. The intervention and the couples training curriculum were based on two earlier IPV prevention programs, Journeys of Transformation developed by Promundo, CARE Rwanda, and RWAMREC (Slegh et al. 2013), and also *SASA!* (Abramsky et al. 2014), initially developed in Uganda. These programs were adapted for the rural Rwandan context through extensive testing and piloting (Stern and Nyiratunga 2017).[12]

The intervention involved delivering 22 weekly group workshops over five months for 640 couples in 49 treated villages. Couples who were members of Village Savings and Loan Associations (VSLAs) that had been selected for the study were invited to volunteer for the couples training. Those who were then randomly selected to participate met weekly for 3-4 hour group discussions with, on average, 19 other selected VSLA couples from their own or nearby treated villages. Each group was led by two trained facilitators, one male and one female, from the Rwandan NGO RWAMREC.

[12] The curriculum adaptation, testing and piloting was funded by the UK Department for International Development and conducted by CARE Rwanda, RWAMREC, and the Rwanda Women's Network with the support of international IPV specialists and a research team affiliated with the What Works to prevent violence consortium (Stern and Nyiratunga 2017). It was informed by early insights from Promundo and Slegh et al. (2013). From 2014-2018, a research team led a study of the complete package of components: opinion leaders training, women's safe spaces, couples training, and couples' activism, with results reported in Chatterji et al. (2020) and Dunkle, Stern and Heise (2020). This first delivery of the curriculum, and the holistic package, was called *Indashyikirwa 1*. The study design of *Indashyikirwa 1* involved randomization at sector level (2 administrative units above villages), with 14 treated sectors receiving all four components, and 14 control sectors receiving none. In this paper, we present the results of a second study, of *Indashyikirwa 2*, which focuses only on evaluating the couples training component. The curriculum was adapted from lessons from the development, piloting and implementation of *Indashyikirwa 1*. Given the popularity of becoming a community activist observed in *Indashyikirwa 1*, and in order to test a potentially more cost-effective model of delivering community activism, the curriculum was altered slightly from *Indashyikirwa* 1 by condensing the community activism curriculum into 2 additional sessions delivered to all trained couples.
In addition to the above key design differences in the interventions of *Indashyikirwa 1 and Indashyikirwa 2,* as highlighted above, there are also differences between our study designs and analysis that preclude a direct comparison of results. Here, we summarize the key study differences. First, the couples sample in Dunkle, Stern and Heise (2020) was selected using a non-blinded recruitment process where couples were recruited to participate in a study (control sectors) or the intervention (treatment sectors). The resulting groups are different on a number of observables, including baseline IPV prevalence. This selection could bias treatment effect estimates given potential differences in unobservables. Second, in addition to measuring impact on couples who attended the training, the study also measured spillovers of the intervention to surrounding community members. Their representative community sample, unlike the couples sample, was recruited in the same way in the treatment and control sectors and thus is less likely to be biased. Chatterji et al. (2020) document a suggestive negative spillover effect (an increase in IPV) on the community sample in *Indashyikirwa 1*, which accords with the negative spillovers we observe in our study. In their community sample, the program caused an increase in sexual IPV (p=0.06) and caused suggestive but not statistically significant increases in emotional and physical IPV (p=0.43 and 0.13 respectively) for the 'spillover' random community sample who did not participate in the couples program (Chatterji et al. 2020). However, a further methodological difference is that standard errors are not clustered at the unit of randomization- the sector- as is considered appropriate for clustered randomized trial designs in economic research (Abadie et al. 2017). The authors note they do not cluster standard errors at the unit of randomization citing statistical power concerns given the intervention had 14 treated sectors and 14 control sectors. However, this risks violating the stable unit treatment value assumption if both the treatment regressors and the errors are correlated within clusters. If standard errors were clustered at the sector level, p-values would likely increase.

The sessions involved facilitated discussions and reflection on concepts including types and uses of power; gender and sex; rights; communication and conflict resolution skills; the benefits of non-violence and gender equality; managing triggers of violence including alcohol use; gender roles in the household; foundations of healthy relationships; shared decision-making; and positive parenting. Towards the end of the curriculum, couples were introduced to activism. The curriculum covered supporting IPV survivors in the community to minimize their shame or stigma, and encouraging couples to identify tangible actions to become 'agents of change' promoting and modeling gender equitable relationships in their communities. Couples were given take-home activities after each session to practice and apply what they had learned. The 22 sessions are listed in Figure A2. Over several years, the curriculum had been developed, piloted, and implemented during *Indashyikirwa 1* before being revised and updated based on lessons learned. In 2018, the revised version, *Indashyikirwa 2,* which included additional sessions supporting all couples to be activists, and a session focusing on positive parenting, was delivered by the same NGOs for the Government of Rwanda.

Couples were provided with a nominal stipend to cover the cost of attendance and implied foregone earnings for attending the workshop (2,500 RWF or approximately $2.5 USD per individual per session attended). Given the low formal employment in the rural sample, this amount was likely to be greater than compensation for costs related to participation.[13] An independent local survey company visited the couples training sessions and collected individual attendance data for each participant.

In applying the identity model to our setting, the IPV prevention couples training program could represent an attempt to change three factors: the social category one chooses to be part of (e.g. a progressive man or woman), the relative status of social categories (e.g. improving women's relative social status), and the prescriptions associated with the social categories of an ideal man or woman (e.g. 'a man undertakes housework and makes decisions jointly with his spouse', and 'a woman can also initiate sex with her partner and decide when she wants to use contraception'). If successful, a policy that achieved such changes would decrease men's (and women's) losses in identity from the real or perceived threat to identity posed by women's empowerment or men's more egalitarian behavior. Such a program may decrease the identity loss for women and men when men take on traditionally 'women's work,' when men spend more time at home and less at the bar, and when women violate prescriptions by initiating or refusing sex with her husband. Ultimately, if successful, such programs may lead to a virtuous cycle of behavior change in trained individuals and positive externalities in their communities, and a reduction in IPV in equilibrium. This could represent a scalable and cost-effective IPV prevention program. However, if such

[13] The anticipation of payment, while not discussed or promoted in the sample recruitment sensitization lottery sessions, potentially increased interest in the intervention and led to over-subscription. In multiple instances, people who were not genuine couples tried to sign up for the program with someone other than their spouse, indicating that there was high interest in the program and the promise of likely financial incentives increased this interest. We used several checks to identify non-authentic couples at several stages- the sensitization/application stage, baseline, during the couples training, and endline.

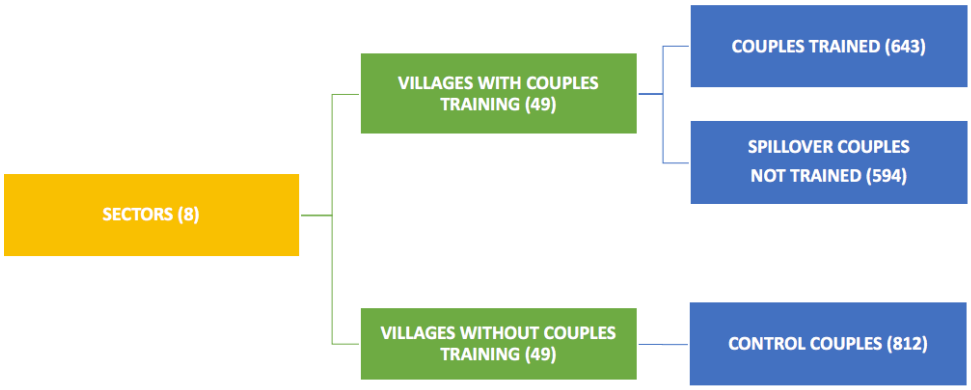
programs do not successfully change gendered social categories and prescriptions, or if the changed behaviors of program participants provokes a negative reaction in others, then the identity loss could provoke an IPV backlash in both participants and the broader community.

2.5 Experimental Design and Sampling

The intervention used VSLAs as the entry point to identify couples.[14] Trained couples were selected from existing members of CARE VSLAs, each of which is associated with a village near a member's home. Most VSLA group members live in the same village as the VSLA meeting point, though some members live in nearby villages. Villages are located below the administrative unit of cells, sectors, districts, and provinces. This study took place within Rwanda's Eastern Province.

In order to measure spillovers to non-trained couples living in the same village, randomization took place at two levels: first, villages within the sample were assigned to treatment or control groups. Within treated villages, couples were then randomly assigned to treatment or control groups. In the rest of this paper, we refer to the three types of couples as: control couples (couples from control villages), spillover couples (control couples in treatment villages), and treated couples (couples in treated villages assigned to get the training).

Figure 1: Experimental Design



The sampling frame was chosen as follows. Within Eastern Province, and in partnership with province and district authorities, the Government of Rwanda selected four districts and eight sectors understood to have high rates of IPV and with a high concentration of VSLAs that the intervention could be delivered through. This selection was done using CARE's data on VSLA activity, and MIGEPROF's consultation with provincial authorities on IPV prevalence. Within these 8 sectors, existing CARE VSLAs were identified within 128 villages. From these 128 villages, the researchers ruled out 30 villages with fewer than

[14] The VSLA entry point was originally in recognition of the need for a program to address the issue of financially controlling behaviors exhibited by some husbands of female VSLA members (Stern and Nyiratunga 2017).

3 VSLAs, and villages that had VSLA meeting points that were very close to the VSLA meeting points of another village. This was done to increase the distance between control and treatment villages, minimizing the risk of contamination. There is no rule preventing residents of one village from participating in VSLAs based in another village. This could lead to non-compliance with village-level treatment status, where a respondent sampled in a treated village VSLA actually lives in a control village. We provide descriptive statistics on compliance and relevant robustness checks later in the paper.

In the selected eight sectors and 98 remaining villages, VSLAs were then selected for inclusion in the sampling frame. In villages with three VSLAs exactly, all VSLAs were selected. In villages with more than three VSLAs, VSLAs in the village were ranked by the number of members, and the three VSLAs with the greatest number of couples were selected for the sample, to maximize the probability of having sufficient subscription to the intervention.

All members of selected CARE VSLAs were then informed about the training opportunity and invited to a sensitization session to get basic details on the intervention described as a group training on "how to improve relationships in the family" and then invited to participate in public lotteries with other interested couples in their VSLA. Importantly, this was carried out in the same way in what we later identified as treatment and control villages, ensuring that respondents were blind to their treatment status during recruitment and the baseline survey.

To apply for the couples training, participants in the VSLA lotteries had to fulfill the following eligibility criteria:

1. Married or cohabiting with a partner and been married or cohabiting for at least 12 months.
2. The participant or his or her partner is an existing member of the VSLA and has been for at least 3 months.
3. The participant and his or her partner are at least 18 years old.
4. The participant and his or her partner are willing and available to participate in the training, and to commit to attending the 22 weekly sessions.
5. The couple has not applied to another VSLA lottery.

Based on the sampling frame of sectors, villages, VSLAs, and couples, the researchers conducted the random assignment of the villages and couples per the following process. The number of couples to be treated was determined in advance by MIGEPROF in consultation with the research team based on the project budget (640 couples). Public VSLA lotteries were held in all sample VSLAs so that all interested, eligible couples were given a lottery ranking from 1 to a maximum of 30 (the maximum number of members in any VSLA). The research team used Stata to randomly assign study villages to treatment or control groups.

All VSLAs within treatment villages were then treatment VSLAs, and within control villages, all VSLAs and all couples were 'pure controls.'

Based on their lottery ranking, the research team assigned couples within the treatment villages to either treatment or control groups, with the couples drawn first from the public lottery first assigned to treatment. The first 60% of couples from each VSLA who were drawn from their VSLA lottery were assigned to treatment, reaching the maximum of 640 treated couples. The treatment assignment was stratified on the sector and the number of applicants in the lottery. Within the 98 study villages, the final sample included 324 VSLAs and 2,042 couples.

In order to assess the external validity and targeting of the couples training program, a household listing of all the couples in the 98 study villages was collected and a representative sample of 686 couples was randomly selected and interviewed in the baseline survey. A comparison of this community sample and the sample of VSLA applicants suggests that VSLA couples are very similar to randomly selected non-VSLA couples on key observables, though VSLA couples appear marginally younger, more educated, more economically vulnerable, and slightly more conservative on gender issues (see Table A2 in the appendix).

3. Data and estimation

3.1 Data collection

Data collection included two rounds of panel surveys: a baseline survey from November 2017 – February 2018 prior to the start of the intervention, which was delivered from May – August 2018. The endline survey took place from February – March 2019, six to eight months after the intervention.[15]

Of the 2,042 study couples interviewed at baseline, 98.3% of women and 97.5% of men were re-interviewed at endline. The attrition rate did not differ by treatment arm. The main reasons couples could not be re-interviewed at endline were that they or their spouse had passed away (10 respondents), they could not be found, they had moved to another country, or they were in jail (6 male respondents). Figure A1 in the appendix shows the couples sample composition and attrition by treatment arm. In addition to the couples survey, village leaders from each 5-person village leadership committee were interviewed about their gender attitudes, village infrastructure, and awareness of the program.

Women's and men's surveys covered numerous topics including household demographics and socioeconomic characteristics, time use, economic activity, household decision-making,

[15] A World Bank retrospective qualitative process review took place in January 2020 to collect information on the intervention implementation.

wellbeing, gender attitudes, beliefs about community gender norms, experience of IPV (women and men), likelihood of becoming violent in hypothetical scenarios (men), use of violence against children, knowledge of Rwandan gender law, and respondents' social networks. Respondents also played incentivized behavioral games to understand household cooperation (using a spouse dictator game) and baseline norms (using the Krupka and Weber (2013) coordination game). The surveys were administered by enumerators of the same sex as the respondent, in private settings where interviews could not be overheard. Most surveys were conducted in empty school classrooms, village offices or halls, church buildings, or other private areas.[16] At endline, we also measured men's and women's height, weight and women's middle-upper arm circumference. The baseline consent form made it clear that refusal to participate in the survey did not affect respondents' eligibility for the intervention.

3.2 Violence measurement

IPV was measured using the international WHO Violence Against Women instrument and administered following the WHO protocol on ethical guidelines for conducting research on women's experience with IPV (WHO 2016). The ethical protocols we adopted are outlined in Appendix B. The standardized WHO instrument, the most common IPV data collection instrument, asks multiple behaviorally specific questions on a range of abusive acts, a technique shown to maximize disclosure (Ellsberg et al. 2001). Following the approach in the literature, we focused on three types of IPV: women's recent experience of emotional (2 questions), physical (7 questions), and sexual violence (3 questions). At endline, for each act of violence, women were asked if their current husband had done this in the last 6 months (to avoid overlap with the end of the intervention). At baseline, the recall period was 12 months. Women were asked how often their current husband had perpetrated each type of violence over the past six (or twelve) months, with answer options being 'never', 'once', 'a few times', and 'often'.[17]

From these questions, we created three binary indicator variables for the three primary outcome measures, as specified in our pre-analysis plan registered on the AEA Trial Registry (Alik-Lagrange et al. 2017). Indicator variables were coded as one if the woman reported experiencing at least one act within each category of violence. If a woman refused to answer one of the violence act questions but had answered yes to another act within that same category of violence, the binary indicator was coded as one. If she refused one question and

[16] When classrooms were not available, surveys were conducted in smaller offices, empty churches or outside but strict privacy protocols were adhered to, including the requirement that interviews pause if someone else was within earshot or eyesight.

[17] Examples of a question from each category include, for emotional violence: "how often has your husband said or done something to humiliate you in front of others in the past 6 months"; for physical violence: "how often has your husband punched you in the past 6 months"; and for sexual violence "in the past 6 months, how often has your husband physically forced you to have sexual intercourse with him even when you did not want to."

all other answers within that violence category were either zeros or refuse, the overall indicator was coded as missing.[18] In the appendix, we show that our main results are robust to coding all refuse-to-answer as 'yes' and all as 'no'.

To increase privacy and minimize the risk of reporting bias, we measured the gender attitude and IPV questions using a survey method called audio-computer assisted self-interview (ACASI). ACASI allows respondents to listen to the pre-recorded questions through headphones and respond by selecting their answers (for example, the corresponding colored square) on the touchscreen of a tablet. After demonstrating how to use the technology, enumerators stand a small distance from the respondent and ensure that privacy is maintained. ACASI is designed to minimize reporting bias by increasing privacy and reducing the embarrassment respondents may feel about disclosing a sensitive answer face-to-face with an enumerator (Cullen 2020). [19] The most sensitive survey questions were asked using ACASI, including: the primary IPV questions (about women's experience of IPV, as well as two questions asked of men about IPV against men), men's likelihood of becoming violent in hypothetical scenarios, men's alcohol consumption, and most of gender attitude questions. ACASI modules were tailored for a low literacy and numeracy population and tested through several pilots.

We also measured impacts on a range of secondary outcomes that are suggestive of mechanisms. For the conceptually-linked outcomes measured with a number of individual survey questions, we constructed summary indices as the most efficient weighted average of a set of outcomes, following the procedure proposed by Anderson (2008), and outlined in our pre-analysis plan. This addresses the problem of multiple inference by reducing the number of hypotheses tested. For others, where most intuitive or appropriate, we either summed the number of statements respondents 'agreed' with in a category (for example, opinion about the number of situations in which a man beating his wife would be acceptable), or we report the raw scores on the individual survey question.

3.3 Data description

Baseline sample characteristics and balance across the treatment groups are shown in Table A1. Baseline balance across intervention arms was achieved for key observable characteristics, with no normalized differences between arms greater than 0.25 standard deviations, a suggested rule of thumb indicating good balance (Imbens and Rubin 2015). The average

[18] A coding error meant that one emotional and one physical violence question asked at baseline was only asked of a third of endline respondents, so this question is dropped when creating the emotional and physical violence endline dummies. However the baseline question- correctly asked of all female respondents- is included in the respective baseline violence dummies to improve precision of our ANCOVA treatment effect estimates given its autocorrelation with endline variables.

[19] Cullen (2020) discusses ACASI's implementation during baseline in more detail, and finds that ACASI does result in marginally greater reporting of violence compared to the standard face-to-face survey method, though still much lower prevalence than under the list experiment. Given the list experiment's greater variance, we chose not to use it to measure the outcome variables in this study given concerns about being underpowered to detect treatment effects.

age of the men and women in the sample is 40 and 36 years respectively. Men have an average of 4.7 years of education compared to 4.5 years for women. Half of the sample is in a legal marriage (the rest cohabit and live as if married). The average number of children is 3.7. Reported IPV at baseline was high. In the 12 months preceding the baseline survey emotional, physical and sexual IPV prevalence was 49%, 40% and 42% respectively.

3.4 Program take-up and compliance to treatment assignment

The intervention had high take-up, both on the extensive and intensive margins. Given project budget limitations, the Government was only able to treat 60% of the couples who applied through the public lottery to part of the program, suggesting high demand. Further, 95% of couples who applied for and were given a place in the couples training attended the training. The average number of sessions that enrolled couples attended was 21.3 out of the 22 total sessions.

While there was nearly full compliance to treatment assignment at individual level, information on households' dwelling location revealed that some couples assigned to control villages actually lived in treatment villages, and that some spillover and treated couples were not resident in the treatment villages. This resulted from some couples joining VSLAs associated with a particular village, but living in nearby villages (as noted above, registration to the lotteries was based on VSLA membership). In Rwanda's context of high population density and no formal restriction on VSLA participation dependent on dwelling location, these cases were anticipated, and we allowed for maximal feasible buffer zones between treatment and control villages at the sampling stage. As a result, we minimized the extent of this partial compliance.[20] We provide robustness checks to examine the implications of this issue, and show that results hold, both in magnitude and precision.

3.5 Econometric specification

We exploit the two-level random assignment of the couples and villages to identify the direct impact on the treated couples, and the impact on spillover couples. We estimate intention-to-treat (ITT) effects on our primary outcomes using an ANCOVA model. This approach adds precision and controls for potential imbalances (McKenzie 2012)[21]:

$$Y_{i,v}^{Post} = \theta Y_{i,v}^{Pre} + \beta T_{i,v} + \delta S_{i,v} + X'_{i,v}\,\gamma + \epsilon_{i,v}$$

Where $Y_{i,v}^{Post}$ and $Y_{i,v}^{Pre}$ are the baseline and endline outcome observations for individual i in village v, $T_{i,v}$ is an indicator variable equal to 1 if the couple was assigned to treatment,

[20] For more than 70% of the pure control couples, no treated couple lives within 600 meters of their dwelling, a sizable distance considering the high number of valleys, hills and lakes in the sectors of the study. This share goes up to approximately 90% for a radius of 200 meters.

[21] The specification and the variable definitions were described in a study pre-analysis uploaded on the RCT registry AEARCTR-0002282 prior to the research team receiving treatment-identifying endline data.

$S_{i,v}$ is an indicator variable equal to one if the couple was assigned as a spillover couple, as associated with couple i's VSLA in village v, and $X_{i,v}$ is a vector of baseline or time invariant controls and randomization strata fixed effects. In all specifications shown below, we present a more parsimonious model without any pre-experiment controls other than baseline IPV and strata. Our identifying assumption, $E[\epsilon_{i,v} \left| Y_{i,v}^{Pre}, T_{i,v}, S_{i,v}, X_{i,v}^{'} \right] = 0$, is motivated by the random assignment of villages and couples, and is supported by the balance tests shown in Table A1.

In this model, β is the estimated ITT effect on treatment couples, and δ the treatment effect on spillover couples. To test whether β and δ are equal, we conduct Wald tests of equality and report the p-values. Standard errors of the estimates are clustered at the village level to reflect the level of randomization. For the three primary IPV outcomes and treated and spillover coefficients, we account for multiple hypothesis testing by following the procedure proposed by Benjamini and Hochberg (1995) and presented in Anderson (2008). For some mechanisms and secondary outcomes only measured at endline, we estimate ITT impacts using a single difference model.

4. Results

4.1 Primary outcomes: Large, unintended increases in IPV

We find that six months after the program ended, the intervention had substantially increased women's self-reported experience of IPV. For treated couples, the intervention caused a 7 percentage point, 5 percentage point and 10 percentage point increase in emotional, physical and sexual violence, respectively. For spillover couples, the intervention caused a 7 percentage point increase in emotional violence, and even larger increases in physical and sexual violence, of 11 and 17 percentage points respectively. Relative to the control group, these impacts represent increases in emotional, physical and sexual IPV for treated couples of 11%, 15% and 24% respectively. For spillover couples, the increase is equivalent to 11%, 33% and 40%. The differences in impact between treated and spillover couples are statistically significant at the 1% level for both physical and sexual violence, confirming the program had larger negative effects for spillover couples. These effects are shown in Table 1, with women in the treated group in Column 2 and the spillover group in Column 3.

We also find negative impacts on other violence-related outcomes, including an increase in husbands' controlling behaviors (for spillover women), an increase in women's reported injuries from sexual violence; an increase in the perception of number of IPV incidents in the village; and a very small increase in the share of spillover women where enumerators see physical signs of injury. IPV against men also increased. Trained men are 8 percentage points more likely to report that their wife emotionally abused them in the preceding 6 months, while spillover men are 7 percentage points more likely to report that their spouse

used physical violence against him. These effects are shown in Table A3 and Table A4 in the appendix.

Table 1: Treatment effects on primary IPV outcomes

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple: p-value	(5) N
Emotional violence	0.66	0.07***	0.07***	0.76	1,974
		(0.03)	(0.02)		
		[0.01]	[0.01]		
Physical violence	0.33	0.05*	0.11***	0.01	1,960
		(0.02)	(0.02)		
		[0.05]	[0.00]		
Sexual violence	0.42	0.10***	0.17***	0.01	1,972
		(0.03)	(0.02)		
		[0.00]	[0.00]		

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in (rounded brackets). Sharpened q-values are shown in [square brackets]. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of direct treatment and spillover effects, and differences between effects on trained and spillover couples. Outcomes are listed on the left. Column 1 shows the control group mean. Column 2 shows average treatment effect, standard errors and sharpend q-values calculated between treated couples and pure control couples. Column 3 shows average treatment effects for spillover couples compared to pure control couples. Sharpened q-values were calculated to adjust for multiple hypothesis testing using the Bonferroni and Hochberg correction method (1995). Column 4 reports p-values for the difference between effects on trained and spillover couples. The unit of observation is individual level, and is calculated using a dummy aggregating whether a wife reported experiencing any of each of the forms of emotional, physical or sexual violence. We control for baseline emotional, physical or sexual violence in the associated ANCOVA regression, as well as a dummy for whether the baseline response was missing.

4.2 Secondary outcomes: Worse impacts for spillover than treated couples

We also consider the intervention's impact on a number of secondary outcomes related to relationship quality, wellbeing, women's decision-making, and economic indicators. Most of these statistically significant impacts are negative, and overall are worse for spillover couples than treated couples. For example, marital satisfaction decreases for spillover men and women (no impact on treated couples). There is no impact on overall relationship quality or women's input in decision-making. There is an increase in spillover men's and women's depression, and men drink alcohol to excess more often. There is an increase in husbands' infidelity, as reported by spillover women. Further, both treated and spillover women are more likely to report that their last marital disagreement escalated into physical violence, and that their partner was drunk before their last quarrel. The exceptions to the generally negative results are a slight increase in treated men's knowledge of Rwandan gender-related laws and in the number of tokens they gave their wife in an incentivized dictator game. These effects are shown in Table 2, and Table A5 and Table A6 in the appendix.

Table 2: Treatment effects on other wellbeing and relationship quality indicators

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Wellbeing and relationship dynamics					
Women: overall satisfied with marital life	0.95***	-0.02	-0.04***	0.11	1,987
	(0.01)	(0.01)	(0.01)		
Men: overall satisfied with marital life	0.98***	0.00	-0.02**	0.01	1,980
	(0.00)	(0.01)	(0.01)		
Women: relationship quality - standardized summary index	-0.01	-0.02	-0.08	0.31	2,042
	(0.04)	(0.06)	(0.06)		
Men: relationship quality - standardized summary index	-0.01	0.01	-0.08	0.13	2,042
	(0.04)	(0.06)	(0.06)		
Women: depression - standardized summary index	-0.00	-0.02	0.10**	0.27	2,042
	(0.03)	(0.04)	(0.04)		
Men: depression - standardized summary index	-0.01	0.03	0.14***	0.00	2,042
	(0.02)	(0.04)	(0.04)		
Women: tokens given to husband in dictator game (out of 10)	7.09***	0.12	-0.07	0.88	1,976
	(0.13)	(0.18)	(0.18)		
Men: tokens given to wife in dictator game (out of 10)	7.55***	0.43**	0.04	0.14	1,978
	(0.13)	(0.17)	(0.18)		
Men: drink alcohol to excess sometimes	0.34***	0.05*	0.03	0.40	2,042
	(0.02)	(0.03)	(0.03)		
Women: Husband drinks alcohol sometimes, often or frequently	0.24***	0.03	0.08***	0.03	2,007
	(0.02)	(0.09)	(0.08)		
Women: last marital disagreement escalated into physical violence	0.11***	0.09***	0.07***	0.00	1,978
	(0.01)	(0.02)	(0.02)		
Women: partner was drunk before last quarrel	0.35***	0.05*	0.07**	0.01	1,965
	(0.02)	(0.03)	(0.03)		
Women: partner cheats on you	0.12***	0.02	0.07***	0.02	1,734
	(0.01)	(0.02)	(0.02)		
Women: Husband spent money set aside for household on himself	0.37***	0.06**	0.09***	0.44	2,042
	(0.02)	(0.03)	(0.03)		
Women: input in decision-making - standardized summary index	-0.00	0.09	-0.10	0.01	2,042
	(0.04)	(0.06)	(0.06)		

Overall, there are very few economic impacts, though both treated and spillover couples reported a slight increase in the total number of assets the household owns, as shown in Table A5. There is no statistically significant impact on women's share of household assets, or on men's or women's total savings, though the savings point estimates imply a potential increase in savings.[22]

Table A7 shows impacts on respondents' beliefs about the gender attitudes and behaviors of others in their community. Men and women were asked what share of men in their community hold inequitable gender attitudes and perpetrated IPV. These 'descriptive

[22] This is notable considering the $2.5 per person per week stipend transferred to trained couples and the fact that 42% of women and 33% of men reported saving some of it, 20% of women and 23% of men reported using it to buy livestock, and zero people reported lending the money to someone else. Fewer than 15% of men or women reported using the money to fund their transport to the training location. However, this stipend size is less than a standard small cash transfer, including the one documented in McIntosh and Zeitlin (2018) that had minimal economic impacts, so minimal impact in our setting is perhaps not surprising.

norms', or second order beliefs, changed for all groups: treated and spillover men and women. The intervention caused treated women and men to believe that fewer men think wives should obey husbands even if the wife disagrees- a positive sign. However, the intervention also *increased* all groups' beliefs about the pervasiveness of physical and sexual IPV in their communities. This community-wide updating in the belief that IPV is widespread may indicate that individuals are updating their beliefs after they observe or otherwise know that IPV has increased, or it may reflect greater awareness and salience of the issue generated by the program's presence in their village.[23] We also find that treated and spillover men believe that general non-IPV physical violence has increased in the past 6 months, and that there have also been more instances of physical IPV in their village in the past 6 months (Table A4). Together, these outcomes support the conclusion that in treated villages, there is a broad understanding that IPV is more common in the community.

5. Mechanisms

5.1 Women's progressive aspirations and attitudes, and men's challenged identity

In this section, we discuss the qualitative and quantitative evidence on potential mechanisms that may be driving the increase in IPV. We conclude the section by discussing robustness checks on our primary results.

Suggestive evidence indicates that the IPV increase may have been caused by a village-wide male backlash against the real or perceived threat to traditional gender identities posed by the program's messages about gender equality. These gender equality messages, formalized in Rwanda over the past 15 or so years through legal changes, are far more progressive than the gender attitudes held by many individuals. The backlash against such messages appears to have been partially mitigated in treated couples, with these men adopting some new progressive attitudes and behaviors encouraged in the training. Yet quantitative data and qualitative reports suggest how deeply the program's messages were felt and how widely they reached in communities.

Qualitative reports

Qualitative reports from an unpublished independent ex-post implementation review of the program suggest that, according to one facilitator, men "didn't like that their hold on power was being challenged. Women were excited that men's power was being challenged. They

[23] On average, all groups still substantially underestimate (or under-report their estimate of) the prevalence of IPV in their village, when we compare these beliefs to the average share of women who report experiencing IPV in our sample, including the IPV mean in our baseline representative village sample.

felt like they were finally learning the reality of the situation" (World Bank 2020).[24] The curriculum started by engaging people in a discussion on power, and on the differences between biological sex and cultural prescriptions of gender. Facilitators noted "Most of the couples thought that gender and sex were the same thing. Once people realized the difference between gender and sex, it changed the dynamics of the class, and discussions tended to improve." Another noted "It was difficult for men to accept that gender is not biological, that it's culture and patriarchy that give more power to men." Facilitators observed men's resistance to the idea of shared domestic responsibilities because it would violate 'natural' prescriptions of men's and women's tasks: "Men would say, 'In our culture, it is the man who pays the [bride price]. Are you now saying that women can also pay the [bride price]?'"

Qualitative reports also suggest that through the program, some trained men did eventually change their attitudes and adopted new behaviors encouraged in the curriculum, such as taking on some domestic or care work or reducing their drinking at bars. However, facilitators noted that some men who did so were shamed or became the subject of rumors that they had taken on "women's work" and "were not real men" and had been bewitched by their wives. According to the training facilitators, this made other men wary of changing, and made early adopters of new behaviors revert back in order to avoid being shamed by their community. One program participant noted "If a man is very serious about the program, he can say, 'It's fine if my wife bewitches me. I don't care, as long as I don't die.' But most of the men reverted back to their old behavior and are doing rough things to their families."

Reports also suggest how such impacts may have become so widely disseminated in communities. Facilitators and leaders noted that in general, men were more resistant to the program, and that many women in the community observed changes in their peers who were involved in the program and wanted to know how they could become involved. Further, in treated villages, village leaders encouraged couples participating in the training to speak about what they learned at their villages' regular compulsory community-wide meetings, as well as informally with their friends and religious communities.[25] [26] Local government

[24] The rapid ex-post implementation review found that the program was broadly implemented as planned. It also made recommendations for ways to improve any future delivery of such programs, for example, increasing training and support to facilitators, and having smaller group sizes to make group numbers more manageable.

[25] Leaders and facilitators noted that couples shared information about the program during informal discussions at home with their neighbors, at the bar, in church, or at the market; during VSLA meetings; during weekly village level 'evening dialogues,' which all local residents above the age of 18 are required to attend and which local leaders often attend; during 'positive parenting' meetings at the village level, which all parents attend on a monthly basis; during compulsory whole-of-village *umuganda* community service gatherings held every month; and during *isibo* meetings, which are groupings of 20 families beneath the village level that meet on a weekly basis. In these community gatherings, couples most often shared testimonials about how changes to their own behavior—like making shared economic decisions—had improved their household finances and their marriages rather than focusing on communicating specific concepts like definitions of IPV (World Bank 2020).

[26] Rwanda has the continent's highest population density. In small villages with an average of ~120 households, some of which had up to ~25 households participating in the intervention, this was a visible program. 67% of spillover men and 69% of spillover women report personally knowing treated couples, and almost half of these spillover couples report talking to the trained couples about the content of the curriculum.

officials were very supportive of the program, potentially aided by the inclusion of a requirement that officials address IPV, in their *imihigo*, a Rwandan system of government performance contracts for all leaders. During village meetings, anecdotal reports suggest that some program messages of gender equality were met with public resistance, particularly the notion of a woman's right to refuse and initiate sex with her husband, which in some cases was so controversial that community members verbally spoke out against this notion in the whole-of-village meetings. One facilitator observed that in the community fora, men "often think that this thing of gender has just come to bring problems to the man and to take away his rights." Another facilitator noted possible consequences of participants introducing ideas from the program in community settings without adequate explanation: "When you start talking about [gender-based violence] in a local community and you don't adequately explain the issue, they'll think that you're just stripping us of our power and property. Local communities can [exact] revenge if they are introduced to these ideas without fully understanding them."

Quantitative results

These qualitative reports - suggestive of a mechanism of backlash against threatened identity, men's loss of power, and shaming of those who try to change behaviors - are also reflected in the quantitative data. Table 3 and Table 4 below and Table A8 in the appendix show results related to gender identities, threats to these identities, and related heterogeneous treatment effects.

First, we find that women in both treated and spillover arms report substantially improved gender attitudes and reduced acceptance of IPV. Treated women report the largest reduction in the number of gender inequitable attitudes held, and in the number of situations in which they believe wife beating is justified. Given treated women spent 22 weeks discussing issues related to gender equality, it is unsurprising that the impact on their gender attitudes is larger than for spillover women.

While treated men also report an improvement in these two types of attitudes, spillover men report the opposite, with an *increase* in the number of reasons they believe wife beating is justified, and a suggestive, but not statistically significant, increase in the number of gender inequitable attitudes they hold. This treated-spillover men divergence in impact is also evident in the number of hypothetical situations in which men say they would be likely to use IPV. Treated men say they would be likely to perpetrate violence against their wife in *fewer* situations than control men, while spillover men say they are likely to become violent in more.

Table 3: Backlash mechanisms: gender attitudes

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Backlash mechanism: Gender attitudes					
Women: No. of gender inequitable attitudes	1.83	-0.40***	-0.10**	0.00	2,008
	(0.02)	(0.04)	(0.04)		
Women: No. of situations IPV is justified	2.18	-0.60***	-0.23***	0.00	1,732
	(0.05)	(0.08)	(0.08)		
Women: No. of reasons a wife refusing sex is justified	1.9	-0.19***	-0.14**	0.46	1,815
	(0.04)	(0.06)	(0.06)		
Men: No. of gender inequitable attitudes	1.61	-0.27***	0.10	0.00	1,989
	(0.04)	(0.05)	(0.06)		
Men: No. of situations IPV is justified	0.82	-0.22***	0.30***	0.00	1,640
	(0.06)	(0.06)	(0.08)		
Men: No. of reasons a wife refusing sex is justified	2.03	-0.10*	-0.08	0.81	1,746
	(0.04)	(0.06)	(0.06)		
Men: No. of hypothetical situations he would be violent in	3.46	-0.41**	0.96***	0.00	1,990
	(0.16)	(0.19)	(0.23)		

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in rounded brackets. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of direct treatment and spillover effects, and differences between effects on trained and spillover couples. Outcomes are listed on the left. Column 1 shows the control mean; column 2 the average treatment effect for treated couples, column 3 the treatment effect for spillover couples. Column 4 reports p-values for the difference between effects on trained and spillover couples.

The only domain of attitudes where there is a worsening for all four groups of respondents- treated men and women and spillover men and women- was attitudes regarding a wife's right to refuse sex with her husband, though the negative impact was not statistically significant for spillover men. All groups believe a wife is justified in refusing sex with her husband in *fewer* situations than the control group. This result, while not obvious from the improvement in women's other gender attitudes, does accord with the qualitative reports of a negative and public reaction against the program's messages regarding sex, particularly a wife's right to refuse and initiate sex. Training facilitators described the theme of sexual relations as among the most controversial in the curriculum and particularly provoking resistance among couples. As one facilitator noted, "there's a commonly held belief that a man should initiate sex. Men were resistant to the idea that women can initiate sex and that men don't just have a right to sex whenever they want it" (World Bank 2020).

In addition to causing more progressive women's attitudes in the two domains of gender roles and acceptability of IPV, the program also increased women's aspirations for agency, and made treated women self-identify as less conservative about gender issues. The program caused women to want more input in decision-making. There was a 6 and 9 percentage point increase in the share of treated and spillover women, respectively, who, in an ideal world, would want more input in decision-making regarding household expenses. There was also a 5 percentage point increase in the share of treated women who want more input in decision-making about childbearing. Treated women (and men) also disagree more with the statement that they 'see [themselves] as more conservative about gender and family issues' than most other people in their village, a reduction of 10 percentage points for women and 16 percentage points for treated men. These changes are shown in Table 4.

To corroborate the threatened identity and backlash mechanism, we would expect evidence of greater friction between wives with their new progressive attitudes, aspirations and identities, and their husbands who use violence in retaliation for their threatened identity. Results in Table 4 support this, showing an increase in *men's* desire for agency and also a widening in the gap between husbands' and wives' gender attitudes. Spillover men report that in an ideal world, they also want more input in decision-making about household expenses, an increase of 3 percentage points. The intervention also increased the difference between husbands' and wives' gender attitudes within couples. The table shows that the average difference in the number of progressive gender attitudes held by a wife and her husband increased by 0.23 reasons for spillover couples and 0.12 reasons for treated couples. Similarly, the difference in the number of additional progressive attitudes a wife has than her husband regarding IPV not being justified also increased, by 0.55 reasons for spillover couples and 0.48 reasons for treated couples. There was no impact on the gap between husbands' and wives' attitudes on the acceptability of a wife refusing sex.

The larger impact on the attitude gap between spillover couples than treated couples may indicate why the IPV backlash was greater for spillover than treated women- there was a larger identity threat felt by spillover men (though while the point estimates differ, the differences between the arms are not statistically significant). It also highlights that while treated men report generally improved attitudes (except on sexual violence), they still did not change as much as their wives did, potentially indicating some friction in marriages even where both parties are becoming more progressive.

As well as identity and agency aspiration differences between treated and spillover men, results on time use in Table 4 also indicate that the intervention spurred some new progressive gender behaviors in treated men. In a detailed 24-hour recall time diary, treated men (as well as their wives) reported spending more time in the day before the endline survey doing domestic and care work, which is traditionally seen as a woman's responsibility.[27] [28] This positive indication of men's behavior change combined with anecdotal reports that men who adopted new behaviors were subject to rumors and social sanctions from others may suggest both why IPV still occurred but then was lower for treated than spillover men. The new conservative identity-defying attitudes and behaviors adopted by treated men may have caused them anxiety and disutility due to violating gender prescriptions and negative reactions of their communities, consistent with the identity model's externality predictions. However it still indicates that six months after the training, treated men adopted and retained some new progressive attitudes, and behaviors. Further, these mixed effects on treated couples- with increased IPV but improved attitudes and some behaviors- could also

[27] This question was asked using a recall breakdown of activities done in each 30 minutes of the preceding 24 hours the day before the survey. It was based on the A-WEAI instrument.

[28] There was a couples training session on the benefits of positive parenting and the benefits to children when both parents are actively engaged in and responsible for their children's development.

be driven by average effects masking heterogeneous treatment effects for particular subgroups. Ultimately, the program appears to have caused a net average increase in IPV, but a larger increase for spillover couples who faced the same threat to traditional identity as treated men but without their new progressive attitudes and behaviors.

Table 4: Backlash mechanism: aspirations and identity

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Backlash mechanism: Aspitations and identity					
Women: want more decision-making input re. household expenses	0.21	0.06***	0.09***	0.23	1,987
		(0.02)	(0.02)		
Men: want more decision-making input re. household expenses	0.07	-0.00	0.03*	0.05	1,978
		(0.01)	(0.02)		
Women: want more decision-making input on childbearing	0.13	0.05**	0.01	0.09	1,775
		(0.02)	(0.02)		
Men: want more decision-making input on childbearing	0.05	0.01	0.02	0.52	1,728
		(0.01)	(0.02)		
Women: Agree that they hold more conservative attitudes about gender roles than most others in their village	0.56	-0.10***	0.00	0.00	2,042
		(0.03)	(0.03)		
Men: Agree that they hold more conservative attitudes about gender roles than most others in their village	0.51	-0.16***	-0.05	0.00	2,042
		(0.03)	(0.03)		
Attitude Gap: No. of additional progressive gender attitudes wife holds than husband	-0.21	0.12*	0.23***	0.17	1,986
		(0.07)	(0.08)		
Attitude Gap: No. of additional progressive views wife holds than husband about IPV not being justified	-1.34	0.48***	0.55***	0.59	1,420
		(0.11)	(0.12)		
Attitude Gap: No. of additional progressive views wife holds than husband about the acceptability of a wife refusing sex	-0.11	-0.12	-0.10	0.90	1,585
		(0.10)	(0.10)		
Women: minutes on care/domestic work as primary activity yesterday	270.52	17.83**	13.63	0.65	2,042
		(8.26)	(8.41)		
Men: minutes on care/domestic work as primary activity yesterday	41.63	12.34***	5.70	0.20	2,042
		(4.37)	(4.38)		

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in rounded brackets. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of direct treatment and spillover effects, and differences between effects on trained and spillover couples. Outcomes are listed on the left. Column 1 shows the control mean; column 2 the average treatment effect for treated couples, column 3 the treatment effect for spillover couples. Column 4 reports p-values for the difference between effects on trained and spillover couples.

Further evidence of treated men's publicly visible behavior change, and the role of the state in promoting the program's gender equality message, is seen in the large increase in marriage formalization in Table A5. Of treated couples who were not formally married at baseline, there was a 36 percentage point increase in marriage by endline, with no impact for spillover couples. Mass wedding ceremonies are not uncommon in Rwanda as the government has promoted marriage formalization to protect women's rights. One facilitator noted that almost all of the couples in her classes who had been in informal marriages got married in a joint marriage legalization ceremony at the end of the program, which was enabled by local

leaders. Another facilitator discussed the rationale for formalizing marriages: "If a woman is legally married, she's more confident in terms of challenging her husband. If a man is in an informal marriage, he's more likely to claim ownership over everything in the home. It imposes some limit on men when marriages are formal/legal."

To further explore the identity backlash mechanism, Table A8 shows non-causal, descriptive relationships between treatment effects on IPV and indicators of women's changed or men's challenged identity. The table shows heterogeneous treatment effects by endline characteristics, so appropriate caution should be applied and results not interpreted causally. We find results broadly consistent with the hypothesis that IPV impacts would be greatest amongst couples where men were most threatened by challenges to traditional gender identities. This table shows that, while few of these correlations are statically significant, descriptively, the largest IPV treatment effects were in couples where: women also desired more input in decision-making; where women most strongly disagreed that they were more conservative about gender issues than most others in their village; and where men believed they *were* more conservative about gender than most others in their village.

Table A9 shows heterogeneous treatment effects by baseline measures of identity and social norms. While possibly underpowered, the results suggest that individual baseline gender attitudes do not strongly determine the magnitude of treatment effects, though baseline indicators of conservative village gender norms, as calculated by the mean number of gender inequitable attitudes and reasons IPV is justified held by other surveyed men in the village, is positively associated with a larger treatment effect on emotional IPV for treated individuals.

If the mechanism for the IPV backlash was men's threatened identity, we might also expect treatment effects to be greatest amongst couples where there was no IPV at baseline. Men who were already perpetrating IPV should face little *additional* motivator to use IPV to assuage threatened identity, and they and their wives may be less likely to adopt new progressive, identity-threatening attitudes. For men not perpetrating IPV at baseline, they may feel a greater loss in identity and greater returns from using IPV to assuage an existing (or acquired) reputation as being more progressive. Consistent with this, we find that treatment effects are greatest for couples where women did not report experiencing IPV at baseline. Table A10 and Table A11 show that spillover effects in particular are largely driven by couples where women did not report IPV at baseline. For treated couples, there is no marginal difference in impact by baseline IPV experience. For women who did report IPV at baseline, Table A10 also shows that there is an increase in escalation in the severity of

the type of violence women report experiencing between baseline and endline, also consistent with the mechanism of threatened identity.[29]

Finally, Table A12 and Table A13 show heterogeneous treatment effects by several baseline indicators of a woman's outside options. Table A12 shows that sexual IPV impacts are greatest for spillover women whose partner paid brideprice, potentially indicating the wealth transfer reduced her outside options by conferring a sense of 'ownership' that may have been particularly threatened by the program's messages of gender equality (Anderson 2007; Lowes and Nunn 2017).[30] However we find few statistically significant heterogeneous treatment effects when we test along other pre-specified baseline indicators of women's outside options, including number of children, years of education, wife's relative education, spouse age difference, childhood witnessing of IPV, belief about divorce permissibility, parents and relatives living nearby, number of close friends, wife's savings and asset ownership, and husbands' depression and alcohol consumption.

Table A12 also shows that the intervention caused a larger increase in IPV for spillover couples who got married to or partnered with their spouse in the 8 years prior to the genocide compared to the women who got married in the 8 years after. This regression controls for baseline IPV experience, the number of traumatic experiences the husband faced during the conflict, as well as spouses' age and baseline gender attitudes. Overall, women who married shortly before the genocide have lower rates of IPV compared to those married after, in the severely constrained marriage market, but this protective factor is more than 'undone' by the intervention which increases IPV substantially more among this group. This result is consistent with the notion that current-day IPV is affected by women's earlier outside options on the marriage market after the considerable loss of life for marriage-age men during the genocide (La Mattina 2017). It is possible these pre-1994 marriages are different on another unobserved dimension, or that these men are particularly threatened by challenged identities. The period of eight years was selected to optimize for a large enough sample size but close enough to the 1994 conflict; results are robust if this time frame is expanded.

[29] For women who experienced emotional IPV at baseline but who did not experience physical or sexual IPV, the intervention caused an escalation in the severity of the type of IPV they experienced at endline, to physical and/or sexual IPV. A similar escalation is observed for spillover women who experienced physical IPV at baseline (but not sexual IPV), with the intervention increasing their reported experience of sexual IPV at endline.

[30] Men who paid bride price are not more conservative than others. Descriptively, men who paid bride price hold fewer gender inequitable attitudes and believe IPV is justified for fewer reasons than men who did not pay brideprice.

5.2 Alternative explanations

Changing descriptive norms

An alternative explanation to the IPV treatment effect being caused by a backlash from men's threatened identity is that the program changed descriptive norms about the prevalence of IPV in communities. As discussed above, the program increased village-wide beliefs that IPV is common, and in doing so, it may have removed a constraint on husbands' behavior by easing their fears of social sanctions and providing them a license to perpetrate IPV.

While most people significantly underestimated the prevalence of sexual and physical IPV in their communities at baseline and endline, the intervention increases these beliefs. Spillover men think sexual IPV is 2.3 percentage points more common than the control group believes, and treated men believe it is 8.6 percentage points more common. If a similar mechanism is in operation as was found in a study on gender norms and female labor market participation in Saudi Arabia (Bursztyn, González, and Yanagizawa-Drott 2018), then updating underestimates of social norms about prevalent IPV- which this large-scale, government-funded program has done- could cause an increase in IPV if it is coordinating behavior around a newly exposed norm of high IPV prevalence.

Data suggest that the intervention was widely visible, with many spillover couples reporting knowing someone who was treated, and also speaking to them about the intervention. However, if updating beliefs on IPV prevalence was the main driver of the IPV increase, we would expect the largest treatment effect from those who updated their beliefs the most– treated couples. However, we find the opposite. Spillover couples have the largest IPV treatment effect but update their beliefs about IPV prevalence less than treated couples. Further, when we consider heterogeneous treatment effects by a (potentially endogenous) measure of how much men's beliefs about IPV prevalence increased between baseline and endline (calculated by the difference between endline and baseline standardized weighted indices on perception of village norms), we find no evidence of a difference by this change. Shown in Table A13, we see that all point estimates on the interaction term are negative or close to zero and non-significant. Changing beliefs about norms do not appear to drive the effect.

Resurfaced trauma and alcohol consumption

In bringing up potentially difficult or uncomfortable issues for some people, for example difficulties in their marriage, childhood memories of witnessing IPV between parents, alcoholism, or past treatment of women, the program may have inadvertently caused old traumas to resurface, which may in turn increase stress, arguments, and IPV. In Table A13, we present heterogeneous treatment effects by men's baseline excessive alcohol consumption, number of negative experiences the husband had during the 1994 genocide, depression,

childhood witnessing IPV, and men's and women's reports of men's excessive alcohol consumption. We find no evidence suggestive of pre-experiment high levels of trauma or alcohol consumption being associated with larger negative treatment effects on IPV. In fact, there is a reduction in emotional IPV reported by the wives of men who drank alcohol excessively at baseline.

Different networks and information channels for women and men

Anecdotally, men and women indicated it was relatively uncommon to have close friendships and casual social interactions with people of the opposite sex. This suggests most information is likely to go through separate social networks and information channels for men and women. At baseline, women reported having more friends with whom they talk about their marital relationship than did their husbands. This could imply a higher likelihood that spillover women have more opportunities to hold more personal discussions about the program's messages regarding gender equality and women's rights than spillover men. This difference in in-depth program message diffusion between men and women may create tensions between spillover spouses, where wives might be ready for change, while husbands might not. We test for heterogeneous treatment effects by baseline measures of the number of close friends that respondents have in the village (see Table A9). These results show no differential treatment effect by the number of close friends a man reports, and only a slightly greater treatment effect on physical IPV for spillover women the greater the number of close friends she listed at baseline. This result is also consistent with a backlash against threatened identity, given that more socially connected spillover women may adopt more progressive attitudes and aspirations, and thus may be perceived as more threatening to traditional gender identities.

Expressive IPV and taste for violence

Existing economic theory on IPV would imply that when there is no shift in husbands' and wives' material resources (ruling out instrumental violence), the only reason IPV would increase would be expressive violence from men who have a preference or 'taste' for violence and have the constraints removed. If expressive IPV from men's taste for violence, rather than threatened identity, is the cause of the increase, then we might expect IPV treatment effects to be greatest amongst men who at baseline express more conservative gender attitudes. However, results in Table A9 show there is no difference in impact by men's baseline gender attitudes and belief the IPV is justified. If expressive violence were the channel, we might also expect that women's baseline outside options would protect her from the IPV backlash, however, we found very little evidence of this, some of which is shown in Table A12 and Table A13. The expressive model of violence also fails to explain the cause of the large negative spillover effects on the community. Overall, an expressive model of IPV is broad enough to encapsulate what is observed in this context, but the motivations of 'taste'

do not appear to explain our results as well as threatened identity, and thus appear less useful in our context.

Frustration with didactic or top-down reforms

An alternative explanation for the IPV increase is that men were frustrated with being 'told how to live their lives' by NGOs or government officials, and they took this frustration out on their wives. Unlike identity, this mechanism might suggest IPV is an emotional reaction to frustration rather than discomfort with threatened gender identity in particular. In a state that has taken significant strides to improve gender equality over a very short period of time, and where legislative changes to women's rights are far more progressive than many people's personal attitudes, and where the government has promoted marriage formalization ceremonies to guarantee women these rights, men may resent new top-down interventions promoting these rights that they may feel impinge their freedoms. If this were the case, we might not expect to see the same results if a) the program were not funded or promoted by government and through official community meetings, or b) if it were delivered in a different context where people feel more able to disregard the messages they are given.

We cannot rule out the backlash being caused by frustration with perceived didactic or top-down programs (nor whether this mechanism is operating in concert with the identity threat). However, the extent of the gender-specific impacts of this program implies a friction specific to gender rather than general frustration with feeling 'pushed' into something. Further, given participation in the training program was voluntary and over-subscribed, it seems unlikely people felt compelled to attend, particularly spillover couples who did not attend the training. We cannot identify the extent to which the backlash is caused by the 'messenger' being a state (though the couples program was implemented by local NGOs on behalf of the state). When we consider heterogeneous treatment effects by whether the villages had above or below median conservative norms at baseline, perhaps where there may be more receptivity to the government-endorsed messages of gender equality, we find no evidence suggesting lower backlash in more progressive villages.

5.3 Robustness checks on primary outcomes

In Table A16 and Table A17 we present the treatment effect estimates on our primary IPV outcomes under different specifications. In all cases, under different estimation specifications including single difference, post-double selection lasso controls (Belloni, Chernozhukov, and Hansen 2014), probit, and treatment-on-the-treated instrumenting treatment with assignment, there is a negligible difference on the magnitude and precision of coefficient estimates. In Table A15, we also present results after coding all women's refuse-to-answer an IPV question as 'yes', and also all refuse-to-answer as 'no', to assess whether differential refusal rates drive the results. We again observe negligible differences. We also assess impact on whether a respondent refused to answer at least one of the questions comprising each of the

three overall IPV primary outcome questions. As shown in Table A14, we find that the only refusal rate that is affected by treatment is for physical violence, with trained women 2 percentage points less likely to refuse to answer, and spillover women 3 percentage points more likely to refuse. There is no treatment effect on the refusal rate for emotional or sexual IPV questions.

To assess the robustness of our main estimates to potential non-compliance from treated couples living in control villages, we assess several additional estimates. First, we exclude from the estimation sample the pure control couples with at least one treated or spillover couple living within various distances of their household (200, 400 600, 800 meters). We also exclude from our sample both the 383 respondents in the first days of data collection who were not asked whether they lived in the same village that their VSLA was associated with (and thus we cannot confirm whether they live in the village used to assign treatment status), and the additional 529 respondents who confirmed they live in a different village to the VSLA-village of assignment. Primary IPV results broadly hold across all of these specifications and sample restrictions, though there are some changes in the magnitude of the treatment effects. We find that excluding pure control couples who actually live close to spillover or treated couples results in treatment effect estimates larger than in our main specification, suggesting that our main ITT estimates generate lower bounds on the true treatment effects. These results are shown in Figure A3, Table A19, and Table A20.

6. Discussion: Truth or reporting

In this section, we present evidence suggesting that the increase in violence likely reflects a true increase in violence rather than just an increase in propensity to report it.

There are three main reasons why the intervention may cause treated or spillover women to report IPV more than control women, regardless of any true impact on IPV: (1) strategic reporting to increase the probability of getting selected for potential future programs (and the associated travel stipend); (2) an expanded understanding of what the IPV question terms mean; and (3) reduced stigma or increased empowerment to report one's experience of IPV. If these reporting effects are correlated with treatment and dominate the true IPV impacts of the intervention, we will be estimating biased treatment effects. We consider evidence on each category below.

6.1 Strategic reporting?

Couples may believe there is an incentive to report IPV more (or less) to increase their chances of getting access to an extended or future program, which may be a desirable outcome given the program included a small travel stipend to cover the costs of training attendance. If people were reporting strategically, we might expect respondents to answer other survey questions in ways that make them appear as favorable candidates for a publicly

funded program, for example, reporting a reduction in self-reported alcohol consumption. However, we find the opposite. Over several questions about alcohol consumption, treated and spillover men and treated and spillover women all report that husbands drink more alcohol or drink alcohol to excess more often. This is shown in Table A21.

Second, we might also expect couples who are reporting strategically to claim that they are poorer, and thus more in need of the travel stipend. Yet as discussed above and shown in Table A5, we observe almost no economic impacts of the program, and where there is an impact, it indicates marginally increased asset ownership for both treated and spillover couples.

Further, if couples were coordinating to report strategically, we might expect husbands and wives to report similarly on the key gender issues that the program is known to address, yet as discussed above, the program increases the gap in treated and spillover men's and women's gender attitudes. Relatedly, if there is a shared village-wide understanding of what type of reporting might increase access to the program, we might expect all groups (treated and control men and women) to report in a similar way. Yet we find different directions of impacts between spouses, and between treated and spillover couples, including on similar issues (for example improved gender attitudes for some groups, worsened refusing sex acceptance for all groups). Further, couples would have recalled their participation in the random public lottery used to assign treatment. Qualitative reports also note that village leaders recall that we used village random lotteries to select couples rather than targeting the program at couples known to have IPV.

Finally, if couples or women were reporting IPV strategically, we might have expected differential IPV reporting at baseline by a couple's rank in the public lottery (which they understood was later going to be used to assign them to treatment or control group), or by their baseline knowledge of whether the program was about gender and family relationships (as opposed to another issue such as agriculture). However, we find that there is no difference in baseline IPV reporting by a couples' rank in the public lottery. At baseline, 83% of women believed the program was about gender and families. There is a small, marginally significant positive correlation between physical IPV reporting by women's knowledge of whether the program was about gender and families rather than about another issue (though for emotional and sexual IPV, this relationship is not significant and had near-zero point estimates). These results are shown in Table A22 and Table A23.

6.2 Expanded understanding of IPV questions?

Second, the intervention may expand treated and spillover women's understanding of what the terms used in the violence questions mean or include. For example, 'slapping', 'choking', 'force' or 'degrading' might mean different things to women who have undergone the training compared to women in the control group. If an expanded understanding were driving

the IPV impacts, it seems likely that the group with the most substantial engagement with discussions on women's rights and what constitutes IPV- treated women who reflect on these issues for 22 weeks- would have the most expanded understanding, and thus the greatest increase in IPV reporting. However, the greatest IPV increases are for spillover women, even though their only exposure to the program's messages of women's rights and what constitutes IPV are through second-hand information and occasional community-wide meetings.[31]

Second, if the IPV increase is driven by an expanded understanding of the IPV questions, we would expect the largest increases to be found for the questions that use more subjective words or concepts compared to more objective questions, which should be interpreted in the same way by treated, spillover and control women. However, we find IPV increases across many individual IPV questions, both objective and subjective (in both Kinyarwanda and English interpretations). Impact by individual IPV question is shown in Table A24.

Finally, we would also expect that if IPV impacts were driven by an expanded understanding, the group with the greatest exposure to the new terms would report the greatest increase in the number of *types* of violence experienced per category, and perhaps also the greatest increase in frequency of experiencing violence. However, as shown in Table A25, we find that while there is an increase in the frequency and number of types of violence reported by both treated and spillover women, the increase is greatest for spillover women.

6.3 Reduced stigma/ more empowered to report?

Finally, treated and spillover women might be more comfortable reporting IPV due to reduced reporting stigma or a sense of empowerment.[32] If greater comfort in reporting were driving treatment effects, then we would likely see other indications that women are more empowered to report IPV or are otherwise more willing to discuss it. However, results shown in Table A26 and Table A27 demonstrate that there is little evidence that women are more comfortable reporting IPV. While treated women are marginally less likely to refuse to answer a physical IPV question, we find that spillover women are marginally *more* likely to refuse. There is no reduction in the refusal rate on emotional or sexual IPV questions. Table A27 also shows that while treated women are 6 percentage points more likely to visit a safe space and to speak to someone about violence that they experienced, they are no more likely to report IPV that they witnessed to local leaders. Treated men used safe spaces at equal rates as treated women, so their use perhaps does not reflect a highly stigmatized activity, and indeed, it was advertised as an available service for couples facing marital difficulties

[31] Qualitative reports also suggest that what couples communicated during the community-wide meetings was more about testimonials about what they learned and changed in their own relationship rather than informing their neighbors about the definitions and foundations of concepts like gender and IPV.

[32] Note that this 'empowered' reporting can be considered as reflecting a positive impact of the training, in a context where underreporting is a recognized constraint to victims' identification and access to relevant services.

during the couples training. There is no evidence that spillover women, who reported the highest rates of IPV, feel reduced stigma about reporting or are more empowered to report IPV.

6.4 Other evidence on truth or reporting

Other general evidence that may suggest the IPV impact is real rather than reporting is discussed here. First, women's self-reports of IPV experience are positively correlated with potential objective indicators and other indicators that literature and theory suggest would be correlated with true IPV. These include whether a woman was distressed during the interview and was referred to the independent study counsellor, her decision to then speak to the counsellor, and, for physical IPV, if the enumerator reported that she had visible signs of injury. While these are small numbers of women, there are positive correlations. These correlations coefficients are shown in Table A28 and Table A29.

In addition, it was common for some of our male sample to travel to another area to undertake agricultural labor several nights a week, and we find treatment effects differ by this baseline proxy of how much time spouses spend together. In total, 230 men at baseline reported working away from home for 4 consecutive nights over the preceding 3 months. Given IPV requires husbands to be present to perpetrate it, IPV is likely to increase the more time spouses spend together. If the IPV increase is a true increase, we would expect it to be lower for couples where at baseline, the husband worked away from home (he is both around less, and perhaps witnessing fewer instances threatening to his identity). If the increase reflects increased reporting, there should be no differential treatment effect by husbands' baseline presence at home. This argument primarily applies to spillover couples who did not have their regular work routine affected by the need to be at training during a weekday. Indeed, the results in Table A30 show that the treatment effects on spillover couples are primarily driven by the non-migrating husband households, with physical and sexual IPV in spillover couples much lower for couples where husbands were migrating for work at baseline. There is no difference for treated couples where husbands migrated, possibly because husbands had to be present for the mid-week training and thus were less likely to travel for work. Further, the wives of men who migrate for work during the week reported experiencing fewer types of sexual IPV, and less frequent IPV than women whose husbands did not travel for work.

Finally, the pattern of treatment effects throughout the 90-minute survey on questions related to sexual IPV, such as women's experience of sexual IPV, an increase in sexual IPV injuries, an increase in the belief about its prevalence in the community, worsening of all groups' attitudes regarding a wife's right to refuse sex, and an increase in the last marital argument being about sex, collectively lend credence to the truth of an impact on women's experience of sexual IPV.

7. Conclusion

Given that one in three women globally is a victim of IPV in her lifetime, it is important for policymakers and researchers to better understand the causes of IPV and to identify scalable policies to prevent it. This paper presents some of the first evidence from a randomized control trial of a couples-based IPV prevention program. It is also one of the few two-level randomized control trials designed to identify spillovers of a social policy intervention.

This paper showed that an intensive 22-week couples discussion group designed to prevent IPV had large, unintended, negative impacts for treated couples and for spillover couples living in treated villages. The intervention caused large increases in women's reported experience of emotional, physical and sexual IPV, and the negative impacts were greatest for spillover couples. We showed evidence suggesting that the IPV treatment effect, while based on women's self-reports, likely reflects a true IPV increase rather than an increase in reporting. We find suggestive evidence that the gender program may have caused an upending of traditional gender identities, spurring a broad community-wide IPV backlash that dominated any potentially positive effects of the training program.[33]

We find evidence suggesting that the backlash against the intervention's messages of women's empowerment appears to have been disseminated across treated communities. At village-wide meetings, concepts around women's rights and marital relations were discussed, after which there were reports of a verbal backlash against a woman's right to refuse and initiate sex with her husband. Reports also indicate that men who adopted new progressive behaviors were shamed. In the quantitative data, we find that both treated and spillover women become more empowered, aspiring for more input in decision-making and holding more progressive gender attitudes. Meanwhile, the program caused an increase in the gap between wives' and husbands' progressive gender attitudes, potentially indicating greater frictions between spouses.

Additionally, our paper makes a methodological contribution, being one of the few studies of a targeted social program that measures spillovers to non-participants in the community. Our results of larger spillover effects than effects on direct beneficiaries has important implications for both health prevention and other social programs that seek to, or may unintentionally, have an impact beyond direct recipients. Our results provide a cautionary tale, demonstrating that even social interventions that primarily work with individuals can cause substantial negative spillovers within communities. Our results could also be interpreted more positively- if the forces driving our large spillovers could be harnessed for good, even targeted programs could have broad community impacts, substantially increasing programs'

[33] Other studies have also found evidence of an IPV backlash from interventions in conservative contexts that improve women's bargaining power or otherwise threaten conservative gender identities, including secondary education reforms, social protection programs, and business training programs (Bulte and Lensink 2019; Angelucci 2008; Erten and Keskin 2018).

reach and cost-effectiveness. Future avenues of research could focus on identifying the mechanisms for such spillovers, including considering how to limit them when identity threats create negative externalities.

To our knowledge, no intervention has been found to cause such a large increase in IPV, let alone an IPV prevention program. Yet this program was carefully designed for Rwanda's context, building on best practice and other gender norms interventions being delivered around the world.[34] Programs similar to this are now being considered or scaled up by governments and NGOs in Sub-Saharan Africa and other regions. This paper therefore has important practical implications for policymakers and other actors considering delivering such programs. It suggests that caution should be applied, additional research conducted, and high frequency monitoring data collected before scaling up such sensitive identity-challenging interventions. In particular, such programs should be cautious in the way that messages of gender equality and women's rights are promoted in communities.

Researchers and policymakers should prioritize investigation of potentially less risky, more cost-effective alternatives to intensive couples-based IPV prevention programs, such as gender equality programs targeted to adolescents. Younger individuals learning these messages may not yet have formed rigid identities, and boys may feel they have less to lose from the empowerment of a hypothetical future wife, and should be less likely to backlash (Dhar, Jain, and Jayachandran 2018). Other non-'gender' interventions such as improving education, improving employment opportunities, or rural electrification that expands access to phones and media technology also have the potential to reduce IPV through non-gender identity threatening, cost-effective means.

In addition to considering alternative policies that may reduce IPV, future research should consider the best delivery channels and tone of messages that seek to change prevailing identities, and likely avoid what could be perceived as didactic approaches. Other research could usefully consider the importance of the gap between law and attitudes, and of political context for determining the impact of gender equality messages on threatened identity and IPV. Further, although our findings are of particular interest for developing countries and gender programs, the behavioral effects of threatened identity are likely to have wider relevance, and may apply in many other contexts and for other issues such as race or national identity.

[34] The curriculum for the couples training program was carefully adapted and piloted over several years to be made appropriate for the Rwandan context, and was based on key IPV prevention programs used in other low and middle income settings over the past several decades.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. 'When Should You Adjust Standard Errors for Clustering?' w24003. http://www.nber.org/papers/w24003.ack.

Abbott, Pamela, and Dixon Malunda. 2016. 'The Promise and the Reality: Women's Rights in Rwanda'. *African Journal of International and Comparative Law* 24 (4): 561–81. https://doi.org/10.3366/ajicl.2016.0173.

Abramsky, Tanya, Karen Devries, Ligia Kiss, Janet Nakuti, Nambusi Kyegombe, Elizabeth Starmann, Bonnie Cundill, Leilani Francisco, Dan Kaye, and Tina Musuya. 2014. 'Findings from the SASA! Study: A Cluster Randomized Controlled Trial to Assess the Impact of a Community Mobilization Intervention to Prevent Violence against Women and Reduce HIV Risk in Kampala, Uganda'. *BMC Medicine* 12 (1): 122.

Aizer, A. 2010. 'The Gender Wage Gap and Domestic Violence.' *The American Economic Review* 100 (4): 1847–59. https://doi.org/10.1257/aer.100.4.1847.

Aizer, A. 2011. 'Poverty, Violence, and Health the Impact of Domestic Violence during Pregnancy on Newborn Health'. *Journal of Human Resources* 46 (3): 518–38. http://jhr.uwpress.org/content/46/3/518.short.

Akerlof, George A, and Rachel E Kranton. 2000. 'Economics and Identity'. *The Quarterly Journal of Economics* 115 (3): 715–53.

Alesina, A, P Giuliano, and N Nunn. 2013. 'On the Origins of Gender Roles: Women and the Plough'. *The Quarterly Journal of Economics* 128 (2): 469–530. https://academic.oup.com/qje/article-abstract/128/2/469/1943509.

Alesina, Alberto, Benedetta Brioschi, and Eliana La Ferrara. 2020. 'Violence Against Women: A Cross-cultural Analysis for Africa'. *Economica*, May. https://doi.org/10.1111/ecca.12343.

Alik-Lagrange, Arthur, Claire Cullen, Muthoni Ngatia, and Julia Vaillant. 2017. 'Preventing Intimate Partner Violence: Impact Evaluation of MIGEPROF Couples Training for IPV Prevention in Eastern Rwanda'. *AEA RCT Registry*. https://www.socialscienceregistry.org/trials/2282/history/19047.

Anderson, Michael L. 2008. '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): 1481–95. https://doi.org/10.1198/016214508000000841.

Anderson, Siwan. 2007. 'The Economics of Dowry and Brideprice'. *Journal of Economic Perspectives* 21 (4): 151–74. https://doi.org/10.1257/jep.21.4.151.

Angelucci, Manuela. 2008. 'Love on the Rocks: Domestic Violence and Alcohol Abuse in Rural Mexico'. *Journal of Economic Analysis & Policy Contributions*. Vol. 8. http://www.bepress.com/bejeap/vol8/iss1/art43.

Asiedu, Elizabeth, Claire Branstette, Babulal Neepa Gaekwad, and Malokele Nanivazo. 2018. 'The Effect of Women's Representation in Parliament and the Passing of Gender Sensitive Policies'. In *AEA Paper Session: Issues on African Development I, 5 January*. Philadelphia.

Baird, S, JA Bohren, C McIntosh, and B Özler. 2014. 'Designing Experiments to Measure Spillover Effects'. 6824. Policy Research Working Paper. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-6824.

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco. 2019. 'Entertainment, Education, and Attitudes Toward Domestic Violence'. In *AEA Papers and Proceedings*, 109:133–37.

Baranov, Victoria, Lisa Cameron, Diana Contreras Suarez, and Claire Thibout. 2020. 'Theoretical Underpinnings and Meta-Analysis of the Effects of Cash Transfers on Intimate Partner Violence in Low- and Middle-Income Countries'. *Journal of Development Studies*. Routledge. https://doi.org/10.1080/00220388.2020.1762859.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. 'Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India'. *Science* 335. https://doi.org/10.1126/science.1211180.

Becker, G. 1973. 'A Theory of Marriage: Part I'. *Journal of Political Economy* 81 (4): 813–46. https://doi.org/10.1086/260084.

———. 2009. *A Treatise on the Family*. Harvard university press. https://books.google.com/books?hl=en&lr=&id=Vnr8w6HwiAAC&oi=fnd&pg=PR7&ots=B4Gnk82U8X&sig=TkjesMqpTmlBS1sSDH6RK8ZUvx8.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. 'High-Dimensional Methods and Inference on Structural and Treatment Effects'. *Journal of Economic Perspectives* 28 (2): 29–50. https://doi.org/10.1257/jep.28.2.29.

Benabou, R, and J Tirole. 2011. 'Laws and Norms'. w17579. *National Bureau of Economic Research*. https://www.nber.org/papers/w17579.

Benjamini, Yoav, and Yosef Hochberg. 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. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bhalotra, S, U Kambhampati, S Rawlings, and Z Siddique. 2020. 'Intimate Partner Violence: The Influence of Job Opportunities for Men and Women'. WPS9118. *World Bank Economic Review*. Policy Research Working Paper. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9118.

Bicchieri, C, E Dimant, S Gächter, and D Nosenzo. 2020. 'Observability, Social Proximity, and the Erosion of Norm Compliance'. 8212. CESifo Working Paper . http://hdl.handle.net/10419/216608www.econstor.eu.

Bloch, Francis, and Vijayendra Rao. 2002. 'Terror as a Bargaining Instrument: A Case Study of Dowry Violence in Rural India'. *American Economic Review* 92 (4): 1029–43. https://doi.org/10.1257/00028280260344588.

Bowles, Samuel. 1998. 'Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions'. *Journal of Economic Literature* 36 (1): 75–111.

Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise. 2018. 'A Mixed-Method Review of Cash Transfers and Intimate Partner Violence in Low- and Middle-Income Countries'. *The World Bank Research Observer* 33 (2): 218–58. https://doi.org/10.1093/wbro/lky002.

Bulte, Erwin, and Robert Lensink. 2019. 'Women's Empowerment and Domestic Abuse: Experimental Evidence from Vietnam'. *European Economic Review* 115 (June): 172–91. https://doi.org/10.1016/j.euroecorev.2019.03.003.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2018. 'Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia'. https://home.uchicago.edu/~bursztyn/Misperceived_Norms_2018_06_20.pdf.

Carrell, Scott E, Mark Hoekstra, and Elira Kuka. 2018. 'The Long-Run Effects of Disruptive Peers'. *American Economic Review* 108 (11): 3377–3415. https://doi.org/10.1257/aer.20160763.

Chatterji, Sangeeta, Erin Stern, Kristin Dunkle, and Lori Heise. 2020. 'Community Activism as a Strategy to Reduce Intimate Partner Violence (IPV) in Rural Rwanda: Results of a Community Randomised Trial'. *Journal of Global Health* 10 (1). https://doi.org/10.7189/jogh.10.010406.

Cools, Sara, and Andreas Kotsadam. 2017. 'Resources and Intimate Partner Violence in Sub-Saharan Africa'. *World Development* 95 (July): 211–30. https://doi.org/10.1016/J.WORLDDEV.2017.02.027.

Cullen, Claire. 2020. 'Method Matters: Underreporting of Intimate Partner Violence in Nigeria and Rwanda'. WPS 9274. World Bank Policy Research Working Paper Series. Washington D.C. http://documents.worldbank.org/curated/en/730981591707408517/Method-Matters-

Underreporting-of-Intimate-Partner-Violence-in-Nigeria-and-Rwanda.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. 'Persuasion: Empirical Evidence'. *Annu. Rev. Econ.* 2 (1): 643–69.

Devries, K. M., J. Y. T. Mak, C. García-Moreno, M. Petzold, J. C. Child, G. Falder, S. Lim, et al. 2013. 'The Global Prevalence of Intimate Partner Violence Against Women'. *Science* 340 (6140). http://science.sciencemag.org/content/340/6140/1527.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2018. 'Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India'. http://faculty.wcas.northwestern.edu/~sjv340/reshaping_gender_attitudes.pdf.

DHS. 2016. 'Rwanda Demographic and Health Survey 2014-15'. https://dhsprogram.com/pubs/pdf/FR316/FR316.pdf.

Dimant, Eugen. 2019. 'Contagion of Pro- and Anti-Social Behavior among Peers and the Role of Social Proximity'. *Journal of Economic Psychology* 73 (August): 66–88. https://doi.org/10.1016/j.joep.2019.04.009.

Doyle, Kate, Ruti G. Levtov, Gary Barker, Gautam G. Bastian, Jeffrey B. Bingenheimer, Shamsi Kazimbaya, Anicet Nzabonimpa, et al. 2018. 'Gender-Transformative Bandebereho Couples' Intervention to Promote Male Engagement in Reproductive and Maternal Health and Violence Prevention in Rwanda: Findings from a Randomized Controlled Trial'. *PLOS ONE* 13 (4): e0192756. https://doi.org/10.1371/journal.pone.0192756.

Duflo, Esther. 2012. 'Women Empowerment and Economic Development'. *Journal of Economic Literature* 50 (4): 1051–79.

Dunkle, K, E Stern, and L Heise. 2020. 'Impact of the Indashyikriwa Training Curriculum for Couples on Intimate Partner Violence in Rwanda: Cluster Randomized Control Trial'.

Efferson, Charles, Sonja Vogt, Amy Elhadi, Hilal El Fadil Ahmed, and Ernst Fehr. 2015. 'Female Genital Cutting Is Not a Social Coordination Norm'. *Science* 349 (6255). http://science.sciencemag.org/content/349/6255/1446.full.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2019. 'General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya'. 26600. National Bureau of Economic Research. https://www.nber.org/papers/w26600.

Ellsberg, Mary, Lori Heise, Rodolfo Pena, Sonia Agurto, and Anna Winkvist. 2001. 'Researching Domestic Violence Against Women: Methodological and Ethical Considerations'. *Studies in Family Planning* 32 (1): 1–16. https://doi.org/10.1111/j.1728-4465.2001.00001.x.

Erten, Bilge, and Pinar Keskin. 2018. 'For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey'. *American Economic Journal: Applied Economics* 10 (1): 64–105. https://doi.org/10.1257/app.20160278.

Eswaran, Mukesh, and Nisha Malhotra. 2011. 'Domestic Violence and Women's Autonomy in Developing Countries: Theory and Evidence'. *Canadian Journal of Economics/Revue Canadienne d'économique* 44 (4): 1222–63. https://doi.org/10.1111/j.1540-5982.2011.01673.x.

Farmer, Amy, and Jill Tiefenthaler. 1997. 'An Economic Analysis of Domestic Violence'. *Review of Social Economy*. Taylor & Francis, Ltd. https://doi.org/10.2307/29769903.

Fearon, James, and Anke Hoeffler. 2014. 'Benefits and Costs of the Conflict and Violence Targets for the Post-2015 Development Agenda Post-2015 Consensus'. Post-2015 Consensus. https://pdfs.semanticscholar.org/8810/64e403f1c0d7d469eb1a02cdeb631089ca1d.pdf.

García-Moreno, Claudia, Cathy Zimmerman, Alison Morris-Gehring, Lori Heise, Avni Amin, Naeemah Abrahams, Oswaldo Montoya, Padma Bhate-Deosthali, Nduku Kilonzo, and Charlotte Watts. 2015. 'Addressing Violence against Women: A Call to Action'. *The Lancet* 385 (9978): 1685–95.

Green, D, A Wilke, and J Cooper. 2018. 'Silence Begets Violence: A Mass Media Experiment to Prevent Violence against Women in Rural Uganda'.

https://www.povertyactionlab.org/sites/default/files/publications/Silence-Begets-Violence_Green_et_al_January2018.pdf.

Green, Donald P., Anna M. Wilke, and Jasper Cooper. 2020. 'Countering Violence Against Women by Encouraging Disclosure: A Mass Media Experiment in Rural Uganda'. *Comparative Political Studies*, April, 001041402091227. https://doi.org/10.1177/0010414020912275.

Green, E, Christopher Blattman, Julian Jamison, and Jeannie Annan. 2015. 'Women's Entrepreneurship and Intimate Partner Violence: A Cluster Randomized Trial of Microenterprise Assistance and Partner Participation in Post-Conflict Uganda'. *Social Science and Medicine* 133: 177–88. https://doi.org/10.1016/j.socscimed.2015.03.042.

Guiso, Luigi, Paola Sapienza, Luigi Zingales, Alberto Alesina, George Akerlof, Gary Becker, Alberto Bisin, et al. 2006. 'Does Culture Affect Economic Outcomes?' *The Jounnal of Economic Perspectives* 20 (2): 23–48.

Gupta, Jhumka, Kathryn L Falb, Heidi Lehmann, Denise Kpebo, Ziming Xuan, Mazeda Hossain, Cathy Zimmerman, Charlotte Watts, and Jeannie Annan. 2013. 'Gender Norms and Economic Empowerment Intervention to Reduce Intimate Partner Violence against Women in Rural Côte d'Ivoire: A Randomized Controlled Pilot Study'. *BMC International Health and Human Rights* 13 (1). https://doi.org/10.1186/1472-698X-13-46.

Hansén, J. 2017. 'Political Gender Quota in Rwanda'. http://lup.lub.lu.se/student-papers/record/8918957.

Haushofer, Johannes, James Reisinger, and Jeremy Shapiro. 2015. 'Your Gain Is My Pain: Negative Psychological Externalities of Cash Transfers'.

Haushofer, Johannes, Charlotte Ringdal, Jeremy P Shapiro, and Xiao Yu Wang. 2019. 'Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya'. w25627. NBER Working Paper Series.

Heise, L, and M Hossain. 2017. 'Measuring Intimate Partner Violence'. London. http://strive.lshtm.ac.uk/system/files/attachments/STRIVE Intimate Partner Violence Measurement Brief.pdf.

Hidrobo, Melissa, John Hoddinott, Amber Peterman, Amy Margolies, and Vanessa Moreira. 2014. 'Cash, Food, or Vouchers? Evidence from a Randomized Experiment in Northern Ecuador'. *Journal of Development Economics* 107: 144–56. https://doi.org/10.1016/j.jdeveco.2013.11.009.

Hossain, Mazeda, Cathy Zimmerman, Ligia Kiss, Tanya Abramsky, Drissa Kone, Monika Bakayoko-Topolska, Jeannie Annan, Heidi Lehmann, and Charlotte Watts. 2014. 'Working with Men to Prevent Intimate Partner Violence in a Conflict-Affected Setting: A Pilot Cluster Randomized Controlled Trial in Rural Côte d'Ivoire'. *BMC Public Health* 14 (1): 339. https://doi.org/10.1186/1471-2458-14-339.

Imbens, GW, and DB Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press. http://biosun01.biostat.jhsph.edu/.

Jayachandran, Seema. 2015. 'The Roots of Gender Inequality in Developing Countries'. *Annu. Rev. Econ* 7: 63–88.

Jensen, Robert, and Emily Oster. 2009. 'The Power of TV: Cable Television and Women's Status in India'. *Quarterly Journal of Economics* 124 (3): 1057–94. https://doi.org/10.1162/qjec.2009.124.3.1057.

Karing, Anne. 2018. 'Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone'.

Kerr-Wilson, Alice, Andrew Gibbs, Erika McAslan Fraser, Leane Ramsoomar, Anna Parke, Hussain M A Khuwaja, and Rachel Jewkes. 2020. 'A Rigorous Global Evidence Review of Interventions to Prevent Violence against Women and Girls'. Pretoria. www.whatworks.co.za.

Koenen, Karestan C, Terrie E Moffitt, Avshalom Caspi, Alan Taylor, and Shaun Purcell. 2003. 'Domestic Violence Is Associated with Environmental Suppression of IQ in Young Children'.

Development and Psychopathology 15: 297–311. https://doi.org/10.1017.S0954579403000166.

Krupka, EL, and RA Weber. 2013. 'Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?' *Journal of the European Economic Association* 11 (3): 495–524. https://academic.oup.com/jeea/article-abstract/11/3/495/2300029.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. 'Soap Operas and Fertility: Evidence from Brazil'. *American Economic Journal: Applied Economics* 4 (4): 1–31. https://doi.org/10.1257/app.4.4.1.

Lowes, S, and N Nunn. 2017. 'Bride Price and the Wellbeing of Women'. In *Towards Gender Equity in Development* , edited by S Anderson. https://ideas.repec.org/p/unu/wpaper/wp2017-131.html.

Lowes, Sara, Nathan Nunn, James A Robinson, Jonathan Weigel, Anne Degrave, James Diderich, Muana Kasongo, et al. 2017. 'The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom'. *Econometrica* 85: 1065–91.

Lybbert, Travis J., and Bruce Wydick. 2018. 'Poverty, Aspirations, and the Economics of Hope'. *Economic Development and Cultural Change* 66 (4): 709–53. https://doi.org/10.1086/696968.

Mani, Anandi, and Emma Riley. 2019. 'Social Networks, Role Models, Peer Effects, and Aspirations'. 120. WIDER Working Paper. https://doi.org/10.35188/UNU-WIDER/2019/756-9.

Mannell, Jenevieve, Iran Seyed-Raeisy, Rochelle Burgess, and Catherine Campbell. 2018. 'The Implications of Community Responses to Intimate Partner Violence in Rwanda'. Edited by Andrew R. Dalby. *PLOS ONE* 13 (5): e0196584. https://doi.org/10.1371/journal.pone.0196584.

Mattina, Giulia La. 2017. 'Civil Conflict, Domestic Violence and Intra-Household Bargaining in Post-Genocide Rwanda'. *Journal of Development Economics* 124: 168–98. https://doi.org/10.1016/j.jdeveco.2016.08.001.

Mcintosh, Craig, and Andrew Zeitlin. 2018. 'Benchmarking a Child Nutrition Program against Cash: Experimental Evidence from Rwanda'. *Poverty-Action.Org*. https://www.poverty-action.org/sites/default/files/publications/Benchmarking.pdf.

McKenzie, D. 2012. 'Beyond Baseline and Follow-up: The Case for More T in Experiments'. *Journal of Development Economics* 99 (2): 210–21. http://www.sciencedirect.com/science/article/pii/S030438781200003X.

Miguel, Edward, and Michael Kremer. 2004. 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities'. *Econometrica* 72 (1): 159–217. https://doi.org/10.1111/j.1468-0262.2004.00481.x.

Oh, Suanna. 2019. 'Does Identity Affect Labor Supply?'

Polavarapu, Aparna. 2011. 'Procuring Meaningful Land Rights for the Women of Rwanda'. *Yale Human Rights and Development Journal Article*. Vol. 14. http://digitalcommons.law.yale.edu/yhrdlj/vol14/iss1/3.

Roy, Shalini, Melissa Hidrobo, John Hoddinott, and Akhter Ahmed. 2018. 'Transfers, Behavior Change Communication, and Intimate Partner Violence: Post-Program Evidence from Rural Bangladesh'. *The Review of Economics and Statistics* 101 (5): 865–77. https://doi.org/10.1162/rest_a_00791.

Slegh, Henny, Gary Barker, Augustin Kimonyo, Prudence Ndolimana, and Matt Bannerman. 2013. '"I Can Do Women's Work": Reflections on Engaging Men as Allies in Women's Economic Empowerment in Rwanda'. *Gender & Development* 21 (1): 15–30.

Stern, Erin, and Ritha Nyiratunga. 2017. 'A Process Review of the Indashyikirwa Couples Curriculum to Prevent Intimate Partner Violence and Support Healthy, Equitable Relationships in Rwanda'. *Social Sciences* 6 (2): 63. https://doi.org/10.3390/socsci6020063.

Stevenson, B., and J. Wolfers. 2006. 'Bargaining in the Shadow of the Law: Divorce Laws and Family Distress'. *The Quarterly Journal of Economics* 121 (1): 267–88. https://doi.org/10.1093/qje/121.1.267.

Tauchen, Helen V., Ann Dryden Witte, and Sharon K. Long. 1991. 'Domestic Violence: A Nonrandom Affair'. *International Economic Review* 32 (2): 491. https://doi.org/10.2307/2526888.

Tur-Prats, Ana. 2019. 'Unemployment and Intimate-Partner Violence: A Gender-Identity Approach'. https://repositori.upf.edu/handle/10230/32636.

Vaillant, J, E Koussoubé, D Roth, R Pierotti, M Hossain, and K Falb. 2020. 'Engaging Men to Transform Inequitable Gender Attitudes and Prevent Intimate Partner Violence: A Cluster Randomised Controlled Trial in North and South Kivu'. *BMJ Global Health* 5 (5). https://gh.bmj.com/content/5/5/e002223.abstract.

WHO. 2012. 'Understanding and Addressing Violence against Women: Intimate Partner Violence'. Geneva. http://apps.who.int/iris/bitstream/10665/77432/1/WHO_RHR_12.36_eng.pdf.

———. 2013. 'Global and Regional Estimates of Violence against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence'. Geneva.

———. 2016. 'Ethical and Safety Recommendations for Intervention Research on Violence against Women'. Geneva. http://www.who.int/reproductivehealth/publications/violence/intervention-research-vaw/en/.

World Bank. 2020. 'Implementation of Indashyikirwa 2.0'. Unpublished rapid ex-post implementation review.

Appendix

Table A1: Baseline Sample Characteristics and Balance across 3 groups

		(1) Control couples		(2) Spillover couples		(3) Treated couples	Normalized differences		
VARIABLE	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3
Baseline IPV rates									
Experienced any emotional violence in last 12m	798	0.47	583	0.48	625	0.51	-0.03	-0.08	-0.05
		[0.02]		[0.02]		[0.02]			
Experienced any physical violence in last 12m	793	0.40	575	0.38	619	0.40	0.05	-0.01	-0.05
		[0.02]		[0.02]		[0.02]			
Experienced any sexual violence in last 12m	795	0.45	575	0.40	618	0.44	0.09	0.01	-0.08
		[0.02]		[0.02]		[0.02]			
Demographic characteristics									
Husband's age	812	40.05	594	40.45	636	40.39	-0.04	-0.04	0.01
		[0.45]		[0.46]		[0.47]			
Wife's age	812	36.07	594	36.84	635	36.36	-0.08	-0.03	0.05
		[0.41]		[0.41]		[0.41]			
Husband's years of education	812	4.79	594	4.57	636	4.67	0.07	0.04	-0.03
		[0.12]		[0.13]		[0.13]			
Wife's years of education	812	4.57	594	4.52	636	4.44	0.02	0.04	0.03
		[0.14]		[0.13]		[0.16]			
Husband has other wives	812	0.03	594	0.02	636	0.03	0.05	-0.03	-0.08
		[0.01]		[0.01]		[0.01]			
In a civil or legal marriage	812	0.51	594	0.47	636	0.51	0.08	-0.01	-0.10
		[0.03]		[0.03]		[0.03]			
No. of children	812	3.68	594	3.80	635	3.72	-0.06	-0.02	0.04
		[0.09]		[0.10]		[0.09]			
Wife can read	812	0.77	594	0.75	636	0.74	0.05	0.08	0.02
		[0.02]		[0.02]		[0.02]			
Ever participated in rally related to family violence	812	0.63	594	0.61	636	0.61	0.04	0.04	0.01
		[0.02]		[0.02]		[0.02]			
No. of gender inequitable attitudes wife expresses	811	2.03	593	1.93	635	2.05	0.11	-0.02	-0.13
		[0.03]		[0.04]		[0.04]			
No. of gender inequitable attitudes husband expresses	811	1.47	594	1.57	634	1.47	-0.11	0.00	0.11
		[0.03]		[0.05]		[0.04]			
No. of situations wife thinks refusing sex is justified	646	2.44	482	2.34	512	2.38	0.07	0.04	-0.03
		[0.06]		[0.06]		[0.06]			
No. of situations husband thinks wife refusing sex is justified	674	2.56	484	2.54	511	2.59	0.02	-0.02	-0.03
		[0.05]		[0.07]		[0.05]			
Men's likelihood of perpetration violence- index	812	-0.00	594	-0.01	636	-0.01	0.05	0.03	-0.01
		[0.01]		[0.01]		[0.01]			

Notes: The normalized differences reflect differences in the means across the groups. Standard errors are clustered at the village level. Fixed effects using sector and number of program applicant strata are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A2: Baseline differences between VSLA couple applicant sample and random community couple sample

	(1) Random Community Couple sample		(2) VSLA Program Applicant Couple sample		(3) T-test Difference	(3) Normalized Difference
VARIABLE	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)	(1)-(2)
Experienced emotional violence in past 12 months	674	0.49	2006	0.48	0.01	0.02
	[98]	[0.02]	[98]	[0.01]		
Experienced physical violence in past 12 months	666	0.40	1987	0.39	0.00	0.01
	[98]	[0.02]	[98]	[0.01]		
Experienced sexual violence in past 12 months	658	0.45	1988	0.43	0.02	0.03
	[98]	[0.02]	[98]	[0.01]		
Number of reasons wife thinks IPV is justified	546	2.52	1676	2.30	0.22**	0.12
	[98]	[0.09]	[98]	[0.06]		
Number of reasons husband thinks IPV is justified	553	1.14	1663	0.96	0.18**	0.13
	[98]	[0.07]	[98]	[0.03]		
Number of reasons wife thinks a wife refusing sex is justified	537	2.35	1640	2.39	-0.05	-0.03
	[98]	[0.06]	[98]	[0.03]		
Number of reasons husband thinks a wife refusing sex is justified	543	2.53	1669	2.56	-0.03	-0.02
	[98]	[0.06]	[98]	[0.03]		
Number of gender inequitable attitudes woman expresses	684	2.00	2039	2.01	-0.00	-0.00
	[98]	[0.03]	[98]	[0.02]		
Number of gender inequitable attitudes respondent expresses	683	1.54	2039	1.50	0.04	0.04
	[98]	[0.03]	[98]	[0.02]		
Men's age	686	43.22	2042	40.27	2.95***	0.28
	[98]	[0.41]	[98]	[0.29]		
Wife's age	686	38.98	2041	36.38	2.59***	0.27
	[98]	[0.38]	[98]	[0.26]		
Wife reports husband drinks too much	426	1.05	1243	0.94	0.12*	0.11
	[98]	[0.05]	[98]	[0.03]		
Wife's years of education	686	3.89	2042	4.51	-0.62***	-0.21
	[98]	[0.14]	[98]	[0.09]		
Husband's years of educaiton	686	4.10	2042	4.69	-0.59***	-0.19
	[98]	[0.14]	[98]	[0.08]		
In a civil or legal marriage	686	0.49	2042	0.50	-0.01	-0.02
	[98]	[0.02]	[98]	[0.02]		
A member of your household skipped meal in last week	686	0.45	2042	0.40	0.05**	0.09
	[98]	[0.02]	[98]	[0.02]		

Figure A1: Sample by treatment arm throughout the study

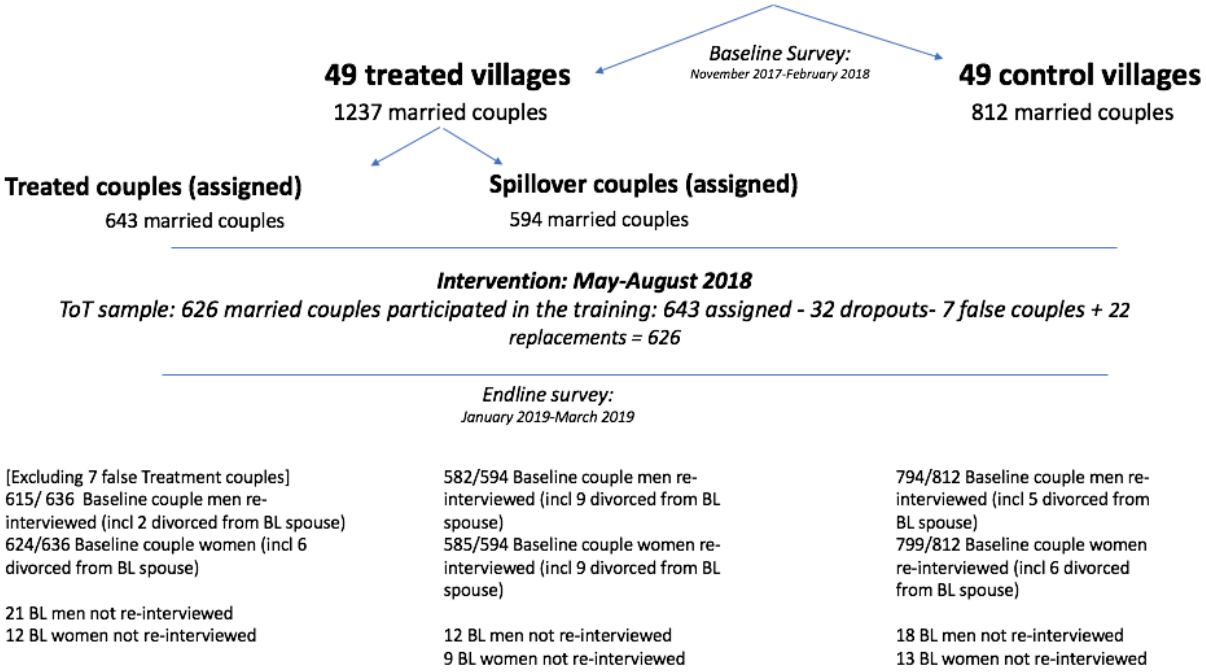


Figure A2: Indashyikirwa 2 couples curriculum

Session #1: Starting the Journey Together
- Part 1: Who We Are and Why We're Here
- Part 2: Great Expectations
- Part 3: Creating a Safe Space
- Part 4: Introducing the Take-Home Exercise

Session #2: How We Change
- Part 1: Stages of Change
- Part 2: Motivators and Barriers to Change

Session #3: It's All about Power
- Part 1: Types of Power
- Part 2: Power and *Indashyikirwa*

Session #4: Power in Our Lives
- Part 1: Our Experiences of Power
- Part 2: Who Has Power?

Session #5: G is for Gender
- Part 1: Gender Lifelines
- Part 2: The 24-hour Day

Session #6: Rights and Reality
- Part 1: Human Rights and Reality
- Part 2: What the Law Says

Session #7: GBV- The Basics
- Part 1: What is Gender-Based Violence
- Part 2: Effects of GBV

Session #8: Understanding 'Power Over'
- Part 1: The New Planet
- Part 2: 'Power Over' in Our Relationships

Session #9: Gender, Power, and Sexuality
- Part 1: Double Standards
- Part 2: The Truth about Our Sexuality

Session #10: Common Triggers of GBV in Couples
- Part 1: Identifying Triggers
- Part 2: Understanding Triggers

Session #11: The Children are Our Future
- Part 1: The Impact of GBV in Couples on Children
- Part 2: Leaving a Legacy of Non-Violence

Session #12: Pausing for Reflection
- Part 1: Participant Teach-Back
- Part 2: Personal Reflection

Session #13: What Makes a Healthy Relationship?
- Part 1: Characteristics of a Healthy Relationship
- Part 2: Feelings into Actions

Session #14: Building the Foundations of a Healthy Relationship
- Part 1: Sharing Positive Time
- Part 2: The Power of Communication

Session #15: Managing Triggers of GBV
- Part 1: What We Say and How We Say It
- Part 2: Conversation in a Conflict

Session #16: Balancing Economic Power
- Part 1: Economic Power in Our Relationships
- Part 2: The Gender and VSL 'Fishbowl'

Session #17: Reducing Excessive Use of Alcohol
- Part 1: Throwing the Drunk Ball
- Part 2: Strategies for Reducing Alcohol Consumption

Session #18: Reflecting on Our Journey So Far
- Part 1: Participant Teach-Back
- Part 2: Personal Reflection

Session #19: Our Community, Our Responsibility
- Part 1: Circles of Influence
- Part 2: Support or Interference?

Session #20: Providing Empowering Response
- Part 1: Moving Beyond Shame and Stigma
- Part 2: Fostering 'Power Within'

Session #21: Simple Actions, Big Impact
- Part 1: Instruct, Inform or Question?
- Part 2: Starting the Conversation

Session #22: Committing to Change!
- Part 1: Roles of an Activist
- Part 2: Making Commitments

Table A3: Treatment effects on IPV-related outcomes

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
IPV-related outcomes					
Women: experienced husband's controlling behaviour in last 6m	0.59***	0.03	0.10***	0.01	1,919
	(0.02)	(0.03)	(0.03)		
Women: reported having bruises, sprains or other minor injuries from IPV	0.35***	0.04	0.04	0.26	733
	(0.03)	(0.04)	(0.04)		
Women: reported having injuries from sexual IPV	0.19***	0.08**	0.09***	0.00	926
	(0.02)	(0.03)	(0.03)		
Men: experienced emotional IPV from wife in past 6months	0.60***	0.08***	0.04	0.02	1,965
	(0.02)	(0.03)	(0.03)		
Men: experienced physical IPV from wife in past 6months	0.29***	0.00	0.07***	0.09	1,959
	(0.02)	(0.03)	(0.03)		

Table A4: Other violence-related indicators

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Other measures of violence					
Men: how many times has a man or woman been badly beaten in your village in the past 6 months?	2.42***	0.53**	0.70**	0.01	1,991
	(0.16)	(0.22)	(0.31)		
Men: Number of physical violence incidents against women in village in past 6 months	1.88***	0.44**	0.47*	0.91	1,991
	(0.14)	(0.19)	(0.25)		
Enumerator: woman shows signs of physical pain or injury	-0.00	0.00	0.01**	0.01	2,008
	(0.00)	(0.00)	(0.00)		
Women: Used physical violence against children in past 6 months	0.65***	-0.04	-0.00	0.22	2,042
	(0.02)	(0.03)	(0.03)		

Table A5: Economic outcomes and legal knowledge questions

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Economic activity and legal knowledge					
Household bought land in past 12 months	0.16***	0.03	-0.03	0.01	2,008
	(0.01)	(0.02)	(0.02)		
Number of types of assets household owns (out of 6)	3.08***	0.09	0.15**	0.07	2,008
	(0.05)	(0.07)	(0.07)		
Total share of hh assets that wife personally owns	0.30***	0.01	-0.02	0.24	2,007
	(0.01)	(0.02)	(0.01)		
Men: bought tobacco in past 7 days	0.08***	-0.00	-0.01	0.94	1,991
	(0.01)	(0.01)	(0.02)		
Men: bought alcohol in past 7 days	0.48***	0.04	0.03	0.2	1,991
	(0.02)	(0.03)	(0.03)		
Women: number of correct gender law knowledge questions	1.00***	-0.01	-0.02	0.63	1,923
	(0.02)	(0.03)	(0.03)		
Men: number of correct gender law knowledge questions	1.01***	0.08**	-0.02	0.03	1,938
	(0.02)	(0.04)	(0.03)		
Couple got married in a civil ceremony in past 12 months	0.51***	0.36***	0.05	0.00	179
	(0.07)	(0.08)	(0.12)		

Notes: Strata fixed effects included. Standard errors clustered at village level. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of direct treatment and spillover effects, and differences between effects on trained and spillover couples. Outcomes are listed on the left. Column 1 shows the control mean. Column 2 shows the mean, standard error and p-value for the basic treatment effect calculated between treated couples and control couples in control villages. Column 3 shows treatment effects for spillover couples compared to pure control couples. Column 4 reports p-values for the difference between effects on trained and spillover couples. The unit of observation is individual level.

Table A6: Economic outcomes

VARIABLES	Current account balance - w	Current account balance - m	Personally own land - w	Total no. of assets household owns	Total no. of assets wife personally owns	Ran non-farm business in past week - w	Worked as an employee for a wage - w	Worked as unpaid apprentice past week - w	Involved in a revenue generating activity past 12m - w	Involved in a revenue generating activity past 12m - m
Treated couple	17,397.45	22,490.33	0.00	0.31*	0.12	0.00	-0.01	0.05**	-0.01	-0.01
	(15,441.15)	(18,738.85)	(0.03)	(0.16)	(0.12)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Spillover couple	4,743.78	-5,870.29	-0.01	0.30*	-0.00	-0.01	0.00	0.02	-0.03	-0.01
	(12,505.40)	(16,700.20)	(0.03)	(0.17)	(0.12)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Control mean	47,880.24***	13,634.56***	0.80***	5.81***	1.79***	0.24***	0.47***	0.22***	0.46***	0.39***
	(8,548.37)	(11,963.61)	(0.02)	(0.12)	(0.08)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Observations	278	567	1,964	2,042	2,042	2,008	2,006	2,008	2,008	1,991
R-squared	0.10	0.05	0.04	0.04	0.05	0.04	0.02	0.03	0.04	0.02
Treated vs. spillover	0.44	0.17	0.77	0.97	0.33	0.57	0.49	0.23	0.43	0.82

Table A7: Second order beliefs: impact on beliefs about IPV prevalence

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Second order beliefs about community norms - what share of men in the village do you believe...					
Women: think wife should obey husband	65.98***	-3.13**	-1.52	0.06	2,007
	(0.89)	(1.45)	(1.42)		
Women: get jealous when wife speaks with other men	57.23***	0.45	1.34	0.46	2,004
	(1.01)	(1.53)	(1.57)		
Women: kick/drag/beat up wife	32.45***	3.77***	1.43	0.02	2,005
	(0.86)	(1.32)	(1.35)		
Women: physically force wife to have sex with them	26.68***	8.72***	3.67**	0.00	1,893
	(1.02)	(1.65)	(1.63)		
Men: think wife should obey husband	55.05***	-3.11**	-0.29	0.15	1,985
	(0.93)	(1.49)	(1.44)		
Men: get jealous when wife speaks with other men	45.87***	0.46	1.02	0.58	1,982
	(1.15)	(1.76)	(1.61)		
Men: kick/drag/beat up wife	18.88***	5.01***	2.55***	0.00	1,986
	(0.65)	(1.06)	(0.88)		
Men: physically force wife to have sex with them	14.03***	8.63***	2.26**	0.00	1,847
	(0.73)	(1.34)	(1.11)		

Notes: Strata fixed effects included. Standard errors clustered at village level. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of direct treatment and spillover effects, and differences between effects on trained and spillover couples. Outcomes are listed on the left. Column 1 shows the control mean. Column 2 shows the mean, standard error and p-value for the basic treatment effect calculated between treated couples and control couples in control villages. Column 3 shows treatment effects for spillover couples compared to pure control couples. Column 4 reports p-values for the difference between effects on trained and spillover couples. The unit of observation is individual level.

Table A8: Exploratory descriptive analysis on mechanism: heterogeneous treatment effects by endline characteristics

OUTCOME	(1) Treated couple	(2) Treated couple x characteristic	(3) Spillover couple	(4) Spillover couple x characterisitc
Endline characteristic: Women: want more decision-making input on household expenses				
Emotional IPV	0.07**	-0.07	0.07**	-0.07
	(0.03)	(0.05)	(0.03)	(0.05)
Physical IPV	0.03	0.01	0.07***	0.07
	(0.02)	(0.06)	(0.02)	(0.06)
Sexual IPV	0.08***	0.05	0.12***	0.12**
	(0.03)	(0.05)	(0.03)	(0.05)
Endline characteristic: Women: want more decision-making input on childbearing				
Emotional IPV	0.06*	0.02	0.05*	-0.01
	(0.03)	(0.06)	(0.03)	(0.06)
Physical IPV	0.02	0.08	0.10***	-0.00
	(0.03)	(0.08)	(0.03)	(0.09)
Sexual IPV	0.07**	0.20***	0.15***	0.08
	(0.03)	(0.07)	(0.03)	(0.08)
Endline characteristic: Women: Believe they hold more **progressive** gender attitudes than most others in their				
Emotional IPV	0.05	0.07	0.02	0.13**
	(0.03)	(0.05)	(0.03)	(0.05)
Physical IPV	0.07*	-0.03	0.10***	0.04
	(0.04)	(0.05)	(0.03)	(0.05)
Sexual IPV	0.13***	-0.07	0.20***	-0.07
	(0.03)	(0.05)	(0.03)	(0.06)
Endline characteristic: Men: Believe they hold more **conservative** gender attitudes than most others in their village				
Emotional IPV	0.07*	0.02	0.08**	-0.02
	(0.04)	(0.05)	(0.04)	(0.05)
Physical IPV	0.01	0.10*	0.08**	0.07
	(0.04)	(0.05)	(0.04)	(0.05)
Sexual IPV	0.09***	0.02	0.17***	-0.00
	(0.04)	(0.05)	(0.04)	(0.05)

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in brackets. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of the treatment and spillover effects interacted with endline variables associated with a backlash against women's empowerment and men's identity. The relevant endline characteristic is listed above each set of regressions for the 3 violence outcome variables. Columns 1 and 3 show the coefficient for treated and spillover couples respectively, and columns 2 and 4 show the interaction coefficient. We control for baseline emotional, physical or sexual violence in the associated ANCOVA regression.

Table A9: Heterogeneous treatment effects by baseline characteristics

OUTCOME	(1) Treated couple	(2) Treated couple x characteristic	(3) Spillover couple	(4) Spillover couple x characterisitc
Backlash mechanism: Heterogeneous treatment effects by baseline identity and perceived norms				
Baseline characteristic: Men's No. of baseline gender inequitable attitudes				
Emotional IPV	0.07***	-0.01	0.06***	0.03
	(0.03)	(0.03)	(0.02)	(0.03)
Physical IPV	0.05**	0.01	0.11***	0.00
	(0.02)	(0.03)	(0.02)	(0.03)
Sexual IPV	0.10***	0.04	0.16***	0.01
	(0.03)	(0.03)	(0.02)	(0.03)
Baseline characteristic: Men's No. of baseline reasons believe IPV is justified				
Emotional IPV	0.08***	-0.00	0.06**	0.01
	(0.03)	(0.02)	(0.03)	(0.02)
Physical IPV	0.03	-0.02	0.10***	-0.04
	(0.03)	(0.02)	(0.03)	(0.02)
Sexual IPV	0.10***	-0.01	0.16***	-0.01
	(0.03)	(0.02)	(0.03)	(0.02)
Baseline characteristic: Village mean men's No. of baseline gender inequitable attitudes				
Emotional IPV	0.06***	0.12	0.06**	0.06
	(0.02)	(0.12)	(0.02)	(0.12)
Physical IPV	0.04*	-0.06	0.11***	-0.13
	(0.02)	(0.10)	(0.02)	(0.13)
Sexual IPV	0.10***	-0.05	0.16***	0.10
	(0.03)	(0.13)	(0.02)	(0.15)
Baseline characteristic: Village mean men's No. of baseline reasons believe IPV is justified				
Emotional IPV	0.07***	0.17**	0.07***	0.09
	(0.02)	(0.07)	(0.02)	(0.07)
Physical IPV	0.04*	-0.03	0.11***	0.06
	(0.02)	(0.06)	(0.02)	(0.07)
Sexual IPV	0.10***	-0.06	0.16***	-0.04
	(0.03)	(0.08)	(0.02)	(0.07)
Baseline characteristic: Village mean baseline emotional IPV prevalence				
Emotional IPV	0.07***	-0.03	0.06***	0.12
	(0.02)	(0.22)	(0.02)	(0.20)
Baseline characteristic: Village mean baseline physical IPV prevalence				
Physical IPV	0.05**	-0.06	0.12***	-0.00
	(0.02)	(0.21)	(0.02)	(0.24)
Baseline characteristic: Village mean baseline sexual IPV prevalence				
Sexual IPV	0.10***	-0.04	0.17***	0.01
	(0.03)	(0.22)	(0.02)	(0.19)
Baseline characteristic: No. of close friends in village that women listed				
Emotional IPV	0.08***	0.00	0.07***	0.00
	(0.03)	(0.01)	(0.02)	(0.01)
Physical IPV	0.05*	0.01	0.11***	0.01*
	(0.02)	(0.01)	(0.02)	(0.01)
Sexual IPV	0.10***	-0.01	0.17***	0.00
	(0.03)	(0.01)	(0.03)	(0.01)
Baseline characteristic: No. of close friends in village that men listed				
Emotional IPV	0.07***	0.01	0.07***	0.00
	(0.03)	(0.01)	(0.02)	(0.01)
Physical IPV	0.05**	0.01	0.12***	-0.00
	(0.02)	(0.01)	(0.02)	(0.01)
Sexual IPV	0.10***	-0.00	0.17***	-0.00
	(0.03)	(0.01)	(0.03)	(0.01)

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in brackets. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of the treatment and spillover effects interacted with baseline variables. The relevant baseline characteristic is listed above each set of regressions for the 3 violence outcome variables. Columns 1 and 3 show the coefficient for treated and spillover couples respectively, and columns 2 and 4 show the interaction coefficient. We control for baseline emotional, physical or sexual violence in the associated ANCOVA regression.

Table A10: Heterogeneous treatment effects by baseline violence experience (by each violence type) and escalation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Endline emotional IPV		Endline physical IPV		Endline sexual IPV		Endline physical &/or sexual IPV	Endline sexual IPV
VARIABLES	No: BL emotional violence	Yes: BL emotional violence	No: BL physical violence	Yes: BL physical violence	No: BL sexual violence	Yes: BL sexual violence	Yes: BL emotional violence	Yes: BL physical violence
Spillover couple	0.11***	0.02	0.17***	0.02	0.21***	0.10***	0.08**	0.06***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.02)
Treated couple	0.08**	0.06*	0.04	0.05	0.11***	0.10***	0.06**	0.03
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)
Observations	1,012	929	1,167	744	1,094	833	1,163	969
R-squared	0.07	0.02	0.06	0.04	0.05	0.04	0.00	0.01

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in brackets. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of the treatment and spillover effects by baseline experience of violence. Column 1 shows treatment effects on emotional IPV for women who report not having experienced emotional IPV at baseline. Column 2 shows emotional IPV impacts for women who reported experiencing emotional IPV at baseline. Columns 3-4 and 5-6 report the equivalent for physical and sexual IPV respectively. Columns 7 and 8 show 'escallation' of violence type from baseline to endline. Column 7 shows impact on physical and/or sexual IPV for women who at basleine reported experiencing emotional IPV. Column 8 shows sexual IPV impact for women who at baseline reported physical IPV. The unit of observation is individual level, and is calculated using a dummy aggregating whether a wife reported experiencing any of each of the forms of emotional, physical or sexual violence.

Table A11: Heterogeneous treatment effects by baseline any violence experience

	(1)	(2)	(3)
VARIABLES	Experienced at least 1 form of emotional violence in last 6m	Experienced at least 1 form of physical violence in last 6m	Experienced at least 1 form of sexual violence in last 6m
Spillover couple	0.19***	0.21***	0.28***
	(0.06)	(0.06)	(0.05)
Treated couple	0.05	0.00	0.13***
	(0.06)	(0.04)	(0.05)
Experienced any violence 12m before baseline	0.20***	0.10***	0.16***
	(0.04)	(0.03)	(0.04)
Treated couple x any violence at baseline	0.02	0.05	-0.04
	(0.06)	(0.05)	(0.05)
Spillover couple x any violence at baseline	-0.16**	-0.13**	-0.15**
	(0.06)	(0.06)	(0.07)
Baseline experienced at least 1 form of emotional violence in last 12m	0.14***		
	(0.02)		
Baseline Experienced at least 1 form of physical violence in last 12m		0.29***	
		(0.03)	
Baseline Experienced at least 1 form of sexual violence in last 12m			0.28***
			(0.02)
Observations	1,912	1,885	1,905
R-squared	0.10	0.13	0.14

Table A12: Heterogeneous treatment effects by baseline characteristics

OUTCOME	(1) Treated couple	(2) Treated couple x characteristic	(3) Spillover couple	(4) Spillover couple x characterisitc
Heterogeneous treatment effects by baseline-reported variables				
Baseline characteristic: Married in the 8yrs before 1994 (compared to married in 8yrs after)				
Emotional IPV	0.13**	-0.13	0.07	0.10
	(0.05)	(0.09)	(0.05)	(0.09)
Physical IPV	0.04	0.04	0.06	0.27***
	(0.05)	(0.09)	(0.05)	(0.09)
Sexual IPV	0.11**	0.01	0.12**	0.23**
	(0.05)	(0.08)	(0.05)	(0.11)
Baseline characteristic: Bride price was paid				
Emotional IPV	0.08**	-0.03	0.03	0.09*
	(0.03)	(0.05)	(0.03)	(0.05)
Physical IPV	0.07**	-0.07	0.09**	0.07
	(0.03)	(0.05)	(0.04)	(0.06)
Sexual IPV	0.12***	-0.05	0.11***	0.13**
	(0.03)	(0.05)	(0.04)	(0.05)

Table A13: Heterogeneous treatment effects by baseline norms and mental wellbeing indicators

OUTCOME	(1) Treated couple	(2) Treated couple x characteristic	(3) Spillover couple	(4) Spillover couple x characterisitc
Heterogeneous treatment effects by alternative mechanisms				
Interaction variable: Endline - baseline unequal norms perception index				
Emotional IPV	0.07***	-0.04*	0.06***	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Physical IPV	0.05*	0.01	0.11***	-0.01
	(0.02)	(0.03)	(0.02)	(0.02)
Sexual IPV	0.10***	-0.01	0.16***	-0.01
	(0.03)	(0.02)	(0.02)	(0.02)
Baseline characteristic: No. of men's negative experiences during the 94 conflict				
Emotional IPV	0.07***	-0.02	0.07***	-0.02
	(0.03)	(0.02)	(0.02)	(0.02)
Physical IPV	0.05*	0.02	0.11***	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Sexual IPV	0.10***	-0.00	0.17***	-0.01
	(0.03)	(0.02)	(0.02)	(0.02)
Baseline characteristic: Men's depression index				
Emotional IPV	0.07***	-0.05	0.06***	-0.04
	(0.03)	(0.04)	(0.02)	(0.04)
Physical IPV	0.05*	0.00	0.11***	-0.00
	(0.02)	(0.03)	(0.02)	(0.04)
Sexual IPV	0.10***	-0.02	0.17***	-0.05
	(0.03)	(0.04)	(0.02)	(0.04)
Baseline characteristic: Men's childhood witnessing physical IPV against his mother				
Emotional IPV	0.07*	-0.00	0.06**	0.02
	(0.04)	(0.05)	(0.03)	(0.05)
Physical IPV	0.04	0.02	0.12***	-0.01
	(0.03)	(0.05)	(0.03)	(0.05)
Sexual IPV	0.10***	0.01	0.17***	-0.00
	(0.03)	(0.06)	(0.03)	(0.06)
Baseline characteristic: Men's self-reported excessive alcohol consumption				
Emotional IPV	0.09***	-0.05	0.10***	-0.10*
	(0.03)	(0.05)	(0.03)	(0.05)
Physical IPV	0.04	0.03	0.10***	0.05
	(0.03)	(0.06)	(0.03)	(0.06)
Sexual IPV	0.11***	-0.01	0.18***	-0.03
	(0.03)	(0.06)	(0.03)	(0.06)
Baseline characteristic: Women's reports of husband's excessive alcohol consumption				
Emotional IPV	0.12***	-0.12**	0.11***	-0.12**
	(0.03)	(0.05)	(0.03)	(0.05)
Physical IPV	0.05**	-0.02	0.16***	-0.12*
	(0.03)	(0.05)	(0.03)	(0.06)
Sexual IPV	0.12***	-0.06	0.19***	-0.06
	(0.03)	(0.06)	(0.03)	(0.06)

Robustness checks

Table A14: Differential refusal rates to answer any question under each violence type, controlling for baseline refusal to answer

VARIABLES	(1) Endline emotional vio: refused at least 1 ques	(2) Endline physical vio: refused at least 1 ques	(3) Endline sexual vio: refused at least 1 ques
Treated couple	-0.00	-0.02*	-0.00
	(0.01)	(0.01)	(0.01)
Spillover couple	0.01	0.03**	0.01
	(0.01)	(0.01)	(0.01)
Baseline emotional vio: refused at least 1 ques	0.15***		
	(0.05)		
Baseline physical vio: refused at least 1 ques		0.19***	
		(0.05)	
Baseline sexual vio: refused at least 1 ques			0.17***
			(0.04)
Observations	2,042	2,042	2,042
R-squared	0.04	0.06	0.04

Notes: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A15: Recoding missing refuse-to-answer violence questions as 'yes' and 'no'

VARIABLES	(1) Emotional vio: all refuses as 'yes'	(2) Emotional vio: all refuses as 'no'	(3) Physical vio: all refuses as 'yes'	(4) Physical vio: all refuses as 'no'	(5) Sexual vio: all refuses as 'yes'	(6) Sexual vio: all refuses as 'no'
Treated couple	0.07***	0.08***	0.04	0.06**	0.10***	0.10***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Spillover couple	0.07***	0.06**	0.11***	0.08***	0.15***	0.15***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Observations	2,008	1,995	2,008	2,008	2,008	2,008
R-squared	0.03	0.04	0.03	0.02	0.03	0.03
Strata fixed effects	X	X	X	X	X	X
ANCOVA baseline controls						
PDS controls						

Table A16: Additional specifications: Single difference, with controls, and probit

	Single difference			Post-double selection lasso controls			Probit		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
VARIABLES	Emotional violence	Physical violence	Sexual violence	Emotional violence	Physical violence	Sexual violence	Emotional violence	Physical violence	Sexual violence
Treated couple	0.08***	0.04*	0.10***	0.07**	0.05**	0.10***	0.22***	0.14**	0.28***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.08)	(0.07)	(0.07)
Spillover couple	0.07***	0.10***	0.15***	0.07***	0.11***	0.17***	0.20***	0.34***	0.46***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.07)	(0.07)	(0.07)
Baseline emotional violence				0.21***			0.64***		
				(0.02)			(0.07)		
Basleine physical violence					0.32***			0.88***	
					(0.02)			(0.07)	
Baseline sexual violence						0.32***			0.87***
						(0.02)			(0.06)
Observations	1,974	1,960	1,972	1,941	1,911	1,927	1,941	1,911	1,927
R-squared	0.03	0.02	0.03	0.09	0.13	0.13			
Strata fixed effects	X	X	X	X	X	X	X	X	X
ANCOVA baseline controls				X	X	X	X	X	X
PDS controls				X	X	X			

Table A17: Additional specifications: IV treatment-on-the-treated

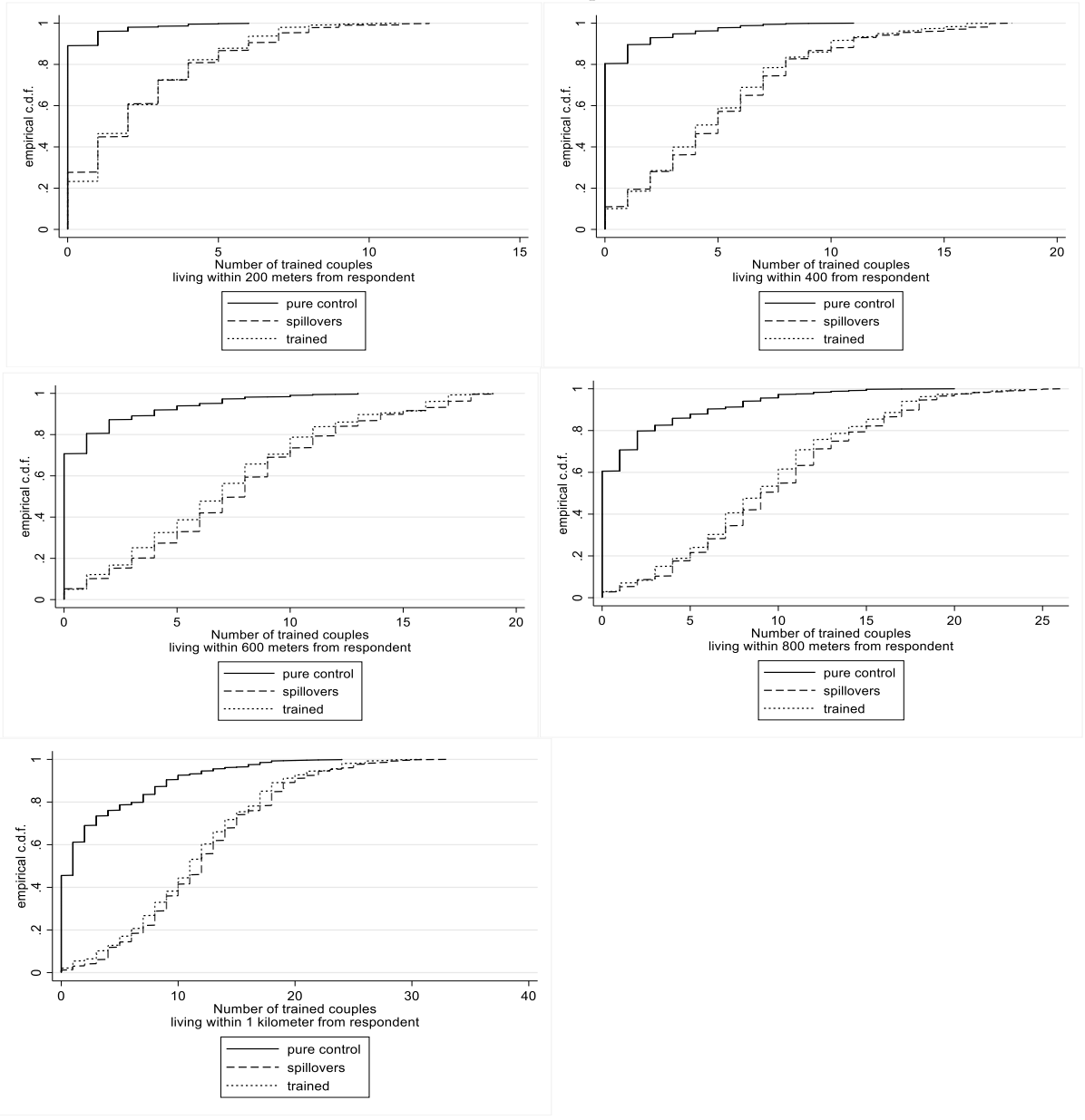
	(1)	(2)	(3)	(4)	(5)	(6)
	First stage	Second stage	First stage	Second stage	First stage	Second stage
VARIABLES	Treated couple compliers	Emotional violence	Treated couple compliers	Physical violence	Treated couple compliers	Sexual violence
Spillover couple	0.04***	0.06***	0.04***	0.10***	0.04***	0.15***
	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.03)
Treated couple assignment	0.96***		0.96***		0.96***	
	(0.01)		(0.01)		(0.01)	
Treated couple compliers		0.08***		0.05*		0.10***
		(0.03)		(0.03)		(0.03)
Constant	-0.05**	0.64***	-0.05*	0.33***	-0.05*	0.55***
	(0.03)	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)
Observations	1,974	1,974	1,960	1,960	1,972	1,972
R-squared	0.89	0.03	0.89	0.02	0.89	0.03
Strata fixed effects	X	X	X	X	X	X
ANCOVA baseline controls						
PDS controls						

Table A18: Correlation matrix of IPV self-reported experience, signs of injury, attitudes, and husband's likelihood of perpetrating IPV

	Emotional violence	Physical violence	Sexual violence	No. of reasons husband likely to be violent	Wife was physically violent to husband	Wife emotionally abused husband	Wife had signs of physical pain or injury	Women: No. of reasons IPV is justified	Women: No. of gender inequitable attitudes	Men: No. of reasons IPV is justified	Men: No. of gender inequitable attitudes
Emotional violence	1										
Physical violence	0.39***	1									
Sexual violence	0.37***	0.41***	1								
No. of reasons husband likely to be violent	0.16***	0.21***	0.17***	1							
Wife was physically violent to husband	0.13***	0.22***	0.13***	0.36***	1						
Wife emotionally abused husband	0.16***	0.15***	0.10***	0.27***	0.32***	1					
Wife had signs of physical pain or injury	0.03	0.06*	0.05	0	0.07**	0.04	1				
Women: No. of reasons IPV is justified	-0.01	0.14***	0.03	0.04	0.06*	-0.02	-0.04	1			
Women: No. of gender inequitable attitudes	-0.04	0.07*	0	0.04	0.04	-0.01	-0.02	0.41***	1		
Men: No. of reasons IPV is justified	0.08**	0.16***	0.06*	0.33***	0.28***	0.08**	-0.01	0.14***	0.10***	1	
Men: No. of gender inequitable attitudes	0.05	0.14***	0.06*	0.19***	0.21***	-0.02	0.01	0.15***	0.14***	0.36***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure A3: How 'pure' are pure control couples? Number of treated couples living within various distances of couples.



- 89% of pure control couples do not have any trained couple neighbors within 200meters
- 80% of pure control couples do not have any trained couple neighbors within 400meters
- 71% of pure control couples do not have any trained couple neighbors within 600meters
- 60% of pure control couples do not have any trained couple neighbors within 800meters
- 45% of pure control couples do not have any trained couple neighbors within 1000meters

Table A19: Treatment effects excluding respondents who live in unconfirmed or different village to VSLA-village of treatment assignment

	(1)	(2)	(3)
	Emotional	Physical	Sexual
VARIABLES	IPV	IPV	IPV
Treated couple	0.11***	0.07**	0.08**
	(0.03)	(0.03)	(0.04)
Spillover couple	0.10***	0.13***	0.16***
	(0.03)	(0.04)	(0.04)
Observations	1,079	1,060	1,078
R-squared	0.09	0.10	0.12

Table A20: Treatment effects excluding control couples living near treated or spillover couples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Emotional IPV				Physical IPV				Sexual IPV			
	200m	400m	600m	800m	200m	400m	600m	800m	200m	400m	600m	800m
VARIABLES	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion	exclusion
Treated couple	0.084***	0.073**	0.065**	0.062**	0.061**	0.058**	0.056**	0.072***	0.105***	0.110***	0.112***	0.105***
	(0.027)	(0.029)	(0.030)	(0.030)	(0.024)	(0.026)	(0.026)	(0.026)	(0.025)	(0.027)	(0.027)	(0.028)
Spillover couple	0.047	0.052*	0.053*	0.051*	0.129***	0.139***	0.129***	0.144***	0.188***	0.179***	0.176***	0.173***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.030)	(0.028)	(0.026)	(0.026)	(0.029)	(0.028)	(0.027)	(0.028)
Observations	1552	1660	1646	1594	1539	1646	1631	1578	1545	1654	1639	1587
Adjusted R-squared	0.066	0.061	0.063	0.061	0.128	0.121	0.122	0.125	0.123	0.123	0.127	0.128

Truth or reporting analysis

Table A21: Increase in men's reported alcohol consumption across all groups

VARIABLES	(1) No. of times respondent had one drink in past month - m	(2) Drank alcohol to excess in past month? - m	(3) Partner drinks too much - w	(4) Was partner drunk before last quarrel or fight - w
Treated couple	0.25	0.05*	0.06**	0.05*
	(0.41)	(0.03)	(0.02)	(0.03)
Spillover couple	0.86**	0.03	0.11***	0.07**
	(0.36)	(0.03)	(0.03)	(0.03)
Observations	2,042	1,239	2,030	1,965
R-squared	0.01	0.03	0.02	0.02
Treated vs. spillover	0.13	0.43	0.09	0.62

Standard errors in parentheses, clustered at village level

*** p<0.01, ** p<0.05, * p<0.1

Table A22: Differential baseline IPV reporting by baseline lottery rank

VARIABLES	(1) Baseline: Experienced emotional violence in last 12m	(2) Baseline: Experienced physical violence in last 12m	(3) Baseline: Experienced sexual violence in last 12m
Treated couple	0.05*	0.02	-0.04
	(0.03)	(0.03)	(0.03)
Spillover couple	0.02	0.01	-0.05
	(0.03)	(0.04)	(0.04)
Couples lottery rank (1-30)	0.00	-0.01	0.01
	(0.01)	(0.01)	(0.01)
Spillover couple x lottery rank	-0.00	0.00	-0.01
	(0.01)	(0.01)	(0.01)
Treated couple x lottery rank	0.01	0.02	-0.02
	(0.01)	(0.01)	(0.01)
Observations	2,001	1,982	1,983
R-squared	0.00	0.00	0.00
Treated vs. spillover p-value	0.55	0.27	0.16

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in brackets. *** p<0.01, ** p<0.05, * p<0.1. The bottom row shows p-value results from a t-test of the difference in coefficients between the treatment dummy interaction coefficients.

Table A23: Differential baseline IPV reporting by baseline knowledge of whether the program was about gender

VARIABLES	(1) Baseline: Experienced emotional violence in last 12m	(2) Baseline: Experienced physical violence in last 12m	(3) Baseline: Experienced sexual violence in last 12m
At baseline: knew training was about families and gender	-0.05	0.07*	0.03
	(0.05)	(0.04)	(0.04)
Constant	0.53***	0.38***	0.38***
	(0.10)	(0.11)	(0.05)
Observations	944	934	935
R-squared	0.04	0.04	0.04

Table A24: Impact by IPV question (subjective and objective)

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N	(6) More subjective than objective interpretation in kinyarwanda? (Y/N)
Impact on each violence outcome: In past 6 months, husband has						
publically humiliated respondent	0.30***	0.12***	0.11***	0.67	1,962	Y
	(0.01)	(0.03)	(0.02)			
insulted respondent	0.60***	0.06**	0.07***	0.89	1,966	N
	(0.02)	(0.03)	(0.02)			
slapped	0.24***	0.05*	0.09***	0.10	1,977	N
	(0.02)	(0.02)	(0.02)			
twisted respondent's arm	0.15***	0.04**	0.04**	0.96	1,968	N
	(0.01)	(0.02)	(0.02)			
punched respondent	0.14***	0.07***	0.09***	0.37	1,974	N
	(0.01)	(0.02)	(0.02)			
kicked or beaten respondent	0.15***	0.05**	0.09***	0.07	1,981	N
	(0.01)	(0.02)	(0.02)			
strangled or burned respondent	0.09***	-0.01	0.00	0.58	1,969	N
	(0.01)	(0.02)	(0.02)			
threatened to attack with a weapon or knife	0.06***	0.02	0.03**	0.61	1,972	N
	(0.01)	(0.01)	(0.01)			
forced respondent to have sex when she did not want to	0.35***	0.08***	0.14***	0.05	1,970	N
	(0.02)	(0.03)	(0.03)			
respondent forced to have sex because was afraid him	0.27***	0.05**	0.13***	0.00	1,982	N
	(0.02)	(0.02)	(0.03)			
forced respondent to do degrading sex act	0.15***	0.04*	0.04*	0.93	1,962	Y
	(0.01)	(0.02)	(0.02)			

Table A25: Increase in frequency and number of types of IPV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Frequency of each type of violence: sometimes or often experienced...									Number of types of violence		
VARIABLES	slapped	arm twisted	punched	beaten	strangled	attacked with a weapon	forced to have sex	forced to have sex from fear	forced to have degading sex	types of emotional violence	types of physical violence	types of sexual violence
Treated couple	0.04**	0.01	0.03*	0.03**	0.00	0.02*	0.04	0.04	-0.00	0.18***	0.21**	0.17***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.05)	(0.09)	(0.06)
Spillover couple	0.09***	0.03*	0.04**	0.07***	0.03**	0.03***	0.07***	0.10***	0.02	0.16***	0.32***	0.30***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.09)	(0.06)
Constant	0.20***	0.07	0.08	0.04*	-0.01	-0.02**	0.21***	0.17***	0.11*	0.84***	0.85***	1.04***
	(0.06)	(0.05)	(0.06)	(0.03)	(0.01)	(0.01)	(0.05)	(0.02)	(0.06)	(0.18)	(0.20)	(0.07)
Observations	1,977	1,968	1,974	1,981	1,969	1,972	1,970	1,982	1,962	2,008	2,008	2,008
R-squared	0.03	0.02	0.04	0.03	0.02	0.02	0.02	0.03	0.02	0.03	0.03	0.03
Treated vs. spillover	0.00	0.15	0.01	0.00	0.09	0.00	0.01	0.00	0.35	0.71	0.23	0.02

Table A26: Endline refuse-to-answer at least 1 question by violence category

VARIABLES	Endline emotional vio: refused at least 1 ques	Endline physical vio: refused at least 1 ques	Endline sexual vio: refused at least 1 ques
Treated couple	-0.00	-0.02*	-0.00
	(0.01)	(0.01)	(0.01)
Spillover couple	0.01	0.03**	0.01
	(0.01)	(0.01)	(0.01)
Observations	2,042	2,042	2,042
R-squared	0.04	0.06	0.04
Treated vs. spillover p-value	0.21	0.00	0.54

Notes: Strata fixed effects included. Standard errors are clustered at village level, shown in brackets. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of the endline refusal rate by type of violence. The unit of observation is individual level, and is calculated using a dummy aggregating whether a wife refused to answer any question about experiencing any of each of the forms of emotional, physical or sexual violence. The bottom row shows p-value results from a t-test of the difference in coefficients between the treated and spillover couple

Table A27: Treatment effects on indicators of stigma and empowerment to report

VARIABLES	(1) Control mean	(2) Effect on treated couples	(3) Effect on spillover couples	(4) Treated vs spillover couple p-value	(5) N
Women: Visited a safe space in the past 6 months	0.06*	0.07**	-0.02	0.00	2,008
	(0.03)	(0.03)	(0.03)		
Women: Spoken with others about physical or sexual violence suffered in past 6 months	0.26***	0.06*	0.02	0.34	1,586
	(0.08)	(0.03)	(0.03)		
Women: (in a hypothetical scenario) would you report IPV that you witnessed to local leaders?	0.71***	0.03	0.01	0.25	2,003
	(0.06)	(0.02)	(0.02)		
Men: Visited a safe space in the past 6 months	0.11***	0.04**	-0.01	0.1	1991
	(0.01)	(0.02)	(0.02)		

Notes: Strata fixed effects included. Standard errors clustered at village level. *** p<0.01, ** p<0.05, * p<0.1. Results show OLS estimates of direct treatment and spillover effects, and differences between effects on trained and spillover couples. Outcomes are listed on the left. Column 4 reports p-values for the difference between effects on trained and spillover couples. The unit of observation is individual level.

Table A28: Correlation coefficients between IPV self-reported experience and gender attitudes

	Emotional violence	Physical violence	Sexual violence	No. of reasons husband likely to be violent	Wife was physically violent to husband	Wife emotionally abused husband	Wife had signs of physical pain or injury	Women: No. of reasons IPV is justified	Women: No. of gender inequitable attitudes	Men: No. of reasons IPV is justified	Men: No. of gender inequitable attitudes	N
Emotional violence	1											1385
Physical violence	0.39***	1										
Sexual violence	0.37***	0.41***	1									
No. of reasons husband likely to be violent	0.16***	0.21***	0.17***	1								
Wife was physically violent to husband	0.13***	0.22***	0.13***	0.36***	1							
Wife emotionally abused husband	0.16***	0.15***	0.10***	0.27***	0.32***	1						
Wife had signs of physical pain or injury	0.03	0.06*	0.05	0	0.07**	0.04	1					
Women: No. of reasons IPV is justified	-0.01	0.14***	0.03	0.04	0.06*	-0.02	-0.04	1				
Women: No. of gender inequitable attitudes	-0.04	0.07*	0	0.04	0.04	-0.01	-0.02	0.41***	1			
Men: No. of reasons IPV is justified	0.08**	0.16***	0.06*	0.33***	0.28***	0.08**	-0.01	0.14***	0.10***	1		
Men: No. of gender inequitable attitudes	0.05	0.14***	0.06*	0.19***	0.21***	-0.02	0.01	0.15***	0.14***	0.36***	1	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A29: Correlations between women's IPV experience and potentially objective indicators

	Emotional violence	Physical violence	Sexual violence	No. of reasons husband likely to be violent	Wife had signs of physical pain or injury	Wife was referred to counsellor	Wife chose to speak to counsellor
Emotional violence	1						
Physical violence	0.40**	1					
Sexual violence	0.56***	0.42**	1				
No. of reasons husband likely to be violent	0.16	0.23	0.25	1			
Wife had signs of physical pain or injury (n=6)	-0.08	0.09	0.01	-0.04	1		
Wife was referred to counsellor	0.26	0.24	0.14	0	-0.01	1	
Wife chose to speak to counsellor	0.32*	0.03	0.06	0.06	-0.08	0.44***	1

Table A30: Heterogeneous treatment effects by baseline husband working away from home

VARIABLES	(1) Endline emotional violence	(2) Endline physical violence	(3) Endline sexual violence
Spillover couple	0.07*	0.14***	0.20***
	(0.04)	(0.04)	(0.04)
Treated couple	0.07**	0.06*	0.11**
	(0.04)	(0.03)	(0.04)
Baseline: Husband worked in a different sector for 4 consecutive nights over past 3 months	-0.04	0.12**	-0.01
	(0.07)	(0.05)	(0.05)
Treated couple x Husband regularly works away from home	0.06	-0.02	0.05
	(0.08)	(0.08)	(0.08)
Spillover couple x Husband regularly works away from home	0.01	-0.18**	-0.15*
	(0.10)	(0.08)	(0.08)
Observations	1,206	1,197	1,201
R-squared	0.08	0.14	0.14
Treated vs. spillover	0.56	0.05	0.02

Appendix B – Summary of ethical context and protocols

Context

This study was conducted following several guidelines on the ethical research of IPV (Heise and Hossain 2017; Ellsberg et al. 2001; WHO 2016). We received ethics approval for the study from Oxford University and the Rwandan National Ethics Committee, and received a Research Visa from the National Institute of Statistics, Rwanda.

The Government of Rwanda partnered with the World Bank to identify and test a promising IPV prevention model- a gender transformative couples training program. If it was found to be effective, the program could form the basis of a scaled-up program that could be delivered to many more couples in the country. The couples program tested here was based on a more intensive program that was tested and promising results have since been presented in two papers (Dunkle, Stern, and Heise 2020; Chatterji et al. 2020). The *Indashyikirwa 1* couples curriculum, 90% of which is shared with Indashyikirwa 2, was carefully adapted, extensively piloted, and revised to be appropriate for the context. The positive early anecdotal findings of the other Indashyikirwa study, and the very positive findings of a similar IPV prevention program for new parents in Rwanda (Doyle et al. 2018) – noting the significant differences in design and analysis between these studies and ours– suggested the program had great promise. Before scaling up a more cost-effective version of the program that condensed one element of the program- couples activism- into the couples training curriculum delivered to ALL trained couples rather than a subset, the Government, supported by the World Bank, decided to evaluate it first and measure anticipated positive spillover effects of the program to the community. The results of this study were intended to inform any decision on scaling up the program. Similar couples-based IPV prevention programs are being delivered in other countries.

Study Protocols Summary

Consent

- Easy to understand, read aloud to respondents to not disadvantage illiterate respondents
- Made clear that: there was no penalty for refusing to participate in the survey (they would still be eligible for the program); their data was private and no one except the survey company would know their name or any other identifying information; that identifying data was kept only to follow up with them for endline; and identifying data would be kept separate from their answers to survey answers.
- Asked whether they had been coerced into answering the survey or participating in the program
- Consent forms had contact details of the research team and ethics committee in case anyone had follow-up questions.

Enumerators

- Trained in conducting sensitive gender-based violence (GBV) surveys and understanding GBV causes
- Interviews held by enumerators the same sex as respondents
- Almost all had previous experience conducting surveys related to GBV in Rwanda

- Never interviewed anyone they knew or in villages they were from
- Provided with support and debriefing from the GBV counsellors assisting with data collection to minimize the risk of vicarious trauma

Data collection

- IPV questions asked using ACASI, which increases privacy and made respondents feel more comfortable refusing to answer
- Made clear anyone could refuse any questions and end the interview at any time
- Men not asked about their perpetration of violence against women (asked about likelihood of becoming violent in set of hypothetical situations)
- Interviewed in private settings where they could not be overheard
- Consent form and introduction to the community as a survey about families and health, not about IPV or women.

Referrals and support

- Counsellors with experience supporting survivors of trauma and GBV were present at data collection sites during data collection. The counsellors were available to speak with respondents who got distressed during the interview, and could follow up with women and men afterwards to help them access services they wanted.
- Most survey sectors had women's safe spaces or One Stop Centers to support survivors of IPV with emergency shelter, medical, legal, and justice support services.
- Respondents were offered small information sheets about the nearest support services – health, medical, legal- for anyone including themselves or community members who might want help.

Data privacy

- No organization (Research team, World Bank, Government) except the survey firm had any identifying data linking respondents to their answers. Encryption key not shared outside a small team at the survey firm.
- Answers stored on encrypted data collection platform. Uploaded daily and removed from password-protected tablets.

3

Method Matters: Underreporting of Intimate Partner Violence in Nigeria and Rwanda

Method Matters: Underreporting of Intimate Partner Violence in Nigeria and Rwanda

Claire Cullen*

Abstract

We analyze the magnitude and predictors of misreporting on intimate partner and sexual violence in Nigeria and Rwanda. We randomly assigned respondents to answer questions using one of three survey methods: an indirect method (list experiment) that gives respondents anonymity; a direct, self-administered method that increases privacy; and the standard, direct face-to-face method. In Rwanda, intimate partner violence rates were double, and in Nigeria, up to 39 percent greater when measured using the list method than direct methods. Misreporting was associated with indicators often targeted in women's empowerment programs such as gender norms and female employment and education. These results suggest that standard survey methods may generate significant underestimates of the prevalence of intimate partner violence and biased correlations and treatment effect estimates.

JEL: J12, J16, K42, O15
Keywords: Gender, Domestic violence, Measurement

*Cullen: Blavatnik School of Government, University of Oxford, claire.cullen@bsg.ox.ac.uk. Acknowledgements: I thank Julien Labonne, Stefan Dercon, Arthur Alik-Lagrange, Noam Angrist, Diana Arango, Dara Kay Cohen, Cheryl Doss, Patricia Fernandes, Markus Goldstein, Harald Hinkel, Summer Lindsey, Anandi Mani, Mũthoni Ngatia, Sreelakshmi Papineni, Rachael Pierotti, Simon Quinn, Diana Contreras Suarez, Alex Twahirwa, Julia Vaillant, Marc Witte, and seminar participants at the World Bank Africa Region Gender Innovation Lab, Oxford CSAE, IFPRI, UC Berkeley, Yale EGEN, Australian Gender Economics Workshop, Kigali Impact Hub research series, and SVRI for helpful comments and suggestions. I would also like to thank the respondents for generously giving their time, and the staff at Laterite Ltd in Rwanda and TNS Global in Nigeria, and Victoria Isika, Jean Paul Ngabonziza, Maritetou Sanogo and Oluwatoyin Zakariya for their superb research assistance and data collection work. The impact evaluations through which this data was collected were conducted by the World Bank Africa Region Gender Innovation Lab and were funded by the World Bank's Nordic Trust Fund, the Swiss Development Cooperation and the Government of Rwanda, and in Nigeria, by USAID and the Umbrella Facility for Gender Equality, a World Bank Group multi-donor trust fund. The views presented in this paper are those of the author and do not represent those of the World Bank or its member countries. All remaining errors are my own. This study was approved by the Oxford University Central University Research Ethics Committee, HML IRB, and the Rwandan National Ethics Committee.

1 Introduction

It is challenging to accurately measure sensitive attitudes, behaviors, and experiences such as political preferences, prejudice, risky sex, and intimate partner violence (IPV). Often, researchers and policy makers must rely on self-reported survey data, yet such data may not be accurate. For sensitive topics in particular, there are many reasons why survey respondents may be reluctant to tell the truth, including a fear of being judged, endangered, or legally penalized (Blair, Coppock and Moor, 2018). If misreporting bias is significant and systematic, researchers and policy makers risk over or under-estimating the prevalence and welfare costs of sensitive issues, and drawing inaccurate conclusions about their causes and potential solutions.

In this paper, we compare self-reported sexual and intimate partner violence rates across three different survey methods. We assess whether survey methods that increase respondent privacy and anonymity affect IPV reporting compared to the status quo of direct face-to-face interviews.[1] In Rwanda and Nigeria, we randomly assigned women to answer questions about their experience of emotional, physical and sexual violence using one of three methods: an indirect method (list experiment) and two direct survey methods: face-to-face questions asked by an enumerator, or audio computer-assisted self-interview (ACASI) on an electronic tablet. Compared to the direct face-to-face method, the list experiment method provides additional confidentiality as respondents do not directly disclose their experience of violence. Instead, they give a response that allows researchers to calculate average prevalence across the whole sample, essentially making respondents' answers anonymous (Blair and Imai, 2012). The ACASI method affords more privacy than face-to-face methods as respondents answer questions themselves by listening to pre-recorded questions in a headset and then selecting their answer on a tablet.

Results suggest that estimated IPV prevalence varies depending on the survey method used to measure it. The most widely used method, face-to-face interview, resulted in the lowest IPV rates, followed by the more private ACASI method, and finally, the anonymous list method with the highest. In Rwanda, IPV rates were double, and in Nigeria, up to 39 percent greater when measured using the list method compared to the direct methods. There was no statistically significant difference between the list and direct methods for rates

[1] In more than 60 developing countries, Demographic and Health Surveys (DHS) have included a standardized domestic violence module where enumerators directly ask women about their experience of a set of specific emotional, physical, and sexual abuse episodes by an intimate partner. Using the face-to-face method, the WHO estimates that approximately one in three women worldwide has experienced IPV in her lifetime (García-Moreno et al., 2013). IPV prevalence varies widely, though rates are generally highest in Africa (Devries et al., 2013; Global Burden of Disease, 2017). Researchers collecting IPV survey data take steps to minimize the risk of underreporting, for example by instituting ethics protocols about privacy during interviews, asking multiple questions to give respondents several opportunities to disclose violence, and using specially-trained female enumerators to conduct women's interviews (Heise and Hossain, 2017; WHO, 2016, 2001). However, it is still assumed that face-to-face data collection may yield underestimates given the sensitivity.

of non-partner sexual violence in Rwanda or sexual IPV in Nigeria. This lack of difference for sexual violence could be explained by a lack of statistical power in Rwanda, and a Nigerian context where sexual IPV was not significantly stigmatized.

We also find that misreporting is systematic and correlated with characteristics often theorized as causes of IPV and targeted in women's empowerment programs. The characteristics that correlate differently with IPV depending on the measurement method used include indicators of women's education and employment, community gender norms, vulnerability, and marital relationship quality. For example, in Rwanda, when IPV was measured using the direct methods, there was no relationship between IPV and conservative community gender norms. However, when IPV was measured using the list method, we find a large, positive relationship between IPV and conservative norms. This suggests that women living in the most conservative villages underreport IPV when asked directly and that the list method's anonymity (but not ACASI's privacy) makes them comfortable enough to report it. Also, in both Rwanda and Nigeria, the education-IPV relationship switches depending on how IPV is measured. There is suggestive evidence of a positive relationship between IPV and women's education when IPV is measured directly, but a negative relationship when using the list method. These results suggest that the standard, direct methods currently used in most surveys risk generating significant underestimates of IPV prevalence. Further, social scientists risk estimating biased correlations and, if treatment is correlated with misreporting, estimating biased treatment effects and drawing inaccurate conclusions about IPV's causes and the effectiveness of interventions to prevent it.[2]

This paper contributes to the growing social science literature studying alternative survey methods that may reduce misreporting (Karlan and Zinman, 2012; Blattman et al., 2016; Chuang et al., 2018; Blair, Coppock and Moor, 2018; Lépine et al., 2020), including on IPV measurement (Agüero and Frisancho, 2017; Peterman et al., 2017; Joseph et al., 2017; Bulte and Lensink, 2019; Agüero et al., 2020).

Recent studies on IPV measurement have compared respondents' experience of violence across the face-to-face and list experiment methods. They found mixed results, with one study in Peru finding the same rates of IPV between face-to-face and list methods (Agüero and Frisancho, 2017) and others in Vietnam, India, and Nigeria finding a difference (Bulte and Lensink, 2019; Joseph et al., 2017; Banerjee, La Ferrara, and Orozco as cited in World

[2]A further motivation for assessing IPV measurement methods is that the status quo of asking direct face-to-face questions places a high emotional burden on both respondents and enumerators. WHO guidelines thus recommend particular protocols for IPV research, for example, ensuring respondents have access to referral services (WHO, 2016). If an alternative measurement method was identified that was less biased, logistically easier and cheaper, and more ethical than the status quo, then it would likely be adopted by more researchers and contribute to new research on IPV. This would improve our understanding of the causes of IPV and may broaden the range of interventions that could be used to reduce it (Peterman et al., 2017).

Bank, 2017).[3] However, in several of these studies, respondents were asked about their IPV experience using two methods within the same survey about similar, but slightly different, violence acts. This potentially limits comparability across methods. To avoid this issue, we conducted one of the first IPV measurement experiments that randomly assigns respondents to answer questions using only one method rather than asking respondents two overlapping questions using different methods. Also, this paper aims to contribute to the limited evidence on who misreports on sensitive questions and why, the implications of misreporting for bias, and the trade-offs between different survey methods that seek to address misreporting.

This is the first paper to compare the prevalence of sexual and intimate partner violence using three survey methods that offer respondents differing levels of privacy and anonymity, and to compare across two countries, and for women and men.[4] It is also one of the first papers to study IPV misreporting in Africa and to demonstrate that misreporting is substantial in some contexts, varies by the type of violence, and is correlated with characteristics commonly targeted in women's empowerment interventions.

The rest of this paper is organized as follows. Section 2 reviews the literature on the measurement of sensitive topics and IPV. Section 3 outlines the experimental design and describes the sample and data. Section 4 presents the results, and 5 concludes.

2 Survey methods to measure sensitive topics

There is robust evidence that people underreport the truth about sensitive topics when they are directly asked in a survey. Such research compares self-reported survey answers to observed or independently verifiable outcomes such as tax records, receipt of welfare benefits, drug prescriptions for mental illness, and voting in an abortion law referendum (Bharadwaj, Pai and Suziedelyte, 2017; Meyer and Mittag, 2019; Gottschalk and Huynh, 2010; Murray-Close and Heggeness, 2018; Rosenfeld, Imai and Shapiro, 2016).

However, reliable, unbiased observational data are rare for many sensitive topics, including IPV.[5] This is particularly true in developing countries, where there is limited available administrative data from the medical or legal services that IPV survivors may use. Further, where these services are available, the people able or choosing to access them are likely not

[3] A related study by Agüero et al. (2020) in rural Peru compares women's direct self-reported experience of IPV with a listing of IPV survivors by female community leaders. They found that community leaders significantly underreport IPV rates compared to women's own self-reports.

[4] In Rwanda, we asked men about their perpetration of emotional violence using one of the three methods. The ACASI method was only tested in Rwanda.

[5] Tourangeau and Yan (2007) define sensitive questions as those a respondent deems intrusive; where they fear the threat of disclosure and the possible consequences of giving a truthful response; or where there are social norms that imply there are socially desirable and undesirable responses.

representative of the broader population of those experiencing IPV.[6]

As a result of these administrative data limitations and concerns about underreporting on face-to-face surveys, researchers are considering alternative survey methods to minimize misreporting bias. Two popular alternative methods, ACASI and the list experiment, address slightly different motivations for misreporting. The ACASI method is primarily designed to overcome misreporting driven by shame when responding face-to-face to an enumerator, while the list experiment addresses misreporting driven by a preference for anonymity or fear of repercussions if one's answer becomes known, as well as shame. Table 1 shows these methods by their approach to questioning and administration.

Table 1: Method comparisons by questioning method and administration

	Question & answer through enumerator	Question & answer through tablet
Direct questioning method	Face-to-face (status quo)	ACASI
Indirect questioning method	List experiment	

2.1 More Privacy: Audio Computer-Assisted Self-Interview

ACASI is a method of data collection where respondents listen to pre-recorded questions through headphones and respond by selecting their answers (for example, the corresponding colored square) on a touchscreen or keypad.[7] After demonstrating how to use the technology, enumerators stand a small distance from the respondent and ensure that privacy is maintained.

ACASI eliminates the embarrassment that respondents may feel about disclosing a sensitive answer face-to-face with an enumerator, potentially minimizing underreporting (Falb et al., 2016; Jarlais et al., 1999; Jaya, Hindin and Ahmed, 2008; Langhaug, Sherr and Cowan, 2010). If shame is a key driver of underreporting, then reporting under ACASI should be greater than face-to-face. ACASI may also reduce enumerator-associated variance by providing a more uniform approach to data collection.

Several experimental studies have compared the prevalence of sensitive health behaviors between ACASI and face-to-face survey methods. Most found that respondents reported sensitive health behaviors and experiences, including IPV, at higher rates under ACASI than the face-to-face method, though the evidence is mixed (Fay et al., 1998; Jaya, Hindin and Ahmed, 2008; Langhaug, Sherr and Cowan, 2010; Mensch et al., 2008; Phillips et al., 2010; Fincher et al., 2015). In one of the few method comparisons with an objective indicator, in

[6]Palermo, Bleck and Peterman (2014) analyzed data from 24 developing countries DHSs and found that only 7 percent of women who report experiencing IPV made a report to a formal source that would then be captured in administrative data.

[7]Figure A1 in the appendix shows ACASI being used.

Brazil, Mensch et al. (2008) compared sexually transmitted infection bio-markers with face-to-face and ACASI self-reports of risky sexual behavior. They found more underreporting when respondents were interviewed face-to-face than with ACASI.

2.2 More Confidentiality and Anonymity: List Experiment

Researchers have also used the list experiment to improve data accuracy when measuring sensitive behaviors and attitudes.[8] The standard list experiment works as follows: respondents are randomly assigned to one of two groups to answer the list question. In the 'non-sensitive' or 'control' group, enumerators read respondents a set of three non-sensitive statements and ask the respondent to tell them how many but not which statements are true for the respondent, with possible answers between zero and three. Meanwhile, the 'sensitive' or 'treatment' group is read the same list of statements, with the addition of a fourth sensitive statement. They are given the same instructions, to say how many, between zero and four statements, are true. Given the sensitive and non-sensitive list groups are randomly assigned, the two groups are, in expectation, equivalent and the only difference is the inclusion of the sensitive statement. The prevalence of the sensitive item across the whole sample is thus the difference in the mean number of items reported by the two groups.

Unlike with the ACASI and face-to-face methods, anonymity is assured because the lists are designed in such a way that very few people in the list sensitive group have all four statements as either true or not true.[9] This way, a respondent is not revealing their true answer to the sensitive item- to the enumerator or anyone with data access. This could mitigate underreporting by respondents concerned about consequences if someone were to find out their answers, or those who doubt the confidentiality of their data, or who find it easier to truthfully admit sensitive answers when asked indirectly.

Ex-ante, it is not clear whether shame or fear would be the biggest driver of misreporting on IPV survey questions. Assuming that responding in the affirmative is stigmatized, if the shame from answering a question face-to-face dominates other motivations for underreporting, then reporting is likely to be greatest under ACASI. If concerns about being identified or fear of potential repercussions dominate, then reporting would likely be greatest under the list experiment.

There are potential drawbacks to the list experiment. One challenge is that, by design, it

[8]For example, Miller (1984) developed the list method to study illicit drug use. Others have used it to study sensitive topics including prejudice in the United States (Kuklinski, Cobb and Gilens, 1997), political preferences in Lebanon (Corstange, 2009), support for militant groups in Afghanistan (Blair, Imai and Lyall, 2014), risky sexual behavior (Chuang et al., 2018; Jamison, Karlan and Raffler, 2013; Lépine et al., 2020), support for female genital cutting (De Cao and Lutz, 2018), vote buying (Gonzalez-Ocantos et al., 2012), and use of microfinance loans (Karlan and Zinman, 2012).

[9]This can be done through careful piloting and selecting a set of statements that are all plausible but where the co-occurrence of all being true or untrue for any one person is unlikely. See Glynn (2013) and Corstange (2009) for a discussion and suggestions.

cannot be used to identify individuals who respond affirmatively to the sensitive question of interest, potentially limiting its use where individuals need to be referred to support services, for example. Also, with list experiment answers ranging from zero to four, it typically has greater variance than direct questioning methods (but also likely lower misreporting bias because of respondent anonymity). Thus list experiments may reduce bias at the cost of efficiency (Blair and Imai, 2012).[10]

Recently, researchers have begun using the list method to measure IPV and other forms of gender-based violence. In several studies, IPV rates are compared between the face-to-face and list methods. The study most closely related to this paper, conducted in Lima, Peru, found no difference in overall IPV prevalence by survey method (Agüero and Frisancho, 2017); studies in other contexts did find a difference. In Sri Lanka, Traunmüller, Kijewski and Freitag (2019) found the self-reported experience of wartime sexual violence (not IPV) to be higher under the list method than face-to-face, and in India, Joseph et al. (2017) found higher rates of IPV under the list method than face-to-face.[11]

In these studies, respondents were either a) randomly assigned to answer questions about the same experience of IPV using the direct face-to-face method or the list method (Agüero and Frisancho, 2017; World Bank, 2017), or b) all respondents were asked about their IPV experience using both methods within the same survey- either with similar, but slightly different, question wording (Bulte and Lensink, 2019) or the same question wording (Joseph et al., 2017).[12] One issue with approach b) is that being asked the same question in two different ways may bias responses to the second question. We take approach a), asking individuals each question only once to minimize this risk, and we also test three methods with varying degrees of privacy and anonymity to provide additional insights into the reasons why people misreport.

[10]Blair, Coppock and Moor (2018) illustrate this trade-off in their review of the literature comparing the list method versus direct questioning experiments in social psychology and political science.

[11]Aside from the Nigeria study (Banerjee, La Ferrara and Orozco, 2019; World Bank, 2017), one other study in Africa has used the list method to measure IPV. Peterman et al. (2017) measured the prevalence of physical violence using the list method in Zambia. They found rates of IPV similar to those reported in the Zambian DHS, which used the face-to-face method, though the samples are not directly comparable.

[12]For example, in Bulte and Lensink's (2019) Vietnam study, the sensitive list method group responded to a list that included the sensitive statement "I am regularly hit by my spouse," then all respondents later in the survey were also asked "how often did your husband push, slap, beat or hit you during the last 6 months" with response options never, rarely, sometimes, often, very often, or refuse. With this design, the questions may not be directly comparable given potential differences in interpretation and biased responses to the second question. Similarly, Joseph et al. (2017) first asked all respondents using the list method, with the sensitive statement being: "At least one woman member of my household has faced physical aggression from her husband anytime during her life," followed by all respondents directly being asked: "Has at least one woman member of your household faced physical aggression from her husband anytime during her life?" Further, this latter study departs from IPV research guidance, which advises asking about specific, unambiguous episodes of violence that minimize the risk of subjective interpretation (Heise and Hossain, 2017). The survey was also administered at the household level implying that respondents could be men or women, potentially including IPV perpetrators.

2.3 Misreporting and truth

Tourangeau and Yan (2007) argue that for sensitive questions where responding in the affirmative is taboo, the survey method that finds the highest prevalence is closest to truth. In this paper, particularly given the stigma surrounding the issue of IPV in the study contexts, we assume that in most cases, the method with the highest prevalence is closest to truth and that most misreporting is underreporting. In rare cases, people may have incentives to overreport, for example, if a respondent believed she would have a greater chance of accessing a program if she reports experiencing IPV. However, it seems likely that for most sensitive, stigmatized topics such as the perpetration and experience of IPV, people are more likely to under than overreport.

Without objective, non-self-reported data, we cannot, ultimately, confirm which method is closest to truth. However, the two alternative survey methods tested in this paper are designed to address several reasons why people might underreport when interviewed with the face-to-face method- shame and fear. Further, the alternative methods are unlikely to introduce new reasons for people to report IPV that they have not experienced. This is particularly true for the list experiment, where answers are anonymous by design, hence individuals have no incentive to overreport given they cannot be followed-up with to provide aid or program services. Therefore, it seems likely that the method with the highest reporting indicates the method that best addresses respondents' reasons for misreporting and is closest to truth.

This assumption is supported by empirical evidence from the only study we identified that compares an observed sensitive behavior with self-reported behavior measured using direct and indirect methods including the list experiment. Rosenfeld, Imai and Shapiro (2016) found that when they were asked directly, people underreported their past voting behavior on a sensitive abortion referendum in Mississippi. However, rates were higher and closer to truth when they were asked using the list experiment. No method yielded rates higher than the true 'socially undesirable' outcome of voting against tightened abortion restrictions. This study suggests that indirect methods such as the list experiment, which mask a respondent's true answer, can reduce misreporting bias.[13]

Overall, evidence on the measurement of sensitive topics suggests that misreporting appears to be highly issue and context dependent (Blair, Coppock and Moor, 2018; Langhaug, Sherr and Cowan, 2010; Tourangeau and Yan, 2007), presumably affected by factors such as the interview setting, cultural norms, security and safety, and confidence in data privacy. Therefore, this paper primarily relates to misreporting about intimate partner and sexual

[13]It also demonstrates that experiments that compare reporting rates across methods can put a sign and bounds on the likely misreporting bias of standard survey methods.

violence, and in contexts similar to those studied here.

3 Experimental design

3.1 Context, data & design

Rwanda and Nigeria have high rates of IPV compared to the global average. According to data gathered using the direct face-to-face method, 45 percent of Rwandan women and 31 percent of Nigerian women report having experienced physical or sexual IPV in the past 12 months (Global Burden of Disease, 2017). This compares to 13 percent of women in the United States and United Kingdom, 42 percent in Kenya, 39 percent in India, and 31 percent in South Africa. In terms of IPV predictors, poorer, less educated women report IPV at higher rates in Rwanda, while richer, more educated women report more IPV in Nigeria (DHS, 2016, 2014).

Both experiments described below were conducted in accordance with WHO guidelines on the safe and ethical collection of data on violence against women (WHO, 2001, 2016).

3.1.1 Rwanda

In Rwanda, the measurement experiment was conducted on a sample of 2,728 heterosexual couples who were participating in the baseline survey of an IPV prevention study. To be eligible for the intervention, couples had to be aged 18 years or over and married or cohabiting for at least 12 months. The study took place in Eastern Province, in eight sectors selected by the Government of Rwanda for high rates of IPV and a high concentration of Village Savings and Loan Associations (VSLAs) (which the intervention was to be delivered through).[14] The sample comprised mostly poor rural households, with women's mean age of 37 and a median of 4 years of schooling.

For the measurement experiment, individual respondents were randomly assigned to answer two experimental questions using one of the three survey methods: face-to-face, ACASI, or list experiment. The sample was equally split across the three groups.

[14]Sectors are administrative units two levels below province, and 2 levels above village. Within the selected sectors, existing VSLAs were identified in 128 villages. The researchers ruled out 30 villages with fewer than three VSLAs each, and villages that were very close. This was done to minimize the risk of cross-village contamination for the separate randomized controlled trial, detailed in the RCT pre-analysis plan (Alik-Lagrange et al., 2017). Within the 98 study villages, the final sample included 2,042 intervention-eligible couples where at least one partner was a member of a VSLA, and 686 couples who were community members selected at random from a village household roster and who were ineligible for the intervention. VSLA couples' treatment status was assigned by the research team only after baseline surveys were completed. Assignment was based on their ranking in a public lottery process that they applied to. The intervention, called 'Indashyikirwa 2', is a weekly couples group discussion that encourages couples to question traditional gender norms and build communication skills. Data were collected from December 2017 to February 2018.

The two experimental questions were asked approximately one hour into the survey, immediately after ACASI-administered questions from the modified Conflict Tactics Scale about all respondents' experiences of IPV in the past 12 months (for women), and the likelihood of becoming violent in a set of hypothetical circumstances (for men).[15] Respondents were then asked the experimental questions according to the method they were assigned. The questions were:

Women:

1. In the past 12 months, has your husband pushed or thrown you against something like a wall or piece of furniture?

2. Has someone other than your current husband ever forced you to have sex with them when you did not want to?

Men:

1. In the past 12 months, have you tried to limit your wife's contact with her family?

2. In the past 12 months, have you threatened to hurt your wife or someone close to her?

The surveys were administered by enumerators the same sex as the respondent, in private settings where interviews could not be overheard. Most surveys were conducted in empty school classrooms, with one enumerator and one respondent per classroom.[16] The empty classroom interview sites afforded a more private setting than is achieved in many DHS and other household survey settings. Therefore, if a lack of privacy causes women to underreport, then the prevalence differences seen in this more private study setting between face-to-face and the two more private methods are likely to be smaller than the differences that might be observed in standard interview settings.

3.1.2 Nigeria

In Nigeria, the measurement experiment was conducted during the endline survey wave of a cluster randomized control trial studying the effects of the Feed the Future Nigeria

[15]When all respondents reached the main section on IPV (which was to be asked through ACASI), they were shown how to use the ACASI module on the tablet and then instructed through practice questions. After the enumerator was satisfied that the respondent could use ACASI, respondents were left to answer the module while the enumerator stood approximately 5 meters away. Over 95% of female respondents and 98% of male respondents got the first ACASI practice question correct. After being assisted again, 99% of male and female respondents were able to answer the ACASI module in a way that maintained their privacy but with some support, for example, if they could not touch the tablet screens, they listened in the headphones and told the enumerator which button to select for them. The IPV ACASI questions asked of all respondents will be used in the randomized controlled trial. For men, the two experimental questions were the only IPV perpetration questions asked in the survey. The survey themes and order of questions are listed in Table A1 and Table A2.

[16]When classrooms were not available, surveys were conducted in empty offices, churches, or outside but strict privacy protocols were adhered to, including the requirement that interviews pause if someone else was within earshot or eyesight.

Livelihoods Project, a multi-component cash transfer, nutrition, and agriculture program.[17] The survey took place in Northwest Nigeria on the border region with Niger. The sample consisted of ultra-poor households, with approximately 30 percent in polygamous marriages, 84 percent identifying as Muslim, and 15 percent of women self-reporting as literate.

The eligibility criteria for the measurement experiment were that the respondent was female, over 17 years of age, and married and living in the same household as a male. In total, 2,817 women met these criteria and were randomly assigned to answer three experimental questions about their experience of IPV using either the face-to-face or list method. Assignment to the method groups was stratified on the same geographic and poverty-related strata used to assign respondents to treatment arms for the randomized control trial.

The three IPV questions used for the measurement experiment were either directly from or combined several violence questions from the DHS.[18] The experimental questions were:

1. In the past 12 months, has your husband said or done something to humiliate you in front of others?

2. In the past 12 months, has your husband slapped you, pushed you, shaken you, or thrown something at you?

3. In the past 12 months, has your husband physically forced you to have sexual intercourse with him when you did not want to?

The respondents were interviewed one-on-one by a female enumerator in a private setting, usually in the household compound. Enumerators were instructed to pause if anyone came within earshot of the interview, not proceeding until privacy was assured.[19]

3.2 The measurement experiment

We assess misreporting bias by comparing rates of IPV and sexual violence across three different survey methods: face-to-face, list experiment, and ACASI (with the latter only used in Rwanda). Given that assignment to the survey method was random, we assume that the only difference in reported violence between the groups is due to the method used.

[17]The study evaluated the impacts of a package of livelihoods interventions, unconditional cash transfers, and a mentoring program, and was implemented by Catholic Relief Services (Bastian, Goldstein and Papineni, 2017). The cash transfers and most livelihoods programming ended in 2017, approximately 12 months before these data were collected.

[18]Female respondents were first asked about their experience of IPV using the list method (asked either the sensitive or non-sensitive list, depending on her randomly assigned group), followed by direct face-to-face IPV questions. The direct questions included the three measurement experiment questions (asked only to those in the face-to-face assigned group), and five face-to-face questions asked of all respondents (to be used as outcome measures for the main randomized trial).

[19]Occasionally this was difficult to achieve in busy household compounds, for example, when household members came to the house to collect items, causing enumerators to pause interviews. Data were collected from May-July 2018.

3.2.1 Estimation: Calculating prevalence under each survey method

Prevalence under the ACASI and face-to-face methods, where the sensitive item is asked directly, is simply the mean proportion of respondents who said they had experienced violence.

Under the list method, prevalence of the sensitive statement is calculated by taking the difference in the mean number of statements reported as true by the two groups (Blair and Imai, 2012). There are two approaches to administering the list to the non-sensitive group. Under the original approach, which we used in Nigeria, respondents are read the same list, bar the fourth sensitive statement, and asked to report how many of the three statements apply to them. Under the alternative design proposed by Corstange (2009), which we used in Rwanda, the non-sensitive group is read out each non-sensitive statement and asked directly whether or not each was true for them, yes or no. Individuals' responses were then summed to create an equivalent count of three non-sensitive statements per respondent, from which prevalence can be estimated in the same way, by calculating the difference in means.[20]

To help respondents keep track of their answers, they were advised to put their fist behind their back or under their hijab out of sight of the enumerator and to straighten a finger each time a true statement was read out. Once all statements had been read out, respondents showed the enumerator their hand to count the number of straightened fingers.[21]

Prevalence under the list method was calculated using a multivariate regression model, following the approach of Agüero and Frisancho (2017). The OLS regression estimated was as follows,

$$Y_i = \alpha + \rho T_i + \epsilon_i$$

where ρ is the coefficient of interest, the difference in the mean number of items reported across the two groups, or prevalence. Y_i is the number of statements on the list that are true for individual i, α is the mean number of items for the non-sensitive group, and T_i is assignment to the list method treatment group.[22] The list non-sensitive group comprised the share of the sample that answered the experimental questions using ACASI or face-to-face.

[20] We used Corstange's approach in Rwanda because of early concerns about statistical power for conducting multivariate analysis, and this alternative approach allows an associated estimator that can be more efficient than OLS. Given the different construction of the lists, this alternative design potentially poses the risk that respondents do not answer in the same way across the sensitive and non-sensitive lists.

[21] In Rwanda, respondents also put their hands behind their backs and transferred a stone from one hand to the other each time a true statement was read out.

[22] Power calculations indicated that the sample in Nigeria was large enough to detect a difference of approximately 6 percentage points between the groups, and in Rwanda, approximately 5 percentage points, with a significance level of 0.05, power of 0.8, and base prevalence of 20 percent.

Table 2 shows the method assignment and group size.[23]

Table 2: Method assignment group size

	List experiment	Face-to-face	ACASI
Answer type	Indirect	Direct	Direct
Rwanda			
N: women	873	918	937
N: men	877	909	942
Nigeria			
N: women	1,405	1,412	

Overall, both samples were well balanced on key observable characteristics. Results in Tables A5 to A8 show balance across the groups assigned to the list sensitive and list non-sensitive groups, and in Rwanda, also across the three groups (ACASI versus face-to-face versus list sensitive). While several of the raw differences are statistically significant at conventional levels, the differences do not appear to be meaningfully significant, they are within the bounds of what might be expected given the number of variables tested for balance, and no normalized differences are greater than 0.25 standard deviations, a suggested rule of thumb indicating good balance (Imbens and Rubin, 2015). In additional specifications, we estimate prevalence with regressions that include controls selected using post-double selection LASSO (Belloni, Chernozhukov and Hansen, 2014). This does not meaningfully change results.

4 Results

4.1 Method matters: IPV reporting differs by the survey method used

This paper assesses whether there is a difference in reported rates of intimate partner and sexual violence between the three survey methods. A difference in prevalence implies misreporting under at least one of the methods.[24]

In Table 3, columns 1-3 show the prevalence of sexual and intimate partner violence by survey method in Rwanda and Nigeria. Columns 4-6 show the difference in prevalence

[23]Table A3 shows the wording of the questions by method that were asked in Rwanda. The non-sensitive questions that were asked in the list experiments are shown in Table A4. These non-sensitive questions were selected from a larger pool of piloted questions following design advice discussed in Glynn (2013), and informed by questions used by Peterman et al. (2017) and Agüero and Frisancho (2017).

[24]As discussed above, we assume that in most cases, the method with the highest reported prevalence is closest to truth. This assumption is based on the results of several studies that compare observed truth with self-reported sensitive behavior, and an assumed lack of incentive to overreport IPV, particularly under the list experiment.

between pairs of methods.[25] The p-values for these prevalence difference tests were adjusted for multiple hypothesis testing. Sharpened q-values were computed using the approach proposed by Benjamini and Hochberg (1995) and Anderson (2008).[26] Results are also shown graphically in Figure 1.

Results demonstrate that the survey method used to measure sexual and intimate partner violence can substantially affect reporting. Overall, the most widely used survey method, face-to-face interviewing, resulted in the lowest reported IPV rates, followed by ACASI, and finally the list method with the highest. Women's reports of experiencing physical IPV in Rwanda were double when measured using the indirect, anonymous list method (20.6 percent) compared to the two direct methods (9.3 percent with face-to-face and 10.3 percent with ACASI). The differences between the list method and each of the direct methods were statistically significant, but the difference between ACASI and face-to-face prevalence was not.

In Nigeria, emotional and physical IPV rates were 33 and 39 percent greater respectively when measured using the list method (emotional violence: 39.7 percent; physical violence: 26.6 percent) compared to face-to-face (emotional violence: 29.9 percent; physical violence: 19.2 percent). The difference between list and face-to-face emotional violence prevalence was statistically significant at the 5 percent level. For physical violence, the difference was statistically significant at the 10 percent level. There was no statistically significant difference between direct and indirect rates of non-partner sexual violence in Rwanda or sexual IPV in Nigeria. Reported prevalence for these sexual violence questions was between 5.1 to 8.8 percent in Rwanda and 26 to 28.1 percent in Nigeria.

For men in Rwanda, the same pattern across methods was observed, with 13.3, 32.6, and 45.1 percent of men admitting they had tried to limit their wife's contact with her family, with the differences all significant at the 1 percent level. These large differences between the methods suggest that, for one reason or another, men believe that they should report that they do not limit their wife's contact with her family. The added privacy of ACASI addresses some of this underreporting, but even more is addressed by the list method's anonymity. For the second men's question about threatening behavior, prevalence was much lower and although the pattern held, the differences were not statistically significant.

Overall, these results suggest that women's experience and men's perpetration of emotional and physical IPV may be substantially underreported when measured using face-to-face or ACASI methods. In Rwanda, where we tested three methods, the smaller differences

[25]Non-response rates were extremely low in both contexts. For all methods, refusal rates were no greater than 0.84% for women and 2.4% for men. Non-response rates are shown in Table A16.

[26]The q-values control for the false discovery rate for the 12 tests for difference in means that we calculate in Rwanda and for the three in Nigeria. All list method estimates were calculated with heteroskedastic-consistent robust standard errors because the error variance necessarily depends on the list group assignment (Blair and Imai, 2012).

Table 3: Prevalence of sexual and intimate partner violence by survey method

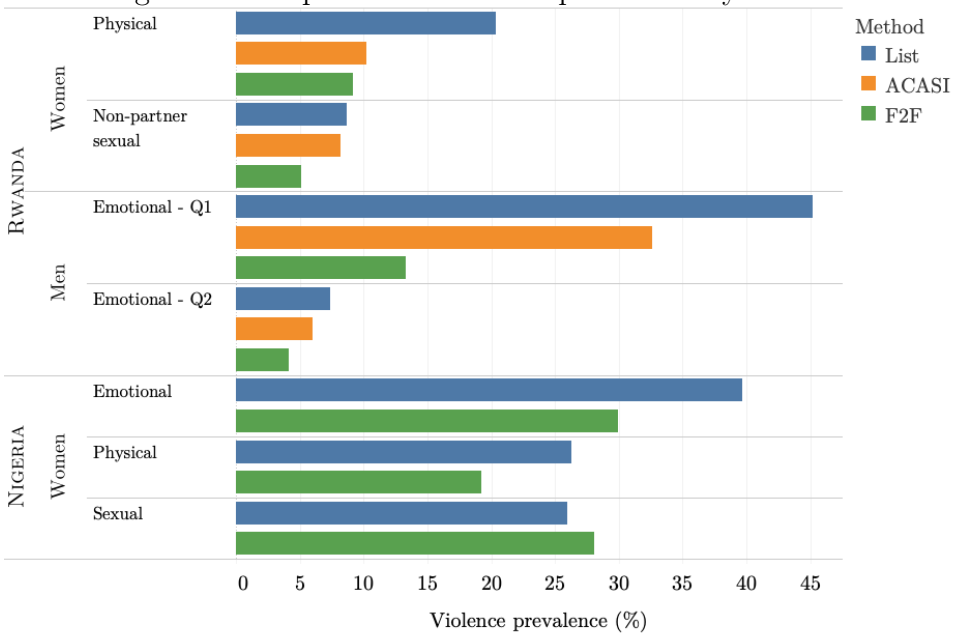
				Difference between pairs of methods		
	(1)	(2)	(3)	(4)	(5)	(6)
Rwanda	Face-to-Face mean	ACASI mean	List method prevalence (difference in means)	List-F2F (3)-(1)	List-ACASI (3)-(2)	ACASI-F2F (2)-(1)
Women						
Q1: husband pushed/ thrown you against something in past 12 months	0.093***	0.103***	0.206***	0.113***	0.103**	0.01
	(0.010)	(0.010)	(0.035)	[0.007]	[0.014]	[0.552]
N	918	914	2,728			
Q2: someone other than current husband ever forced you to have sex when didn't want to	0.051***	0.082***	0.088***	0.037	0.006	0.031**
	(0.007)	(0.009)	(0.034)	[0.443]	[0.872]	[0.017]
N	917	914	2,727			
Men						
Q1: tried to limit wife's contact with her family in past 12 months	0.133***	0.326***	0.451***	0.318***	0.125***	0.193***
	(0.011)	(0.016)	(0.038)	[0.001]	[0.007]	[0.001]
N	909	877	2,728			
Q2: threatened to hurt wife or someone close in past 12 months	0.041***	0.059***	0.071**	0.03	0.012	0.018
	(0.007)	(0.008)	(0.036)	[0.542]	[0.816]	[0.113]
N	909	926	2,728			
Nigeria						
In the past 12 months, has your husband:						
Q1: said or done something to humiliate you in front of others	0.299***		0.397***	0.098**		
	(0.012)		(0.032)	[0.015]		
N	1,394		2,817			
Q2: slapped you, pushed you, shaken you, or thrown something at you	0.192***		0.266***	0.074*		
	(0.011)		(0.043)	[0.083]		
N	1,394		2,817			
Q3: physically forced you to have sex even when you did not want to	0.281***		0.260***	-0.017		
	(0.012)		(0.031)	[0.551]		
N	1,381		2,817			

Note: The left column shows the violence variables. Columns 1, 2 and 3 show prevalence by method. Standard errors are shown in parentheses. Columns 4, 5 and 6 shows differences in prevalence between pairs of methods. For the test of difference in means, Benjamini Hochberg (1995) adjusted q-values are displayed in square brackets, from a chi-square test of the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. Nigeria regressions are adjusted for sampling stratification. List estimates show robust standard errors.

between ACASI and face-to-face prevalence suggest that it is primarily the list method's anonymity rather than ACASI's privacy that makes people more likely to report IPV. If shame were the biggest driver of misreporting, it seems likely that we would have observed larger differences between ACASI and face-to-face prevalence, and smaller differences between list and ACASI.

If the difference in prevalence between the list and direct methods is a proxy for fear, shame or social desirability bias, then these results also suggest that questions about emotional and physical IPV are more sensitive than sexual violence, at least in Nigeria where the experiment had more statistical power. This surprising result could potentially be explained by the contexts. In Nigeria, a quarter of the women interviewed using both the list

Figure 1: Comparison of violence prevalence by method



and the face-to-face methods reported experiencing sexual IPV in the past year. Qualitative interviews and focus groups suggested that there was less stigma associated with sexual IPV than physical IPV. Participants indicated that religious leaders cautioned male followers not to use physical violence against their wives, but that sexual IPV was viewed differently and was often not considered violence. Some may also consider sexual violence as resulting from a husband's passion rather than anger, potentially making it less shameful to report than physical or emotional violence, which were sometimes seen as disciplining errant wives.

In Rwanda, however, qualitative interviews indicated that sexual violence was highly sensitive. There, the lack of difference between face-to-face and list prevalence of lifetime non-partner sexual violence may be due to truly lower prevalence and a lack of statistical power. Alternatively (or in addition), non-partner sexual violence may be so stigmatized that the mere mention of it in the list of statements is enough to make women underreport. However, given there was a difference between ACASI and face-to-face prevalence for this question, this suggests that some women are underreporting non-current partner sexual violence when interviewed face-to-face, and that the additional privacy of ACASI addresses some of their reasons for underreporting. This ACASI-face-to-face difference for the sexual violence question contrasts with the lack of difference between these methods for the physical IPV question.

Potentially, these question and method differences in Rwanda could be due to the different perceived stakes for women who confirm having experienced IPV- which is illegal and

husbands could be fined and jailed if found guilty- versus confirming experience of non-current partner sexual violence - which may feel shameful to admit face-to-face but does not carry the perceived risk that one's husband might be jailed. This would imply that answer anonymity improves truthful reporting for sensitive questions where the perceived legal repercussions may be large, but methods like ACASI that improve privacy do not help address such misreporting. Instead, ACASI's added privacy may primarily increase truthful reporting on sensitive questions when there is no risk of negative legal repercussions if one's answer is discovered.

4.2 Characteristics of misreporting

In this section, we consider whether misreporting is random, and if not, identify the characteristics that predict it. To do this, we assess correlations between respondent characteristics and IPV, as reported under each of the survey methods. When the IPV-characteristic coefficient differs across methods, this implies systematic misreporting under at least one of the methods that is associated with the characteristic. This analysis provides suggestive evidence on the type of person that might misreport, potential reasons for misreporting, and the risks of drawing biased descriptive and causal inferences from such data.

In this section, we limit analysis to women's experience of IPV given that there may be systematically different reasons for women to misreport on non-partner sexual violence questions and given that men's self-reported perpetration is generally considered less reliable.

4.2.1 Estimation: Characteristics associated with misreporting

For those assigned to the face-to-face or ACASI groups, we assess the characteristics of those reporting IPV by regressing women's self-reported experience of IPV against the variable of interest, such as years of education, as below.

$$Y_i = \alpha + \delta x_i + \epsilon_i$$

For the list method, we regress the number of list items reported by an individual against an indicator variable for whether they were in the list sensitive group, as well as the variable of interest and the interaction term. This regression model is shown below,

$$Y_i = \alpha + \rho T_i + \gamma x_i + \delta(T_i.x_i) + \epsilon_i$$

where Y_i is the number of true statements for individual i, T_i is assignment to the sensitive

list method group, and δ is the parameter of interest.[27] If there is no (or consistent) misreporting across the methods, the relevant coefficient for each method should be equal. If there are differences, this implies that the relationship between IPV and the characteristic is sensitive to the measurement method, and we may be estimating biased descriptive and causal relationships. For example, if we find that IPV rates are lower for educated women when we use face-to-face data, but higher for educated women when we use list method data, then we can conclude that misreporting bias is correlated with women's education. To better understand the characteristics of those who misreport, we also compare mean IPV prevalence for respondents above and below the characteristic median (for numeric variable characteristics) or by their indicator variable characteristics.

4.2.2 Results: Respondent characteristics predict misreporting

We analyze misreporting by respondent characteristics that fall into three overlapping categories. These characteristics include: 1) those most likely to motivate respondents to misreport (to understand why people misreport IPV); 2) variables typically targeted in IPV prevention interventions (to evaluate risks of estimating biased treatment effects); and 3) characteristics frequently cited as IPV risk factors (to assess potential challenges to IPV-related theory that has been informed by analysis of face-to-face survey data). The first two categories of characteristics include perceptions of village gender norms; poverty and vulnerability; marital relationship quality; and personal gender attitudes. Additional IPV risk factors considered include women's education and literacy, age, labor force participation and other indicators of women's bargaining power, where theory and empirical evidence are mixed on the relationship with IPV (Bobonis, González-Brenes and Castro, 2013; Buller et al., 2018; Erten and Keskin, 2018; Heise and Kotsadam, 2015; Haushofer et al., 2019; Eswaran and Malhotra, 2011).

Tables 4 and 5 show the relationships between IPV and respondent characteristics by survey method. The first 3 (6) columns show the relevant regression coefficient of interest for Rwanda (Nigeria), and the last 3 columns show the results of chi-squared tests of the coefficient differences between pairs of methods. Each coefficient shown in the tables was estimated in a separate regression. These tables show several significant differences in coefficients across the methods, implying that a number of characteristics are associated with systematic misreporting of IPV. These results are also shown graphically in Figures A2 and A3. We also show the differences in mean IPV prevalence for respondents above and below

[27] This approach follows that of Holbrook and Krosnick (2010), Coutts and Jann (2011), and Agüero and Frisancho (2017). While linear regression may be less efficient than other multivariate statistical estimators proposed for list experiments by Corstange (2009), Imai (2011), Blair and Imai (2012), and Imai, Park and Greene (2015), it is simpler to interpret and compare relevant coefficients across methods and specifications for a descriptive analysis.

the median of the characteristic in Tables A9 and A10 and graphically in Figures A4 and A5.

This subsection presents the results of descriptive analysis and as such, usual caution should be applied when interpreting results given that omitted variables or other sources of bias may be driving observed correlations. Given this is exploratory analysis only, we do not adjust the p-values of tests of equivalence for multiple hypothesis testing.[28]

Table 4: Rwanda- correlations between IPV and respondent characteristics across methods

	Q: husband pushed or thrown you			Method differences: P-values from tests of differences		
	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(3)-(2)
Variable	List	F2F	ACASI	List-F2F	List-ACASI	ACASI-F2F
Vulnerability index	0.23***	0.06***	0.05*	0.07	0.05	0.65
	(0.08)	(0.02)	(0.03)			
Unequal village gender norms index	0.37***	0.03	0.04	0.00	0.00	0.75
	(0.10)	(0.03)	(0.03)			
Relationship quality index	0.03	-0.07***	-0.04***	0.06	0.17	0.18
	(0.05)	(0.01)	(0.01)			
Gender engagement index	-0.10**	-0.02*	-0.01	0.11	0.07	0.48
	(0.04)	(0.01)	(0.01)			
Progressive gender attitudes index	0.01	-0.01	-0.03	0.75	0.56	0.47
	(0.06)	(0.02)	(0.02)			
Bargaining power measures						
Legally married	-0.05	0.01	-0.03	0.49	0.83	0.21
	(0.07)	(0.02)	(0.02)			
Wife's age	0.00	0.00	0.00	0.43	0.34	0.68
	(0.00)	(0.00)	(0.00)			
Wife owns land	0.03	0.00	-0.03	0.72	0.49	0.36
	(0.08)	(0.02)	(0.02)			
Wife's years of education	0.02	-0.00	-0.01*	0.13	0.06	0.27
	(0.01)	(0.00)	(0.00)			
Wife in paid work last week	-0.15**	-0.00	0.03	0.04	0.02	0.33
	(0.07)	(0.02)	(0.02)			

Note: Standard errors are shown in parentheses (List method estimates show robust standard errors). *** p<0.01, ** p<0.05, * p<0.1. Columns 1-3 show the relevant covariate coefficients from separate regressions. Columns 4-6 show the p-values from a chi-squared test of the differences between these coefficients. The vulnerability index includes household economic shock experience, husband's trauma score, husband's alcohol consumption, and whether any household member has skipped a meal in the past week. The unequal gender norms index includes incentivized and unincentivized questions about perceived unequal gender attitudes held by others in their village. The relationship quality index includes measures of marital trust, empathy and communication. The gender engagement index includes questions on the wife's previous participation in rallies and workshops about family violence and gender equality. The gender attitudes index includes questions about respondent beliefs about gender roles and the acceptability of IPV.

These results show that in both countries, vulnerable women report IPV at lower rates when they are asked about their experience using direct methods than when interviewed using the list method. Specifically, in Rwanda, a one standard deviation increase in a

[28]For conceptually linked variables (for example, questions about relationship quality such as trust and empathy in the marriage), we generated weighted standardized summary indices (Anderson, 2008) to minimize the false discovery rate. For other indicators of interest where there is less theoretical coherence between the variables, we report coefficients on the raw variable. For instance, it is not clear that the following indicators of women's bargaining power are measures of the same construct, nor whether they would each induce a respondent to underreport IPV more or less: education, asset ownership, age, marital status, and engagement in paid employment.

Table 5: Nigeria- correlations between IPV and respondent characteristics across methods

	Q1: emotional violence		Q2: physical violence		Q3: sexual violence		Method differences: P-values from tests of differences		
	(1)	(2)	(3)	(4)	(5)	(6)	(1)-(2)	(3)-(4)	(5)-(6)
Variable	List	F2F	List	F2F	List	F2F	Q1	Q2	Q3
Vulnerability index	0.02	-0.08***	0.09	-0.05***	0.03	-0.05**	0.14	0.05	0.25
	(0.06)	(0.02)	(0.07)	(0.02)	(0.07)	(0.02)			
Unequal village gender norms index	0.06	0.08***	0.09*	0.05***	0.08*	0.05***	0.61	0.43	0.50
	(0.04)	(0.02)	(0.05)	(0.02)	(0.04)	(0.02)			
Relationship quality index	-0.07	-0.21***	-0.02	-0.21***	-0.04	-0.27***	0.11	0.04	0.01
	(0.08)	(0.03)	(0.09)	(0.02)	(0.08)	(0.03)			
Public speaking confidence index	-0.02	-0.08***	0.01	-0.03***	0.00	-0.01	0.11	0.34	0.68
	(0.04)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)			
Progressive gender attitudes index	-0.12*	-0.13***	-0.23***	-0.09***	-0.16**	-0.16***	0.91	0.11	0.95
	(0.07)	(0.03)	(0.08)	(0.03)	(0.07)	(0.03)			
Bargaining power measures									
Wife's age	0.00	0.00**	0.00	0.00	0.00	-0.00	0.57	0.49	0.59
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)			
Wife reads and writes	0.12	-0.08**	0.00	-0.12***	0.11	-0.04	0.03	0.20	0.12
	(0.08)	(0.03)	(0.10)	(0.02)	(0.09)	(0.03)			
Polygamous: wife number (1-4)	-0.03	0.02	0.06	0.03	0.13	-0.02	0.56	0.74	0.13
	(0.08)	(0.04)	(0.10)	(0.03)	(0.09)	(0.03)			
Wife brought assets to marriage	-0.04	-0.18***	0.03	-0.13***	-0.02	-0.09***	0.04	0.03	0.34
	(0.06)	(0.02)	(0.07)	(0.02)	(0.07)	(0.02)			

Note: Standard errors are shown in parentheses (List method estimates show robust standard errors). All regressions are adjusted for sampling stratification. *** p<0.01, ** p<0.05, * p<0.1. Columns 1-6 show the relevant covariate coefficients from separate regressions. Columns 7-9 show the p-values from a chi-squared test of the differences between these coefficients. The vulnerability index includes measures of food poverty and household economic insecurity, and wife's stress. The unequal village gender norms index includes beliefs about unequal gender attitudes held by others in the village. The relationship quality index includes measures of trust, satisfaction, empathy and communication in the marriage. The public speaking confidence index includes measures of willingness to speak up in public. The progressive gender attitudes index includes measures of personal views on the acceptability of IPV and on gender roles.

woman's vulnerability index was associated with a 6 percentage point increase in reported experience of IPV when violence was measured face-to-face, a 5 percentage point increase when measured using ACASI, and a 23 percentage point increase when measured using the list method. In Nigeria, vulnerable women were also more likely to report physical IPV under the list method than face-to-face. For emotional and sexual IPV, the same pattern existed but the differences were not statistically significant. Table A9 shows that, in Rwanda, as we might expect, the difference between IPV reporting when it was measured using the list and direct methods was driven by women above median vulnerability. There was no IPV reporting difference across the list and direct methods for the least vulnerable women.

More vulnerable women may face greater personal safety or economic security risks if they report that their husband perpetrates IPV and, for example, he discovered her answer or was jailed as a result. For such women, the list experiment's anonymity may provide them with additional reassurance and safety, making them more willing to report IPV. These results demonstrate that the relationship between vulnerability and IPV can be very different if estimated using anonymous list experiment IPV data. In Rwanda, it also demonstrates that

ACASI's added privacy does not address the reasons vulnerable women are misreporting.

In Rwanda, we also observe a similar misreporting pattern based on a respondent's perception of unequal village gender norms, which could proxy for shame about IPV or fear of potential repercussions if a woman's survey answer were to be discovered. When a woman perceived that village gender norms were unequal, she was more likely to report IPV under the list method, where her answer is anonymous, than she was under the face-to-face or ACASI methods. Specifically, in Rwanda, a one standard deviation increase in a woman's perception of unequal village gender norms was associated with a 3 percentage point increase in the probability of experiencing IPV when violence was measured face-to-face, a 4 percentage point increase when measured using ACASI, and a 37 percentage point increase when using the list method. However, the positive conservative norms-IPV relationship was only statistically significant when IPV was measured using the list method. In Rwanda, the differences between the list and direct methods were statistically significant; in Nigeria, they were not.[29] Table A9 shows that the gender norms-IPV relationship difference between the methods was driven by the respondents who believed their village had the most conservative norms. There was no IPV reporting difference across the list and direct methods for respondents who believed their village had progressive norms. As with vulnerability, these results demonstrate that the relationship between community gender norms and IPV can be very different if estimated using anonymous list experiment data.

The Rwandan result challenges an alternative hypothesis about norms and IPV misreporting which posits that in villages with highly unequal gender norms, where IPV is normalized, women may feel more able to openly report IPV. If that were the case, we should have observed no difference between the coefficients for face-to-face-estimated IPV and the two more private or anonymous methods. Instead, in Rwanda at least, it appears that only the anonymous method, but not ACASI, made vulnerable women and women living in conservative communities comfortable enough to report IPV.

Greater marital relationship quality, measured using questions about satisfaction, empathy, and trust in marriage, was also associated with underreporting. In both countries, women who reported greater relationship quality were more likely to report IPV when interviewed using the list method than face-to-face; the method differences were statistically significant on three out of the four IPV questions considered. A potential reason for women

[29]In Nigeria, the lack of IPV reporting difference across the methods by perceived norms may be due to the different gender norms questions that comprise the index in the two countries. The Nigerian norms index included only unincentivized questions about perceptions of other villagers' attitudes regarding the acceptability of women working or traveling outside the home. In Rwanda, the norms index focused on IPV-specific norms and included both unincentivized questions about perceptions of village IPV prevalence and others' attitudes, as well as four incentivized norms vignettes based on an adapted version of the Krupka-Weber norms coordination game (Krupka and Weber, 2013). In this game, respondents were paid if they selected the modal answer in their village about whether described behaviors about IPV and women leaving violent husbands were considered socially acceptable or unacceptable in their village. This incentive-compatible measure was intended to overcome social desirability bias.

who report being in a more trusting, supportive relationship to underreport IPV when asked directly (but not indirectly) is because the direct question forces her to confront an uncomfortable truth about her marriage, or she may feel disloyal by directly admitting her husband has been violent. As Tables 4 and 5, and the associated figures A2 and A3 show, when IPV is measured using the list experiment, there is no relationship between IPV and relationship quality; it is only when using direct methods that we find a negative relationship between IPV and marital quality.

Several of these results highlight a potential measurement bias issue for the growing number of policies and interventions being assessed on their IPV impact. For example, an intervention that improves relationship quality- as several IPV prevention programs seek to do- may have no impact on actual IPV rates. However, it may lead to a reduction in a wife's willingness to directly report IPV for the reasons discussed above. This would bias treatment effect estimates, suggesting a reduction in IPV that was, in fact, an artefact of the measurement method used. Alternatively, if a policy led to a worsening in community gender norms, IPV outcomes that were measured using direct methods may not detect a resulting actual increase in IPV if women are less willing to report it. In both of these cases, list experiment-measured IPV may be less prone to reporting bias and more likely to reflect true treatment effects.

We also examined whether characteristics that are proposed as causes or common correlates of IPV remain stable or at least with the same sign on the relationship, regardless of the method used, or whether the nature of the relationship depends on the measurement method. Across different indicators of women's intrahousehold bargaining power, we find some evidence of systematic misreporting of IPV, though no clear pattern across different indicators. For example, in Rwanda, we find a strong relationship between employment and IPV misreporting. Results in Table A9 show that this reporting effect is driven by a large difference in reporting for women who are not engaged in paid employment. Similar to the norms and vulnerability results, this may suggest that women with the weakest bargaining power or outside options underreport the most when asked directly about their IPV experience. Again, the list experiment appears to make these women more willing to report IPV.

Potentially most concerning, we observe some opposite-signed relationships, where the sign on the relationship between IPV and the respondent's characteristic is positive for one measurement method and negative for another method. In Rwanda, these variables include wife's years of education, relationship quality, and engagement in paid work, and in Nigeria, wife's vulnerability, literacy, and assets brought into the marriage. For example, in both Rwanda and Nigeria, there is suggestive evidence of a positive relationship between IPV and

women's education when IPV is measured directly, but a negative relationship when using the list method. Specifically, literate Nigerian women are 12 percentage points more likely to experience emotional IPV if it is measured using the list method, and 8 percentage points less likely if IPV is measured face-to-face. This is similar to results in Peru and India, where researchers found women's education to be positively correlated with IPV when measured using the list method but negatively when measured face-to-face (Agüero and Frisancho, 2017; Joseph et al., 2017). The opposite-signed IPV-education associations imply opposing theories about women's bargaining power as a cause of IPV, highlighting the importance of minimizing measurement bias to better understand IPV causes, theory, and prevention.

Overall, the sensitivity of the relationships between IPV and key bargaining power and other indicators to the survey method used is concerning. It suggests that caution should be applied when using direct IPV measures to understand descriptive and causal relationships as misreporting may bias correlations and treatment effect estimates. The anonymity of the list experiment may minimize this reporting bias.

4.3 Robustness checks

In the section below, we report the results of robustness checks on the list method, followed by checks on the overall analysis.

4.3.1 Checks on list method assumptions

There are two assumptions that underpin the validity of statistical analyses of list experiments, and which, if they hold, should result in an unbiased difference-in-means estimator (Blair and Imai, 2012). The first assumption is that of 'no design effect', meaning that the inclusion of the sensitive item in the list does not affect answers to non-sensitive items. The second assumption is that of 'no liars', or that the answers given for the sensitive item are truthful.

Blair and Imai (2012) propose a likelihood ratio test to assess the probability that the first assumption is violated. The test is based on the assumption that the inclusion of a sensitive statement should not change the sum of affirmative answers to the non-sensitive statements, and that cumulative proportions of different respondent types should be non-negative (Bulte and Lensink, 2019). If one of the proportions is negative, then the no design effect assumption has likely been violated. The rejection of the null hypothesis of no design effect using Blair and Imai's statistical test, made available on the open-source software R (Blair and Imai, 2011), would suggest there was a design effect. The test returns a Bonferroni-corrected minimum p-value of the test.

In Nigeria, the test returned a minimum p-value of 1, suggesting there was no design effect for the three list questions. In Rwanda, where the list non-sensitive group was directly asked each non-sensitive question, there is statistically significant evidence of a design effect for three out of the four questions, suggesting that group assignment affected answers to the non-sensitive questions. There was no design effect for the first men's question.

What then is the potential direction of bias introduced by this assumption violation in Rwanda? In the three cases in Rwanda where this assumption was violated, the design effect is likely to have resulted in downwards-biased prevalence estimates for the list method. This is because the tests that it failed, assessing whether the cumulative proportion of responses in the non-sensitive group was greater than the proportion in the sensitive list group, were for the share of respondents reporting that zero items were true for them. More respondents in the list sensitive group reported zero than in the non-sensitive group. This implies that the mean in the list sensitive group is likely to be downwards biased, implying the same for the difference-in-means estimate and the resulting IPV prevalence.[30] Further, the one Rwanda question that did not fail the no design effects test showed the greatest difference in prevalence between the list and other methods. It therefore seems plausible that, if biased, the three Rwandan list experiments that failed the no design effects test produced downwards biased prevalence estimates. The associated test p-values and proportions are shown in Tables A11 and A12.

There is no explicit test for the second 'no liars' assumption underpinning the list experiment. However, the most obvious threat to this assumption is when a respondent's anonymity for their answer to the sensitive statement is compromised. This 'ceiling effect' exists when a respondent's true answer is affirmative for all of the statements, thus their affirmative response to the sensitive statement will no longer be hidden. However, the proportion of respondents answering with the maximum number of items (three and four for the non-sensitive and sensitive list groups, respectively) is within the range of other studies.[31] Further, the presence of a ceiling effect likely indicates downwards biased estimates because other respondents would be likely to have understated their true number of 4 affirmative answers in order to conceal their affirmative sensitive answer (Blair and Imai, 2012).

[30] One study by Flavin and Keane (2009) evaluates whether the design of the non-sensitive list affects responses. This study found that asking the non-sensitive group directly about the non-sensitive items resulted in a higher mean number of items than when the non-sensitive questions were asked using the traditional 'how many of the following items' approach. As a result, the direct approach to asking the non-sensitive questions resulted in downwards-biased prevalence estimates, though it should be noted that their 'non-sensitive' items were somewhat sensitive. Further, a recent paper by Chuang et al. (2018) found that having relatively sensitive 'non-sensitive' statements on the list appears to mask the sensitive item of interest, and results in lower variance and higher prevalence estimates. In Rwanda and Nigeria, we used relatively innocuous non-sensitive list items, which, if results from the Chuang et al. measurement experiment on risky sexual behavior hold, then it again seems plausible that these list experiment results may be downwards biased.

[31] The proportion of respondents in the non-sensitive list group reporting three items as true is never greater than 15.3 percent (for the women's physical violence question in Rwanda), and the proportion of respondents in the sensitive list group reporting the full number of items is never greater than 5.4 percent (for the women's physical violence question in Nigeria).

4.3.2 Checks on overall approach

The two key assumptions underpinning the unbiasedness of the list experiment estimator have parallels for the experiment as a whole. While there is no direct test for these assumptions holding for the ACASI and face-to-face methods, there are some indicators of whether or not they were violated.

First, we checked balance on survey questions asked of all respondents after the experimental violence questions. If respondents were differentially affected by one survey method, perhaps indicating a design effect, we may observe them answering subsequent survey questions differently. We found very few statistically significant differences in respondent answers despite the many tests for significance. None of these differences are materially significant.

Second, we analyzed the main list method prevalence regressions with enumerator fixed effects and controls selected using post-double-selection LASSO (Belloni, Chernozhukov and Hansen, 2014). We found no notable differences in coefficients or significance. These results are shown in Tables A13 to A15.

Third, although a general 'no liars' assumption is untestable, we compared refusal rates across methods. This could indicate the relative risk of misreporting for each method. The refusal rate with ACASI, where all one has to do to skip a question is touch a button, may be a lower-bound on the share of people who might have lied if they were interviewed face-to-face and would be uncomfortable asking an enumerator to skip the question. Indeed, we find that ACASI, has the highest refusal rate out of the methods, though it is still low, at 0.84 percent for women. Table A16 shows refusal rates by question and method.

Additional indications that these results are robust are supported by the broadly similar patterns of misreporting by survey method that were observed over different countries, genders, violence types, perpetration and victimization questions, and list method designs.

5 Conclusion

This paper provided some of the first evidence that the prevalence of intimate partner violence varies substantially depending on the survey method used to measure it. Our results suggest that in some contexts, standard survey methods are likely to result in significant underestimates of the prevalence and welfare costs of IPV. The most widely used method, face-to-face interview, resulted in the lowest prevalence estimates, followed by the more private ACASI method. Finally, the list method, which allows respondents to report anonymously, resulted in the highest prevalence. In Rwanda, physical IPV rates were double, and in Nigeria, 39 percent greater when measured using the list method compared to the direct

methods.

This paper also found that misreporting was systematic. The women who underreported when interviewed directly (compared to when interviewed with the list method) tended to be more vulnerable, not engaged in paid work, living in conservative communities, and more educated. These results demonstrate that misreporting was correlated with indicators that theory suggests are risk factors for IPV and with variables often targeted in women's empowerment programs. Our results show non-trivial patterns of misreporting, such that gradients on some IPV-predictor relationships changed substantially, and some signs flipped, depending on the method used to measure IPV.

As other studies have found, direct measures of IPV can produce biased treatment effect estimates, for example, finding programs caused a reduction in IPV if measured face-to-face, but an increase if measured with the list experiment (Bulte and Lensink, 2019). Given the growing number of studies measuring program impacts on IPV, our results further highlight the risks of collecting measurement error-prone data for generating unbiased correlations and treatment effect estimates. Given that most IPV-related theory and prevention interventions are informed by face-to-face survey data, misreporting bias could have significant implications for our understanding of the prevalence of IPV, what causes it, and how to prevent it.

Our three-method comparison in Rwanda provides additional insights into potential reasons for misreporting. There were strikingly few differences between ACASI-measured and face-to-face-measured prevalence and IPV-predictor relationships. This suggests that women in Rwanda were no more comfortable reporting IPV with ACASI's added privacy, yet the list method's anonymity made many more women willing to report. This may suggest that on questions about illegal behaviors where respondents fear a risk of legal sanctions or repercussions from their answer, for example, in places where physical and sexual IPV are criminalized, then only anonymous survey methods may induce some respondents to report.[32]

Our results also have implications for the growing literature that uses face-to-face IPV data from multiple countries to draw inferences about IPV prevalence, predictors, and causes (Devries et al., 2013; García-Moreno et al., 2013; Cools and Kotsadam, 2017; Jewkes et al., 2017; Alesina, Brioschi and La Ferrara, 2020; Heise and Kotsadam, 2015). Our finding of significant and systematic underreporting for some types of IPV in our two rural African contexts contrasts with findings from Lima, Peru, where researchers found no overall difference between face-to-face and list method prevalence (Agüero and Frisancho, 2017). These

[32]The similar face-to-face and ACASI reporting patterns may also be affected by the strict private interview setting in Rwanda. In a busier, more typical survey setting where privacy from curious onlookers and other household members is harder to enforce, we may have found that women would be even less willing to report IPV face-to-face, and ACASI may have made more of a difference. See Appendix B for a further discussion on IPV measurement method trade-offs.

divergent results indicate that reporting bias is, unsurprisingly, strongly affected by contextual factors. Given significant, systematic face-to-face misreporting in some contexts but not others, our results imply that caution should be used when aggregating and analyzing face-to-face IPV data from diverse settings, lest context-specific misreporting bias results.

Policy makers and social scientists measuring sensitive topics could use similar randomized survey experiments to compare IPV reporting rates between the list experiment and direct methods to identify levels of stigma, the types of people most likely to misreport, and the risk of generating biased treatment effects. For example, social scientists researching the impact of interventions on IPV might use such an experiment to sign potential reporting bias in their treatment effect estimates. Also, given Demographic and Health Surveys' large sample sizes and use in multi-country analysis, perhaps such surveys should measure several IPV questions using the list method as well as the face-to-face method. This would allow data users to assess the likely extent of reporting bias by context, allowing them to make adjustments to inferences about true levels and correlates of IPV.[33]

Future research could be conducted in other contexts and on other types of violence and sensitive topics to better understand what drives misreporting. Other potential areas for additional research include studying how to best design and administer the list experiment to minimize bias, and developing alternative measurement methods for efficient causal research on IPV which, according to face-to-face survey data, is estimated to affect one in three women globally (García-Moreno et al., 2013).

[33]DHS data on IPV cover more than 60 developing countries and are cited in more than 5,000 papers on google scholar. If DHS IPV data are measured with substantial, systematic, and context-specific bias as this paper suggests is likely, this has significant implications for the use of such data.

References

Agüero, Jorge, and Verónica Frisancho. 2017. "Misreporting in Sensitive Health Behaviors and Its Impact on Treatment Effects: An Application to Intimate Partner Violence." IDB Working Paper No. IDB-WP-853.

Agüero, Jorge, Úrsula Aldana, Erica Field, Verónica Frisancho, and Javier Romero. 2020. "Is Community-Based Targeting Effective in Identifying Intimate Partner Violence?" *AEA Papers and Proceedings*, 110(May): 605–09.

Alesina, Alberto, Benedetta Brioschi, and Eliana La Ferrara. 2020. "Violence Against Women: A Cross-cultural Analysis for Africa." *Economica*. May.

Alik-Lagrange, Arthur, Claire Cullen, Mũthoni Ngatia, and Julia Vaillant. 2017. "Preventing intimate partner violence: Impact Evaluation of MIGEPROF couples training for IPV prevention in Eastern Rwanda." AEA RCT Registry. June 28. https://doi.org/10.1257/rct.2282-2.0.

Anderson, Michael L. 2008. "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.

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco. 2019. "Entertainment, education, and attitudes toward domestic violence." *AEA Papers and Proceedings*, 109: 133–37.

Bastian, Gautam Gustav, Markus P. Goldstein, and Sreelakshmi Papineni. 2017. "Are cash transfers better chunky or smooth?: evidence from an impact evaluation of a cash transfer program in northern Nigeria." World Bank Gender Innovation Lab Policy Brief, Washington, D.C.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." *The Review of Economic Studies*, 81(2): 608–650.

Benjamini, Yoav, and Yosef Hochberg. 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.

Bharadwaj, Prashant, Mallesh M. Pai, and Agne Suziedelyte. 2017. "Mental health stigma." *Economics Letters*, 159: 57–60.

Blair, Graeme, Alex Coppock, and Margaret Moor. 2018. "When to Worry About Sensitivity Bias: Evidence from 30 Years of List Experiments." Retrieved from: https://graemeblair.com/papers/sensitivity-bias.pdf. Accessed on 15 May 2020.

Blair, Graeme, and Kosuke Imai. 2011. "List: Statistical Methods for the Item Count Technique and List Experiment. Comprehensive R Archive Network (CRAN)."

Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis*, 20: 47–77.

Blair, Graeme, Kosuke Imai, and Jason Lyall. 2014. "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan." *American Journal of Political Science*, 58(4): 1043–1063.

Blattman, Christopher, Julian Jamison, Tricia Koroknay-Palicz, Katherine Rodrigues, and Margaret Sheridan. 2016. "Measuring the measurement error: A method to qualitatively validate survey data." *Journal of Development Economics*, 120: 99–112.

Bobonis, Gustavo J, Melissa González-Brenes, and Roberto Castro. 2013. "Public Transfers and Domestic Violence: The Roles of Private Information and Spousal The Roles of Private Information and Spousal Control." *American Economic Journal: Economic Policy*, 5(51): 179–205.

Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise. 2018. "A Mixed-Method Review of Cash Transfers and Intimate Partner Violence in Low- and Middle-Income Countries." *The World Bank Research Observer*, 33(2): 218–258.

Bulte, Erwin, and Robert Lensink. 2019. "Women's empowerment and domestic abuse: Experimental evidence from Vietnam." *European Economic Review*, 115: 172–191.

Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban. 2018. "Sex, Lies, and Measurement." Retrieved from: https://web.stanford.edu/~pdupas/CDHS_measurement.pdf. Accessed on 1 May 2019.

Cools, Sara, and Andreas Kotsadam. 2017. "Resources and intimate partner violence in Sub-Saharan Africa." *World Development*, 95: 211–230.

Corstange, Daniel. 2009. "Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT." *Political Analysis*, 17(1): 45–63.

Coutts, Elisabeth, and Ben Jann. 2011. "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)." *Sociological Methods & Research*, 40(1): 169–193.

De Cao, Elisabetta, and Clemens Lutz. 2018. "Sensitive Survey Questions: Measuring Attitudes Regarding Female Genital Cutting Through a List Experiment." *Oxford Bulletin of Economics and Statistics*, 80(5): 871–892.

Devries, Karen M, Joelle YT Mak, Claudia García-Moreno, Max Petzold, James C Child, Gail Falder, Stephen Lim, Loraine J Bacchus, Rebecca E Engell, Lisa Rosenfeld, et al. 2013. "The global prevalence of intimate partner violence against women." *Science*, 340(6140): 1527–1528.

DHS. 2014. "Nigeria Demographic and Health Survey." National Population Commission - NPC/Nigeria and ICF International, Abuja.

DHS. 2016. "Rwanda Demographic and Health Survey 2014-15." National Institute of Statistics of Rwanda (NISR), Ministry of Health (MOH), and ICF International, Kigali and Maryland.

Erten, Bilge, and Pinar Keskin. 2018. "For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey." *American Economic Journal: Applied Economics*, 10(1): 64–105.

Eswaran, Mukesh, and Nisha Malhotra. 2011. "Domestic violence and women's autonomy in developing countries: theory and evidence." *Canadian Journal of Economics/Revue canadienne d'économique*, 44(4): 1222–1263.

Falb, Kathryn, Sophie Tanner, Khudejha Asghar, Samir Souidi, Stan Mierzwa, Asham Assazenew, Theresita Bakomere, Pamela Mallinga, Katie Robinette, Woinishet Tibebu, and Lindsay Stark. 2016. "Implementation of Audio-Computer Assisted Self-Interview (ACASI) among adolescent girls in humanitarian settings: feasibility, acceptability, and lessons learned." *Conflict and Health*, 10(1): 32.

Fay, R E, C F Turner, A D Klassen, J H Gagnon, J. H. Pleck, and F. L. Sonenstein. 1998. "Prevalence and patterns of same-gender sexual contact among men." *Science*, 280(5365): 867–873.

Fincher, Danielle, Kristin VanderEnde, Kia Colbert, Debra Houry, L. Shakiyla Smith, and Kathryn M. Yount. 2015. "Effect of Face-to-Face Interview Versus Computer-Assisted Self-Interview on Disclosure of Intimate Partner Violence Among African American Women in WIC Clinics." *Journal of Interpersonal Violence*, 30(5): 818–838.

Flavin, Patrick, and Michael Keane. 2009. "How Angry am I? Let Me Count the Ways: Question Format Bias in List Experiments."

García-Moreno, C, C. Pallitto, K. Devries, H. Stöckl, C. Watts, and N. Abrahams. 2013. "Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence." Geneva:World Health Organization.

Global Burden of Disease. 2017. "Global Burden of Disease Study 2016: Health-related Sustainable Development Goals (SDG) Indicators 1990-2030." Seattle, United States: Institute for Health Metrics and Evaluation (IHME).

Glynn, Adam N. 2013. "What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment." *Public Opinion Quarterly*, 77(S1): 159–172.

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio, and David W. Nickerson. 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science*, 56(1): 202–217.

Gottschalk, Peter, and Minh Huynh. 2010. "Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error." *Review of Economics and Statistics*, 92(2): 302–315.

Haushofer, Johannes, Charlotte Ringdal, Jeremy P Shapiro, and Xiao Yu Wang. 2019. "Income changes and intimate partner violence: Evidence from unconditional cash transfers in Kenya." NBER Working Paper No. 25627.

Heise, Lori, and Andreas Kotsadam. 2015. "Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys." *The Lancet Global Health*, 3(6): 332–340.

Heise, Lori, and Mazeda Hossain. 2017. "Measuring intimate partner violence." STRIVE, London.

Holbrook, A. L., and J. A. Krosnick. 2010. "Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method's Validity." *Public Opinion Quarterly*, 74(2): 328–343.

Imai, Kosuke. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association*, 106(494): 407–416.

Imai, Kosuke, Bethany Park, and Kenneth F Greene. 2015. "Using the Predicted Responses from List Experiments as Explanatory Variables in Regression Models." *Political Analysis*, 23: 180–196.

Imbens, Guido., and Donald B. Rubin. 2015. *Causal inference for statistics, social, and biomedical sciences : an introduction.* Cambridge University Press.

Jamison, Julian C, Dean Karlan, and Pia Raffler. 2013. "Mixed method evaluation of a passive mHealth sexual information texting service in Uganda." NBER Working Paper No. 19107.

Jarlais, Don C Des, Denise Paone, Judith Milliken, Charles F Turner, Heather Miller, James Gribble, Qiuhu Shi, Holly Hagan, and Samuel R Friedman. 1999. "Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial." *The Lancet*, 353(9165): 1657–1661.

Jaya, Michelle J Hindin, and Saifuddin Ahmed. 2008. "Differences in young people's reports of sexual behaviors according to interview methodology: a randomized trial in India." *American Journal of Public Health*, 98(1): 169–74.

Jewkes, Rachel, Emma Fulu, Ruchira Tabassam Naved, Esnat Chirwa, Kristin Dunkle, Regine Haardörfer, Claudia Garcia-Moreno, et al. 2017. "Women's and men's reports of past-year prevalence of intimate partner violence and rape and women's risk factors for intimate partner violence: A multicountry cross-sectional study in Asia and the Pacific." *PLoS medicine*, 14(9).

Joseph, George, Syed Usman Javaid, Luis Alberto Andres, Gnanaraj Chellaraj, Jennifer L Solotaroff, and S Irudaya Rajan. 2017. "Underreporting of gender-based violence in Kerala, India: an application of the list randomization method." World Bank Policy Research Working Paper 8044.

Karlan, Dean S, and Jonathan Zinman. 2012. "List randomization for sensitive behavior: An application for measuring use of loan proceeds." *Journal of Development Economics*, 98(1): 71–75.

Krupka, EL, and RA Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. "Racial Attitudes and the New South." *The Journal of Politics*, 59(2): 323–349.

Langhaug, Lisa F., Lorraine Sherr, and Frances M. Cowan. 2010. "How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries." *Tropical Medicine & International Health*, 15(3): 362–381.

Lépine, Aurélia, Carole Treibich, Cheikh Tidiane Ndour, Khady Gueye, and Peter Vickerman. 2020. "HIV infection risk and condom use among sex workers in Senegal: evidence from the list experiment method." *Health Policy and Planning*, 35(4): 408–415.

Mensch, Barbara S., Paul C. Hewett, Richard Gregory, and Stephane Helleringer. 2008. "Sexual Behavior and STI/HIV Status Among Adolescents in Rural Malawi: An Evaluation of the Effect of Interview Mode on Reporting." *Studies in Family Planning*, 39(4): 321–334.

Meyer, Bruce D, and Nikolas Mittag. 2019. "Using linked survey and administrative data to better measure income: implications for poverty, program effectiveness, and holes in the safety net." *American Economic Journal: Applied Economics*, 11(2): 176–204.

Miller, Judith D. 1984. "A New Survey Technique For Studying Deviant Behavior." PhD diss. George Washington University.

Murray-Close, M, and M.L. Heggeness. 2018. "Manning up and womaning down: How husbands and wives report their earnings when she earns more." US Census Bureau Social, Economic, and Housing Statistics Division Working Paper.

Palermo, T., J. Bleck, and A. Peterman. 2014. "Tip of the Iceberg: Reporting and Gender-Based Violence in Developing Countries." *American Journal of Epidemiology*, 179(5): 602–612.

Peterman, Amber, Tia M. Palermo, Sudhanshu Handa, and David Seidenfeld. 2017. "List randomization for soliciting experience of intimate partner violence: Application to the evaluation of Zambia's unconditional child grant program." *Health Economics*, 27(3): 622–628.

Phillips, Anna E, Gabriella B Gomez, Marie-Claude Boily, and Geoffrey P Garnett. 2010. "A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries." *International Journal of Epidemiology*, 39(6): 1541–1555.

Rosenfeld, Bryn, Kosuke Imai, and Jacob N. Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science*, 60(3): 783–802.

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive questions in surveys." *Psychological Bulletin*, 133(5): 859–883.

Traunmüller, Richard, Sara Kijewski, and Markus Freitag. 2019. "The Silent Victims of Sexual Violence during War: Evidence from a List Experiment in Sri Lanka." *Journal of Conflict Resolution*, 63(9): 2015–2042.

WHO. 2001. "Putting women's safety first: Ethical and safety recommendations for research on domestic violence against women." World Health Organization, Geneva.

WHO. 2016. "Ethical and safety recommendations for intervention research on violence against women." World Health Organization, Geneva.

World Bank. 2017. "Experimental Evaluation of MTV Shuga: Changing social norms and behaviours with entertainment education." World Bank Development Impact Evaluation (DIME). Retrieved from: http://pubdocs.worldbank.org/en/438421467236362785/Entertainment-Edu-workshop-Flyer-6-3-16.pdf.

A Appendix: Figures and Tables

Figure A1: ACASI screen and administration



Table A1: Rwanda: survey structure

Consent form
Demographics
Housing, assets & income
Labour & agriculture
Household decion-making
Gender attitudes & relationship quality
Personal attitudes about IPV (ACASI)
Experience of IPV (ACASI)- women; Likelihood of perpetrating violence (ACASI)- men

2 experimental sexual violence + IPV questions		
ACASI	*Face-to-face*	*List method*
2 direct ACASI questions	2 direct face-to-face questions	2 sensitive list questions
6 direct non-sensitive questions	6 direct non-sensitive questions	

Beliefs about community IPV attitudes and behaviour
Social networks
Behavioural games

Table A2: Nigeria: survey structure

Consent form
Household roster- demographics
Business & income
Labour
Expenditure
Assets & housing
Nutrition & food security
Plot & crop roster
Happiness, shocks & family history
Savings and loans
Gender attitudes & relationship quality
3 experimental IPV questions
Face-to-face — *List method*
3 non-sensitive sets of list questions — 3 sensitive sets of list questions
Women's empowerment in agriculture
3 direct experimental questions
5 direct IPV questions
Anthropometrics measures

Table A3: Rwanda- physical violence question wording by method

Sensitive list group	Non-sensitive list method group	
List experiment	Face-to-face	ACASI
Of the following 4 things, how many HOW MANY of them have happened to you in the past 12 months?	1. In the past 12 months, has your partner pushed or thrown you against something like a wall or piece of furniture?	1. If your partner has pushed or thrown you against something like a wall or piece of furniture in the past 12 months, press the blue square. If he has not, press the red square. If you do not want to answer this question, press the star.
1. You have taken care of a sick family member	2. Have you taken care of a sick family member in the past 12 months?	2. Have you taken care of a sick family member in the past 12 months?
2. You have gone to a health worker to talk about your health	3. Have you gone to a health worker to talk about your health in the past 12 months?	3. Have you gone to a health worker to talk about your health in the past 12 months?
3. You forgot to return a borrowed item	4. Have you forgotten to return a borrowed item in the past 12 months?	4. Have you forgotten to return a borrowed item in the past 12 months?
4. Your husband has pushed or thrown you against something like a wall or piece of furniture.		
[0,1,2,3, or 4]	[yes/no]	[yes/no]

Table A4: Non-sensitive questions

Rwanda - Has this happened to you in the past 12 months/ ever?
Women Q1
Have you taken care of a sick family member in the past 12 months?
Have you gone to a health worker and talked about your health in the past 12 months?
Have you forgotten to return a borrowed item in the past 12 months?
Women Q2
Have you ever borrowed a neighbor's tool for farm work?
Have you ever travelled to another village for a week or more for work?
Have you ever used a phone to transfer mobile phone credit to a friend?
Men Q1
Have you gone to a health worker and talked about your health in the past 12 months?
Have you forgotten to return a borrowed item in the past 12 months?
Have you talked to a friend about their health troubles in the past 12 months?
Men Q2
Have you borrowed a neighbor's tool for farm work in the past 12 months?
Have you traveled to another village for work for a week or more in the past 12 months?
Have you used a phone to transfer mobile phone credit to a friend in the past 12 months?
Nigeria - HOW MANY of the following 3 things have happened to you in the past 12 months?
Q1
You purchased a fan
You went to Abuja with a family relative
You participated in a marriage celebration in a neighborhood household
Q2
You gave a gift to another woman who gave birth in the village
You talked to a health worker about your health
You talked to the village leader
Q3
You participated in a name giving ceremony in a neighborhood household
You went to Kano
You took care of a sick family member who was unable to care for themselves

Table A5: Rwanda: women- balance across 3 survey method groups

		(1)		(2)		(3)	(4)	(5)	(6)	F-test for
		ACASI		Face-to-face		List	Normalized difference			joint
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)	orthogonality
Age	935	36.858	919	37.513	870	36.733	-0.07	0.01	0.08	0.189
		[0.309]		[0.329]		[0.335]				
Age at marriage/living with partner	934	21.301	918	21.788	870	21.320	-0.109	0.00	0.10	0.03**
		[0.134]		[0.158]		[0.145]				
In a civil/legal marriage	935	0.491	919	0.491	870	0.500	0.00	-0.02	-0.02	0.91
		[0.016]		[0.016]		[0.017]				
Wife's years of education	935	4.443	919	4.405	870	4.199	0.01	0.08	0.07	0.18
		[0.102]		[0.098]		[0.097]				
Number of biological children	935	3.775	919	3.804	869	3.776	-0.01	0.00	0.01	0.951
		[0.072]		[0.073]		[0.076]				
Wife worked as a paid employee in past week	935	0.391	919	0.402	870	0.399	-0.02	-0.02	0.01	0.9
		[0.016]		[0.016]		[0.017]				
Savings (RWF)	935	23433.950	919	29867.790	870	25093.729	-0.05	-0.026	0.04	0.438
		[1363.894]		[5617.442]		[2741.997]				
One spouse is a member of a VSLA	935	0.748	919	0.749	870	0.748	0.00	0.00	0.00	0.999
		[0.014]		[0.014]		[0.015]				
Wife worked in agriculture in past week	935	0.898	919	0.899	870	0.891	0.00	0.03	0.03	0.824
		[0.010]		[0.010]		[0.011]				
Husband ever drinks alcohol	932	0.572	915	0.532	867	0.602	0.08	-0.06	-0.14	0.01**
		[0.016]		[0.017]		[0.017]				
Husband has other wives/partners	934	0.040	918	0.035	867	0.040	0.03	0.00	-0.03	0.803
		[0.006]		[0.006]		[0.007]				

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A6: Rwanda: women- balance across 2 list method groups

		(1)		(2)	(3)	(4)
		List non-sensitive		List sensitive	Normalized	F-test for
		group		group	difference	joint
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)	orthogonality
Age	1854	37.182	870	36.733	0.046	0.264
		[0.226]		[0.335]		
Age at marriage/living with partner	1852	21.542	870	21.320	0.051	0.219
		[0.104]		[0.145]		
In a civil/legal marriage	1854	0.491	870	0.500	-0.018	0.656
		[0.012]		[0.017]		
Wife's years of education	1854	4.424	870	4.199	0.075	0.067*
		[0.071]		[0.097]		
Number of biological children	1854	3.790	869	3.776	0.006	0.878
		[0.052]		[0.076]		
Wife worked as a paid employee in past week	1854	0.396	870	0.399	-0.005	0.905
		[0.011]		[0.017]		
Savings (RWF)	1854	26623.108	870	25093.729	0.014	0.739
		[2868.367]		[2741.997]		
One spouse is a member of a VSLA	1854	0.748	870	0.748	-0.000	0.993
		[0.010]		[0.015]		
Wife worked in agriculture in past week	1854	0.899	870	0.891	0.026	0.534
		[0.007]		[0.011]		
Husband ever drinks alcohol	1847	0.552	867	0.602	-0.101	0.015**
		[0.012]		[0.017]		
Husband has other wives/partners	1852	0.037	867	0.040	-0.016	0.693
		[0.004]		[0.007]		

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A7: Rwanda: men- balance across 3 survey method groups

Variable	N	(1) ACASI Mean/SE	N	(2) Face-to-face Mean/SE	N	(3) List Mean/SE	Normalized difference (1)-(2)	(1)-(3)	(2)-(3)	F-test for joint orthogonality
Age	940	40.895	907	41.183	877	40.958	-0.03	-0.006	0.02	0.83
		[0.351]		[0.348]		[0.357]				
Age at marriage/living with partner	940	25.367	907	25.696	877	25.613	-0.06	-0.04	0.01	0.48
		[0.188]		[0.207]		[0.206]				
Household size (exclusing self and partner)	940	3.491	907	3.492	877	3.486	-0.000	0.00	0.00	1.00
		[0.063]		[0.066]		[0.063]				
Husband's years of education	940	4.566	907	4.605	877	4.420	-0.01	0.05	0.06	0.41
		[0.096]		[0.106]		[0.106]				
Household member skipped meal last week	940	0.417	907	0.396	877	0.431	0.04	-0.03	-0.07	0.31
		[0.016]		[0.016]		[0.017]				
Paid bride price	940	0.373	907	0.399	877	0.391	-0.05	-0.04	0.02	0.51
		[0.016]		[0.016]		[0.016]				
Worked as paid employee in past week	940	0.583	907	0.579	877	0.577	0.01	0.01	0.00	0.97
		[0.016]		[0.016]		[0.017]				
Number of traumatic experiences in conflict (max 7)	939	3.265	905	3.255	877	3.282	0.01	-0.01	-0.02	0.93
		[0.050]		[0.050]		[0.052]				
Contraception use	938	0.755	907	0.766	875	0.759	-0.03	-0.01	0.02	0.84
		[0.014]		[0.014]		[0.014]				

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A8: Nigeria: balance across 2 survey method groups

Variable	(1) Face-to-face N	(1) Face-to-face Mean/SE	(2) List sensitive group N	(2) List sensitive group Mean/SE	Normalized difference (1)-(2)	F-test for joint orthogonality
Wife can read and write	1412	0.159	1405	0.132	0.08	0.04**
		[0.010]		[0.009]		
Number of children	1412	4.674	1405	4.722	-0.02	0.60
		[0.067]		[0.065]		
Age	1412	35.118	1405	35.011	0.01	0.80
		[0.298]		[0.288]		
In polygamous marriage	1412	0.414	1405	0.396	0.04	0.35
		[0.013]		[0.013]		
In paid work last week	1412	0.004	1405	0.010	-0.07	0.07*
		[0.002]		[0.003]		
Household experienced accidents/disasters past year	1412	3.000	1405	3.070	-0.06	0.09*
		[0.030]		[0.029]		
Current value of household assets (Naira)	1412	22689.122	1405	21154.032	0.05	0.18
		[846.079]		[755.367]		
Husband's age when first married	1410	22.994	1403	24.243	-0.06	0.09*
		[0.159]		[0.728]		
Number of gender equitable attitudes held	1412	2.505	1405	2.528	-0.04	0.30
		[0.016]		[0.016]		
Locus of control	1412	10.794	1405	10.631	0.06	0.14
		[0.077]		[0.078]		

Notes: The value displayed for F tests are F statistics for the differences in means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure A2: Rwanda- relationships between physical IPV and respondent characteristics across survey methods

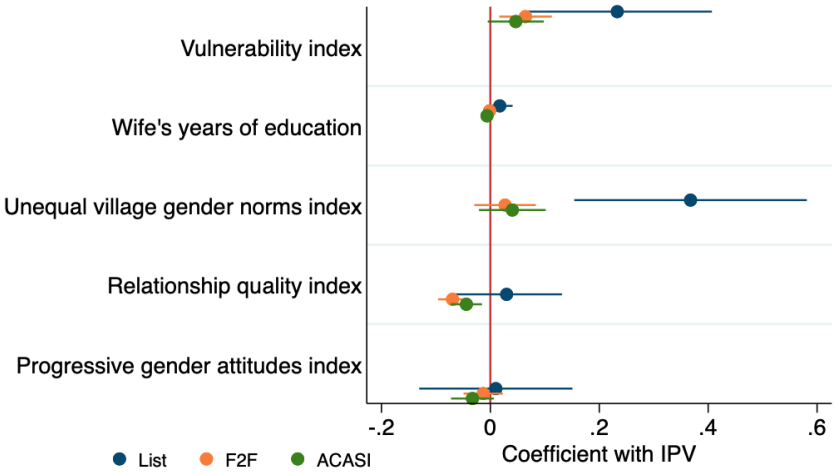


Figure A3: Nigeria- relationships between IPV and respondent characteristics across survey methods

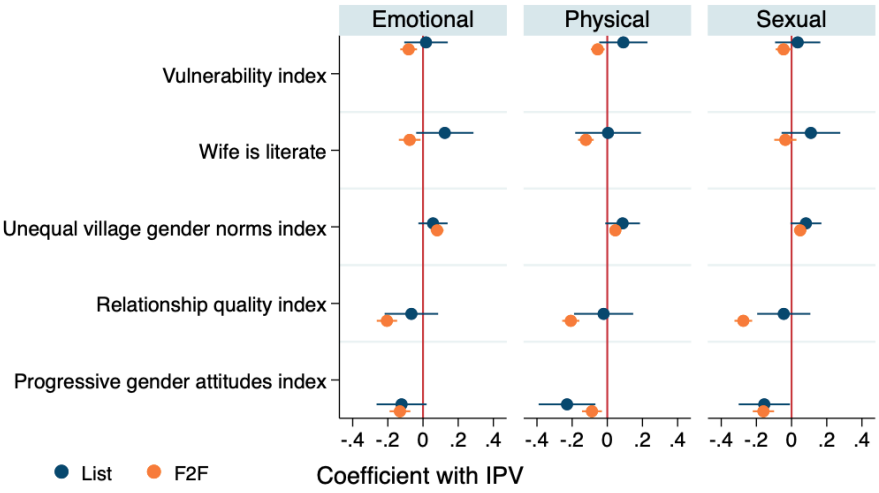


Table A9: Rwanda: Difference in mean physical IPV prevalence across methods and by respondent characteristics

	Below median			Above median		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	List	F2F	ACASI	List	F2F	ACASI
Vulnerability index	0.12	0.08	0.10	0.28	0.10	0.11
	(0.05)	(0.01)	(0.01)	(0.04)	(0.01)	(0.01)
Unequal village gender norms index	0.09	0.09	0.08	0.31	0.10	0.12
	(0.05)	(0.01)	(0.01)	(0.04)	(0.01)	(0.02)
Relationship quality index	0.20	0.14	0.13	0.21	0.05	0.07
	(0.05)	(0.02)	(0.02)	(0.05)	(0.01)	(0.01)
Progressive gender attitudes index	0.21	0.10	0.12	0.21	0.08	0.09
	(0.05)	(0.01)	(0.02)	(0.05)	(0.01)	(0.01)
Gender engagement index	0.34	0.12	0.11	0.15	0.08	0.10
	(0.06)	(0.02)	(0.02)	(0.04)	(0.01)	(0.01)
Wife's years of education	0.20	0.10	0.13	0.21	0.09	0.08
	(0.05)	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)
Wife's age	0.19	0.08	0.10	0.22	0.11	0.11
	(0.05)	(0.01)	(0.01)	(0.05)	(0.01)	(0.01)
Indicator variables	Not legally married			Legally married		
Legally married	0.23	0.09	0.12	0.18	0.10	0.09
	(0.05)	(0.01)	(0.02)	(0.05)	(0.01)	(0.01)
	Wife not in paid work			Wife in paid work		
Wife in paid work last week	0.27	0.09	0.09	0.11	0.09	0.12
	(0.04)	(0.01)	(0.01)	(0.05)	(0.02)	(0.02)

Notes: This table shows the difference in IPV prevalence by a number of repondent characteristics for the physical IPV question. The left column lists the respondent characteristics correlated with IPV. Columns 1-6 show the mean IPV prevalence by respondent characteristic (above or below median for continuous variables, or by binary characterisics). Columns 1 and 4 were estimated using a regression on the number of list items reported for the restricted sample of those above and below median. Standard errors are shown in parentheses.

Figure A4: Rwanda- Above and below median: difference in mean physical IPV prevalence across methods by respondent characteristics

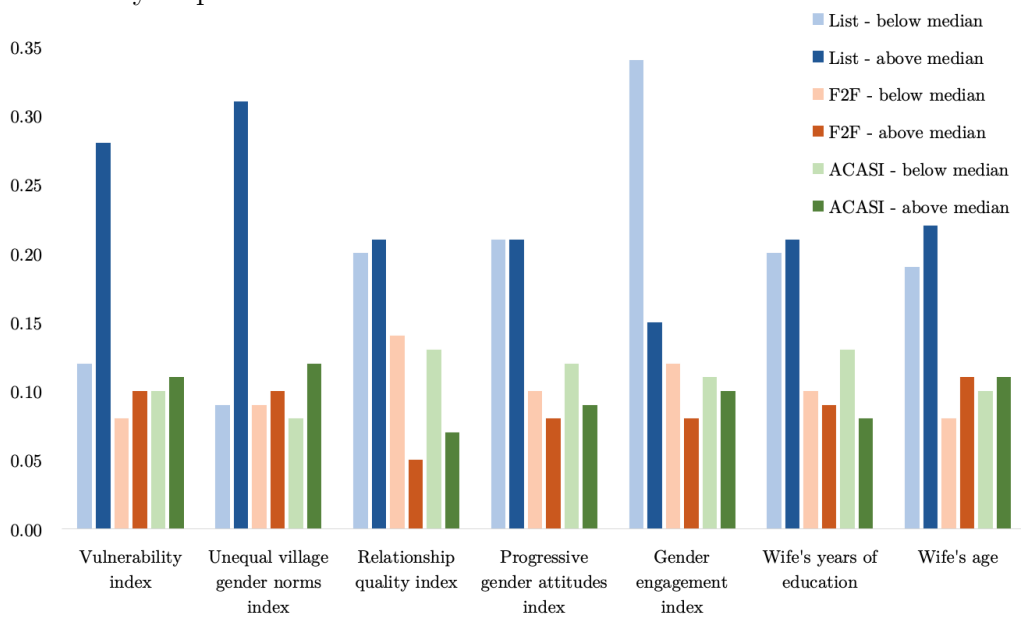


Table A10: Nigeria: Difference in mean IPV prevalence across methods and by respondent characteristics

	Emotional IPV				Physical IPV				Sexual IPV			
	Below median		Above median		Below median		Above median		Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variable	List	F2F	List	F2F	List	F2F	List	F2F	List	F2F	List	F2F
Vulnerability index	0.43	0.33	0.36	0.26	0.27	0.20	0.26	0.19	0.27	0.30	0.25	0.26
	(0.04)	(0.02)	(0.05)	(0.02)	(0.05)	(0.01)	(0.05)	(0.01)	(0.04)	(0.02)	(0.05)	(0.02)
Unequal village gender norms index	0.40	0.24	0.39	0.36	0.24	0.15	0.29	0.23	0.24	0.27	0.28	0.30
	(0.04)	(0.02)	(0.04)	(0.02)	(0.05)	(0.01)	(0.05)	(0.02)	(0.05)	(0.02)	(0.05)	(0.02)
Relationship quality index	0.44	0.39	0.36	0.22	0.29	0.27	0.24	0.13	0.28	0.39	0.25	0.19
	(0.04)	(0.02)	(0.05)	(0.01)	(0.05)	(0.02)	(0.05)	(0.01)	(0.05)	(0.02)	(0.05)	(0.01)
Progressive gender attitudes index	0.41	0.35	0.38	0.25	0.30	0.24	0.23	0.14	0.28	0.35	0.24	0.21
	(0.04)	(0.02)	(0.04)	(0.02)	(0.05)	(0.02)	(0.05)	(0.01)	(0.05)	(0.02)	(0.05)	(0.02)
Wife's age	0.36	0.27	0.43	0.32	0.21	0.18	0.31	0.21	0.23	0.28	0.29	0.27
	(0.04)	(0.02)	(0.04)	(0.02)	(0.05)	(0.01)	(0.05)	(0.01)	(0.05)	(0.02)	(0.05)	(0.02)
Indicator variables	Wife illiterate		Wife literate		Wife illiterate		Wife literate		Wife illiterate		Wife literate	
Wife reads and writes	0.38	0.31	0.48	0.24	0.27	0.21	0.24	0.09	0.25	0.29	0.33	0.25
	(0.03)	(0.01)	(0.08)	(0.00)	(0.04)	(0.01)	(0.09)	(0.00)	(0.04)	(0.01)	(0.08)	(0.00)
	Least confident		Most confident		Least confident		Most confident		Least confident		Most confident	
Confidence in speaking up in village	0.41	0.34	0.38	0.21	0.25	0.21	0.29	0.15	0.27	0.29	0.25	0.27
	(0.04)	(0.02)	(0.05)	(0.02)	(0.05)	(0.01)	(0.06)	(0.02)	(0.04)	(0.01)	(0.05)	(0.02)
	Polygamous		Monogamous		Polygamous		Monogamous		Polygamous		Monogamous	
Marriage type	0.36	0.32	0.43	0.28	0.17	0.21	0.33	0.18	0.19	0.31	0.31	0.26
	(0.05)	(0.02)	(0.04)	(0.02)	(0.06)	(0.02)	(0.05)	(0.01)	(0.05)	(0.02)	(0.04)	(0.02)
	No assets to marriage		Brought assests		No assets to marriage		Brought assests		No assets to marriage		Brought assests	
Wife brought assets to marriage	0.42	0.40	0.38	0.22	0.25	0.27	0.27	0.14	0.28	0.33	0.25	0.24
	(0.00)	(0.02)	(0.04)	(0.01)	(0.00)	(0.02)	(0.04)	(0.01)	(0.00)	(0.02)	(0.04)	(0.02)

Notes: This table shows the difference in IPV prevalence by a number of repondent characteristics, for each of the 3 IPV questions asked in Nigera. The left column lists the variables correlated with IPV along the three types of violence. Columns 1-12 show the mean IPV prevalence by respondent characteristics (above or below median for continuous variables, or by binary characterisics). Coefficients in columns 1, 3, 5, 7, 9, and 11 were estimated using a regression on the number of list items reported. Standard errors are shown in parentheses.

Figure A5: Nigeria- difference in mean IPV prevalence by respondent characteristics: above and below median (left) and indicator variables (right)

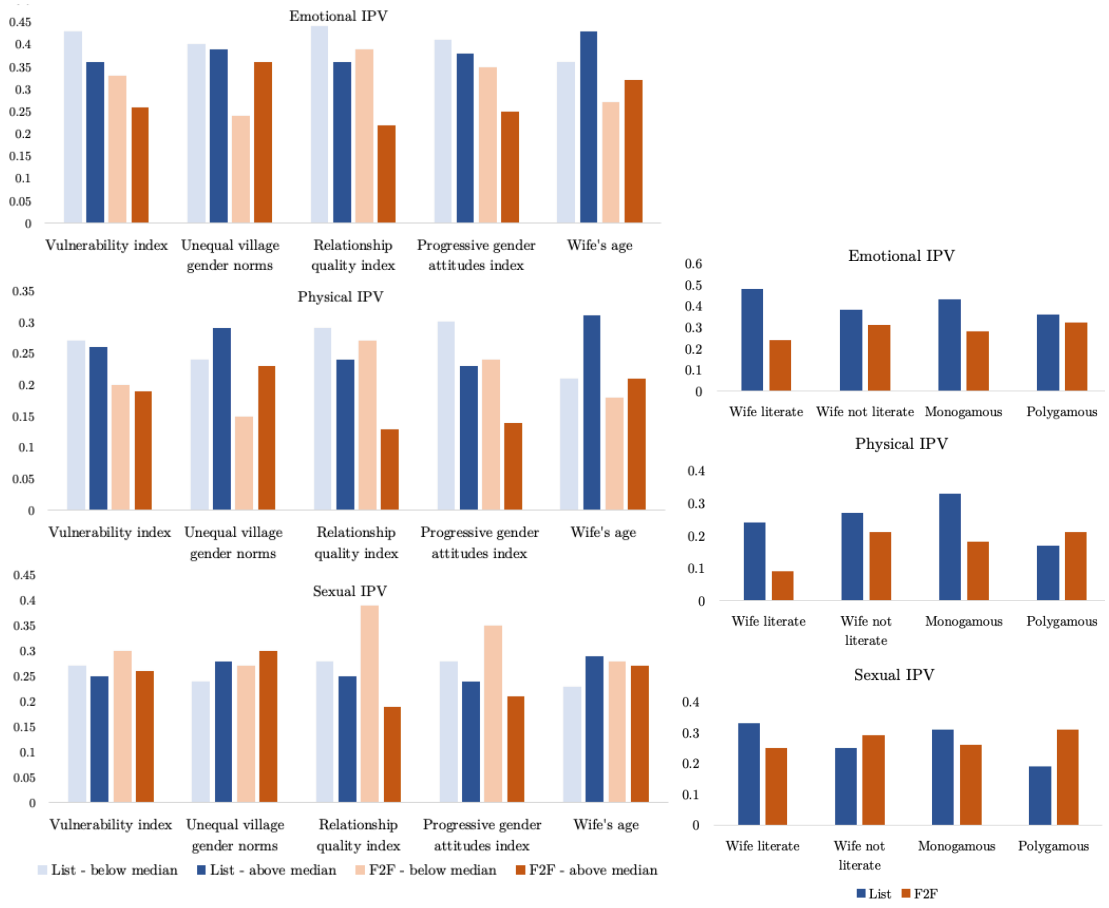


Table A11: Rwanda- No design effects test and proportions responding with each option

	Women (% of respondents)				Men (% of respondents)			
	Q1		Q2		Q1		Q2	
Response	Non-sensitive	Sensitive	Non-sensitive	Sensitive	Non-sensitive	Sensitive	Non-sensitive	Sensitive
0	3.67	6.76	13.69	19.84	8.97	8.55	11.24	16.88
1	28.03	18.9	60.22	43.81	39.22	21.32	46.14	32.5
2	53.05	43.99	22.26	28.78	44.57	40.82	34.74	38.65
3	15.26	27.84	3.83	6.65	7.24	24.97	7.89	11.29
4		2.52		0.92		4.33		0.68
No design effect p-value:		0.0012		0.0001		1		0.0001

Note: Table shows Bonferroni-corrected minimum P-values from no design effects test. Null hypothesis is of no design effect. Computed using List package on R (Blair & Imai, 2011)

Table A12: Nigeria- No design effects test and proportions responding with each option

	Nigeria (% of respondents)					
	Q1		Q2		Q3	
Response	Non-sensitive	Sensitive	Non-sensitive	Sensitive	Non-sensitive	Sensitive
0	13.81	8.9	12.54	8.61	11.47	11.25
1	59.99	40.64	47.95	38.01	50.14	38.72
2	19.62	37.15	24.86	35.59	32.93	33.67
3	6.59	9.54	14.66	12.38	5.45	13.17
4		3.77		5.41		3.2
No design effect p-value:		1		1		1

Note: Table shows Bonferroni-corrected minimum P-values from no design effects test. Null hypothesis is of no design effect. Computed using List package on R (Blair & Imai, 2011)

Table A13: Rwanda women - robustness checks with controls and enumerator fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Q1: Past year emotional IPV			Q2: Lifetime non-partner sexual violence		
List method prevalence	0.206***	0.203***	0.203***	0.088***	0.102***	0.084**
	(0.035)	(0.033)	(0.035)	(0.034)	(0.031)	(0.033)
Observations	2,728	2,715	2,728	2,727	2,714	2,727
Controls		x			x	
Enumerator fixed effects			x			x

Notes: Controls selected using PDS lasso (Belloni et al., 2014). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A14: Rwanda men - robustness checks with controls and enumerator fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Q1: Limited wife's family contact			Q2:Threatened to hurt wife/someone close		
List method prevalence	0.451***	0.440***	0.442***	0.071**	0.065*	0.067*
	(0.038)	(0.034)	(0.038)	(0.036)	(0.034)	(0.037)
Observations	2,728	2,721	2,728	2,728	2,721	2,728
Controls		x			x	
Enumerator fixed effects			x			x

Notes: Controls selected using PDS lasso (Belloni et al., 2014). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A15: Nigeria women - robustness checks with controls and enumerator fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Q1: Past year emotional IPV			Q2: Past year physical IPV			Q3: Past year sexual IPV		
List method prevalence	0.397***	0.390***	0.415***	0.263***	0.258***	0.271***	0.260***	0.258***	0.275***
	(0.032)	(0.031)	(0.026)	(0.035)	(0.035)	(0.027)	(0.032)	(0.032)	(0.025)
Observations	2,817	2,800	2,817	2,817	2,800	2,817	2,817	2,800	2,817
Controls		x			x			x	
Enumerator fixed effects			x			x			x

Notes: Controls selected using PDS lasso (Belloni et al., 2014). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A16: Refusal rates by question and method

	Face-to-face		ACASI		List	
Rwanda	No. respondents	%	No. respondents	%	No. respondents	%
Women Q1	0	0	23	0.84	0	0
Women Q2	1	0.04	23	0.84	1	0.04
Men Q1	0	0	66	2.41	0	0
Men Q2	0	0	16	0.59	0	0

	Face-to-face		List	
Nigeria	No. respondents	%	No. respondents	%
Q1	5	0.18	0	0
Q2	5	0.18	0	0
Q3	18	0.64	0	0

B Supplementary discussion on IPV measurement

The evidence presented in this paper regarding the substantial and systematic misreporting on IPV in Rwanda and Nigeria suggests several issues for consideration by those planning to measure IPV and other sensitive topics.

What trade-offs should social scientists and policy makers consider when deciding what method to use when measuring self-reported IPV? The face-to-face method is the simplest approach. Unlike the list and ACASI methods, it asks questions in the same format as most other survey questions and thus does not require a lengthy period of explanation by the enumerator. Nor does it require electronic devices or headphones, nor some minimum cognitive ability for respondents to understand more complex question formats. However, as this experiment shows, the face-to-face method likely results in the most underreporting.

Despite expectations, women's self-reported IPV experience did not significantly differ between ACASI and face-to-face methods. This suggests that for researchers whose objective is to reduce misreporting bias in IPV measurement, ACASI may not be worth the additional cost (though there may be ethical, logistical, or other reasons why ACASI may be preferred).

Is the list experiment then the optimal IPV data collection method? This paper suggests that of the methods tested here, the list experiment appears to most minimize underreporting. However, there are a number of challenges associated with its design, administration, and analysis that may make it inappropriate for some purposes or in some contexts. For example, the list method is more time-consuming and cognitively demanding relative to other methods.

Other challenges with the list experiment have been well documented by Blair and Imai (2012); Blair, Imai and Lyall (2014); Blair, Coppock and Moor (2018) and Chuang et al. (2018). However, one particular issue should be highlighted: the list experiment typically has greater variance than direct questioning methods, but also likely lower misreporting bias. Blair, Coppock and Moor (2018) suggest that if the expected misreporting bias is between 5 and 10 percentage points, direct questions are preferred if sample sizes are less than 3,000. In this paper, we found that the bias for IPV measurement was larger than 10 percentage points in some cases, suggesting IPV measurement may be an area where the method's use is justified. However, for those using list experiments to estimate treatment effects, adequately powered studies likely require even larger sample sizes given that list method groups get split again by treatment arms. At the very least, researchers and policy makers can use these

results to help inform their survey method choice.[34]

It is worth noting that for researchers measuring IPV using the multi-question DHS module, an equivalent set of list experiment questions would not allow for the construction of an 'any violence' indicator variable. Traditionally, most studies aggregate women's direct yes/no responses in each category of emotional, physical, and sexual violence to create summary indicator variables. This cannot be done with the list experiment as respondents' direct answers are unknown. Replacing the traditional aggregated approach, which has been recommended because it gives women multiple chances to disclose violence, with list estimates may replace one form of bias (underreporting on individual questions) with different bias (from not being able to create a summary indicator reflecting the extent of violence experienced by each respondent). Further research could seek to establish how these biases relate and their respective impacts on prevalence and correlation estimates, and whether there are methods to simultaneously minimize both.

Finally, while the list method may minimize misreporting bias in correlations and treatment effect estimates caused by fear or shame, it cannot mitigate bias in treatment effects from interventions that change women's understanding or interpretation of terms used in IPV outcome questions such as 'humiliation', 'forced sex', or 'threatening', for example. Yet both of these sources of reporting bias- 'liberating' (from shame and fear) and 'learning' (from a changed interpretation of question wording or broadening awareness)- may confound treatment effect estimates of interventions' impact on IPV. The list method is only likely to minimize the former source of bias while leaving the latter unchanged.

[34]In addition to statistical power concerns with the list experiment, there are the well-documented challenges with designing appropriate non-sensitive list questions to prevent 'floor' and 'ceiling effects' so that respondents' answers are disguised. There are also debates about the optimal types of non-sensitive items, with some evidence suggesting that the non-sensitive items should be somewhat sensitive so that the question of interest is not salient, with associated concerns about the sensitivity of list method estimates to differences in types of non-sensitive questions. It is also possible that ordering of the sensitive question within the list matters, as well as the number of sets of list questions asked. For example, in this study, given each respondent was assigned to either the sensitive or non-sensitive list group, there may be reporting effects for the subsequent lists a respondent was read, with respondents perhaps becoming increasingly suspicious of the method, though in Nigeria, list method prevalence did not substantially decline over the 3 questions on emotional, physical and sexual violence. See Coutts and Jann (2011); Glynn (2013); Blair and Imai (2012); Blair, Coppock and Moor (2018) and Chuang et al. (2018) for a fuller discussion on list experiment trade-offs and design.

4

Can role models and correcting misperceived norms change gender attitudes? Evidence from a survey experiment in Rwanda

Can role models and correcting misperceived norms change gender attitudes? Evidence from a survey experiment in Rwanda

Claire Cullen

Abstract

An individual's perceptions about other peoples' attitudes – whether accurate or not - can influence their own attitudes and behaviors. We provide evidence from rural Rwanda that while most men and women privately do not hold conservative attitudes about household gender roles and intimate partner violence (IPV), most over-estimate the share of others who do. In this survey experiment, we assess whether correcting an individual's over-estimates of others' conservative gender attitudes affects their own gender attitudes and behaviors. We randomly assigned women and men who over-estimate conservative gender norms to receive information updates with the 'true' norm (measured 12 months prior), also varying the type of role model about which people are given information. We measure the impact of this information update on an index of self-reported gender attitudes and behavior in an incentivized spouse dictator game. We find that men's gender attitudes become more progressive and they give more to their spouse after having their misperceptions corrected; there is no average impact for women. On role models, we find that men give more to their spouse after being told about all role models, but our study suggests they may respond most to information about village leaders. For women, there is no significant difference between being updated about men or women role models. However, we find suggestive evidence that women may respond negatively to having their beliefs corrected about other women. Our results indicate a potentially important factor in gender inequality, where misperceived norms affect individuals' own gender attitudes and behaviors.

Keywords: Social norms, beliefs, gender, role models, information, survey experiment, Rwanda

1 Introduction

Gender inequality is one of the most universal and long-standing forms of inequality. Currently, women spend two to ten times more time on unpaid household and care work than men, and one in three women experiences intimate partner violence (IPV) at some point in her lifetime (Ferrant, Pesando and Nowacka, 2014; Palermo, Bleck and Peterman, 2014; WHO, 2013). In some countries, women also face limits on their freedom, mobility, and input in decisions about their family and their own health (Dhar, Jain and Jayachandran, 2018).

One potential driver of persistent and widespread gender inequality is men's and women's preference for behaving in accordance with social norms. There is strong evidence that individuals' decisions may be influenced by their beliefs about others' attitudes and behaviors, and their sense of what is socially sanctioned or penalized (social norms). These social considerations have been shown to affect behavior in numerous settings such as effort in school (Bursztyn and Jensen, 2015), political behavior (Dellavigna et al., 2017), and now, increasingly, on behaviors associated with gender inequality (Banerjee, La Ferrara and Orozco, 2019; Bursztyn, González and Yanagizawa-Drott, 2018; Field et al., 2019; Jayachandran, 2020).

Yet people may not hold accurate beliefs about what others believe and do, particularly in contexts with limited media dissemination or on less openly discussed issues. Recent evidence confirms that individuals misperceive the truth about topics that are sensitive or politically charged, such as immigration rates (Grigorieff, Roth and Ubfal, 2018), the social permissibility of xenophobia (Bursztyn, Egorov and Fiorin, 2019), and other men's attitudes about the acceptability of women's participation in the labor force (Bursztyn, González and Yanagizawa-Drott, 2018). These studies have shown that information that corrects inaccurate beliefs about social norms can cause people's own attitudes and behaviors to shift in a short period of time. Information about others' true preferences may be particularly effective in situations of pluralistic ignorance. In these circumstances, most people privately reject a norm but incorrectly believe that most others accept it and may, therefore, end up following the norm (Bursztyn, González and Yanagizawa-Drott, 2018). In settings where most individuals assume that a behavior or opinion is stigmatized, the stigma may be sustained in equilibrium (Bursztyn, Egorov and Fiorin, 2019). However, as yet, little is known about the role of norms about traditional gender roles and IPV acceptability on an individual's behavior or whether particular role model types are more influential than others.

In this paper, we present the results of an information experiment designed to understand the role of norms and role models on gender attitudes and behavior. The experiment, conducted in rural Eastern Rwanda, involved measuring men's and women's beliefs about the

gender attitudes of other men and women in nearby communities, in two attitude domains: traditional household gender roles and IPV privacy. Specifically, we measured beliefs about what share of others "believe that it is shameful for a woman if her husband cooks or does the housework?" and "believe that violence between husband and wife is a private matter and others should not intervene?". We randomly assigned respondents who over-estimated the share of people with conservative gender norms to receive information about others' 'true' gender attitudes, as measured in survey questions 12 months prior. Individuals were assigned to one of several 'role model' groups or a control group and then asked their beliefs about the gender attitudes of their role model group.[1] We then assess whether correcting norms misperceptions affects an individual's attitudes, using an index of self-reported gender attitudes, and behavior in an incentivized spousal dictator game. This survey experiment took place during the endline of an IPV prevention randomized control trial (RCT). In this paper, we restrict analysis to only individuals in the RCT control arm who live in pure control villages to avoid confounding treatment effects with intervention.

We first show that most people overestimate conservative gender norms while privately holding more progressive gender attitudes.[2] In our survey, 33% of men and 48% of women agreed with the statement: "Violence between a man and a woman is a private matter and others should not intervene," and 10% of men and 43% of women agreed with the statement: "It is shameful for a wife if her husband does the housework."[3] When asked what share of people in nearby communities they believed agreed with those statements, 68% of men and 82% of women over-estimated in at least one of the two domains. People tended to be more accurate at estimating the true norm of others with whom they are more similar. For example, women more accurately perceive the attitudes of other women than of men.

For individuals who over-estimate conservative gender norms, we then evaluate the impact of being provided the 'correct' information. For men, we find that having norms misperceptions corrected increases the progressiveness of their own self-reported gender attitudes by 0.07 standard deviations on average, and by 0.09 standard deviations if they are updated on their misperceptions of both norms. Men also give an additional 0.49 out of 10 tokens

[1] Women were asked about the views of, and given correct information on, either: men or women. Men were asked about the views of, and given correct information on either: male village leaders, or successful male business owners. The control group was asked their beliefs about each role model for their gender, and given information on none.

[2] Regarding the reliability of norms beliefs measurement, Haaland, Roth and Wohlfart (2020) discuss evidence suggesting that unincentivized survey reports strongly correlate with incentivized belief measures and that incentives generally do not have any statistically significant effect on reported beliefs. They conclude that while the evidence on truth-telling in belief measurement is mixed, there is no systematic evidence that incentivized measurement reduces experimenter demand effects, and in some cases, may even lower truth-telling. At baseline, we tested potential social desirability bias in norms reporting by measuring related community gender norms using incentivized and unincentivized questions and found they were positively correlated.

[3] This descriptive data was collected at the baseline survey of an IPV prevention RCT (Alik-Lagrange et al., 2017). The survey experiment was implemented during the endline of this study. This paper restricts analysis of the endline survey experiment to the RCT control arm.

to their spouse in a dictator game, equivalent to approximately $0.08 USD. For women, there was no average treatment effect of having overestimates corrected. However we find that women who over-estimate the share of others with conservative gender attitudes in the domain of gender roles report *worse* personal gender attitudes after having their misperceptions corrected. On role models, we find that men give more to their spouse after being told about all role models, but there is suggestive evidence they may respond most positively to information on village leaders. There is no significant difference in impacts between women updated about men or women role models. However, there is some suggestive evidence that women may respond negatively to having their beliefs corrected about other women.

Given we measure beliefs and attitudes using unincentivized, self-reported survey responses, we might worry about social desirability or another form of reporting bias driving results.[4] However, it is not obvious that reporting bias would necessarily lead to more progressive reporting in such a socially conservative setting. We find that most men and women are willing to report at least some socially conservative personal views. For example, 68% of men in our sample say they would be likely to use IPV against their wife in at least one of the hypothetical situations we pose. However, it is still possible that experimenter demand effects or another form of reporting bias may affect attitude responses. To help mitigate this risk, we also used as an additional outcome measure an incentivized dictator game, played after 10 minutes of unrelated survey questions.

There are three main contributions of this paper. First, we contribute to the growing literature demonstrating that individuals misperceive social norms, and that personal attitudes and economic behavior are influenced by norms misperceptions. Our focus on norms about gender roles and IPV and their impact on attitudes and behavior builds on a growing economic literature on gender norms and economic behavior in low-income countries (Banerjee, La Ferrara and Orozco, 2019; Beaman et al., 2012; Field et al., 2019; Green, Wilke and Cooper, 2020; Jayachandran, 2020) and more broadly on the role of norms information on decision-making (Bursztyn, Egorov and Fiorin, 2019; Bursztyn, González and Yanagizawa-Drott, 2018; Grigorieff, Roth and Ubfal, 2018). Our results - that men change their attitudes and behavior in response to having their norms beliefs corrected - highlight a potentially important margin for promoting social change.

Second, this paper contributes to the economic literature on role models and aspirations (Akerlof, 1997; Angrist, 2020; Del Carpio and Guadalupe, 2018; LaFerrara, 2019; LaFerrara, Chong and Duryea, 2012; Mani and Riley, 2019; Ray, 2006). As yet, there is limited evidence on the types of role models and social referents that may be more or less effective at influencing gender norms. We contribute to this research, showing that people have different

[4]For ethical reasons, we did not correct people who believed others were more *progressive* than the 'true' norm.

perceptions about different groups' beliefs, often holding the most accurate beliefs about those they are most proximate to. Our findings also suggest that individuals respond differently to norms information about different role model types, potentially based on a shared identity (in our case, sex), or those in positions of authority (in our case, village leaders).

Finally, our results contribute to the literature on motivated beliefs, suggesting that some people may self-deceive and be resistant to, or backlash against, attempts to shift their attitudes with evidence (Bénabou and Tirole, 2016; Kahan, 2013; Kappes and Sharot, 2019; Zimmermann, 2020). Our findings suggest that most people do not incorporate the new norm information into updated posterior beliefs. This may indicate that norms beliefs reflect motivated reasoning and thus may be resistant to change. In such cases, advocacy campaigns or government programs designed to shift attitudes and behaviors by encouraging people to coordinate around a new progressive norm or by using evidence to try to persuade people to change behavior may be ineffective or potentially harmful. Overall, this paper builds on the emerging evidence that social norms and other behavioral factors may be an important decision-making friction that contributes to gender inequality.

This paper is structured as follows. Section 2 discusses the context, data, and experimental design. Section 3 presents and discusses the results, and Section 4 concludes.

2 Context and data

2.1 Context

This study took place in Eastern Rwanda during the endline survey of an IPV prevention study, reported in Cullen et al. (2020) (in Chapter 2 of this thesis). The Government of Rwanda has undertaken numerous progressive reforms over the past two decades to improve women's representation in parliament, ability to inherit property, ability to divorce, and to outlaw domestic violence. Yet these progressive reforms have taken place in a region with some of the most conservative gender attitudes and highest rates of IPV in the world (Global Burden of Disease, 2017). For example, in Eastern province, the area where this study took place, 43% of women in our sample believe that a husband is justified in hitting his wife if she refuses to have sex with him, and 45% believe he is justified in hitting her if she burns the food. The comparable share of men agreeing is 16% and 14%, respectively. These conservative gender attitudes are also reflected in very high rates of IPV. In this sample, 44% of women reported having experienced sexual IPV in the previous 12 months.

2.2 Sampling

The survey experiment sample includes married or cohabiting heterosexual couples in Rwanda's Eastern Province. The sample comprised couples from Village Savings and Loan Associations (VSLAs) who had applied to be part of a Government-funded NGO-delivered 5-month weekly couples discussion group designed to prevent IPV. The survey experiment sample included the VSLA members who were participating in an RCT evaluating the IPV prevention program.[5] The survey experiment was administered during the endline of this RCT. We present results only in the RCT control arm to avoid confounding effects with the intervention. This left the 49 control villages from a total of 98 RCT study villages.

Only individuals who over-estimated the share of others with conservative gender attitudes are included in the survey experiment sample (we thus exclude the 137 women and 240 men from the RCT control arm who did not over-estimate either of the two conservative norms), leaving 554 men and 662 women in our sample. During the RCT baseline, Cullen et al. (2020) used household rosters to survey a random sample of community members and found that they were very similar to VSLA members in this sample on observable characteristics, though VSLA couples appear marginally younger, more educated, more economically vulnerable, and slightly more conservative on gender issues. For additional information on sampling, see the main study description and pre-analysis plan (Alik-Lagrange et al., 2017; Cullen et al., 2020).

2.3 Survey experiment

The data for the survey experiment was collected in two stages. The descriptive data about the 'true' norm was collected during the RCT baseline, and the survey experiment itself was conducted 12 months later, during the RCT endline.[6] The baseline 'true norm' data was collected from the full sample, including study-eligible VSLA members in the RCT control, treatment, and spillover arms, as well as randomly sampled community members from village household rosters.

Before the endline survey, women were randomly assigned to one of three role model groups: men role models, women role models, or control; men were assigned to one of four role model groups: men, male village leaders, successful male business owners, and control. Randomization was stratified on the RCT treatment group, village, and baseline responses to incentivized questions about community gender norms.

[5]The RCT is described in detail in Cullen et al. (2020) and included a two-level randomization of villages and individuals into treatment or control groups.

[6]Baseline data was collected between November 2017-February 2018. The baseline sample included the same VSLA couples interviewed for the survey experiment, plus a representative village sample of an additional 700 couples and 200 village leaders from the study villages.

The survey experiment was then conducted during the endline as follows. After individuals had spent approximately 45 minutes answering survey questions regarding demographics, employment, savings, gender attitudes, and IPV for the main RCT outcomes, we administered the survey experiment. First, we measured men's and women's beliefs about the gender attitudes of other men and women in nearby communities. Specifically, we measured beliefs about what share of people in their assigned role model group:

1. "believe that it is shameful for a woman if her husband cooks or does the housework?" and

2. "believe that violence between a husband and wife is a private matter and others should not intervene?"

People assigned to a role model treatment group and who had over-estimated the share of community members holding conservative views when compared to the 'truth' were provided with information on the 'correct' share (which was measured 12 months prior). The control group was asked their beliefs about all role model groups but were not given any information on the true norm. The baseline survey had included questions about the respondent's occupation, income, and whether they were a village leader.[7] This data was used to generate the summary statistics for the types of role models that our qualitative and quantitative research and existing literature suggested may be important for influencing attitudes and behaviors.[8]

We used baseline data on individuals' privately held gender attitudes to calculate the 'true norm' and at endline, to assess the extent to which individuals misperceive norms and to correct overestimates. These descriptive statistics are shown in Table 1. The summary statistics of the true norm that individuals were informed of were as follows:

1. The proportion of [role model type] in villages near here who believe that it is shameful for a woman if her husband cooks or does housework: [Men: 10%; Successful male business owners: 8%; Male village leaders: 9%; Women: 43%].

2. The proportion of [role model type] in villages near here who believe that violence between a husband and wife is a private matter and others should not intervene: [Men: 33%; Successful male business owners: 23%; Male village leaders: 17%; Women: 47%].

[7] Given that the 'true norm' data was collected 12 months prior, it is possible that people discounted the information updates given to them from baseline as 'old information'. However, in the RCT control villages considered in this paper, there was no substantial change in IPV rates or gender attitudes between baseline and endline.

[8] At baseline, we asked men whose relationships they admired, with some of the most popular responses being successful business owners (23%) and community leaders (12%).

The norms information statistics were intentionally kept constant across all 49 villages to avoid sampling variation. While these average sample norms might differ from the local village norms, we find that results are robust across villages with different characteristics.

After treated respondents had received the correct norms information, we measured respondents' gender attitudes on a set of related questions about household gender roles and IPV acceptability and privacy, and their allocation to their spouse in an incentivized dictator game. For ethical reasons, we only updated individuals on the true norm if they over-estimated conservative gender attitudes. We did not correct under-estimates as that may have risked increasing conservative gender attitudes or behaviors. The conclusions we draw are thus limited to the sample of conservative norm over-estimators. In the results section, we discuss how the over-estimator sample differs from under-estimators.

The two norms domains of household gender roles and IPV privacy were selected after baseline data and qualitative work indicated they had large wedges between privately held attitudes and beliefs about others' attitudes. This suggested a potentially important pluralistic ignorance margin to correct beliefs.[9] Further, while these domains are often categorized together under the umbrella of 'gender norms,' individuals often had different attitudes in the two domains, suggesting different factors may drive them. They also posed a minimal risk of unintended harm, unlike updating people on IPV behaviors.

The survey experiment was conducted between January and March 2019 by trained data collectors the same sex as the respondent, in private interview settings. Figure A4 in the appendix summarizes the experiment. The pre-analysis plan was registered on the AEA Trial Registry (Cullen, 2020).

2.4 Outcomes

For respondents who over-estimated conservative gender norms, we assess the treatment effects of being given correct information on four outcome measures- three attitude outcomes and one incentivized behavioral outcome. They are listed below.

1. An overall standardized weighted index (Anderson, 2008) comprised of the 'progressiveness' of the respondent's answers to the nine gender attitudes questions asked after

[9]This private attitude-social norms belief wedge was calculated by comparing respondents' beliefs using the baseline question: "Thinking about all men and women in your village, generally speaking, how many men in your village think violence between husband and wife is a private matter and others should not intervene," with the 'truth' collected at baseline using the average share of respondents agreeing with the statement: "Violence between husband and wife is a private matter and others should not intervene. Do you agree or disagree?". During baseline, the former norms perception question was answered using a likert scale with options of "Almost Nobody," "A few people," "About 25%," "About 50%," "About 75%," "Almost everyone." Therefore, we could not directly calculate a raw score of the wedge between baseline norms perceptions and baseline mean attitudes, but we could infer an indicative, relative size of the gap. For ethical reasons, we did not provide people with updated information on others' IPV-related behaviors, given the potential for adverse consequences. The baseline norms perceptions options were based on the questions and visual aids from Paluck and Shepherd (2016).

they were informed of the correct norm.

We also consider the impact on the two different domains of gender norms- traditional household gender roles and IPV privacy. We consider them separately because norms across the two domains may be quite different and more or less open to change, and respondents may have been given corrected information on zero, one, or both domains:

2. A standardized weighted index of the 'progressiveness' of respondents' answers to the four questions about traditional household gender roles

3. A standardized weighted index of the 'progressiveness' of respondents' answers to the five questions about IPV stigma and acceptability.[10]

We also consider impacts on decisions in an incentivized dictator game played at the end of our survey:

4. Number of 0-10 tokens allocated to the respondent's spouse (each token given to the spouse represents 150 Rwandan Francs, or approximately $0.20 USD).

This game was played approximately 10 minutes after the survey experiment, and after other unrelated survey questions were asked, thus potentially masking the game as being related to the information experiment and minimizing experimenter demand effects. Experimenter demand effects are likely to be lower in tasks with real money at stake than gender attitude questions asked immediately after the norms information provision (Haaland, Roth and Wohlfart, 2020).[11] In our context, the game was primarily intended to provide an incentivized measure of whether men increase sharing with their spouse based on information that others are more gender progressive. We were agnostic about whether an increase in tokens that a wife gives her husband is a positive outcome given it could also indicate women are relinquishing power in household bargaining.

Given that we gathered numerous outcome indicators within each norm domain, we construct Anderson (2008) summary indices as the most efficient weighted average to test whether the intervention had effects on related outcomes. The index weights each outcome using the inverse of their variance-covariance matrix. This estimator provides the most efficient estimation of the treatment effect across related outcomes (Baranov et al., 2020),

[10]These attitude questions were based on piloting, qualitative research, and a combination of sources, including Demographic and Health Surveys, the WHO Multi-Country Study on Women's Health and Domestic Violence, the International Men and Gender Equality Survey, and questions listed in the J-PAL Guide to Measuring Women's and Girls' Empowerment in Impact Evaluations (DHS, 2016; Fleming et al., 2015; Garcia-Moreno et al., 2005; Glennerster, Walsh and Diaz-Martin, 2018; Green, Wilke and Cooper, 2018).

[11]Appendix B includes the study script, experiment questions, and visual aid. Appendix C provides the script and aids used in the spouse dictator game, where individuals, in private, were asked to allocate between their spouse and themselves ten tokens representing real money. This game was based on that in Hoel (2015).

and controls the false discovery rate by collapsing numerous outcomes into one index. Given only two men and seven women refused to answer an outcome question, we do not construct lee bounds for treatment estimates.

We also assess whether respondents updated their posterior beliefs about conservative community norms when provided with the information update. Posteriors were measured using one post-information update question using a 6-point likert scale about whether they believe most people in their village would agree with conservative attitudes expressed in a hypothetical scenario.[12]

This experiment is likely underpowered. We based initial power calculations on baseline estimates of the over-estimation of conservative norms in the full sample (approximately 50% of people over-estimated). We had planned to use the full RCT sample, assuming that the norms experiment treatment effects would not vary substantially by RCT treatment arm, and that there would be small RCT effects, particularly in the spillover arm. However, contrary to expectations, the main RCT had significant effects in both the treatment and spillover arms and also affected how individuals responded to the norms experiment. To avoid confounding this experiment's effects with the RCT, we limit our analysis to the RCT control arm. Power calculations suggested we would have 80 percent power to detect a treatment effect of 0.12 standard deviations in standardized gender attitude scores with a sample of 273 in the intervention arms and 273 in the control arm, and standard deviations of 0.5. This is similar to, or more conservative than effect sizes found in other information experiments (Haaland, Roth and Wohlfart, 2020).

2.5 Estimation strategy

2.5.1 Impact of getting information on others' beliefs

Gender norm domain: household gender roles and IPV privacy

We evaluate the causal effect of being assigned to get an information update by estimating intention-to-treat effects using ordinary least squares. We first regress a standardized weighted index of the respondents' self-reported gender attitudes on an indicator variable of whether the respondent was assigned to any information treatment group (pooling all role model arms). Each regression is restricted to one sex, and to only respondents who over-estimated the proportion of others with conservative gender attitudes in at least one of the two domains. See Specification 1 below,

[12]While this is an imprecise measure of posteriors, it avoids exactly replicating the experiment's norms elicitation question asked a few questions earlier in the survey, which would have appeared repetitive and potentially confused or frustrated respondents and biased responses (Haaland, Roth and Wohlfart, 2020).

$$Y_i = \beta_0 + AnyInformationTreatment * \beta_1 + X_i' + \epsilon_i \quad (1)$$

In equation 1, β_1 is the average treatment effect of being given information on the true gender norm for individuals who over-estimate conservative gender norms. X_i' is a vector of pre-experiment controls selected using the post-double selection LASSO procedure (Belloni, Chernozhukov and Hansen, 2014), as specified in the pre-analysis plan. All regressions also include strata fixed effects used in the random assignment, including RCT treatment group, village, and baseline responses to incentivized questions about community gender norms.

In our first model, the outcome variable is the standardized weighted index for all nine questions. We then estimate impacts of the two separate gender norms domains to assess whether there are different impacts by the type of gender norm one is given information on, and whether there is an additive effect of being informed about both norms. We do this by separately comparing respondents who over-estimate each or both of the domains with their equivalent in the control group. This produces three mutually exclusive categories: only over-estimating conservative household gender norms, only over-estimating IPV privacy norms, or over-estimating both, then interacted with the treatment assignment indicator variable to measure impact.

$$\begin{aligned} Y_i = \beta_0 + OverestimatedGenderRoleNorms.\beta_1 + OverestimatedGenderRoleNorms * Treat\beta_2 \\ + OverestimatedIPVNorms.\beta_3 + OverestimatedIPVNorms * Treat\beta_4 \\ + OverestimatedBoth.\beta_5 + OverestimatedBoth * Treat\beta_6 + X_i' + \epsilon_i \end{aligned} \quad (2)$$

We assess the impact of the information treatment for over-estimators in each of these three categories by considering three standardized weighted index outcomes: the overall composite of the respondents' self-reported gender attitudes; the sub-index of attitudes only related to household gender norms; and the sub-index of attitudes only related to IPV acceptability. This allows us to assess whether those over-estimating and being given information on the different domains respond differently to the general gender attitudes questions, or those specifically related to the questions in that domain.

Role Models

We then assess the differential impact of being given information on different types of role models by regressing the standardized indices of gender attitudes on role model group

of assignment. Specification 3 shows our model for estimating treatment effects on men.

$$Y_i = \beta_0 + MaleVillageLeaders.\beta_1 + Men.\beta_2 + SuccessfulBusinessmen.\beta_3 + X_i' + \epsilon_i \quad (3)$$

In equation 3, β_1 is the average treatment effect for men given information about male village leaders, β_2 is the treatment effect for those given information about men, and β_3 is the treatment effect for those given information about successful male business owners.

We also estimate the equivalent regression for women, as follows,

$$Y_i = \beta_0 + Men.\beta_1 + Women.\beta_2 + X_i' + \epsilon_i \quad (4)$$

where β_1 is the average treatment effect for individuals given information about men, and β_2 is the treatment effect for those given information about women. Given statistical power constraints, we do not interact the norm domain with role model treatment group.

2.6 Descriptive statistics of the over-estimators and treatment groups

Tables A2-A5 in the appendix show descriptive statistics on key observable characteristics in the men's and women's samples for over-estimators versus non-over-estimators, and over-estimators across each of the three domain groups (overestimating conservative household norms, conservative IPV norms, and both). These tables show that overall, the over-estimators are, predictably, quite different from those who did not over-estimate conservative norms. On average, they think of themselves as more traditional about gender issues than others in their village, hold more conservative gender attitudes, are poorer, less educated (men), younger (men), more depressed (men), more likely to drink alcohol (men), more violent (men), and more likely to have experienced IPV (women).

We also compare the characteristics of respondents who overestimate by each norm domain. These tables show that women who over-estimate conservative norms in both domains are more likely than women who only over-estimate on one domain to have experienced IPV, believe rates of IPV and controlling behaviors are higher in their village, and be married to a husband who drinks alcohol. Women who over-estimate on only one domain (either IPV privacy or household gender roles) are much more alike on observables than those over-estimating both. For men, we observe that those over-estimating only on household norms appear more progressive on a number of gender indicators than those who over-estimate on both or only on IPV norms.

Tables A6 and A7 show the equivalent characteristics of women and men who were

assigned to and over-estimate each of the role model categories. When we consider descriptive differences between respondents who over-estimate conservative norms in the different role model groups, surprisingly, few differences exist despite these role models having different gender attitudes. This applies to both women and men respondents. One of the few notable differences is that the men assigned to the village leader group who over-estimate village leaders' conservative attitudes tend to personally hold more conservative gender attitudes and believe themselves to be more gender conservative than others in their village, compared to over-estimators on the other role model types. This could possibly suggest that when we ask about individuals' perceptions of norms, these perceptions are formed by motivated beliefs, that is, people are reporting beliefs that accord with their own attitudes, particularly those choosing to believe that leaders, who are official power-holders in a government promoting gender equality and who tend to hold the most progressive attitudes, actually support their views. There are no differences for women over-estimating about men or women role models on key observable characteristics.

3 Results

3.1 Norms misperceptions: pre-experiment beliefs and attitudes

We start by describing the misperceptions individuals have about gender norms relating to gender roles and IPV privacy. We find that many people over-estimate the share of others holding conservative gender attitudes. Table 1 shows that a majority of people overestimate conservative gender norms while privately holding more progressive gender attitudes. The numbers in the baseline column show the 'true' norm used in the endline experiment information update.[13] The grey rows show the share of respondents over-estimating in each domain, and for each role model. In our survey, 33% of men and 48% of women agreed with the statement: "Violence between a man and a woman is a private matter and others should not intervene," and 10% of men and 43% of women agreed with the statement: "It is shameful for a wife if her husband does the housework." When asked what share of people in nearby communities they believed agreed with those statements, 68% of men and 82% of women over-estimated in at least one of the two domains.

In the appendix, Table A1 also shows the joint distributions of those over-estimating in

[13]We also calculated the true baseline norm of male village leaders from a separate baseline survey administered in each of the 98 villages, where we interviewed two village leaders from each 5-person village leadership group. 9% of male village leaders believe that it is shameful for a woman if her husband cooks or does housework, and 8% believe IPV is a private matter and others shouldn't intervene. We also calculated the mean attitudes of the third male role model group- successful businessmen- using baseline data on our sample who: own a business, have taken out a business loan, and have high savings. Of these men, 8% believe it is shameful for a woman if her husband cooks or does housework, and 23% believe IPV is a private matter.

each gender norm domain. In Figures A1-A3, we also show scatterplots of the relationships between village means of the true norm and beliefs about the norm, both measured at endline.

Table 1: Descriptive data on beliefs and norm perceptions

	(1)	(2)
Women	**Baseline**	**Endline**
	(%)	
True share of women believing it is shameful if husbands do housework	**43**	**36**
Beliefs about share of men who think it's shameful if husbands do housework: mean		49
Beliefs about share of women who think it's shameful if husbands do housework: mean		49
% over-estimating baseline share of men who think it's shameful if husbands do housework		81
% over-estimating baseline share of women who think it's shameful if husbands do housework		55
True share of women believing IPV is private	**48**	**55**
Beliefs about share of men who think IPV is a private matter: mean		51
Beliefs about share of women who think IPV is a private matter: mean		44
% over-estimating baseline share of men who think violence between partners is a private matter		63
% over-estimating baseline share of women who think violence between partners is a private matter		48
N	2042	
Men		
True share of men believing it is shameful if husbands do housework	**10**	**15**
Beliefs about share of men who think it's shameful if husbands do housework: mean		31
Beliefs about share of successful businessmen who think it's shameful if husbands do housework: mean		24
Beliefs about share of male village leaders who think it's shameful if husbands do housework: mean		15
% over-estimating baseline share of men who think it's shameful if husbands do housework		63
% over-estimating baseline share of successful businessmen who think it's shameful if husbands do housework		67
% over-estimating baseline share of male village leaders who think it's shameful if husbands do housework		43
True share of men believing IPV is private	**33**	**40**
Beliefs about share of men who think IPV is a private matter: mean		31
Beliefs about share of successful businessmen who think IPV is a private matter: mean		26
Beliefs about share of male village leaders who think IPV is a private matter: mean		13
% over-estimating baseline share of men who think violence between partners is a private matter		36
% over-estimating baseline share of successful businessmen who think violence between partners is a private		38
% over-estimating baseline share of male village leaders who think violence between partners is a private matter		24
N	2042	

Notes: Table shows descriptive data on sample mean beliefs and attitudes of individuals at baseline and endline surveys. The endline sample is restricted to the RCT control group, the baseline sample included the full RCT sample, which was balanced across treatment arms, and randomly sampled community members. At baseline, beliefs were measured using a likert scale, so the baseline mode is not shown here.

Of course, we cannot rule out that answers about one's own and others' attitudes are influenced by social desirability bias or another form of reporting bias. However, the fact that respondents appear so willing to report the 'socially undesirable' response of expressing conservative gender attitudes in a country where there are many national conversations and compulsory village-level discussions about women's rights somewhat mitigates this concern. We find that most men and women are willing to report at least some socially conservative

personal views. For example, 68% of men in our sample say they would be likely to use IPV against their wife in at least one of the 15 hypothetical situations we pose, and the mean number was 3.5.

The scatterplots in Figures A1-A3 also show that mean gender norms vary substantially by village. They also demonstrate that people hold more accurate beliefs about norms when they are guessing within their own sex; men reasonably accurately perceive other men's attitudes, and women reasonably accurately perceive attitudes of other women. However, there is a negative correlation between women's beliefs about men and men's true beliefs. We should note that in the scatterplots showing the true norm versus the belief about the norm, beliefs were measured regarding the perception of others in 'villages near here' rather than 'this village' (for which the mean was calculated). Therefore, these scatterplots do not compare exact like with like.

3.2 Treatment effects by gender norm domain

We turn to our survey experiment correcting over-estimates about conservative gender norms. Tables A8 and A9 provide evidence that individual characteristics are balanced across the experimental arms, confirming that the randomization was successful.

We first present the average treatment effect of being given corrected information on any gender norm and on different norm domains. Shown in Table 2, we assess impact on the three attitudes indices and incentivized spouse transfers. For men, column 7 shows that being given any corrected norms information caused a 0.07 standard deviation increase in the progressiveness of men's subsequent gender attitudes, significant at the 10 percent level. It also increased the number of tokens men allocated to their spouse in the dictator game, by 0.49 tokens (out of 10), significant at the 10 percent level, as shown in column 11. When we compare treated and control groups within each of the three categories that over-estimators fall into (only over-estimating conservative gender role norms, only over-estimating conservative IPV norms, or overestimating both), we see that the positive average treatment effect for men appears to be primarily driven by the respondents who overestimated both conservative gender norms. Having misperceptions corrected about both norms domains increased the progressiveness of men's gender attitudes by 0.09 standard deviations, and increased the number of tokens they allocated to their spouse by 0.49, both significant at the 10 percent level. These are shown in columns 8 and 12, respectively. This may indicate an additive effect of being given more information, or that these men are different from those over-estimating on only one domain.

For women, there was no overall effect of providing women who over-estimated con-

Table 2: Treatment effects of any norm information, and information on different norm domains

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Women						Men					
VARIABLES	Composite progressive gender attitudes index	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to husband (out of 10)	Tokens given to husband (out of 10)	Composite progressive gender attitudes index	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to wife (out of 10)	Tokens given to wife (out of 10)
Overestimate conservative gender roles only		0.19***	0.25***	0.11		-0.20		-0.04	-0.09	-0.08		0.16
		(0.06)	(0.07)	(0.09)		(0.44)		(0.15)	(0.21)	(0.20)		(0.89)
Treatment x Overestimate gender roles only		-0.10	-0.16**	0.01		0.82*		0.05	0.06	0.08		0.51
		(0.06)	(0.07)	(0.09)		(0.44)		(0.06)	(0.08)	(0.07)		(0.34)
Overestimate IPV acceptability only		0.08	0.09	0.03		1.07		0.00	0.00	0.00		0.00
		(0.10)	(0.12)	(0.15)		(0.76)		(0.00)	(0.00)	(0.00)		(0.00)
Treatment x Overestimate IPV only		0.03	0.08	-0.02		-0.45		-0.14	-0.15	-0.23		0.47
		(0.11)	(0.13)	(0.16)		(0.83)		(0.16)	(0.21)	(0.20)		(0.91)
Overestimate norms in both domains		0.00	0.00	0.00		0.00		-0.11	-0.12	-0.14		0.18
		(0.00)	(0.00)	(0.00)		(0.00)		(0.15)	(0.20)	(0.19)		(0.86)
Treatment x Overestimate both		0.01	0.01	-0.02		-0.11		0.09*	0.05	0.11*		0.49*
		(0.04)	(0.04)	(0.05)		(0.27)		(0.05)	(0.07)	(0.07)		(0.30)
Treatment x Overestimated either category	-0.00				0.19		0.07*				0.49**	
	(0.03)				(0.22)		(0.04)				(0.21)	
Observations	647	647	647	647	645	645	548	548	548	548	545	545
PDS controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of strata	98	98	98	98	98	98	98	98	98	98	97	97

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

servative gender norms with any updated information on true norms, shown in column 1. However, we find suggestive evidence that women who over-estimate conservative gender role norms but not conservative IPV norms respond negatively to the information, suggesting there may be an attitude backlash against progressive norm information. Compared to the equivalent control group, women who get updated on gender role norm information reduce the 'progressivism' of their overall gender attitudes by 0.16 standard deviations, significant at the 10 percent level, as shown in column 3. However, for this same group of women, we also observe an increase in the number of tokens they give their spouse, by 0.82, significant at the 10 percent level, shown in column 6.

The seemingly opposing effects on women's attitudes and spouse allocations contrasts with the consistently positive effects for men on both outcomes. However, women's apparent mixed response to being given progressive norms information may, in fact, reflect consistency in negatively responding to the information. These women may be backlashing against the gender identity-threatening norms information and responding by adopting more conservative attitudes and submissive financial behaviors associated with traditional prescriptions of what 'good women' do. Identity theory might predict such a backlash (Akerlof and Kranton, 2000). Ultimately, given men's existing dominance over household financial resources and decision-making in our context, we anticipated that information on progressive norms could

help address this imbalance by prompting men to allocate more to their wives. We did not necessarily expect information on progressive gender norms to change a woman's financial allocation to her husband.

3.3 Treatment effects by role model

We now evaluate the effect of correcting beliefs about different types of role models. Table 3 shows that for women, there is no average impact of either of the two role models, and the differences between them are not statistically significant (though these effects are unlikely to represent precise nulls). For men, we find that being corrected about village leaders' more progressive attitudes causes a 0.11 standard deviation increase in their progressive personal attitudes, significant at the 5 percent level. This role model information also increases transfers to mens' spouses in the incentivized dictator game, by 0.51 tokens or the equivalent of approximately $0.08 USD. However, the differences in treatment effect estimates across men's different role models are not statistically significant. Men who were corrected about progressive norms of men and businessmen also allocate more to their spouse in the dictator game.

Table 3: Treatment effects of norms information on different role model types

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
VARIABLES	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to husband (out of 10)	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to wife (out of 10)
Women role model treatment	-0.03	-0.02	-0.04	0.03				
	(0.04)	(0.04)	(0.05)	(0.26)				
Men role model treatment	0.01	0.00	0.03	0.30	0.04	0.03	0.04	0.51**
	(0.03)	(0.04)	(0.05)	(0.24)	(0.04)	(0.06)	(0.06)	(0.26)
Successful businessmen role model treatment					0.06	0.06	0.08	0.46*
					(0.04)	(0.06)	(0.06)	(0.26)
Male village leader role model treatment					0.11**	0.06	0.13**	0.51*
					(0.05)	(0.07)	(0.06)	(0.29)
Observations	647	647	647	645	548	548	548	545
PDS controls	YES	YES	YES	YES	YES	YES	YES	YES
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES
Number of strata	98	98	98	98	98	98	98	98
Women vs men p-value	0.29	0.55	0.17	0.29				
Men vs businessmen p-value					0.60	0.62	0.50	0.86
Men vs leader p-value					0.16	0.59	0.21	0.99
Businessmen vs leader p-value					0.36	0.93	0.54	0.86

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

3.4 Posteriors

Now, we turn to results on posterior beliefs. We measured posteriors to assess whether individuals accept the information update and incorporate it into their beliefs. The posteriors question specifically related to beliefs about other people in the respondent's village and their likely opinions about a hypothetical scenario where a husband taking on domestic responsibilities is the subject of gossip by his neighbors. Therefore, our posteriors question most relates to beliefs about norms on traditional household gender roles. It does not reflect, one-to-one, the same belief about gender norms as respondents were updated on. While this difference was intended to ensure the question wording did not directly overlap with earlier beliefs questions, it also represents a noisy indicator of posterior updating from the experiment.

Table 4: Impact of information update on posteriors

	(1)	(2)	(3)	(4)	(5)	(6)
	Women			Men		
VARIABLES	Unlikely that most other people in this village agree with conservative neighbours about household gender roles					
Overestimate conservative gender roles only		0.04			-0.05	
		(0.08)			(0.19)	
Treatment x Overestimate gender roles only		-0.07			0.05	
		(0.08)			(0.07)	
Overestimate IPV acceptability only		0.02			0.00	
		(0.14)			(0.00)	
Treatment x Overestimate IPV acceptability only		-0.07			-0.12	
		(0.15)			(0.19)	
Overestimate conservative norms in both domains		0.00			-0.07	
		(0.00)			(0.18)	
Treatment x Overestimate both		-0.07			-0.02	
		(0.05)			(0.06)	
Treatment x Overestimated either category	-0.07			0.01		
	(0.04)			(0.04)		
Women role model treatment			-0.11**			
			(0.05)			0.01
Men role model treatment			-0.03			(0.05)
			(0.05)			-0.01
Successful businessmen role model treatment						(0.06)
						0.07
Male village leader role model treatment						(0.06)
Observations	647	647	647	548	548	548
PDS controls	YES	YES	YES	YES	YES	YES
Strata FE	YES	YES	YES	YES	YES	YES
Number of strata	98	98	98	98	98	98

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4 shows the impact of the information update on posteriors. We find that the share of people believing it unlikely that most people in their village hold conservative gender attitudes is generally unaffected by being updated on norms information, which we

had told them was based on data gathered a year prior, and based on an average of people surveyed from nearby villages. For women, we actually find that being told about correct gender norms about other women (but not men) results in a *worsening* in posteriors about gender norms. Combined with the negative treatment effect for women over-estimating conservative household gender norms, this negative posterior updating could also indicate that women are 'backlashing' against the corrected norms information. It may also provide evidence for motivated beliefs, suggesting that some women may perceive the norms they wish to believe in, choosing to self-deceive and reject new evidence when it does not accord with their identity. If true, this would suggest that people may be resistant to evidence that runs counter to their attitudes, and that in such cases or for such people, attempts to use evidence to correct pluralistic ignorance may not be effective. It seems unlikely that this group of women informed about women role models, more than any other group, discounted the norm information update as out-of-date or answered the original belief question with a response they did or did not truly believe. For men, there were no impacts on posteriors after being corrected about more progressive gender norms, though standard errors are large and our sample is small, so they are not precisely-estimated nulls.

3.5 Individual survey questions

We also consider whether the treatment effects are driven by any of the nine outcome questions that comprise the composite index. Overall, we find relatively similar, non-significant effects for the individual questions. In the appendix, Table A12 shows results for the few binary individual outcome questions with statistically significant treatment effects. For women, there was a marginally statistically significant reduction in progressive gender attitudes on two individual binary questions, and for men, an increase in progressive attitudes on one.

3.6 Robustness checks

When we estimate treatment effects without using pre-experiment control variables, we find broadly similar point estimates as the model with post-double selection LASSO controls, though most effects are no longer statistically significant. The two remaining statistically significant treatment effects are the negative impacts on attitudes for women who only over-estimate conservative norms about gender roles, and the number of tokens that men allocate to their spouse after being given any information update. These control-free results are shown in Table A13, role model treatment effects are shown in Table A14, and Table A10 shows equivalent impacts on posterior beliefs.

For the role model effects, we find similar positive effects for men in the number of tokens

they allocate their spouse after being updated on any of the role models, but for women, we find statistically significant evidence suggesting that women react negatively to information updates on women role models (but not men role models).

When we consider treatment effects on non-gender attitude questions that were asked later in the survey, potentially indicating lingering experimenter demand effects, we find no impact, as shown in Table A11. There was also no treatment effect on enumerator-observed distress in respondents. The incentivized spouse dictator game was played after these remaining survey questions. This suggests that treatment effects on any dictator game behavior are unlikely to reflect immediate priming of the norms information, which appear to have dissipated by the time it was played.

Overall, the broadly similar results across the models with and without controls, and across self-reported attitudes asked immediately after the information update and incentivized decisions made 10 minutes later, suggests estimates are likely robust. However, future studies should be conducted in larger samples, and with longer-term follow-up and behavioral outcomes outside a study setting to confirm the robustness of our findings.

4 Conclusion

This paper has shown that i) a majority of men and women in Rwanda privately hold progressive attitudes about traditional gender roles and IPV; ii) most people over-estimate the share of others with conservative gender attitudes; iii) after having their misperceptions corrected, men's gender attitudes become more progressive, and they allocate more to their spouse in an incentivized dictator game, while women's attitudes and behavior are unaffected, on average.

We also identified that information about different role models can affect attitudes and behaviors in different ways. We find some suggestive evidence that women given corrected norms information about women (but not about men) may backlash or react negatively. And men appear to most positively respond to information on male village leaders.

Overall, this paper builds on the emerging evidence that beliefs about social norms may be an important decision-making friction that influences individuals' own gender attitudes and behaviors. Our results, while estimated on a sample of over-estimators of conservative gender norms, can contribute to our understanding of the role that social norms play in the persistence of gender inequality, and how information might lift these constraints. Norms information might be particularly effective in rural, less connected, or less democratic societies, where other sources of information dissemination are less available (Bursztyn, González and Yanagizawa-Drott, 2018). It may also be particularly important for polarized or stigmatized

issues such as gender or other identities where people may not openly discuss their privately held views.

Our results may be relevant to policymakers seeking to shift attitudes and behaviors by emphasizing a particular norm, for example, vaccinating children, fathers sharing household responsibilities, voter registration, or mask wearing during a pandemic. Our findings also suggest that individuals' attitudes, sometimes assumed to be immutable underlying preferences, can change when we show people how their attitudes differ from others'. We also find some suggestive evidence of the potential for negative reactions to norms information, with some women appearing to backlash against it. Potentially, advocacy campaigns or social interventions that seek to change attitudes or preferences through norms information or persuasion may inadvertently cause a backlash if they threaten identities, or if beliefs about norms reflect motivated reasoning. Future research could identify the factors determining when backlash poses a risk, and help us better understand the role of motivated beliefs in perpetuating misperceptions of norms.

Future work could also enrich this analysis by assessing longer-term attitude and behavior change after the provision of norms information. It might also consider the impact of gender norms information provision on both over-and under-estimators of conservative norms, as may be more realistic in an information campaign. Further, research to understand the root of conservative gender attitudes and identities, and why different role models matter more could help those designing policy to address gender inequality. Finally, research could also usefully study the relative role of personal attitudes and norms perceptions on behavioral outcomes, and the causal relationships between norms perceptions and personal attitudes.

References

Akerlof, George A. 1997. "Social distance and social decisions." *Econometrica*, 65(5): 1005–1027.

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *The Quarterly Journal of Economics*, 115(3): 715–753.

Alik-Lagrange, Arthur, Claire Cullen, Muthoni Ngatia, and Julia Vaillant. 2017. "Preventing intimate partner violence: Impact Evaluation of MIGEPROF couples training for IPV prevention in Eastern Rwanda."

Anderson, Michael L. 2008. "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.

Angrist, N. 2020. "The Messenger Matters: Experimental Evidence from a Sex Education Intervention."

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco. 2019. "Entertainment, Education, and Attitudes Toward Domestic Violence." Vol. 109, 133–137.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial." *American Economic Review*, 2020(3): 824–859.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India." *Science*, 335.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives*, 28(2): 29–50.

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives*, 30(3): 141–164.

Bursztyn, L, and R Jensen. 2015. "How does peer pressure affect educational investments?" *The Quarterly Journal of Economics*, 130(3): 1329–1367.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2018. "Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia."

Bursztyn, L, G Egorov, and S Fiorin. 2019. "From Extreme to Mainstream: The Erosion of Social Norms."

Cullen, C, A Alik-Lagrange, M Ngatia, and J Vaillant. 2020. "The unintended impacts of an intimate partner vioelnce prevention program."

Cullen, Claire. 2020. "Can role models and correcting beliefs about norms change gender attitudes? Evidence from a survey experiment in Rwanda."

Del Carpio, Lucia, and Maria Guadalupe. 2018. "More Women in Tech? Evidence from a Field Experiment Addressing Social Identity."

Dellavigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2017. "Voting to Tell Others." *Review of Economic Studies*, 84(1): 143–181.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2018. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India."

DHS. 2016. "Rwanda Demographic and Health Survey 2014-15."

Ferrant, Gaëlle, Maria Pesando, and Keiko Nowacka. 2014. "Unpaid Care Work: The missing link in the analysis of gender gaps in labour outcomes." OECD Development Center, Paris.

Field, E. M., R. Pande, N. Rigol, S. G. Schaner, and C. T. Moore. 2019. "On Her Own Account: How Strengthening Women's Financial Control Affects Labor Supply and Gender Norms."

Fleming, Paul J., Jennifer McCleary-Sills, Matthew Morton, Ruti Levtov, Brian Heilman, and Gary Barker. 2015. "Risk Factors for Men's Lifetime Perpetration of Physical Violence against Intimate Partners: Results from the International Men and Gender Equality Survey (IMAGES) in Eight Countries." *PLOS ONE*, 10(3): e0118639.

Garcia-Moreno, C J H, M Ellsberg, L Heise, and C Watts. 2005. *WHO multi-country study on womens health an domestic violence against women.* Geneva:WHO.

Glennerster, Rachel, Claire Walsh, and Lucia Diaz-Martin. 2018. "A Practical Guide to Measuring Women's and Girls' Empowerment in Impact Evaluations." The Abdul Latif Jameel Poverty Action Lab, Boston.

Global Burden of Disease. 2017. "Global Burden of Disease Study 2017: Health-related Sustainable Development Goals (SDG) Indicators 1990-2030." Institute for Health Metrics and Evaluation (IHME), Seattle, United States.

Green, D, A Wilke, and J Cooper. 2018. "Silence Begets Violence: A mass media experiment to prevent violence against women in rural Uganda."

Green, Donald P., Anna M. Wilke, and Jasper Cooper. 2020. "Countering Violence Against Women by Encouraging Disclosure: A Mass Media Experiment in Rural Uganda." *Comparative Political Studies*, 001041402091227.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal. 2018. "Does Information Change Attitudes Towards Immigrants? Representative Evidence from Survey Experiments."

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2020. "Designing Information Provision Experiments."

Hoel, JB. 2015. "Heterogeneous households: A within-subject test of asymmetric information between spouses in Kenya." *Journal of Economic Behavior & Organization*, 118: 123–135.

Jayachandran, Seema. 2020. "Social Norms as a Barrier to Women's Employment in Developing Countries."

Kahan, Dan M. 2013. "Ideology, motivated reasoning, and cognitive reflection." *Judgment and Decision Making*, 8(4): 407–424.

Kappes, A, and T Sharot. 2019. "The automatic nature of motivated belief updating." *Behavioural Public Policy*, 3(1): 87–103.

LaFerrara, Eliana. 2019. "Aspirations, Social Norms, and Development." *Journal of the European Economic Association*, 17(6): 1687–1722.

LaFerrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics*, 4(4): 1–31.

Mani, Anandi, and Emma Riley. 2019. "Social networks, role models, peer effects, and aspirations." UNU WIDER.

Palermo, T., J. Bleck, and A. Peterman. 2014. "Tip of the Iceberg: Reporting and Gender-Based Violence in Developing Countries." *American Journal of Epidemiology*, 179(5): 602–612.

Paluck, E, and H Shepherd. 2016. "Changing climates of conflict: A social network experiment in 56 schools." *Proceedings of the National Academy of Sciences*, 113(3).

Ray, D. 2006. "Aspirations, Poverty, and Economic Change." In *Understanding poverty*. 409–421.

WHO. 2013. "Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence." World Health Organization, Geneva.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review*, 110(2): 337–363.

A Appendix: Figures and Tables

Table A1: Number of over-estimators by gender attitude domain

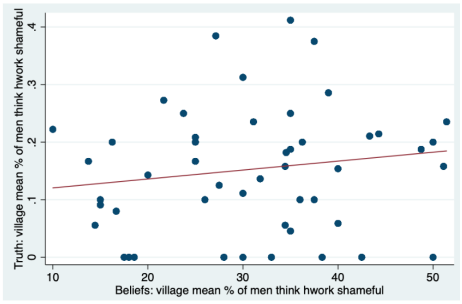
		Women: Overestimated share of men/women believing IPV is private and others shouldn't intervene	
		No	Yes
Women: Overestimated share of men/women believing housework is shameful for men	No	137	68
	Yes	177	417

		Men: Overestimated share of men/leaders/businessmen believing IPV is private and others shouldn't intervene	
		No	Yes
Men: Overestimated share of men/leaders/ businessmen believing housework is shameful for men	No	240	45
	Yes	255	254

Beliefs vs. truth: village mean of perceptions of the norm by village mean of others' actual attitudes

Figure A1: Men's beliefs about men: truth and beliefs by village mean

Housework is shameful for men



IPV is a private matter

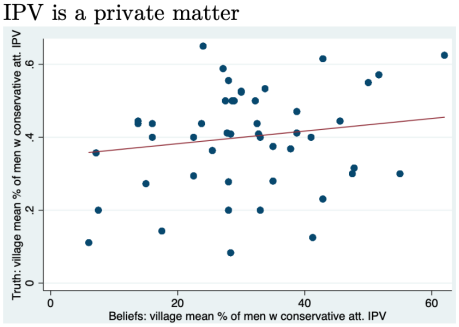
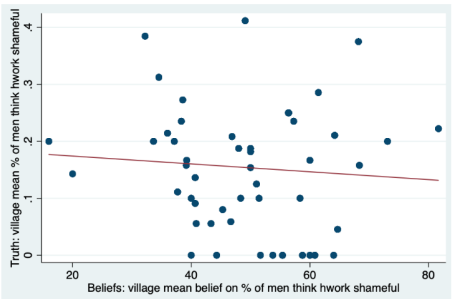


Figure A2: Women's beliefs about men: truth and beliefs by village mean

Housework is shameful for men



IPV is a private matter

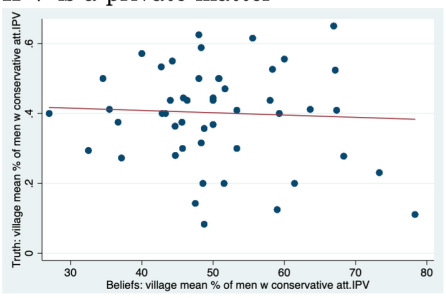


Figure A3: Women's beliefs about women

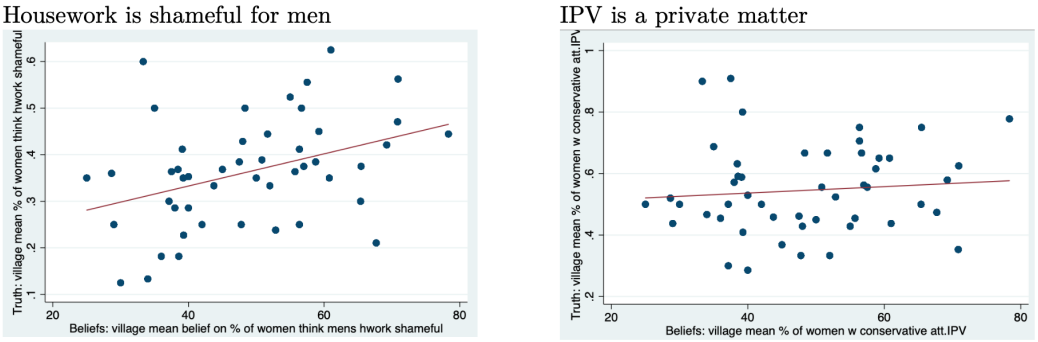


Figure A4: Survey experiment process

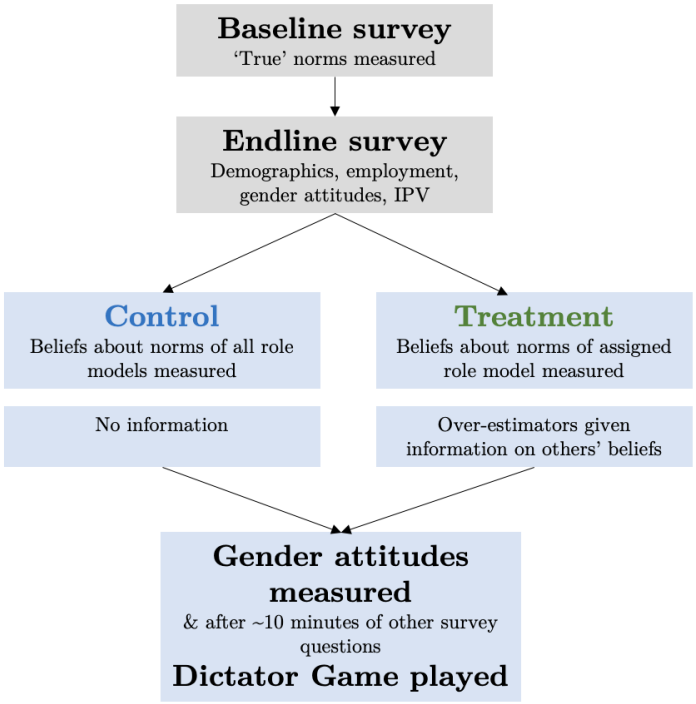


Table A2: Women: Difference between over-estimators and non-over-estimators

VARIABLE	N	(1) Did not overestimate Mean/SE	N	(2) Overestimated Mean/SE	T-test Difference (1)-(2)	Normalized difference
What share of men think wife should obey husband	137	57.88	661	67.81	-9.92***	-0.34
		[2.89]		[1.08]		
What share of men get jealous when wife speaks with other men	136	43.82	660	60.39	-16.57***	-0.55
		[2.58]		[1.13]		
What share of men kick/drag/beat up wife	137	22.34	660	34.73	-12.39***	-0.54
		[1.68]		[0.90]		
Perceive self as more traditional about gender issues than most others in village	150	0.47	662	0.58	-0.11**	-0.23
		[0.04]		[0.02]		
Partner drinks some alcohol	137	0.18	662	0.26	-0.08	-0.19
		[0.03]		[0.02]		
Believe violence between partners is a private matter - endline	137	0.41	662	0.57	-0.17***	-0.33
		[0.04]		[0.02]		
Believe violence between partners is a private matter - baseline	150	0.45	662	0.50	-0.06	-0.12
		[0.04]		[0.02]		
Women: Number of gender inequitable attitudes	137	1.48	662	1.89	-0.41***	-0.47
		[0.07]		[0.03]		
Village mean: Weighted gender norms index	150	-0.05	662	0.01	-0.06	-0.18
		[0.02]		[0.01]		
Age	150	35.13	662	36.28	-1.15*	-0.13
		[0.73]		[0.35]		
Years of education	150	4.83	662	4.51	0.32*	0.11
		[0.23]		[0.11]		
How many children have you given birth to?	150	3.87	662	3.63	0.24	0.11
		[0.20]		[0.08]		
Experienced at least 1 form of emotional violence in last 6m	136	0.60	647	0.67	-0.08	-0.16
		[0.04]		[0.02]		
Experienced at least 1 form of physical violence in last 6m	134	0.32	648	0.33	-0.01	-0.02
		[0.04]		[0.02]		
Experienced at least 1 form of sexual violence in last 6m	135	0.33	651	0.44	-0.11*	-0.22
		[0.04]		[0.02]		
F-test of joint significance (p-value)					0.00***	
F-test, number of observations					758	

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_w are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A3: Men: Difference between over-estimators and non-over-estimators

VARIABLE	N	(1) Did not overestimate Mean/SE	N	(2) Overestimated Mean/SE	T-test Difference (1)-(2)	Normalized difference
What share of men think wife should obey husband	239	51.21	554	56.53	-5.32**	-0.18
		[2.20]		[1.20]		
What share of men get jealous when wife speaks with other men	239	37.95	554	49.44	-11.49***	-0.36
		[2.07]		[1.31]		
What share of men kick/drag/beat up wife	239	10.33	553	22.41	-12.07***	-0.65
		[0.79]		[0.83]		
Perceive self as more traditional about gender issues than most others in village	258	0.41	554	0.55	-0.13***	-0.26
		[0.03]		[0.02]		
Believe violence between partners is a private matter - endline	240	0.32	554	0.44	-0.12***	-0.25
		[0.03]		[0.02]		
Believe violence between partners is a private matter - baseline	258	0.30	554	0.38	-0.08**	-0.17
		[0.03]		[0.02]		
Men: Number of gender inequitable attitudes	240	1.33	554	1.73	-0.40***	-0.36
		[0.06]		[0.05]		
Village mean: Weighted gender norms index	258	-0.06	554	0.03	-0.08	-0.25
		[0.02]		[0.01]		
Age	258	39.76	554	40.18	-0.43	-0.04
		[0.59]		[0.41]		
Years of education	258	5.31	554	4.55	0.77***	0.25
		[0.18]		[0.13]		
Number of hypothetical situations men would likely become violent	240	2.74	554	3.77	-1.03***	-0.27
		[0.23]		[0.17]		
Have moderate to severe depression	258	0.11	554	0.09	0.02	0.07
		[0.02]		[0.01]		
Drank alcohol to excess	258	0.26	554	0.37	-0.11**	-0.23
		[0.03]		[0.02]		
Member of the household skipped a meal in the last week	240	0.15	554	0.19	-0.05**	-0.12
		[0.02]		[0.02]		
F-test of joint significance (p-value)					0.00***	
F-test, number of observations					791	

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_m are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A4: Women: Differences by domain of over-estimation

	(1) Overestimated only unequal household gender norms		(2) Overestimated only conservative IPV privacy norms		(3) Overestimated both		T-test Differences		
VARIABLE	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)
What share of men think wife should obey husband	177	64.58	68	65.29	416	69.59	-0.72	-5.02**	-4.30
		[2.16]		[3.94]		[1.30]			
What share of men get jealous when wife speaks with other men	176	55.97	68	53.38	416	63.41	2.58	-7.45***	-10.03
		[2.13]		[4.10]		[1.39]			
What share of men kick/drag/beat up wife	177	31.64	68	30.29	415	36.77	1.34	-5.13	-6.48
		[1.72]		[3.17]		[1.11]			
Perceive self as more traditional about gender issues than most others in village	177	0.54	68	0.53	417	0.60	0.01	-0.06	-0.07**
		[0.04]		[0.06]		[0.02]			
Partner drinks some alcohol	177	0.20	68	0.16	417	0.29	0.04	-0.10**	-0.13
		[0.03]		[0.04]		[0.02]			
Believe violence between partners is a private matter - endline	177	0.46	68	0.56	417	0.62	-0.10	-0.16***	-0.06
		[0.04]		[0.06]		[0.02]			
Believe violence between partners is a private matter - baseline	177	0.43	68	0.56	417	0.53	-0.13*	-0.10	0.03
		[0.04]		[0.06]		[0.02]			
Women: Number of gender inequitable attitudes	177	1.75	68	1.78	417	1.97	-0.03	-0.22**	-0.19*
		[0.06]		[0.09]		[0.04]			
Village mean: Weighted gender norms index	177	-0.03	68	-0.14	417	0.05	0.11	-0.09	-0.20
		[0.02]		[0.04]		[0.02]			
Age	177	35.04	68	37.13	417	36.67	-2.09	-1.63	0.46
		[0.65]		[1.15]		[0.44]			
Years of education	177	4.83	68	3.81	417	4.49	1.02**	0.34	-0.68
		[0.21]		[0.36]		[0.14]			
How many children have you given birth to?	177	3.45	68	3.78	417	3.69	-0.33	-0.23	0.09
		[0.15]		[0.26]		[0.10]			
Experienced at least 1 form of emotional violence in last 6m	172	0.63	65	0.63	410	0.70	-0.00	-0.07	-0.06
		[0.04]		[0.06]		[0.02]			
Experienced at least 1 form of physical violence in last 6m	173	0.25	66	0.35	409	0.36	-0.09	-0.11*	-0.01
		[0.03]		[0.06]		[0.02]			
Experienced at least 1 form of sexual violence in last 6m	173	0.35	65	0.54	413	0.47	-0.19**	-0.12*	0.07*
		[0.04]		[0.06]		[0.02]			
F-test of joint significance (p-value)							0.36	0.03**	0.47
F-test, number of observations							228	566	462

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_w are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A5: Men: Differences by domain of over-estimation

	(1) Overestimated only unequal household gender norms		(2) Overestimated only conservative IPV privacy norms		(3) Overestimated both		T-test Differences		
VARIABLE	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)
What share of men think wife should obey husband	255	56.55	45	55.78	254	56.65	0.77	-0.10	-0.88
		[1.81]		[4.80]		[1.69]			
What share of men get jealous when wife speaks with other men	255	46.00	45	40.44	254	54.49	5.56	-8.49*	-14.04
		[1.89]		[4.89]		[1.90]			
What share of men kick/drag/beat up wife	254	18.86	45	16.00	254	27.09	2.86	-8.23***	-11.09***
		[1.11]		[3.01]		[1.27]			
Perceive self as more traditional about gender issues than most others in village	255	0.49	45	0.42	254	0.63	0.07	-0.14***	-0.20*
		[0.03]		[0.07]		[0.03]			
Believe violence between partners is a private matter - endline	255	0.30	45	0.64	254	0.54	-0.34***	-0.23***	0.11
		[0.03]		[0.07]		[0.03]			
Believe violence between partners is a private matter - baseline	255	0.33	45	0.42	254	0.44	-0.10	-0.11***	-0.01
		[0.03]		[0.07]		[0.03]			
Men: Number of gender inequitable attitudes	255	1.51	45	1.89	254	1.92	-0.38**	-0.40***	-0.03
		[0.07]		[0.13]		[0.07]			
Village mean: Weighted gender norms index	255	-0.02	45	0.08	254	0.06	-0.10	-0.08	0.02
		[0.02]		[0.05]		[0.02]			
Age	255	39.35	45	42.31	254	40.65	-2.97	-1.30	1.66
		[0.57]		[1.68]		[0.63]			
Years of education	255	4.97	45	3.91	254	4.23	1.06***	0.74***	-0.32
		[0.20]		[0.36]		[0.19]			
Number of hypothetical situations men would likely become violent	255	3.30	45	3.22	254	4.35	0.08	-1.04*	-1.12
		[0.24]		[0.49]		[0.26]			
Have moderate to severe depression	255	0.06	45	0.16	254	0.11	-0.09*	-0.04*	0.05
		[0.02]		[0.05]		[0.02]			
F-test of joint significance (p-value)							0.01***	0.00***	0.01***
F-test, number of observations							299	508	299

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_m are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A6: Women: Differences in over-estimators by role model treatment arm

VARIABLE	N	(1) Overestimate women's conservative attitudes Mean/SE	N	(2) Overestimate men's conservative attitudes Mean/SE	T-test Difference	Normalized difference
What share of men think wife should obey husband	181	69.83	242	67.02	2.81	0.10
		[1.95]		[1.78]		
What share of men get jealous when wife speaks with other men	180	62.28	241	59.17	3.11	0.10
		[2.23]		[1.89]		
What share of men kick/drag/beat up wife	180	36.22	242	35.00	1.22	0.05
		[1.87]		[1.48]		
Perceive self as more traditional about gender issues than most others in village	181	0.62	242	0.59	0.04	0.08
		[0.04]		[0.03]		
Partner drinks some alcohol	181	0.26	242	0.27	-0.01	-0.02
		[0.03]		[0.03]		
Believe violence between partners is a private matter - endline	181	0.59	242	0.60	-0.01	-0.02
		[0.04]		[0.03]		
Believe violence between partners is a private matter - baseline	181	0.52	242	0.53	-0.01	-0.03
		[0.04]		[0.03]		
Women: Number of gender inequitable attitudes	181	1.96	242	1.93	0.03	0.04
		[0.06]		[0.05]		
Village mean: Weighted gender norms index	181	0.03	242	0.00	0.03	0.07
		[0.03]		[0.02]		
Age	181	37.01	242	36.10	0.90	0.10
		[0.64]		[0.59]		
Years of education	181	4.10	242	4.74	-0.64*	-0.22
		[0.20]		[0.19]		
How many children have you given birth to?	181	3.62	242	3.62	0.00	0.00
		[0.16]		[0.13]		
Experienced at least 1 form of emotional violence in last 6m	175	0.70	236	0.67	0.03	0.06
		[0.03]		[0.03]		
Experienced at least 1 form of physical violence in last 6m	176	0.39	238	0.32	0.07	0.15
		[0.04]		[0.03]		
Experienced at least 1 form of sexual violence in last 6m	175	0.50	239	0.41	0.08*	0.17
		[0.04]		[0.03]		
F-test of joint significance (p-value)					0.80	
F-test, number of observations					398	

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_w are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. This sample excludes the control group who were asked their views on all role models

Table A7: Men: Differences in over-estimators by role model treatment arm

	(1) Overestimate businessmens' conservative attitudes		(2) Overestimate mens' conservative attitudes		(3) Overestimate male leaders' conservative attitudes		T-test Differences			Normalized Differences		
VARIABLE	N	Mean/SE	N	Mean/SE	N	Mean/SE	(1)-(2)	(1)-(3)	(2)-(3)	(1)-(2)	(1)-(3)	(2)-(3)
What share of men think wife should obey husband	140	56.21	147	60.48	100	53.00	-4.26	3.21	7.48*	-0.16	0.12	0.26
		[2.23]		[2.31]		[2.87]						
What share of men get jealous when wife speaks with other men	140	53.50	147	48.50	100	46.80	5.00	6.70*	1.70	0.17	0.23	0.06
		[2.41]		[2.59]		[2.92]						
What share of men kick/drag/beat up wife	140	23.50	147	22.72	100	21.40	0.78	2.10	1.32	0.04	0.11	0.07
		[1.65]		[1.65]		[1.80]						
Perceive self as more traditional about gender issues than most others in village	140	0.49	147	0.50	100	0.66	-0.01	-0.17***	-0.16**	-0.02	-0.34	-0.32
		[0.04]		[0.04]		[0.05]						
Believe violence between partners is a private matter - endline	140	0.40	147	0.46	100	0.53	-0.06	-0.13	-0.07	-0.11	-0.26	-0.15
		[0.04]		[0.04]		[0.05]						
Believe violence between partners is a private matter - baseline	140	0.34	147	0.39	100	0.44	-0.06	-0.10*	-0.05	-0.12	-0.21	-0.09
		[0.04]		[0.04]		[0.05]						
Men: Number of gender inequitable attitudes	140	1.61	147	1.76	100	1.96	-0.14	-0.35**	-0.20	-0.13	-0.31	-0.18
		[0.09]		[0.09]		[0.12]						
Village mean: Weighted gender norms index	140	0.04	147	0.03	100	0.03	0.01	0.01	0.00	0.03	0.04	0.01
		[0.03]		[0.03]		[0.03]						
Age	140	40.64	147	40.45	100	39.71	0.19	0.93	0.74	0.02	0.09	0.08
		[0.88]		[0.78]		[0.99]						
Years of education	140	4.77	147	4.81	100	4.08	-0.04	0.69*	0.73	-0.01	0.24	0.25
		[0.26]		[0.26]		[0.26]						
Number of hypothetical situations men would likely become violent	140	4.10	147	3.29	100	4.01	0.81	0.09	-0.72	0.21	0.02	-0.19
		[0.35]		[0.30]		[0.42]						
Have moderate to severe depression	140	0.11	147	0.08	100	0.13	0.03	-0.02	-0.05	0.09	-0.07	-0.16
		[0.03]		[0.02]		[0.03]						
F-test of joint significance (p-value)							0.62	0.02**	0.36			
F-test, number of observations							287	240	247			

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_m are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. This sample excludes the control group who were asked their views on all role models

Table A8: Balance in women's pre-experiment questions

VARIABLE	N	(1) Control- No norms Mean/SE	N	(2) Men norms Mean/SE	N	(3) Women norms Mean/SE	T-test Difference (1)-(2)	T-test Difference (1)-(3)	T-test Difference (2)-(3)
What share of men think wife should obey husband	238	67.06	242	67.02	181	69.83	0.03	-2.78	-2.81
		[1.90]		[1.78]		[1.95]			
What share of men get jealous when wife speaks with other men	239	60.21	241	59.17	180	62.28	1.04	-2.07	-3.11
		[1.83]		[1.89]		[2.23]			
What share of men kick/drag/beat up wife	238	33.32	242	35.00	180	36.22	-1.68	-2.90	-1.22
		[1.42]		[1.48]		[1.87]			
Perceive self as more traditional about gender issues than most others in village	239	0.54	242	0.59	181	0.62	-0.05	-0.08	-0.04
		[0.03]		[0.03]		[0.04]			
Partner drinks some alcohol	239	0.24	242	0.27	181	0.26	-0.03	-0.02	0.01
		[0.03]		[0.03]		[0.03]			
Believe violence between partners is a private matter - endline	239	0.54	242	0.60	181	0.59	-0.05	-0.04	0.01
		[0.03]		[0.03]		[0.04]			
Believe violence between partners is a private matter - baseline	239	0.46	242	0.53	181	0.52	-0.07	-0.05	0.01
		[0.03]		[0.03]		[0.04]			
Women: Number of gender inequitable attitudes	239	1.80	242	1.93	181	1.96	-0.12	-0.15	-0.03
		[0.06]		[0.05]		[0.06]			
Village mean: Weighted gender norms index	239	0.00	242	0.00	181	0.03	0.00	-0.03	-0.03
		[0.02]		[0.02]		[0.03]			
Age	239	35.91	242	36.10	181	37.01	-0.19	-1.09	-0.90
		[0.58]		[0.59]		[0.64]			
Years of education	239	4.59	242	4.74	181	4.10	-0.14	0.49*	0.64*
		[0.18]		[0.19]		[0.20]			
How many children have you given birth to?	239	3.65	242	3.62	181	3.62	0.03	0.03	-0.00
		[0.13]		[0.13]		[0.16]			
Experienced at least 1 form of emotional violence in last 6m	236	0.64	236	0.67	175	0.70	-0.03	-0.06	-0.03
		[0.03]		[0.03]		[0.03]			
Experienced at least 1 form of physical violence in last 6m	234	0.31	238	0.32	176	0.39	-0.01	-0.08	-0.07
		[0.03]		[0.03]		[0.04]			
Experienced at least 1 form of sexual violence in last 6m	237	0.43	239	0.41	175	0.50	0.02	-0.07	-0.08*
F-test of joint significance (p-value)							0.89	0.41	0.51
F-test, number of observations							436	383	381

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_w are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. This sample is restricted to only those who over-estimated, and those in the RCT control group.

Table A9: Balance in men's pre-experiment questions

VARIABLE	N	(1) Control-No norms Mean/SE	N	(2) Businessmen norms Mean/SE	N	(3) Men norms Mean/SE	N	(4) Village leader norms Mean/SE	T-test Difference (1)-(2)	T-test Difference (1)-(3)	T-test Difference (1)-(4)	T-test Difference (2)-(3)	T-test Difference (2)-(4)	T-test Difference (3)-(4)
What share of men think wife should obey husband	167	55.45	140	56.21	147	60.48	100	53.00	-0.77	-5.03	2.45	-4.26	3.21	7.48*
		[2.28]		[2.23]		[2.31]		[2.87]						
What share of men get jealous when wife speaks with other men	167	48.44	140	53.50	147	48.50	100	46.80	-5.06*	-0.06	1.64	5.00	6.70*	1.70
		[2.53]		[2.41]		[2.59]		[2.92]						
What share of men kick/drag/beat up wife	166	21.81	140	23.50	147	22.72	100	21.40	-1.69	-0.91	0.41	0.78	2.10	1.32
		[1.58]		[1.65]		[1.65]		[1.80]						
Perceive self as more traditional about gender issues than most others in village	167	0.56	140	0.49	147	0.50	100	0.66	0.07	0.06	-0.10	-0.01	-0.17***	-0.16**
		[0.04]		[0.04]		[0.04]		[0.05]						
Believe violence between partners is a private matter - endline	167	0.40	140	0.40	147	0.46	100	0.53	-0.00	-0.06	-0.13**	-0.06	-0.13	-0.07
		[0.04]		[0.04]		[0.04]		[0.05]						
Believe violence between partners is a private matter - baseline	167	0.38	140	0.34	147	0.39	100	0.44	0.05	-0.01	-0.06	-0.06	-0.10*	-0.05
		[0.04]		[0.04]		[0.04]		[0.05]						
Men: Number of gender inequitable attitudes	167	1.66	140	1.61	147	1.76	100	1.96	0.05	-0.09	-0.30**	-0.14	-0.35**	-0.20
		[0.08]		[0.09]		[0.09]		[0.12]						
Village mean: Weighted gender norms index	167	0.01	140	0.04	147	0.03	100	0.03	-0.03	-0.02	-0.01	0.01	0.01	0.00
		[0.03]		[0.03]		[0.03]		[0.03]						
Age	167	39.86	140	40.64	147	40.45	100	39.71	-0.78	-0.59	0.15	0.19	0.93	0.74
		[0.71]		[0.88]		[0.78]		[0.99]						
Years of education	167	4.41	140	4.77	147	4.81	100	4.08	-0.36	-0.40	0.33	-0.04	0.69*	0.73
		[0.25]		[0.26]		[0.26]		[0.26]						
Number of hypothetical situations men would likely become violent in	167	3.79	140	4.10	147	3.29	100	4.01	-0.31	0.50	-0.22	0.81	0.09	-0.72
		[0.31]		[0.35]		[0.30]		[0.42]						
Have moderate to severe depression	167	0.06	140	0.11	147	0.08	100	0.13	-0.05**	-0.02	-0.07**	0.03	-0.02	-0.05
		[0.02]		[0.03]		[0.02]		[0.03]						
F-test of joint significance (p-value)									0.21	0.69	0.07*	0.83	0.04**	0.25
F-test, number of observations									306	313	266	287	240	247

Notes: The value displayed for t-tests are the differences in the means across the groups. Fixed effects using variable norms_strata_m are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. This sample is restricted to only those who over-estimated, and those in the RCT control group.

Table A10: Impact of information update on posteriors- no controls

	(1)	(2)	(3)	(4)	(5)	(6)
	Women			Men		
VARIABLES	Unlikely that most other people in this village agree with conservative neighbours about household gender roles					
Overestimate conservative gender roles only		0.04			0.03	
		(0.08)			(0.08)	
Treatment x Overestimate gender roles only		-0.07			0.05	
		(0.08)			(0.07)	
Overestimate IPV acceptability only		0.05			0.16	
		(0.14)			(0.18)	
Treatment x Overestimate IPV acceptability only		-0.11			-0.19	
		(0.15)			(0.19)	
Overestimate conservative norms in both domains						
Treatment x Overestimate both		-0.07			-0.02	
		(0.05)			(0.06)	
Treatment x Overestimated either category	-0.07*			0.01		
	(0.04)			(0.05)		
Women role model treatment			-0.12**			
			(0.05)			0.02
Men role model treatment			-0.04			(0.06)
			(0.04)			-0.03
Successful businessmen role model treatment						(0.06)
						0.06
Male village leader role model treatment						(0.06)
Observations	662	662	662	554	554	554
PDS controls	NO	NO	NO	NO	NO	NO
Strata FE	YES	YES	YES	YES	YES	YES
Number of strata	98	98	98	98	98	98

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Treatment effect on post-experiment survey questions

Women		Men	
VARIABLES	Treatment x Overestimated either category	VARIABLES	Treatment x Overestimated either category
Number of Rwandan gender law questions correct out of 3	-0.06	Number of Rwandan gender law questions correct out of 3	0.03
	(0.05)		(0.05)
Rwandan law makes it too easy to bring a charge of violence against a man	-0.00	Lost job in past 6 months	-0.02
	(0.01)		(0.03)
You are closer to the people in your VSLA than others in your village, on average	-0.05	You are closer to the people in your VSLA than others in your village, on average	0.01
	(0.04)		(0.03)
Overall happiness	-0.06	Overall happiness	0.02
	(0.05)		(0.05)
Controls: do you know anyone who participated in the IPV prevention program?	0.03	Controls: do you know anyone who participated in the IPV prevention program?	-0.01
	(0.04)		(0.04)
Did you visit a community safe space?	-0.05	Did you visit a community safe space?	-0.00
	(0.03)		(0.03)
Enumerator: did the respondent show any signs of distress during the interview	-0.00	Enumerator: did the respondent show any signs of distress during the interview	0.00
	(0.01)		(0.01)

Table A12: Treatment effect on individual questions (limited to binary response questions and those with statistically significant impacts)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Women						Men		
VARIABLES	Agree: Bathing and feeding children are mothers' duties			In a hypothetical situation where a cousin's husband is being violent, you would help her report him rather than advising her not to			In a hypothetical situation where a cousin's husband is being violent, you would help her report him rather than advising her not to		
Overestimate conservative gender roles only		0.24*			0.07			-0.03	
		(0.12)			(0.08)			(0.10)	
Treatment x Overestimate gender roles only		-0.22*			-0.11			-0.03	
		(0.12)			(0.08)			(0.04)	
Overestimate IPV acceptability only		-0.25			-0.16			0.00	
		(0.21)			(0.14)			(0.00)	
Treatment x Overestimate IPV acceptability only		0.37			0.12			-0.11	
		(0.23)			(0.15)			(0.11)	
Overestimate conservative norms in both domains		0.00			0.00			-0.13	
		(0.00)			(0.00)			(0.10)	
Treatment x Overestimate both		-0.06			-0.06			0.09**	
		(0.08)			(0.05)			(0.04)	
Treatment x Overestimated either category	-0.05			-0.06			0.03		
	(0.06)			(0.04)			(0.03)		
Women role model treatment			-0.05			-0.09*			0.01
			(0.07)			(0.05)			(0.03)
Men role model treatment			-0.05			-0.03			0.03
			(0.07)			(0.04)			(0.03)
Successful businessmen role model treatment									0.05
									(0.04)
Male village leader role model treatment									
Observations	647	647	647	647	647	647	554	554	554
PDS controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of strata	98	98	98	98	98	98	98	98	98

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A13: Treatment effects without controls – norm domains

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Women						Men					
VARIABLES	Composite progressive gender attitudes index	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to husband (out of 10)	Tokens given to husband (out of 10)	Composite progressive gender attitudes index	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to wife (out of 10)	Tokens given to wife (out of 10)
Overestimate conservative gender roles only						-0.06						0.31
						(0.46)						(0.41)
Treatment x Overestimate gender roles only		-0.15**	-0.22***	-0.04		0.64		0.04	0.04	0.04		0.40
		(0.07)	(0.08)	(0.10)		(0.46)		(0.07)	(0.09)	(0.08)		(0.36)
Overestimate IPV acceptability only		-0.20*	-0.24*	-0.25		1.01		0.07	-0.02	0.08		-0.36
		(0.12)	(0.14)	(0.18)		(0.79)		(0.16)	(0.22)	(0.20)		(0.87)
Treatment x Overestimate IPV only		0.05	0.08	0.10		-0.27		-0.25	-0.13	-0.32		0.88
		(0.12)	(0.14)	(0.18)		(0.86)		(0.17)	(0.23)	(0.21)		(0.92)
Overestimate conservative norms in both		-0.23***	-0.29***	-0.17*				-0.15**	-0.14	-0.16*		
		(0.07)	(0.08)	(0.10)				(0.07)	(0.10)	(0.09)		
Treatment x Overestimate both		-0.03	-0.03	-0.07		-0.13		0.07	0.00	0.09		0.50
		(0.04)	(0.05)	(0.06)		(0.28)		(0.06)	(0.08)	(0.07)		(0.32)
Treatment x Overestimated either category	-0.04				0.14		0.05				0.51**	
	(0.03)				(0.22)		(0.04)				(0.22)	
Observations	662	662	662	662	654	654	554	554	554	554	551	551
PDS controls	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of strata	98	98	98	98	98	98	98	98	98	98	97	97

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A14: Treatment effects without controls – role models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
VARIABLES	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to husband (out of 10)	Composite progressive gender attitudes index	Progressive attitudes index- gender roles	Progressive attitudes index- IPV acceptability	Tokens given to wife (out of 10)
Women role model treatment	-0.07*	-0.08*	-0.09	-0.03				
	(0.04)	(0.05)	(0.06)	(0.27)				
Men role model treatment	-0.01	-0.02	0.01	0.26	0.04	0.02	0.03	0.47*
	(0.04)	(0.04)	(0.05)	(0.25)	(0.05)	(0.07)	(0.06)	(0.28)
Successful businessmen role model treatment					0.08	0.08	0.09	0.50*
					(0.05)	(0.07)	(0.06)	(0.28)
Male village leader role model treatment					0.04	-0.01	0.05	0.59*
					(0.06)	(0.08)	(0.07)	(0.31)
Observations	662	662	662	654	554	554	554	551
Number of strata	98	98	98	98	98	98	98	97
Women vs men p-value	0.11	0.18	0.10	0.28				
Men vs businessmen p-value					0.38	0.44	0.32	0.94
Men vs leader p-value					0.94	0.65	0.72	0.72
Businessmen vs leader p-value					0.46	0.25	0.58	0.77

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix B

Survey instruments The survey experiment was administered at the end of a 70-minute survey on gender beliefs, household roles, time use, women's role in decision-making, business and agriculture, gender attitudes and beliefs, and IPV experience. Towards the end of this survey, women and men were asked the following set of questions based on their randomly assigned survey experiment group. Women were randomly assigned to group A, B, or control. Men were randomly assigned to group B, C, D, or control.

Beliefs:

Script for enumerator: *I will ask you questions about what your neighbors think. Think about all the people in villages near here. I will say a statement now. Tell me how many people in villages near here you think will agree to this statement. In other words, if you were to talk to ten people from villages near here, how many of them would agree?* (Enumerator: if they don't understand: use the visual aid and ask: out of 10 [x], how many of them believe...)					
	Group A: Women (women only)	**Group B: men (both women and men)**	**Group C: successful male business owners (men only)**	**Group D: male village leaders (men only)**	**Control (both women and men)**
1	What share of women in villages near here believe that it is shameful for a woman if her husband cooks or does housework?	What share of men in villages near here believe that it is shameful for a woman if her husband cooks or does housework?	What share of successful male business owners in villages near here believe that it is shameful for a woman if her husband cooks or does housework?	What share of male village leaders in villages near here believe that it is shameful for a woman if her husband cooks or does housework?	Asked about beliefs regarding each of the groups for respondent's sex [Men: men, male business owners, male village leaders. Women: Women, men]
2	What share of women in villages near here believe that violence between husband and wife is a private matter and others should not intervene?	What share of men in villages near here believe that violence between husband and wife is a private matter and others should not intervene?	What share of successful male business owners in villages near here believe that violence between husband and wife is a private matter and others should not intervene?	What share of male village leaders in villages near here believe that violence between husband and wife is a private matter and others should not intervene?	

Information update:

Information provided only if over-estimated the share of people holding gender unequal views: *Thank you very much for answering these questions! We will now go over your answers, and we will tell you the correct answer to these questions.*					No information to control group
3	You estimated that about X percent of women in villages near here believe that it is shameful for a woman if her husband cooks or does housework. According to the last interview that we conducted a year ago, 43% of women or a bit more than 4 out of 10 women, in villages near here believe that it is shameful for a woman if her husband cooks or does housework.	You estimated that about X percent of men in villages near here believe that it is shameful for a woman if her husband cooks or does housework. According to the last interview that we conducted a year ago, 10% of men or one out of ten men, in villages near here believe that it is shameful for a woman if her husband cooks or does housework.	You estimated that about X percent of successful male business owners in villages near here believe that it is shameful for a woman if her husband cooks or does housework. According to the last interview that we conducted a year ago, 8% of male business owners or a bit less than 1 out of ten male business owners, in villages near here believe that it is shameful for a woman if her husband cooks or does housework.	You estimated that about X percent of male village leaders in villages near here believe that it is shameful for a woman if her husband cooks or does housework. According to the last interview that we conducted a year ago, 9% of male village leaders or a bit less than 1 out of 10 male village leaders, in villages near here believe that it is shameful for a woman if her husband cooks or does housework.	
4	You estimated that X percent of women in villages near here believe that violence between husband and wife is a private matter and others should not intervene. According to the last interview that we conducted a year ago, 47% of women or a bit less than 5 out of 10 women in villages near here believe that violence between husband and wife is a private matter and others should not intervene.	You estimated that X percent of men in villages near here believe that violence between husband and wife is a private matter and others should not intervene. According to the last interview that we conducted a year ago, 33% of men, or a bit more than 3 out of ten men in villages near here believe that violence between husband and wife is a private matter and others should not intervene.	You estimated that X percent of successful male business owners in villages near here believe that violence between husband and wife is a private matter and others should not intervene. According to the last interview that we conducted a year ago, 23% of successful male business owners or a bit more than 2 out of ten male business owners in villages near here believe that violence between husband and wife is a private matter and others should not intervene.	You estimated that X percent of male village leaders in villages near here believe that violence between husband and wife is a private matter and others should not intervene. According to the last interview that we conducted a year ago, 17% of male village leaders or a bit less than 2 out of 10 male village leaders in villages near here believe that violence between husband and wife is a private matter and others should not intervene.	

Outcomes

We will now ask you some questions about your opinions on men and women in society.		
Q Category (Household = H or IPV)	**Question wording**	**Answer options**
H1 "agree_neighbors"	Suppose a husband in a neighboring household has started helping his wife out more around the house, taking care of the children and helping to prepare food. Some neighbors are saying that he has been bewitched and is no longer a strong man. To what extent would you agree with the neighbors?	1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly Agree
Robustness test: did related beliefs update? "agree_village"	*How likely is it that most other people in this village agree with them?*	*1. Impossible/it would never happen* *2. Very unlikely* *3. Unlikely* *4. Likely* *5. Very Likely*
IPV1 "intervention_improvement"	Suppose a man beats his wife almost every evening. To what extent do you agree that their neighbors should wait for the situation to improve on its own instead of intervening?	1. Should intervene 2. Should wait to improve (coded as likert scale in women's survey)
IPV2 "ipv_action"	Suppose you visit your cousin and she tells you that her husband beat her severely and asks you for help. Suppose there is only one of two actions that you can take. Please tell us which one you would prefer to take. (Green, Wilke, & Cooper, 2018)	1. I would accompany her to the police or local authorities to report the incident 2. I would calm her down and tell her that the situation is bound to get better

	RESPONDENTS RANDONLY ASSIGNED TO GET IPV3A or IPV3B (Green, Wilke, & Cooper, 2018) **Combined to create one variable for main analysis**	
IPV3 A "ipv_report"	Imagine that you are walking back from the garden one day, and through the bushes you see someone from your community beating his wife, who is silently crying and cowering in the corner of their home. Suppose you see a group of people from your community standing nearby, discussing the incident. Would you report it to a local leader?	1. Yes 2. No, I don't feel comfortable meddling in other people's private affairs
IPV3 B	Imagine that you are walking back from the garden one day, and through the bushes you see someone from your community beating his wife, who is silently crying and cowering in the corner of their home. Suppose you know that you are the only who has observed the incident. Would you report it to a local leader?	1. Yes 2. No, I don't feel comfortable meddling in other people's private affairs
H2 "women_role"	The most important role of a woman is to take care of her home and cook for her family	1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly Agree
H3 "mothers_duties"	Bathing and feeding of children are mother's duties.	1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly Agree
H4 "equal_share_chores"	Men and women should share household chores.	1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly Agree
IPV4 "tolerate_ipv"	A woman should tolerate violence to keep her family together	1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly Agree
IPV5 "deserve_beating"	There are situations in which a woman deserves to be beaten	1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly Agree

Visual aid used by the enumerators



Appendix C

Instructions for the Incentivized Spouse Dictator Game:
Consent form read out and consent given or not.
General instructions, based on those given in Hoel (2015):

1. In this module, you will have a chance to win real money. The researchers for this project have arranged for the money. Whatever money you win in these questions will be yours to keep and take home.
2. In this module you will play with your husband. Please note even though you are 'playing it' with them, your husband will be playing this module somewhere else and making their decisions alone and in secret.
3. You will make only secret decisions. This means that no one, not even your spouse, will know what decision you have made. Only the researchers in the World Bank, who do not know your name or where you live, will know about your choices.
4. At the end of this session, we will combine your choices with the choices of your partner's, your stipend for attending today, and we will then add or subtract a random extra amount of money, and together, this will determine how much money you win. This means that your decisions are hidden from everyone and no one else knows what decisions you made for this module.
5. You will then receive your money at the end of today before we leave.
6. We do not want to involve you in anything that you do not understand. For that reason, we are going tell you everything there is to know about each module before we ask you to play. If at any time you decide that you do not wish to participate in a module for any reason, you are free to leave.

Now I would like to ask you a few questions to check your understanding of the instructions. If you don't understand some part, I will just repeat the instructions and that will be fine.

1. Who will be your partner in these modules? [Your partner; anonymous person from my village]
2. Who will be told about the decisions you make today? [No one except researchers at the World Bank, who don't know your name or where you live]
3. Will I ever tell your village elder or the government or NGO about your choices? [No]
4. Will anyone be able to figure out what decisions I make? [No, because we will pay people different amounts based on secret decisions of their spouses and community members, plus their stipend, AND we will also add or subtract a random secret amount, so no one will be able to figure out what decisions anyone else made]

Great! Now that you understand the general rules let's play.

Spouse Decider Module
Note: Respondent will be playing this module with 10 tokens and two labeled tins.
Secret spouse decider module:
Module rules:

1. Your partner in this module is your spouse.
2. Your choices in this module will be secret. I will not tell your spouse how you have played. I will also not tell anyone else in the village, the government or in the NGO how you have played.
3. In this module, I will give you bottle caps that represent real money.
4. You can keep all these bottle caps and money for yourself or you can give some or all to your spouse.
5. The bottle caps you keep are worth 100 RWF for yourself. But if you give them to your spouse, they are worth 150 RWF to your spouse. For example, if you keep 2 bottle caps, they are worth 200 RWF to you. If you give 2 bottle caps, they are worth 300 RWF to your spouse. You can use this sheet to help you understand the value of the bottle caps. This column shows how much the bottle caps are worth if you keep them, and this column shows how much the caps are worth if you give them.
6. Here are 2 labeled tins. This one is labeled 'self' and this one is labeled 'spouse'. Drop the bottle caps you would like to keep in the tin labeled 'self' and the bottle caps you would like to give in the tin labeled 'spouse'.
7. Because this module is secret, your spouse will not know that the money comes from you because they will be given their total payment based on their decisions, your decisions, and the travel stipend, and we will also add or take away a random secret amount for each participant. They will not know if you do give them money, and they will also not know if you do not give them money.

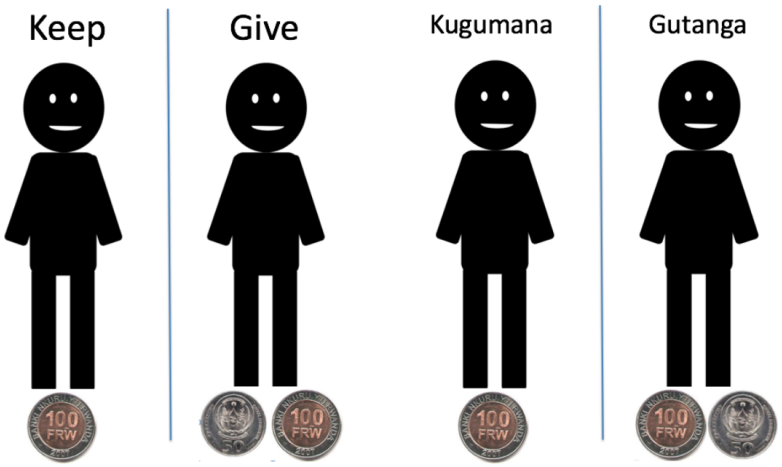
Again, I would like to ask you a few questions to check your understanding of the rules of the secret spouse decider module. Questions for secret spouse module:

1. Here are 7 bottle caps. If you keep them by putting them in the 'self' tin [demonstrate], how much are they worth to you? [700 RWF]
2. If you give the 7 bottle caps by putting them in the 'spouse' tin, how much are they worth to your spouse? [1050 RWF]
3. Who is your partner for this module? [spouse]
4. Will I tell your spouse what you decided? [no]

5. Will your spouse know how much you gave to them in this module? [no]

Ok! Now lets play the module! Here are 10 bottle caps. You can make your choice as you wish.
Now I am going to turn around and look away while you make your decision. Tell me when you are ready and I will come back.
Ok, you have chosen to keep X bottle caps worth X to you and give X bottle caps worth X to your spouse. Is that correct?

Visual Aids



Urupapuro rwerekana agaciro k'imipfundikizo

Imipfundikizo ugumana (RWF)	Imipfundikizo	Imipfundikizo utanga (RWF)
0	0	0
100	1	150
200	2	300
300	3	450
400	4	600
500	5	750
600	6	900
700	7	1050
800	8	1200
900	9	1350
1000	10	1500

5
Conclusions and Policy Implications

In this thesis, I considered how social and behavioral frictions in decision-making contribute to gender inequality in developing economies. I focused on individuals' concerns about social norms, role models, and preserving identity, and examined the role they play in beliefs, attitudes, and behaviors, including intimate partner violence. I used field experiments in Rwanda and Nigeria to provide empirical evidence on these relationships. My three articles each contribute to a different aspect of the question of why gender inequality exists and persists.

In my first paper, I approached the issue of gender inequality by considering the effectiveness of an intervention considered best practice in how to reduce IPV. I presented evidence on the social and behavioral frictions in decision-making that may have contributed to the unexpected IPV backlash. In my second paper, I considered the issue of gender inequality from a measurement lens. I analyzed the magnitude and predictors of misreporting bias when surveying women and men about IPV, and assessed the role of social norms in misreporting. In my third paper, I presented research on misperceptions about social norms and their effect on gender attitudes and related behaviors.

Some limitations of this research include the reliability of the measures as proxies for the ultimate outcomes of interest. For example, in my IPV prevention study, we relied on the standard self-reported survey questionnaires asking women if they had experienced IPV. Also, in my norms information experiment paper, I relied on self-reported attitude questions asked immediately after the information treatment, and incentivized behavior measured 10 minutes later. It would be ideal to use longer-term, independent, or observable outcome measures that are less at risk of experimenter demand effects or reporting bias, and are closer proxies of the true respondent behavior of interest outside a survey setting.

Also, given my need to exclude those in the RCT treatment and spillover arms, my norms experiment was likely underpowered.

There are several policy implications stemming from my research. Given the large, unintended, and community-wide negative impacts of the couples training program in Rwanda, future delivery of similar programs that promote women's equality in conservative settings should proceed with caution. They should be rigorously assessed for their direct and spillover effects, and only proceed once reliable monitoring of potential backlash is possible so that programs can be halted and re-designed if needed. Further, policymakers should consider and evaluate a broader suite of potential interventions that might reduce IPV. For example, alternative policies that may be less threatening to identity and less likely to provoke a backlash might include girls' education, gender attitude programs targeted to adolescents, or edutainment programs. Such interventions may mitigate the risk of backlash if they feel less didactic, are not 'labeled' as women's empowerment programs, do not invoke zero-sum bargaining power concerns in existing couples, or if they change girls' partner preferences, for example. Ultimately, more research is needed on the impact of alternative policies that may be more affordable and effective at preventing IPV.

On measurement, my findings suggest that IPV is likely to be substantially underreported in some contexts, and that reporting bias is systematic. This means that researchers and policymakers risk under-estimating the prevalence and welfare costs of IPV and may draw inaccurate conclusions about its causes and potential solutions. This suggests that policymakers should use survey methods that increase respondent privacy and anonymity to collect IPV prevalence data. Further, policymakers and social scientists should invest in research on alternative, more efficient, reliable IPV data collection tools which might then contribute to new research on IPV. This would ultimately improve our understanding of the causes of IPV and broaden the range of interventions used to tackle it.

My research also suggests that many individuals over-estimate conservative gender norms, and that information on the true norm can change behavior and attitudes. This serves as a proof-of-concept for potential interventions. It indicates that policymakers wanting to promote more gender-egalitarian behaviors could inform people of true progressive norms and potentially spark behavior change. It also suggests that interventions can have more influence if they use particular role model types and that programs should seek to understand the types of role models that have the biggest impact.

Future research on IPV could consider the causes of an IPV backlash, the role of threatened identity, and also how externalities transmit through networks and communities. On social norms, additional research could consider the impact of corrected norms information for those who under-estimate conservative norms to assess potential unintended consequences. It could also seek to identify the role of motivated reasoning in belief

formation about norms. Further research could also assess the impact of norms information on longer-term observable outcomes outside a lab-in-the-field setting. Further, it would be useful to know about the importance of context, in particular, the availability of free and accessible media on the impact of norms information campaigns, and the mechanisms that make particular role models more effective. On measurement, future research on sensitive survey methods could usefully test alternative, efficient survey methods that minimize reporting bias on IPV survey questions, given the large variance and other concerns with the list experiment.

Together, my three papers contribute to our understanding of the measurement of IPV, the social and behavioral factors that contribute to gender inequality, and the effectiveness of interventions to reduce it. My research suggests that individuals' economic behavior is influenced both by their desire to preserve their identity and to behave in accordance with social norms. Identity and norms thus likely contribute to behaviors related to gender inequality and may, therefore, be important to address in order to overcome it.

References

Abramsky, Tanya, Karen Devries, Ligia Kiss, Janet Nakuti, Nambusi Kyegombe, Elizabeth Starmann, Bonnie Cundill, Leilani Francisco, Dan Kaye, and Tina Musuya. 2014. 'Findings from the SASA! Study: A Cluster Randomized Controlled Trial to Assess the Impact of a Community Mobilization Intervention to Prevent Violence against Women and Reduce HIV Risk in Kampala, Uganda'. *BMC Medicine* 12 (1): 122.

Agüero, Jorge, and Verónica Frisancho. 2017. 'Misreporting in Sensitive Health Behaviors and Its Impact on Treatment Effects: An Application to Intimate Partner Violence'. IDB Working Paper No. IDB-WP-853.

Aizer, A. 2010. 'The Gender Wage Gap and Domestic Violence.' *The American Economic Review* 100 (4): 1847–59. https://doi.org/10.1257/aer.100.4.1847.

Akerlof, G.A. 1997. 'Social Distance and Social Decisions'. *Econometrica*, 1005–27. https://elibrary.ru/item.asp?id=8589014.

Akerlof, George A, and Rachel E Kranton. 2000. 'Economics and Identity'. *The Quarterly Journal of Economics* 115 (3): 715–53.

Alesina, A, P Giuliano, and N Nunn. 2013. 'On the Origins of Gender Roles: Women and the Plough'. *The Quarterly Journal of Economics* 128 (2): 469–530. https://academic.oup.com/qje/article-abstract/128/2/469/1943509.

Alesina, Alberto, Benedetta Brioschi, and Eliana La Ferrara. 2020. 'Violence Against Women: A Cross-cultural Analysis for Africa'. *Economica*, May. https://doi.org/10.1111/ecca.12343.

Angrist, N. 2020. 'The Messenger Matters: Experimental Evidence from a Sex Education Intervention'.

Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena. 2020. 'Bride Price and Female Education'. *Journal of Political Economy* 128 (2): 591–641. https://doi.org/10.1086/704572.

Baird, S, JA Bohren, C McIntosh, and B Özler. 2014. 'Designing Experiments to Measure Spillover Effects'. 6824. Policy Research Working Paper. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-6824.

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco. 2019. 'Entertainment, Education, and Attitudes Toward Domestic Violence'. In *AEA Papers and Proceedings*, 109:133–37.

Baranov, Victoria, Lisa Cameron, Diana Contreras Suarez, and Claire Thibout. 2020. 'Theoretical Underpinnings and Meta-Analysis of the Effects of Cash Transfers on Intimate Partner Violence in Low- and Middle-Income Countries'. *Journal of Development Studies*. Routledge. https://doi.org/10.1080/00220388.2020.1762859.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. 'Female

Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India'. *Science* 335. https://doi.org/10.1126/science.1211180.

Bénabou, Roland, and Jean Tirole. 2016. 'Mindful Economics: The Production, Consumption, and Value of Beliefs'. *Journal of Economic Perspectives* 30 (3): 141–64. https://doi.org/10.1257/jep.30.3.141.

Bhalotra, S, U Kambhampati, S Rawlings, and Z Siddique. 2020. 'Intimate Partner Violence: The Influence of Job Opportunities for Men and Women'. WPS9118. *World Bank Economic Review*. Policy Research Working Paper. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9118.

Bloch, Francis, and Vijayendra Rao. 2002. 'Terror as a Bargaining Instrument: A Case Study of Dowry Violence in Rural India'. *American Economic Review* 92 (4): 1029–43. https://doi.org/10.1257/00028280260344588.

Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise. 2018. 'A Mixed-Method Review of Cash Transfers and Intimate Partner Violence in Low- and Middle-Income Countries'. *The World Bank Research Observer* 33 (2): 218–58. https://doi.org/10.1093/wbro/lky002.

Bulte, Erwin, and Robert Lensink. 2019. 'Women's Empowerment and Domestic Abuse: Experimental Evidence from Vietnam'. *European Economic Review* 115 (June): 172–91. https://doi.org/10.1016/j.euroecorev.2019.03.003.

Bursztyn, L, G Egorov, and S Fiorin. 2019. 'From Extreme to Mainstream: The Erosion of Social Norms'. https://pdfs.semanticscholar.org/3151/53ba85aa709828af2464dc1b1e8b4dc92d92.pdf.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2018. 'Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia'. https://home.uchicago.edu/~bursztyn/Misperceived_Norms_2018_06_20.pdf.

Carpio, Lucia Del, and Maria Guadalupe. 2018. 'More Women in Tech? Evidence from a Field Experiment Addressing Social Identity'. 11876. IZA - Institute of Labor Economics Discussion Paper Series. https://www.iza.org/publications/dp/11876/more-women-in-tech-evidence-from-a-field-experiment-addressing-social-identity.

Chatterji, Sangeeta, Erin Stern, Kristin Dunkle, and Lori Heise. 2020. 'Community Activism as a Strategy to Reduce Intimate Partner Violence (IPV) in Rural Rwanda: Results of a Community Randomised Trial'. *Journal of Global Health* 10 (1). https://doi.org/10.7189/jogh.10.010406.

DiTraglia, F. J., C. García-Jimeno, R. O'Keeffe-O'Donovan, and A. Sanchez. 2018. 'Identifying Causal Effects in Experiments with Social Interactions and Non-Compliance'. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3608828.

Doyle, Kate, Ruti G. Levtov, Gary Barker, Gautam G. Bastian, Jeffrey B. Bingenheimer, Shamsi Kazimbaya, Anicet Nzabonimpa, et al. 2018. 'Gender-

Transformative Bandebereho Couples' Intervention to Promote Male Engagement in Reproductive and Maternal Health and Violence Prevention in Rwanda: Findings from a Randomized Controlled Trial'. Edited by Jacobus P. van Wouwe. *PLOS ONE* 13 (4): e0192756. https://doi.org/10.1371/journal.pone.0192756.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2019. 'General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya'. 26600. National Bureau of Economic Research. https://www.nber.org/papers/w26600.

Ericsson, Sanna. 2020. 'Cultural Gender Norms and Neighbourhood Exposure: Impacts on the Gender Gap in Math Cultural Norms and Neighbourhood Exposure: Impacts on the Gender Gap in Math'. https://project.nek.lu.se/publications/workpap/papers/wp20_6.pdf.

Eswaran, Mukesh, and Nisha Malhotra. 2011. 'Domestic Violence and Women's Autonomy in Developing Countries: Theory and Evidence'. *Canadian Journal of Economics/Revue Canadienne d'économique* 44 (4): 1222–63. https://doi.org/10.1111/j.1540-5982.2011.01673.x.

Farmer, Amy, and Jill Tiefenthaler. 1997. 'An Economic Analysis of Domestic Violence'. *Review of Social Economy*. Taylor & Francis, Ltd. https://doi.org/10.2307/29769903.

Field, E. M., R. Pande, N. Rigol, S. G. Schaner, and C. Moore. 2019. 'On Her Own Account: How Strengthening Women's Financial Control Affects Labor Supply and Gender Norms'. No. w26294. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=On+Her+Own+Account%3A+How+Strengthening+Women%27s+Financial+Control+Affects+Labor+Supply+and+Gender+Norms&btnG=.

Green, Donald P., Anna M. Wilke, and Jasper Cooper. 2020. 'Countering Violence Against Women by Encouraging Disclosure: A Mass Media Experiment in Rural Uganda'. *Comparative Political Studies*, April, 001041402091227. https://doi.org/10.1177/0010414020912275.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal. 2018. 'Does Information Change Attitudes Towards Immigrants? Representative Evidence from Survey Experiments'. *SSRN*. https://doi.org/10.2139/ssrn.2768187.

Guiso, Luigi, Paola Sapienza, Luigi Zingales, Alberto Alesina, George Akerlof, Gary Becker, Alberto Bisin, et al. 2006. 'Does Culture Affect Economic Outcomes?' *The Jounnal of Economic Perspectives* 20 (2): 23–48.

Haushofer, Johannes, James Reisinger, and Jeremy Shapiro. 2015. 'Your Gain Is My Pain: Negative Psychological Externalities of Cash Transfers'.

Haushofer, Johannes, Charlotte Ringdal, Jeremy P Shapiro, and Xiao Yu Wang. 2019. 'Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya'. w25627. NBER Working Paper Series.

Hossain, Mazeda, Cathy Zimmerman, Ligia Kiss, T Abramsky, Drissa Kone, Monika Bakayoko-Topolska, Jeannie Annan, Heidi Lehmann, and Charlotte Watts. 2014. 'Working with Men to Prevent Intimate Partner Violence in a Conflict-Affected Setting: A Pilot Cluster Randomized Controlled Trial in Rural Côte d'Ivoire.' *BMC Public Health* 14 (1). https://doi.org/10.1186/1471-2458-14-339.

Jayachandran, Seema. 2015. 'The Roots of Gender Inequality in Developing Countries'. *Annu. Rev. Econ* 7: 63–88.

———. 2020. 'Social Norms as a Barrier to Women's Employment in Developing Countries'. 27449. Cambridge, MA. https://doi.org/10.3386/w27449.

Joseph, George, Syed Usman Javaid, Luis Alberto Andres, Gnanaraj Chellaraj, Jennifer Solotaroff, Sebastian Irudaya Rajan, C García-Moreno, et al. 2017. 'A New Survey Technique For Studying Deviant Behavior'. *Journal of the American Statistical Association*, World Bank Policy Research Working Paper, 39 (1): 211–30. https://doi.org/10.1162/rest.2010.11232.

Kahan, Dan M. 2013. 'Ideology, Motivated Reasoning, and Cognitive Reflection'. *Judgment and Decision Making* 8 (4): 407–24. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2182588.

Kappes, A, and T Sharot. 2019. 'The Automatic Nature of Motivated Belief Updating'. *Behavioural Public Policy* 3 (1): 87–103.

Karing, Anne. 2018. 'Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone'.

LaFerrara, Eliana. 2019. 'Aspirations, Social Norms, and Development'. *Journal of the European Economic Association* 17 (6): 1687–1722. https://academic.oup.com/jeea/article-abstract/17/6/1687/5608602.

LaFerrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. 'Soap Operas and Fertility: Evidence from Brazil'. *American Economic Journal: Applied Economics* 4 (4): 1–31. https://doi.org/10.1257/app.4.4.1.

Lowes, Sara, Nathan Nunn, James A Robinson, Jonathan Weigel, Anne Degrave, James Diderich, Muana Kasongo, et al. 2017. 'The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom'. *Econometrica* 85: 1065–91.

Lybbert, Travis J., and Bruce Wydick. 2018. 'Poverty, Aspirations, and the Economics of Hope'. *Economic Development and Cultural Change* 66 (4): 709–53. https://doi.org/10.1086/696968.

Mani, Anandi, and Emma Riley. 2019. 'Social Networks, Role Models, Peer Effects, and Aspirations'. 120. WIDER Working Paper. https://doi.org/10.35188/UNU-WIDER/2019/756-9.

Miguel, Edward, and Michael Kremer. 2004. 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities'. *Econometrica* 72 (1): 159–217. https://doi.org/10.1111/j.1468-0262.2004.00481.x.

Oh, Suanna. 2019. 'Does Identity Affect Labor Supply?'