WORKING PAPER Nº IDB-WP-1596

Gender-based Research and Interviewer Effects:

Evidence for Latin America and the Caribbean

Matías Guizzo-Altube
Carlos Scartascini

Inter-American Development Bank
Department of Research and Chief Economist

April 2024

Gender-based Research and Interviewer Effects:

Evidence for Latin America and the Caribbean

Matías Guizzo-Altube*
Carlos Scartascini*

* Inter-American Development Bank

Inter-American Development Bank
Department of Research and Chief Economist

April 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Guizzo, Matias.
Gender-based research and interviewer effects: evidence for Latin America and the Caribbean / Matias Guizzo-Altube, Carlos Scartascini.
p. cm. — (IDB Working Paper Series ; 1596)
Includes bibliographical references.
1. Gender mainstreaming-Latin America. 2. Gender mainstreaming-Caribbean Area. 3. Equality-Latin America. 4. Equality-Caribbean Area. 5. Sex discrimination-Latin America. 6. Sex discrimination-Caribbean Area. 7. Social sciences-Methodology. I. Scartascini, Carlos G., 1971- II. Inter-American Development Bank. Department of Research and Chief Economist. III. Title. IV. Series.
IDB-WP-1596

http://www.iadb.org

Abstract

As the measurement of public opinion and attitudes toward gender issues has gained increasing attention in gender policy and research, the accurate design, implementation, and analysis of survey data has become paramount. In this paper, we examine the impact of the gender of the interviewer on survey responses in Latin America and the Caribbean, both individually and in the aggregate, using data from the LAPOP surveys (2012-2018). We also present a formal model that illustrates how the gender of the interviewer may influence responses and derive an optimal adjustment in survey design and analysis to account for this bias. Our findings highlight the substantial influence of social norms on gender opinions, revealing that respondents are more likely to overestimate men's abilities in politics relative to women's and to accept domestic violence when interviewed by men rather than women. Two key actions emerge as a corollary: survey firms should always provide the gender of the interviewer in their data, and researchers should adjust for this bias in their analysis. Without implementing these recommended practices, survey data may not accurately reflect what individuals in the region think about gender issues, potentially skewing behaviors and public policy.

JEL classifications: J16, C83, D91
Keywords: Gender norms, Gender equality, Surveys, Behavioral biases

We are grateful to Liz Zechmeister, Noam Lupu, Dan Gingerich, Ernesto Calvo, and Tommy Murphy for comments on previous versions of the document. The information and opinions presented herein are entirely those of the authors, and no endorsement by the Inter-American Development Bank (IDB), its Board of Executive Directors, or the countries they represent is expressed or implied. IDB management had no involvement in the study design, analysis, or interpretation of the data, in the writing of the report, or in the decision to submit the article for publication.

1 Introduction

In recent decades, the study of gender issues has rapidly gained ground in the social sciences. The number of gender-related articles culled from the social categories of the JSTOR collection has soared from less than 1% in 1970 to more than 25% in 2016.[1] Concurrently, there has been a marked rise in the use of data-driven methodologies in these studies. Survey-based research constitutes nearly 12% of empirical gender-related studies (on average, about 1,000 have been published annually between 2011 and 2016).[2] Claudia Goldin receiving the Nobel Prize in 2023 for her research and contributions to understanding the changing roles of women in the economy is a testament to the prominence of this research.

The meteoric rise of survey-based gender research underscores the critical need to scrutinize the processes by which such data are produced and interpreted. One crucial element often overlooked and underexplored in this process is the potential effect of the gender of the interviewer on respondents' answers, particularly when survey questions touch on gender issues and preferences.

This paper aims to delve into the dynamics of the "gender-of-interviewer effect" and its implications for survey design and data interpretation. We investigate how the gender of the interviewer can lead to biased inference from the data and potentially skew the understanding of gender issues. To address this issue, we first introduce a formal model and provide a theoretical foundation for understanding and mitigating the biases introduced by the gender of interviewers. Subsequently, we quantify the impact of the interviewer's gender on survey responses using data from 26 Latin American and Caribbean countries. We explore, both theoretically and empirically, some nuances of the effect, such as its interaction with the gender of the respondent and with the presence of other people in the room, mainly a spouse.

The significance of the differences is highlighted by the following examples. The surveys analyzed in this paper show that when interviewed by men, respondents are 4.5 percentage points more likely to perceive men as superior political leaders compared to those interviewed by women. Moreover, acceptance of domestic violence is reported to be over five percentage points higher when the interviewer is male. These variations are not trivial and can substantially influence our

[1] See Figure A1 in the Appendix. This ratio surpassed 30% in 2022, but we only report data until 2016 due to an abrupt decline in the number of articles accessible through Constellate since 2017.

[2] In 2019, this percentage rose to around 16%. It is also excluded from the figures for the same reason mentioned in the previous footnote.

understanding of public opinion. For instance, in Guyana, the reported acceptance of domestic violence can fluctuate significantly, ranging from 31% to 36%. This figure could escalate to as high as 59% if all interviews were conducted by men or drop to as low as 24% with exclusively female interviewers. Regarding regional averages, acceptance rates are estimated at 22% for situations involving neglected chores and 33% for infidelity cases. These rates could increase to 27% and 39%, respectively, if all respondents were interviewed by men or decrease to 19% and 29% with all-female interviewers.

Perceptions of social norms have significant consequences in shaping behaviors and public policy. For example, when young married men underestimate the support for women working outside the home by other men, the likelihood that women will work outside the home is lower than when they are correctly informed (Bursztyn et al., 2020). As such, surveys that underestimate or overestimate support for certain behaviors can have large negative welfare consequences. For example, in the case of Guyana highlighted above, household behavior, public policy, and enforcement could differ substantially if people read that surveys indicate that acceptance of domestic violence in the country is 59% compared to 24%.

Because most surveys still do not include the gender of the interviewer in the publicly available datasets, this paper hopes to shed light on an underappreciated source of bias in gender research and provide actionable insights for scholars and practitioners to minimize such bias, thereby enhancing the accuracy and reliability of future studies.

2 Interviewer Characteristics and Potential Biases

Face-to-face interviews are one specific encounter of many within the broader set of social interactions. It is regulated by common social norms, defined as "a rule of behavior such that individuals prefer to conform to it on condition that they believe that (a) most people in their reference network conform to it (empirical expectation), and (b) that most people in their reference network believe they ought to conform to it (normative expectation)" (Bicchieri, 2017, p. 35). Thus, both participants behave not only as respondent and interviewer, but also as all the other roles they play in society based on "who they are" and, more importantly, "who the other person is" in society and how others are expected to behave. Beyond their roles as respondent and interviewer, each individual possesses an identity that transcends these specific interactions. Their personality

traits, behaviors, attitudes, beliefs, and preferences are shaped by their gender, age, ethnicity, nationality, education, occupation, and economic status, among many other characteristics. However, their behavior is also influenced by the other person's characteristics and how those characteristics interact with their own. For instance, a woman may be expected to behave differently when she is surrounded by men rather than other women.

When a questionnaire is developed to measure attitudes or beliefs, this complex system of expectations and norms can potentially distort the expression of the reality that researchers aim to capture, thereby inducing measurement errors. These errors are commonly referred to as "response effects" when they can be attributed to a specific source (Sudman and Bradburn, 1974). One source of biases, referred to as "role-restricted effects," originates from the context-specific behaviors of the interview that are absent in other social interactions. These effects can be mitigated with more precise instructions and training for the interviewer.[3] However, a second type of bias, known as "role-independent effects," is caused by the interviewer's characteristics and the social constructs and expectations associated with these traits and, therefore, cannot be easily overcome. Moreover, these are not properly identifiable biases, as the interplay of different social dynamics can simultaneously hinder the true answers in different directions. Instead, the traditional empirical strategy is to demonstrate the existence of differences in responses across interviewers' characteristics without assuming which is the most "genuine" answer (although many scholars theorize mechanisms and try to argue what they believe to be the "true answer"). If responses differ by interviewers' characteristics, then at least one, if not all, has to be shifted from the "true answer" as a result of this response effect. Gender of interviewer effects are a form of role-independent effects. Even though researchers can manipulate the composition of the pool of interviewers and balance certain characteristics, these effects cannot be completely eliminated, as they result from social phenomena that transcend the interview and affect all other interpersonal interactions.

Even in situations where human interaction is significantly less direct, such as in telephone interviews (Cotter et al., 1982; Groves and Fultz, 1985; Kane and Macaulay, 1993; Catania et al., 1996, among others) or video-enhanced web surveys (Fuchs, 2009), there is evidence that interviewer characteristics can affect survey responses.[4] These characteristics can even impact the rate of response, as choosing not to respond is another valid option (Silber et al., 2021).

[3] See Dijkstra (1987) for an example of the effects of interviewing styles on respondent behavior.
[4] See Davis et al. (2010) for a comprehensive review of role-independent interviewer effects.

Role-independent interviewer effects have been extensively studied over the last century. A prominent example is the surge of studies in the areas of race and gender that emerged in response to the human rights movements in the United States during the 1960s. As researchers used surveys to collect data on attitudes, many discovered that the identity of the interviewer mattered. For instance, in efforts to measure racism and attitudes towards segregation and racial equality, researchers found that respondents were less likely to express racist views to African-American interviewers. Similarly, to a lesser extent, African-American respondents were more likely to disclose experiences of racism (Schuman and Converse, 1971; Cotter et al., 1982; Finkel et al., 1991).

In gender studies, the literature generally indicates that both men and women tend to present themselves as more progressive in terms of gender equality to female interviewers. However, the results are not definitive, as most studies report effects that are weak and inconsistent (Landis et al., 1973; Kane and Macaulay, 1993; Padfield and Procter, 1996; Lipps and Lutz, 2017; Zoch, 2021 cover various dimensions of gender attitudes; Becker et al., 1995; Catania et al., 1996; Heeb and Gmel, 2001; Bignami-Van Assche et al., 2003; Anglewicz et al., 2009; Fuchs, 2009; Liu and Stainback, 2013; Kianersi et al., 2020 focus on questions about sexual behavior, health, and family planning; Huddy et al., 1997; Flores-Macias and Lawson, 2008; Benstead, 2014b; Sundström and Stockemer, 2022 concentrate on opinions about women in politics; and Groves and Fultz, 1985 targets economic attitudes). Research has also examined other interviewer characteristics such as ethnicity (Weeks and Moore, 1981; Reese et al., 1986), social class (Katz, 1942), birthplace (Weinreb, 2006; Anglewicz et al., 2009), and religiosity (Blaydes and Gillum, 2013; Benstead, 2014a).

In explaining the mechanisms that underpin role-independent interviewer effects, two primary hypotheses have been put forward in the literature (Benstead, 2014a). The social distance theory suggests that respondents tailor their responses to match the perceived attitudes and beliefs of the interviewer, a phenomenon that can be influenced by various forms of perceived social distance, such as cultural, racial, or socioeconomic disparities. Alternatively, the power dynamics hypothesis posits that the perceived hierarchical relationship between the interviewer and respondent affects response authenticity; individuals may conceal their true beliefs in the presence of an interviewer perceived as more powerful, while feeling free to be more candid when they see themselves as more or equally powerful. While distinct in their views on respondent honesty, these theories fit within the broader framework of social desirability bias, indicating that respondents may modify their answers

based on anticipated interviewer preferences. This is supported by empirical evidence showing how alterations in perceived power dynamics or social distance can impact response truthfulness. The interaction between these theories highlights the complexity of interviewer effects and underscores the need for a nuanced understanding of how different factors influence survey responses.

The complexities of interviewer selection and the challenge of isolating "true" responses from those influenced by interviewer characteristics have sparked considerable debate around the design and execution of interviews. In response, early research initiatives sought to align interviewer and respondent demographics or even allow respondents to select their interviewers, aiming to reduce potential biases. However, these strategies raised concerns about introducing new forms of selection bias. In an effort to circumvent role-independent interviewer effects and broader issues of social desirability bias, later innovations pivoted towards removing human elements from data collection entirely, favoring technology-driven methods like web surveys. While this approach has gained popularity for its cost-efficiency and ease of deployment, it has drawbacks, particularly in achieving representative samples across diverse populations and ensuring inclusivity for groups with limited internet access. This ongoing dialogue in the field underscores the importance of nuanced, context-sensitive interview methodologies that carefully weigh the benefits and limitations of technology-mediated data collection against the imperative for inclusivity and representativeness in research.

In reviewing the literature, a recurrent observation is that biases associated with interviewer characteristics are often modest in scale and not uniformly statistically significant. This variability in effect strength may stem from the necessity for the interviewer's traits to be distinctly relevant to the survey questions to sway responses. Moreover, discrepancies between the analytical depth of surveys and the distribution of prevailing social norms could further obscure these effects. For example, Flores-Macias and Lawson (2008) highlights how regional differences might attenuate observable impacts, a pattern that resonates with our findings. Our analysis, detailed in the following section, reveals a nuanced landscape across Latin America and the Caribbean, with country-specific variations in the size, direction, and statistical significance of interviewer influences. This underscores the importance of considering regional and cultural contexts in interpreting such effects.

3 A Simple Formalization of the Gender-of-Interviewer Effect

Consider a sample of n respondents of gender $\mu \in \{f, m\}$ who are asked by interviewers of gender $\theta \in \{f, m\}$ about their position on a particular issue. Each individual has a true opinion $v \in \mathbb{R}$ which, without loss of generality, can be ordered to be increasing in a potentially "pro-female" direction, such that the difference between the mean opinions of female and male respondents is non-negative, i.e., $\Delta v \equiv \bar{v}_f - \bar{v}_m \geq 0$. When giving their answers $x \in \mathbb{R}$ to the interviewers, the respondents try to be as faithful as possible to their true personal opinions v, while at the same time trying to appeal to the interviewer, whose opinion the respondent infers from their gender as $\tilde{v}_\theta$.[5] The problem faced by a representative respondent can be formalized as maximizing their utility $u(x)$, given by the function

$$u(x) = -(x - v)^2 - \delta(x - \tilde{v}_\theta)^2,$$

where the parameter $\delta \geq 0$ captures the extent to which the respondent cares about the position of the interviewer. The optimal response x^* that maximizes the respondent's utility is given by

$$x^* = v + \gamma\left(\tilde{v}_\theta - v\right),$$

where $\gamma \equiv \delta/(1 + \delta) \in [0, 1]$ can be interpreted as the weight of the interviewer's position and $1 - \gamma$ as the weight of the individual's actual position in the reported response. A straightforward result from this expression is that, if the respondent has a minimal interest in appealing to the interviewer ($\delta > 0$) and believes that the interviewer has a different position on the issue ($\tilde{v}_\theta \neq v$), then the reported opinion will differ from the respondent's true position. The more the respondent is concerned with pleasing the interviewer and the greater their perceived difference in opinion with the interviewer, the greater the distortion will be in reporting the response.

In this setup, the gender-of-interviewer effect can be defined as the difference between the optimal responses given to male and female interviewers. Taking the response given to a female

[5] The model can be easily modified to include a standard social desirability bias as a third term of the utility with a "socially desirable" opinion. However, we omit this component because it does not affect the results of interest for this paper.

interviewer as the benchmark, the "male interviewer effect" (MIE) would take the form of

$$MIE \equiv x^*(\theta = m) - x^*(\theta = f) = \gamma(\tilde{v}_m - \tilde{v}_f).$$

Given that, in the model, the respondent cannot observe the interviewer's opinion and has to infer it by their gender, the expected interviewer's opinion $\tilde{v}_\theta$ should be the average opinion of people of gender θ. However, people may have a distorted perception of the opinions distribution and, in consequence, they may be miscalculating the average opinion by gender. This phenomenon can be addressed with a distortion parameter $\varphi \geq 0$ in the inference made by the respondent, such that $\tilde{v}_f = \bar{v}_f + \varphi$ and $\tilde{v}_m = \bar{v}_m - \varphi$.[6] For simplicity and for the moment, we assume that the distortion is symmetrical, so the overestimation of $\bar{v}_f$ is equivalent to the underestimation of $\bar{v}_m$ in magnitude. This assumption can be relaxed without affecting the main conclusions. The resulting expression for the male interviewer effect of decomposing the opinions inferred by the respondents would be

$$MIE = -\gamma\left(\Delta v + 2\varphi\right).$$

This result predicts that a minimal distortion in the perceived opinion distributions by gender is enough for the gender-of-interviewer effect to exist, even if there is no real difference in the average opinions by gender, conditional on γ being positive. However, the larger the difference of average opinions by gender and the larger the distortion in respondent's perceptions, the greater the effect will be.

4 Gender-of-Interviewer Effects in Surveys from Latin America and the Caribbean

We employ data from the AmericasBarometer surveys conducted by the Latin American Public Opinion Project (LAPOP) to illustrate the contemporary effects of the interviewer's gender on responses in Latin America and the Caribbean.[7] LAPOP has been conducting this survey on a

[6] See Bordalo et al. (2016) and Bonomi et al. (2021) for a more detailed analysis of stereotypes and distortions in the perception of opinions from the lens of the social identity theory and the self-categorization theory.

[7] Further information about the Barometer available in their webpage https://www.vanderbilt.edu/lapop/about-americasbarometer.php.

biennial basis since 2004, encompassing up to 26 countries in the region.[8] The survey primarily utilizes face-to-face interviews for data collection, with the exception of the 2021 wave, which relied on telephone interviews (addressed separately toward the end of this section). The number of responses in each country and survey wave ranges from 1,000 to more than 4,400, with the average and median responses being 1,677 and 1,507, respectively. LAPOP began documenting the gender of the interviewers in the 2010 wave of the survey. Of the sample of 191,612 respondents interviewed between the 2010 and 2018 waves of the survey, 64.4% were interviewed by female enumerators.

The practice of recording the interviewer's gender almost coincided with the start of including a series of questions related to gender attitudes in the questionnaire (which happened in 2012). Having access to gender-focused questions and the gender of the interviewers provides an opportunity for an in-depth comparative analysis of the effects of the interviewer's gender in various dimensions across the region.

In the survey waves that document the interviewer's gender, we identified eight gender-related questions. Six of these questions measure the level of agreement or approval of the following statements:

- "When there is not enough work, men should have a greater right to jobs than women."
- "The state ought to require that political parties reserve some space on their lists of candidates for women, even if they have to exclude some men."
- "It is justified that the husband hits his wife if she neglects the household chores."
- "It is justified that the husband hits his wife if she is unfaithful."
- "When a mother works outside the home, the children suffer."
- "In general, men are better political leaders than women."

The remaining two questions provide "A man," "A woman," and "It does not matter" as the possible answers:

- "Who do you think would be more corrupt as a politician?"

[8] While some waves of the AmericasBarometer surveys also included Canada and the United States, these countries were excluded from the analysis in this note.

- "If a politician is responsible for running the national economy, who would do a better job?"

As can be observed from the list, the questions could be ranked according to whether they confront men and women and whether they are more or less aligned with existing norms and values in a society. The gender-of-the-interviewer effect could then be stronger for some of these questions than others and more important in some countries than in others.

To evaluate the relevance of this effect, for each question, we selected the most recent wave in which it was included and estimated the male interviewer effect using the linear model in equation 1.[9] Most of the questions offered three or four response options of either agreement or disagreement, without an indifference option, which we recode as binary variables of agreement. The questions on gender quotas and right to work offered response options on an agreement scale of 1 to 7. We standardize these two variables into the interval [0,1], such that 1 represents the maximum level of agreement and 0, the maximum level of disagreement. For the two questions on corruption and ability to manage the economy, we create one binary variable for responding "A man" and "A woman."

In the model, y_i is the answer to question y by respondent i, $MaleInterviewer_i$ is a binary variable that indicates whether individual i was interviewed by a male or a female interviewer, $\mathbf{x}_i$ is a vector of observable characteristics of respondent i that may affect their true or reported gender attitudes, including their own gender, age, education, wealth, marital status, whether they live in a urban or rural area, how important religion is in their life,[10] and whether the spouse of the interviewee was present during the interview in interaction with the respondent's gender,[11] and ε_i captures the variance of y not explained by the model. We estimate β using ordinary least squares; it measures the male interviewer effect.

$$y_i = \alpha + \beta MaleInterviewer_i + \mathbf{x}_i\gamma + \varepsilon_i \tag{1}$$

Figure 1 provides a summary of the estimates of the gender-of-interviewer effect, pooling both male and female respondents. Each line in the figure represents the results for a different question,

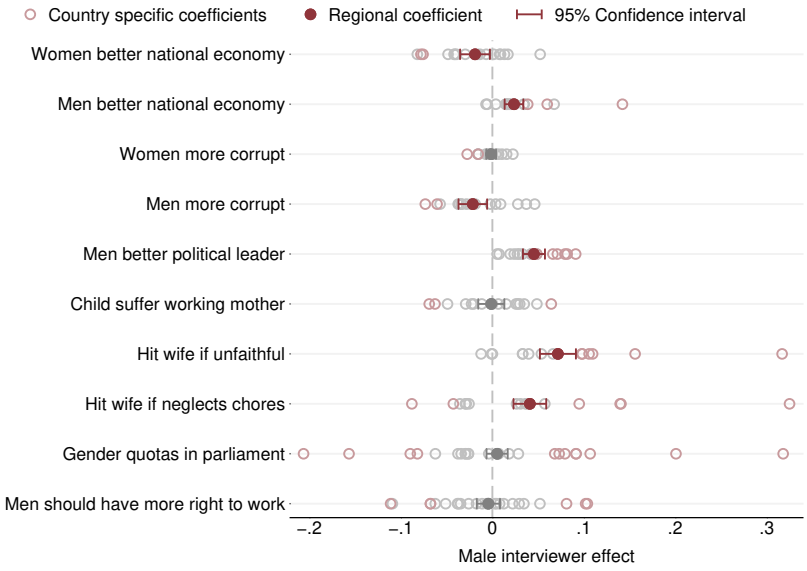
[9] We use data from the 2012 wave for the questions about gender quotas and right to a job, from the 2014 wave for the two questions about the justifiability of a husband hitting his wife, and from the 2018/19 wave for all the other questions. Figure A8 presents the regional results for all available waves.

[10] Ideology is not asked in several countries, comprising about 15% of the 2018 sample and more than 50% of the 2014 sample, so we do not include it as a control variable. However, results are robust to controlling for ideology in the cases where the question was included.

[11] Differential gender-of-interviewer effects by gender of interviewee and presence of spouse are covered towards the end of this section.

plotting the regional and individual country coefficients. As all variables run in a 0 to 1 agreement scale, coefficients can be interpreted as the difference in agreement between those respondents interviewed by men and those interviewed by women. For instance, people interviewed by men are 4.5 percentage points more likely to agree that men are better political leaders than those interviewed by women. There are 10 lines in place of 8 due to the structure of the questions concerning corruption and the management of the national economy, which provides the respondent the option to indicate a gender, as opposed to expressing agreement. Results are robust to corrections of p-values for multiple hypothesis testing (see Table A8 in the Appendix).

Figure 1: Male Interviewer Effects in Countries of Latin America and the Caribbean



Source: AmericasBarometer survey waves 2012, 2014, and 2018/19. Latin American Public Opinion Project (LAPOP), by Vanderbilt University.
Note: The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents, presence of spouse during the interviews, and country fixed effects (only in the regional estimates). Estimates are produced using probability weights and correct standard errors following the survey sample design. See the main text for a full description of the questions.

Several observations can be gleaned from the figure. Firstly, a regional significant effect is discernible for questions that juxtapose men and women in terms of their political capabilities, as well as a question concerning the justification of gender-based violence. When interviewed by a male enumerator, respondents are more likely to assert that men are better at managing the

economy, that men are superior political leaders, and that it is acceptable for a man to resort to physical violence against a woman if she is unfaithful or neglects household chores, as well as less likely to say that women would manage better the national economy or that men are more corrupt than women as politicians. The magnitude of these effects can be substantial. For instance, the reported acceptance of violence is, on average, more than 5 percentage points higher when the interviewer is a man compared to a woman. Conversely, these significant effects are absent when the question does not directly compare men and women, such as in the question about children suffering when mothers work or when discussing quotas and employment rights.

Secondly, these regional effects mask considerable heterogeneity across countries. As we delve into the country-level effects, we observe that these effects differ in both magnitude and direction. In certain countries, the effects are statistically significant in the "pro-female" direction, which runs counter to the regional estimate. Conversely, for questions pertaining to equal job rights, gender quotas, and working mothers, the effects at the regional level are null. However, they are statistically significant in numerous countries, exhibiting both positive and negative directions (e.g., gender quotas do not show a significant regional effect but very strong negative and positive effects—up to 20 and 30 percentage points differences—at the country level). The following subsection, along with Figures A4 through A7 in the Appendix, presents detailed results for each question by country. These findings are consistent with those reported by Flores-Macias and Lawson (2008) for Mexico, where it was suggested that despite the absence of effects at the country level, significant sub-regional effects exist that are neutralized when averaged out.

4.1 Does the Gender of the Respondent Matter Too?

4.1.1 Gender of Respondent in the Model

So far, the model assumes a universal γ (or δ) across all individuals. However, different hypotheses suggest that γ may vary based on both the interviewer's and the interviewee's gender. For example, under the power dynamics hypothesis, respondents are theorized to be more inclined to alter their responses to align with male interviewers due to power imbalances, which can be represented as a higher γ for respondents interviewed by men. In a different direction, the social distance theory posits a higher γ in situations where the genders of the interviewer and interviewee differ. These hypotheses can be represented by indexing this parameter by both respondent's (μ) and

interviewer's (θ) genders as $\gamma_{\mu\theta}$. The power dynamics hypothesis would be accurate in a case with $\gamma_{fm} > \gamma_{ff}$ and $\gamma_{mm} \geq \gamma_{mf} \approx 0$, while the social distance theory would prevail if $\gamma_{fm} > \gamma_{ff}$ and $\gamma_{mf} > \gamma_{mm}$.

A different way in which the effect could be influenced by the gender of the respondent, and one less addressed in the literature, is through φ. It is possible that respondents of different genders have different perceptions of the opinion distributions. For example, it could be possible that people have a better estimate of their own gender's distribution because of gender homophily. If this parameter was indexed by gender as $\varphi_{\mu\theta}$, this would incorporate the possibility of different perceptions of the opinion distributions by the gender of the respondent and by gender of the population.

Incorporating both modifications to the model, the male interviewer effect as a function of the respondent's gender could be written as

$$MIE(\mu) = -\gamma_{\mu f}\left(\Delta v + \varphi_{\mu f} + \varphi_{\mu m}\right) + \Delta\gamma_{\mu}\left(\bar{v}_m - \varphi_{\mu m} - v\right) \text{ with } \Delta\gamma_{\mu} \equiv \gamma_{\mu m} - \gamma_{\mu f}.$$

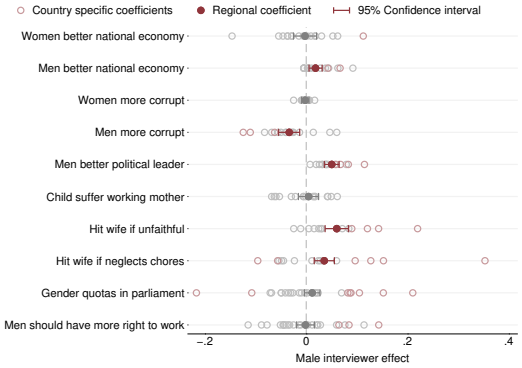
4.1.2 Empirical Results: Gender of Respondent

Figure 2 separates the effects presented in Figure 1 according to the gender of the respondent. Interestingly, many of the results are consistent across groups. For example, both women and men are more willing to tell a male interviewer that violence against women is acceptable, and that men are better political leaders and for managing the economy. Some of the other results differ according to the gender of the respondent. For example, on average, women are less likely to answer that men are more corrupt when in front of men than when in front of another woman. That effect is not significant for men. Males are more willing to say that women are worse for managing the national economy in front of another man than in front of women. The same does not happen for women.

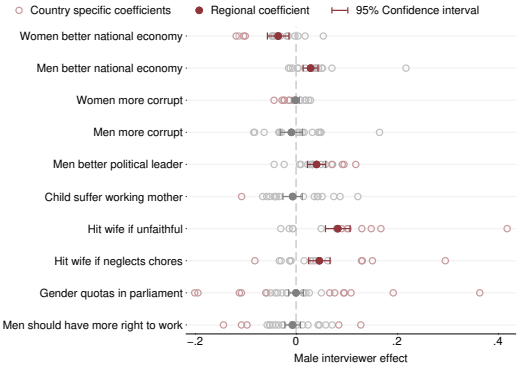
As discussed in the previous section, it is to be expected that, when it is present, the effect goes in the direction that the respondents expect the interviewer to support. However, as these results suggest, even in countries that are culturally similar, social norms can be complex and different enough to induce effects in unexpectedly different directions, and it is not always obvious what respondents expect from interviewers on average.

Figure 2: Male Interviewer Effects by Gender of Respondents

(a) Effects on female respondents



(b) Effects on male respondents



Note: The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents, presence of spouse during the interviews, and country fixed effects (only in the regional estimates). Estimates are produced using probability weights and correct standard errors following the survey sample design. See the main text for a full description of the questions.

Decomposing some of the results by country provides a more detailed perspective on what the regional averages reveal and obscure. Figure 3 depicts two such cases (the full set of questions is provided in Section A.3 of the Appendix). The top row displays responses to the question of whether men are better political leaders. It can be observed that where the male interviewer effect exists, it is positive. Both men and women are more inclined to assert that men are superior political leaders when interviewed by a male.

The bottom row illustrates that effects can vary by country and gender. In Uruguay, Paraguay, and Bolivia, men are less likely to affirm that women would manage the economy more effectively when interviewed by another male. In contrast, in Peru, women are more likely to posit that women would perform better when interviewed by a male than a female.

4.2 An Additional Complexity: Having Others in the Room

4.2.1 Presence of spouse in the model

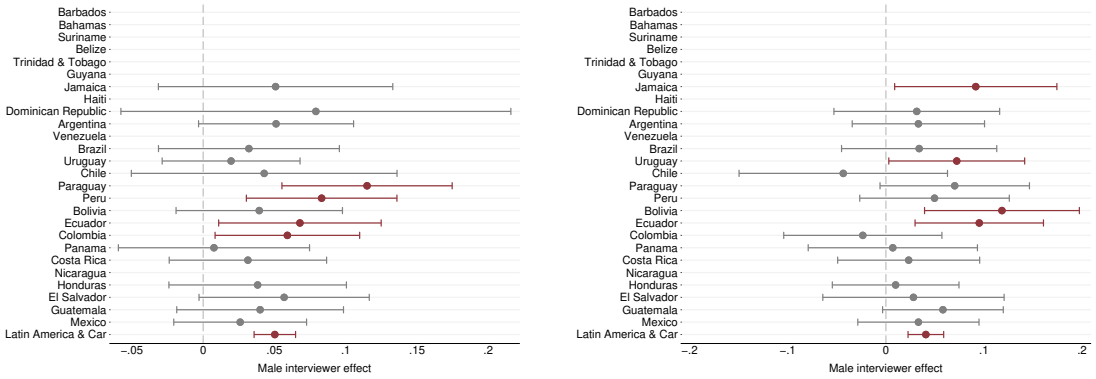
As we argued before, both participants of the interaction behave not only as respondent and interviewer but also as all the other roles they play in society based on who they are, who the other person is in society, and how they are expected to behave by others. The evidence so far

Figure 3: Male Interviewer Effects by Gender of Respondents: Country by Country

(I) Agreement: Men are better political leaders

(a) Female respondents

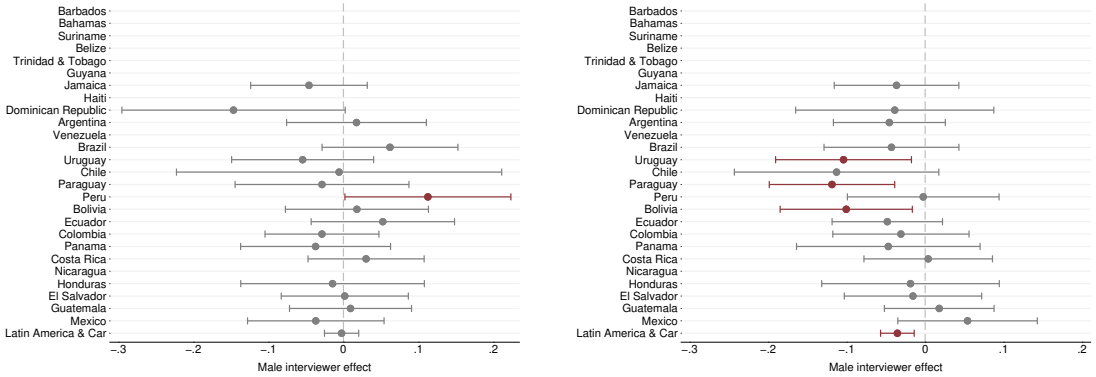
(b) Male respondents



(II) Agreement: Women would manage better the national economy

(c) Female respondents

(d) Male respondents



Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents and the presence of spouse during the interviews. Estimates are produced using probability weights and correct standard errors following the survey sample design. See the main text for a full description of the questions.

suggests that the gender of the interviewer matters and that it may matter differently for male and female respondents. This interaction may be further complicated by the presence of others during the interview, particularly the respondent's spouse. If present, the respondent may be reacting not only to what she/he thinks the interviewer expects but also to what she/he thinks her/his spouse expects.

In the model, the presence of the spouse (or other relevant individual) may be addressed by

adding a third term to the respondent's utility function. Taking $v^s \in \mathbb{R}$ as the spouse's opinion[12] and $\sigma \geq 0$ as the importance for the respondent to show themself in agreement with their spouse, the utility would take the form of

$$u(x) = -(x - v)^2 - \delta(x - \tilde{v}_\theta)^2 - \sigma(x - v^s)^2.$$

For the respondent, the optimal reported opinion now would also try to be close to the spouse's opinion and take the form of

$$x^* = v + \tilde{\gamma}\left(\tilde{v}_\theta - v\right) + \tilde{\gamma}^s\left(v^s - v\right), \text{ with } \tilde{\gamma} \equiv \frac{\delta}{1 + \delta + \sigma} \text{ and } \tilde{\gamma}^s \equiv \frac{\sigma}{1 + \delta + \sigma},$$

where it can be seen that the opinion of the interviewer loses relevance compared to the optimal response in absence of the spouse, given that $\tilde{\gamma} \leq \gamma$.

The expression for the male interviewer effect when the spouse is present would also change its form to

$$MIE = -\tilde{\gamma}\left(\Delta v + 2\varphi\right),$$

being smaller in magnitude than in the absence of the spouse, and even smaller as being in agreement with the spouse's opinion is more relevant to the respondent.

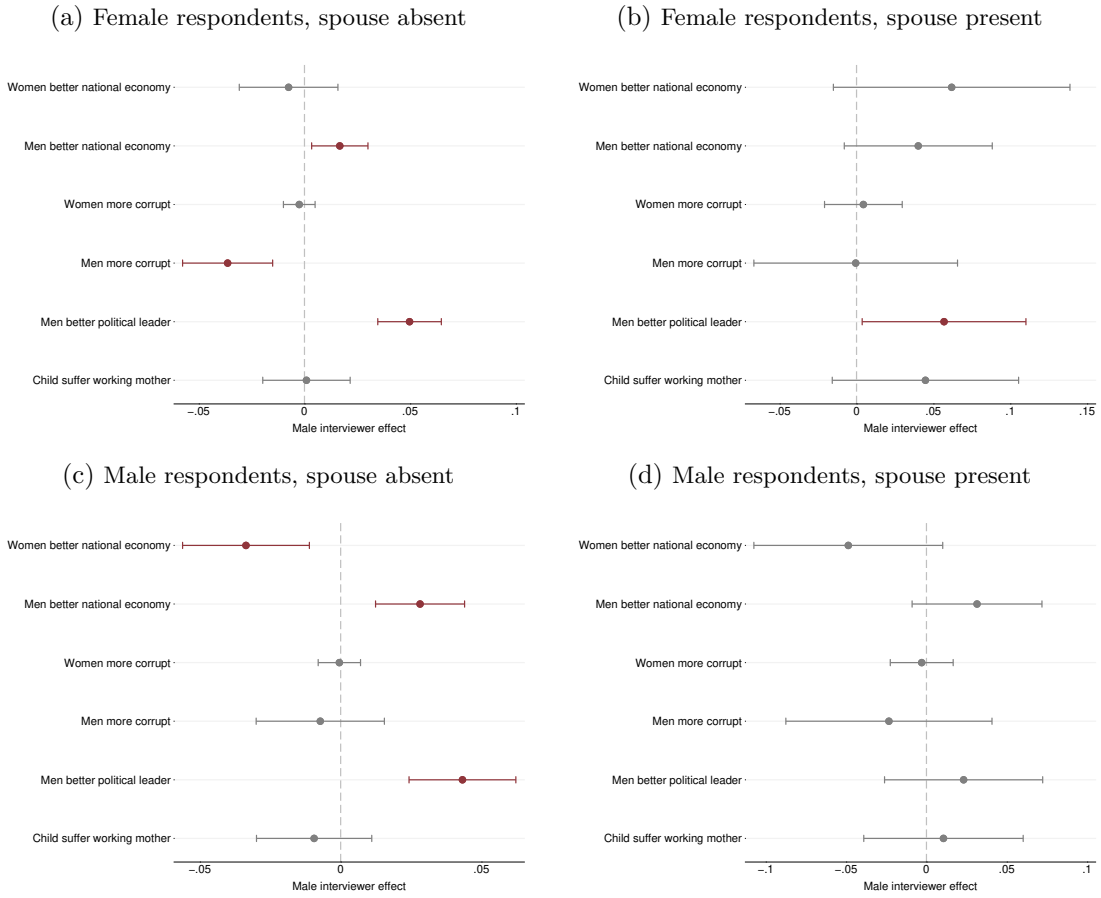
Again, the gender of the respondent may be relevant for this analysis. For example, female respondents may be more concerned with agreeing with their husbands' opinions than male respondents. This would shrink the gender-of-interviewer effect even more for female respondents.

4.2.2 Empirical Results: Presence of Spouse

The data support the relevance of taking the presence of the spouse into account, and they bring even more insights about this phenomenon, as there is a high correlation between the presence of the husband and the interviewer being a man for female respondents. This may indicate a territorial reaction on the part of some husbands that could have a strong influence in the response and even dominate the male interviewer effect. If that was the case, not controlling for the presence of the husband could bias the male interviewer effect estimates.

[12] We assume that the spouse's opinion v^s is known by the respondent, but it can be also a distorted perception. This could be addressed by adding a distortion to the actual value of v^s, but it would not change our main results.

Figure 4: Male interviewer Effects by Presence of Spouse during the Interview



Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents and presence of spouse during the interview. Estimates are produced using probability weights and correct standard errors following the survey sample design. See Section 4 of the main text for a full description of the questions.

The presence of the spouse during the interview was recorded in the 2018/19 wave of the Americas Barometer, which allows us to estimate separate effects for cases where the spouse is present and where he/she is absent. In fact, the presence of the spouse is a relevant factor in the sensitivity of the reported gender attitudes to the gender of the interviewer for both male and female respondents. The male interviewer effect dissipates when the respondent's spouse is present during the interview, as shown in Figure 4. If this means that the spouse presence dominates the gender-of-interviewer effect, and there is a correlation between spouse presence and interviewer gender, then omitting this variable as a control would bias the estimates of male interviewer effects.

It is worth noting, however, that only less than 10% of the interviews (about 2,390 observations) were conducted in the presence of the spouse, which can also affect the significance of the estimates.

4.3 Are Interviewers Randomly Assigned? Face-to-face versus Telephone Surveys

Two important concerns about the consistency of any estimates of gender-of-interviewer effects are that interviewer assignments are not always completely random and that the interviewer's gender may be correlated with other characteristics of the interviewer that play their own role in influencing responses. While the former can be tested and controlled during the design and implementation of a survey, the latter depends mostly on the supply of enumerators and the type of people who self-select into the role.

Anecdotal evidence suggests that there may exist some correlation between interviewer characteristics, for example, due to recent changes in the interviewer labor market. As it has been a traditionally female-dominated occupation, male enumerators tend to be younger and less experienced than female ones, among other potential differences. If there is a relevant effect of at least one of these characteristics on reported gender attitudes, failing to account for these correlations could confound estimates of gender-of-interviewer effects. However, as long as these characteristics do not play a significant role in influencing responses to gender-related questions, omitting them should not bias either the responses themselves or the estimates of the gender-of-interviewer effect. The problem is that this information is typically kept in-house and lost in the research process because it is deemed irrelevant to the research purpose.

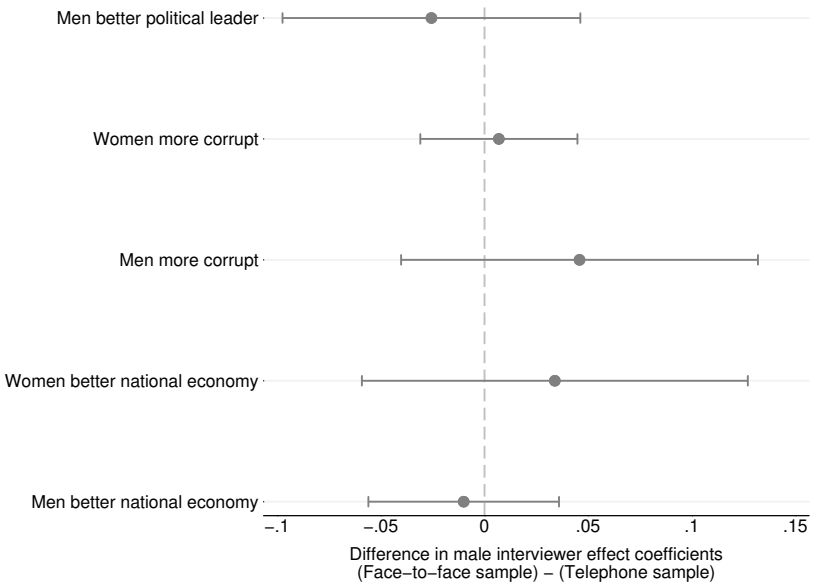
Regarding the dimension of exogeneity of interviewer gender with respect to respondent characteristics, the randomness of interviewer-interviewee matching is key, although it is not always the case. For example, it is common to try to minimize the risk of exposure of interviewers when they are sent into the field. Sometimes, this means assigning certain areas to certain enumerators according to the estimated risk of danger. In many countries around the world, populations living in rural or marginalized urban areas tend to be more conservative. Sending a female enumerator alone to these areas, moreover, may be riskier than sending a male enumerator. If this pattern is strong, it may lead to a correlation between the interviewer's gender and respondent characteristics that follow the lines of this area selection, such as age, education, wealth, religiosity, or family

structure.

Although the demographic characteristics of respondents appear to be generally balanced between those interviewed by male and female enumerators in the LAPOP surveys, as shown in Tables A1 through A6 in the Appendix, we use the 2021 edition of the survey for a brief comparison between a face-to-face and a telephone survey. The 2021 Mexico survey, conducted by telephone, includes the three questions from the 2018 survey that compare men and women in their political competence. Under the assumption that telephone interviews are less likely to have interviewer selection issues than face-to-face interviews[13], male interviewer effects should be less biased in the former. Figure 5 shows that the estimated effects of being interviewed by a male enumerator in the 2021 telephone survey are not statistically different from the same effects in the

[13]Balance tests for all demographic variables included in the 2021 Mexico survey are presented in Table A7 in the Appendix.

Figure 5: Male Interviewer Effects in Face-to-face versus Telephone Surveys in Mexico



Source: AmericasBarometer survey waves 2018/19 and 2021. Latin American Public Opinion Project (LAPOP), by the University of Vanderbilt.
Note: 95% confidence intervals. Point estimates represent the difference between the male interviewer effect estimated from the 2018 face-to-face survey and the effect estimated from the 2021 telephone survey. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for gender, age, education, and wealth of the respondents, and whether they live in rural or urban areas. See Table A7 in the Appendix for the coding of these variables in the 2021 wave of the survey. Estimates are produced using probability weights and correct standard errors following the survey sample design. See the main text for a full description of the questions.

2018 face-to-face survey.

4.4 Discussion: How Well Is Gender Opinion Measured?

4.4.1 Gender Opinion Measurements from LAPOP

A corollary to the existence of gender-of-interviewer effects is that the average attitude measured by a survey will be a function of the composition of the pool of interviewers, which varies significantly by country and over the years, as shown in Figure 6. As an example, Figure 7 shows what the levels of acceptance of domestic violence would be if all interviewers were male or female, in addition to the actual average level of acceptance (Figures A2 and A3 in the Appendix reproduce the figure with the full set of questions). The average opinion of each survey can also be seen as the weighted average of the two extremes according to the gender ratio of the interviewer pool in each country. In Guyana, the most extreme case, reported acceptance of domestic violence ranges from 31% to 36%, but could be as high as 59% if all interviewers were male or as low as 24% if all people were interviewed by female enumerators. In terms of the regional average rate of acceptance, it is estimated to be 22% for cases of a wife neglecting chores and 33% for cases of infidelity. However, these rates could go up to 27% and 39%, respectively, if all respondents were interviewed by men or go down to 19% and 29% if all responses were collected by female interviewers.

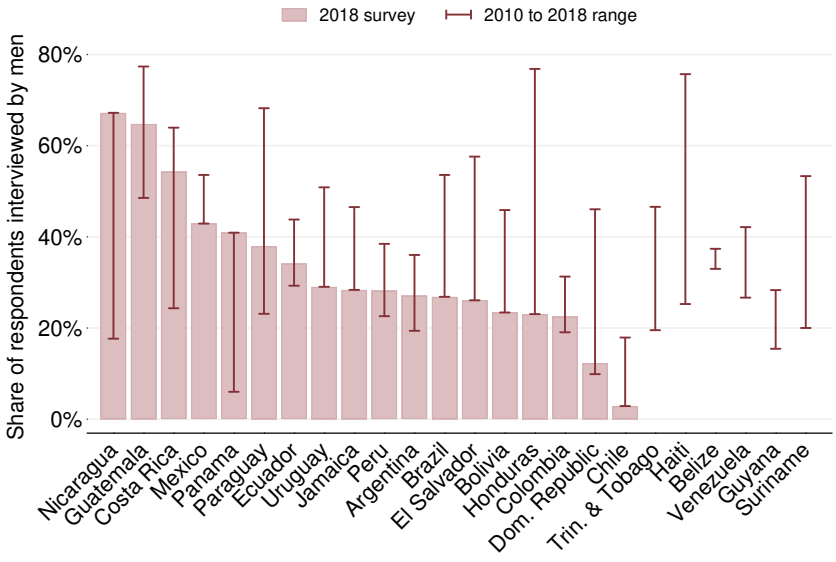
Again, as mentioned before, we cannot identify the "true" acceptance level from this figure. Instead, we can only conclude that survey results can be influenced by the composition of the pool of interviewers. This adds another layer of scrutiny to the survey results, as the people who design and conduct a survey may be manipulating the survey results, voluntarily or not, from very early in the process when they select and assign enumerators to each respondent.

The formal model presented here may help to understand the measurement problems introduced by the gender-of-interviewer effect in combination with the gender composition of the interviewer pool. It can also be used to arrive to an optimal adjustment to the survey responses and help correct the biases generated.

4.4.2 Using the Model to Correctly Estimate Gender Opinions

The individual responses presented before can be easily scaled into aggregate responses from a survey sample. The theoretical aggregate responses may inform the survey design or reweighting

Figure 6: Distribution of Responses by Gender of the Interviewer in the AmericasBarometer

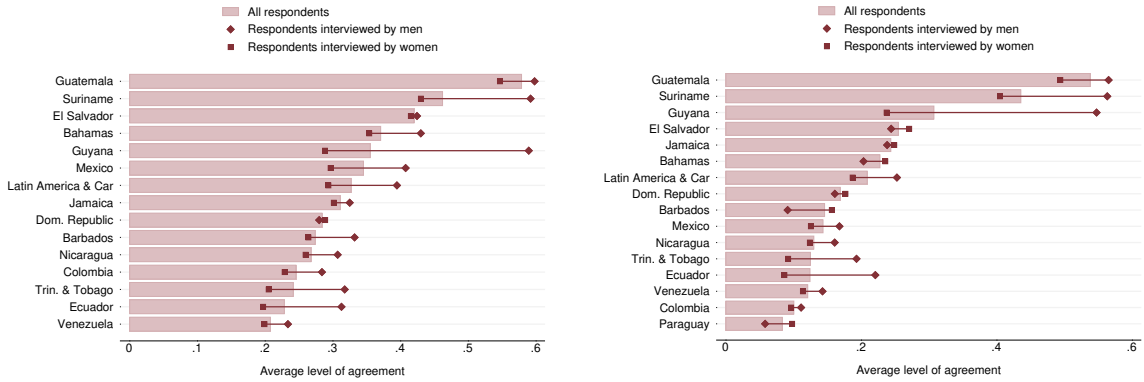


Source: AmericasBarometer survey waves 2010 to 2018/19. Latin American Public Opinion Project (LAPOP), by Vanderbilt University.
Note: Trinidad and Tobago, Haiti, Belize, Venezuela, Guyana, and Suriname were not included in the 2018/19 wave of the survey. The range presented for Haiti, Venezuela, and Guyana is 2010 to 2016 and 2010 to 2014 for Trinidad and Tobago, Belize, and Suriname.

Figure 7: Acceptance of Domestic Violence by Interviewer Gender and Country

(a) Hit wife if she is unfaithful

(b) Hit wife if she neglects the chores



Note: Estimates are produced with probability weights following the survey design. See main text for a full description of the questions.

scheme in order to cancel out individual distortions in the average response. Under the assumption of random assignment of interviewers to interviewees, such that the distribution of v is the same,

both unconditional and conditional on gender, for the targeted population, the entire sample, and the samples interviewed by enumerators of each gender, the average response for female ($\bar{x}_f$) and male ($\bar{x}_m$) interviewees will be given by

$$\bar{x}_f = \bar{v}_f + \gamma \left[\varphi \left(\frac{n_f^f}{n_f} - \frac{n_f^m}{n_f} \right) - \Delta v \frac{n_f^m}{n_f} \right] \quad \text{and} \quad \bar{x}_m = \bar{v}_m + \gamma \left[\varphi \left(\frac{n_m^f}{n_m} - \frac{n_m^m}{n_m} \right) + \Delta v \frac{n_m^f}{n_m} \right]. \quad (2)$$

Under the same assumptions, the overall average response of the survey ($\bar{x}$) will be given by

$$\bar{x} = \left(\frac{n_f}{n} \bar{v}_f + \frac{n_m}{n} \bar{v}_m \right) + \gamma \left[\varphi \left(\frac{n^f}{n} - \frac{n^m}{n} \right) + \Delta v \left(\frac{n_m^f}{n} - \frac{n_f^m}{n} \right) \right], \quad (3)$$

where n_μ^θ is the number of respondents of gender μ interviewed by someone of gender θ, n_μ is the total number of respondents of gender μ, n^θ is the total number of respondents interviewed by gender θ, and n is the total number of people interviewed in the survey.

In this setup, to accurately capture the true average values of v among male, female, and overall pool of respondents using the survey, the researcher could plan the allocation of responses among interviewers of each gender by interviewees' gender. Alternatively, the researcher may reweight responses using the optimal structure post-survey to adjust the estimates. According to equations 2 and 3, the pool of interviewers and interviewees can be stratified to cancel out the biases introduced by both the genuine gender-based differences in opinions (Δv) and the potential distortion in the perception of average opinions (φ). The resulting average responses will be unbiased, meaning that $\bar{x}_f = \bar{v}_f$, $\bar{x}_m = \bar{v}_m$, and $\bar{x} = \bar{v}$, if and only if all of the following conditions are met:

$$\frac{n_m}{n_f} = \frac{N_m}{N_f} \equiv \omega \qquad n_m^m - n_f^f = \left(n_m^f - n_f^m \right) \left(1 + \frac{\Delta v}{\varphi} \right) \quad (4)$$

$$n_f^f = n_f^m \left(1 + \frac{\Delta v}{\varphi} \right) \qquad n_m^m = n_m^f \left(1 + \frac{\Delta v}{\varphi} \right) \quad (5)$$

In the first condition in equation 4, it is necessary that the gender composition of respondents mirrors that of the target population, which is consists of N_m men and N_f women, being their ratio defined as $\omega \equiv \frac{N_m}{N_f}$. This equality guarantees that the average reported opinion in the survey net of distortions, expressed as $\frac{n_f}{n} \bar{v}_f + \frac{n_m}{n} \bar{v}_m$, accurately reflects the true average opinion in the population, $\bar{v}$. Concurrently, the second condition in equation 4 is necessary to cancel out the biases

introduced by Δv and φ in equation 3. Together, these two conditions ensure that the survey's average response serves as an unbiased estimate of the population's average opinion. Furthermore, if both conditions in equation 5 are also met, the average response within each gender group will also be an unbiased estimate of the true average opinion conditional on gender. It is important to note that, while conditions in equation 5 are sufficient for fulfilling the second condition in equation 4, these are not necessary conditions for achieving the equality $\bar{x} = \bar{v}$.

Considering all the aforementioned conditions, the exact number of responses (or weightings) required from each interviewer's and interviewee's gender can be determined such that responses are unbiased measures of real values. These optimal amounts can be expressed in function of the parameters and the survey sample size as follows:

$$n_f^m = n\left(\frac{1}{1+\omega}\right)\left(\frac{1}{2+\frac{\Delta v}{\varphi}}\right) \qquad n_f^f = n\left(\frac{1}{1+\omega}\right)\left(\frac{1+\frac{\Delta v}{\varphi}}{2+\frac{\Delta v}{\varphi}}\right) \tag{6}$$

$$n_m^f = n\left(\frac{\omega}{1+\omega}\right)\left(\frac{1}{2+\frac{\Delta v}{\varphi}}\right) \qquad n_m^m = n\left(\frac{\omega}{1+\omega}\right)\left(\frac{1+\frac{\Delta v}{\varphi}}{2+\frac{\Delta v}{\varphi}}\right) \tag{7}$$

Another interpretation for these expressions is that they represent the optimal proportions for reweighting observations in order to correct biases in the average responses. It is noteworthy that, in scenarios where a distortion in the perceived distribution of values exists (denoted by $\varphi > 0$), mixing genders between respondents and interviewers is the optimal strategy, rather than matching them. This finding holds true not only for gender but also when the types of respondents and interviewers are other significant demographic dimensions that may influence the measurement process, such as race, ethnicity, religion, social class, or age.

4.5 Proposing an Estimator for γ

The ultimate goal of quantifying the male interviewer effect is, however, to demonstrate that people's opinions are influenced by (their perception of) interviewers' expectations, and that this could lead to biases in the average reported opinions. In the model, demonstrating that $\gamma > 0$ is tantamount to showing that there exist a difference in responses by gender of the interviewer. However, the model allows the researcher to go beyond the simple demonstration of differences by estimating the value of γ through the structure of optimal responses. If γ is allowed to vary according to the gender of both interviewer and interviewee, and φ to differ across the gender

distributions, the value of γ can be estimated using the following equation:[14]

$$\hat{\gamma}_{\mu}^{\theta} = \frac{\left(\mathbb{I}_{(\theta=f)} - \mathbb{I}_{(\theta=m)}\right)\left(\bar{x}_{\mu}^{\theta} - \bar{x}_{\mu}\right)}{\mathbb{I}_{(\theta\neq\mu)}\left(\bar{x}_{f} - \bar{x}_{m}\right) + \hat{\varphi}_{\theta}}, \tag{8}$$

where $\hat{\gamma}_{\mu}^{\theta}$ is the estimated sensitivity to the perceived interviewer's opinion, with θ and μ denoting the genders of the interviewer and interviewee, respectively, $\mathbb{I}$ is an indicator function, $\bar{x}_{\mu}^{\theta}$ is the average response of interviewees of gender μ interviewed by someone of gender θ, and $\hat{\varphi}$ is the estimate for φ. For practical estimation of φ, it is imperative to incorporate an additional question to the survey regarding the respondent's perception of the average response to the same question. While the LAPOP does not incorporate such questions, several surveys, such as the one in Bursztyn et al. (2023), do include these questions and allow for the estimation of φ_{θ} as the discrepancy between the reported belief of the average opinion ($\bar{\tilde{v}}_{\theta}$) and the average reported opinion ($\bar{x}_{\theta}$) for each gender θ.

While further research should be done in order to accurately estimate γ and test the various hypotheses on the dynamics that generate gender-of-interviewer effects, the model provides valuable insights into why and how people may alter their answers in response to the interviewer's characteristics and into how these distortions can impact the average estimates in public opinion measurement. In line with the empirical evidence presented above, these theoretical findings advocate for strategies that take into consideration the interviewers' characteristics when designing and executing surveys, as well as in the subsequent data analysis, which is further elaborated in the final recommendations.

5 Conclusions and Recommendations

It has been well documented that interviewer characteristics can significantly impact the data collection process in surveys that involve human interactions, particularly when obtaining information about gender norms and attitudes toward gender equality. As highlighted in the literature and demonstrated in this paper, the interviewer's gender is especially pertinent. Intriguingly, our

[14] Allowing γ and φ to vary according to the gender of interviewer and interviewee would change the optimal proportions n_{μ}^{θ} to $n_{f}^{f} = n\left(\frac{1}{1+\omega}\right)\left(\frac{\gamma_{f}^{f}\varphi_{f}}{\gamma_{f}^{f}\varphi_{f}+\gamma_{f}^{m}(\Delta v+\varphi_{m})}\right)$, $n_{f}^{m} = n\left(\frac{1}{1+\omega}\right)\left(\frac{\gamma_{f}^{m}(\Delta v+\varphi_{m})}{\gamma_{f}^{f}\varphi_{f}+\gamma_{f}^{m}(\Delta v+\varphi_{m})}\right)$, $n_{m}^{m} = n\left(\frac{\omega}{1+\omega}\right)\left(\frac{\gamma_{m}^{m}\varphi_{m}}{\gamma_{m}^{m}\varphi_{m}+\gamma_{m}^{f}(\Delta v+\varphi_{f})}\right)$, and $n_{m}^{f} = n\left(\frac{\omega}{1+\omega}\right)\left(\frac{\gamma_{m}^{f}(\Delta v+\varphi_{f})}{\gamma_{m}^{m}\varphi_{m}+\gamma_{m}^{f}(\Delta v+\varphi_{f})}\right)$.

findings suggest that the direction of this effect is not always straightforward and might depend on subtle nuances in the social norms of a particular community. Therefore, in order to prevent these effects from contaminating the evidence and bias estimation results, researchers can adopt several strategies during the analysis of pre-existing data, coding of new data, and the design of the data collection process.

When using already collected survey data, researchers should first consider the potential for interviewer effects on the variables of interest. For example, in gender studies, if the variable of interest is behavioral or attitudinal, the interviewer's gender is likely to be relevant. If this is the case, they should ascertain the technology used for data collection and whether it involved any human interaction. If information about the interviewers' characteristics is available, they should verify if the interviewers and respondents were randomly matched. Any pattern in the characteristics of respondents and interviewers, such as a significantly higher number of respondents interviewed by men in rural areas than in urban areas, could create an identification problem and obstruct the resolution of response effects. In cases of random assignment of interviewers to respondents, the researcher could opt to reweight the observations to balance the characteristics of the interviewers (based on census data or what they consider to be a fair representation of reality) and to include these characteristics as control variables in multivariate analyses. However, it is crucial to note that reweighting observations when the interviewers were not randomly assigned could induce further biases. Also, when the interviewers' characteristics are not relevant, reweighting can introduce unwanted random noise into the estimations, so the relevance of those characteristics should be confirmed.

For researchers coding new data or designing the data collection process, the same considerations should be made but from a backward induction perspective. If the interviewers' characteristics can significantly influence what the questionnaire intends to measure, then planning the pool of interviewers in advance might be beneficial. In those cases, it may be helpful to add questions on normative expectations to the survey in order to perfect the weightings afterward. Whenever possible, coding every relevant characteristic of the interviewers can also be useful for subsequent quality checks of the data and the consistency of the generated estimates. When feasible, it might also be advisable to opt for technologies that do not involve human interactions, like web surveys, either as the primary method or as a benchmark to compare with face-to-face collected data

(Nederhof, 1985; Kreuter et al., 2008; Langhaug et al., 2010; Oates et al., 2022). This can even be combined with an additional reweighting of respondents to make them representative of the overall population (Grewenig et al., 2023).

In empirical research, trade-offs are an inevitable part of every stage, from selecting data collection methodologies to choosing interviewers, assigning respondents, and analyzing data. Each decision comes with its own set of advantages and disadvantages, necessitating a careful balance to ensure the integrity and consistency of the findings. The paramount goal is to ensure that reported results authentically reflect the true preferences and social norms of the study population. This is crucial, as reporting inaccuracies can significantly influence public behavior and policymaking, with far-reaching implications for societal welfare.

References

Anglewicz, P., Adams, J., Obare, F., Kohler, H.-P., and Watkins, S. (2009). The malawi diffusion and ideational change project 2004-06: Data collection, data quality, and analysis of attrition. *Demographic Research*, 20:503–540.

Becker, S., Feyisetan, K., and Makinwa-Adebusoye, P. (1995). The effect of the sex of interviewers on the quality of data in a nigerian family planning questionnaire. *Studies in Family Planning*, 26(4):233.

Benstead, L. J. (2014a). Does interviewer religious dress affect survey responses? evidence from morocco. *Politics and Religion*, 7(4):734–760.

Benstead, L. J. (2014b). Effects of interviewer–respondent gender interaction on attitudes toward women and politics: Findings from morocco. *International Journal of Public Opinion Research*, 26(3):369–383.

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press.

Bignami-Van Assche, S., Reniers, G., and Weinreb, A. A. (2003). An assessment of the KDICP and MDICP data quality: Interviewer effects, question reliability and sample attrition. *Demographic Research*, Special 1:31–76.

Blaydes, L. and Gillum, R. M. (2013). Religiosity-of-interviewer effects: Assessing the impact of veiled enumerators on survey response in egypt. *Politics and Religion*, 6(3):459–482.

Bonomi, G., Gennaioli, N., and Tabellini, G. (2021). Identity, beliefs, and political conflict. *Quarterly Journal of Economics*, 136(4):2371–2411.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *Quarterly Journal of Economics*, 131(4):1753–1794.

Bursztyn, L., Cappelen, A. W., Tungodden, B., Voena, A., and Yanagizawa-Drott, D. H. (2023). How are gender norms perceived? Working Paper 31049, National Bureau of Economic Research.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*, 110(10):2997–3029.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., and Hauck, W. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, 60(3):345.

Cotter, P. R., Cohen, J., and Coulter, P. B. (1982). Race-of-interviewer effects in telephone interviews. *Public Opinion Quarterly*, 46(2):278.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., and Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, 25(1):14–26.

Dijkstra, W. (1987). Interviewing style and respondent behavior: An experimental study of the survey-interview. *Sociological Methods & Research*, 16(2):309–334.

Finkel, S. E., Guterbock, T. M., and Borg, M. J. (1991). Race-of-interviewer effects in a preelection poll: Virginia 1989. *Public Opinion Quarterly*, 55(3):313.

Flores-Macias, F. and Lawson, C. (2008). Effects of interviewer gender on survey responses: Findings from a household survey in mexico. *International Journal of Public Opinion Research*, 20(1):100–110.

Fuchs, M. (2009). Gender-of-interviewer effects in a video-enhanced web survey: Results from a randomized field experiment. *Social Psychology*, 40(1):37–42.

Grewenig, E., Lergetporer, P., Simon, L., Werner, K., and Woessmann, L. (2023). Can internet surveys represent the entire population? a practitioners' analysis. *European Journal of Political Economy*, 78:102382.

Groves, R. M. and Fultz, N. H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods & Research*, 14(1):31–52.

Heeb, J. L. and Gmel, G. (2001). Interviewers' and respondents' effects on self-reported alcohol consumption in a swiss health survey. *Journal of Studies on Alcohol*, 62(4):434–442.

Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P. J., and Pugliani, P. (1997). The effect of interviewer gender on the survey response. *Political Behavior*, 19(3):197–220.

Kane, E. W. and Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly*, 57(1):1.

Katz, D. (1942). Do interviewers bias poll results? *Public Opinion Quarterly*, 6(2):248.

Kianersi, S., Luetke, M., Jules, R., and Rosenberg, M. (2020). The association between interviewer gender and responses to sensitive survey questions in a sample of haitian women. *International Journal of Social Research Methodology*, 23(2):229–239.

Kreuter, F., Presser, S., and Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5):847–865.

Landis, J. R., Sullivan, D., and Sheley, J. (1973). Feminist attitudes as related to sex of the interviewer. *The Pacific Sociological Review*, 16(3):305–314.

Langhaug, L. F., Sherr, L., and Cowan, F. M. (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.

Lipps, O. and Lutz, G. (2017). Gender of interviewer effects in a multi-topic centralized CATI panel survey. *methods*, data:20 Pages. Artwork Size: 20 Pages Publisher: methods, data, analyses.

Liu, M. and Stainback, K. (2013). Interviewer gender effects on survey responses to marriage-related questions. *Public Opinion Quarterly*, 77(2):606–618.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3):263–280.

Oates, M., Crichton, K., Cranor, L., Budwig, S., Weston, E. J. L., Bernagozzi, B. M., and Pagaduan, J. (2022). Audio, video, chat, email, or survey: How much does online interview mode matter? *PLOS ONE*, 17(2):e0263876.

Padfield, M. and Procter, I. (1996). The effect of interviewer's gender on the interviewing process: A comparative enquiry. *Sociology*, 30(2):355–366.

Reese, S. D., Danielson, W. A., Shoemaker, P. J., Chang, T.-K., and Hsu, H.-L. (1986). Ethnicity-of-interviewer effects among mexican-americans and anglos. *Public Opinion Quarterly*, 50(4):563.

Schuman, H. and Converse, J. M. (1971). The effects of black and white interviewers on black responses in 1968. *Public Opinion Quarterly*, 35(1):44.

Silber, H., Roßmann, J., Gummer, T., Zins, S., and Weyandt, K. W. (2021). The effects of question, respondent and interviewer characteristics on two types of item nonresponse. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(3):1052–1069.

Sudman, S. and Bradburn, N. M. (1974). *Response effects in surveys: a review and synthesis*. Aldine Pub. Co.

Sundström, A. and Stockemer, D. (2022). Measuring support for women's political leadership. *Public Opinion Quarterly*, 86(3):668–696.

Weeks, M. F. and Moore, R. P. (1981). Ethnicity-of-interviewer effects on ethnic respondents. *Public Opinion Quarterly*, 45(2):245.

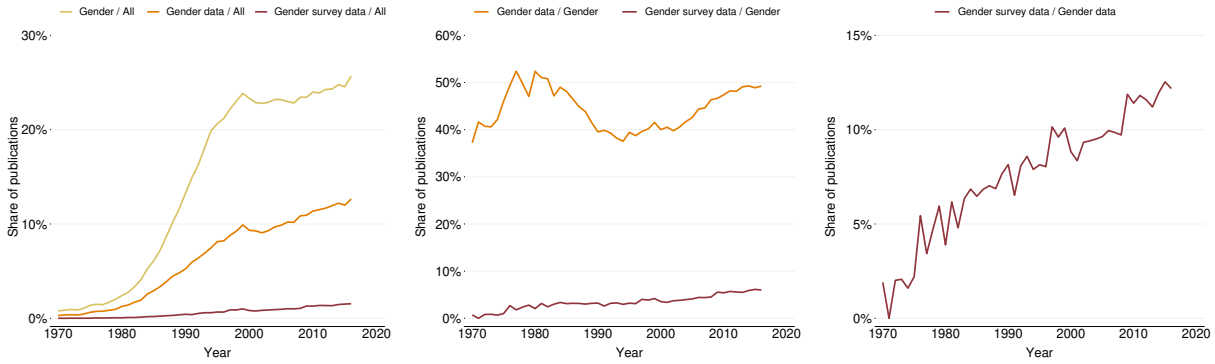
Weinreb, A. A. (2006). The limitations of stranger-interviewers in rural kenya. *American Sociological Review*, 71(6):1014–1039.

Zoch, G. (2021). Gender-of-interviewer effects in self-reported gender ideologies: Evidence based on interviewer change in a panel survey. *International Journal of Public Opinion Research*, 33(3):626–636.

A Appendix

A.1 Gender Research and Use of Survey Data in Social Studies

Figure A1: Gender Research and Use of Survey Data in Social Studies



Note: Data were generated using Constellate, a search engine developed by JSTOR Labs. Ratios are estimated using the total number of articles in selected categories of the JSTOR collection that result from queries with different sets of keywords. The filtered categories are Arts, Business, Economics, Education, History, Law, Linguistics, Philosophy, Political Science, Religion, and Social Sciences. The keywords for each set of articles are (1) none [all articles in the selected categories], (2) "gender," (3) "gender AND data," (4) "gender AND survey data."

A.2 Balance of Demographic Variables

As mentioned in the main text, to correctly identify gender-of-interviewer effects, the pool of interviewers must not be correlated with any other variable that can affect attitudes of the respondents in the questions of interest. We test for balance in the observable demographic variables that are included as control variables in our estimations of gender-of-interviewer effects. Tables A1 through A6 present the average value of each control variable by sex of the interviewer and the difference between these two values for each country in the sample for the 2018/19 edition of the survey. Table A7 presents differences in the demographic variables included in the 2021 Mexico survey by gender of the interviewer. These estimates take into account the survey sample design as they incorporate specific country weights for the country estimates and regional weights for the regional estimates, both provided by LAPOP, as well as correct standard errors following the sample stratification structure.

Table A1: Balance in Share of Female Respondents by Sex of the Interviewer

Country	Female interviewer	Male interviewer	Difference
Mexico	0.52	0.49	-0.03
	(0.02)	(0.02)	(0.02)
Guatemala	0.55	0.50	-0.05*
	(0.02)	(0.02)	(0.03)
El Salvador	0.50	0.51	0.01
	(0.02)	(0.03)	(0.03)
Honduras	0.51	0.49	-0.02
	(0.01)	(0.03)	(0.04)
Nicaragua	0.53	0.49	-0.04**
	(0.02)	(0.02)	(0.02)
Costa Rica	0.51	0.49	-0.01
	(0.02)	(0.02)	(0.02)
Panama	0.54	0.49	-0.05***
	(0.02)	(0.02)	(0.01)
Colombia	0.50	0.54	0.04
	(0.01)	(0.03)	(0.03)
Ecuador	0.50	0.49	-0.01
	(0.02)	(0.02)	(0.01)
Bolivia	0.50	0.49	-0.00
	(0.01)	(0.03)	(0.01)
Peru	0.51	0.49	-0.02***
	(0.02)	(0.02)	(0.01)
Paraguay	0.54	0.43	-0.11***
	(0.02)	(0.02)	(0.03)
Chile	0.51	0.57	0.06
	(0.01)	(0.07)	(0.07)
Uruguay	0.54	0.52	-0.02
	(0.02)	(0.02)	(0.02)
Brazil	0.52	0.49	-0.03**
	(0.02)	(0.02)	(0.01)
Argentina	0.52	0.49	-0.03*
	(0.02)	(0.02)	(0.01)
Dom. Republic	0.51	0.34	-0.17***
	(0.01)	(0.04)	(0.04)
Jamaica	0.53	0.50	-0.02*
	(0.02)	(0.03)	(0.01)

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Table A2: Balance in Age of Respondents by Sex of the Interviewer

Country	Female interviewer	Male interviewer	Difference
Mexico	41.86	42.30	0.44
	(0.56)	(0.67)	(0.78)
Guatemala	39.89	36.87	-3.03***
	(0.67)	(0.47)	(0.63)
El Salvador	39.63	41.32	1.69*
	(0.50)	(0.86)	(0.88)
Honduras	38.25	37.77	-0.48
	(0.47)	(0.86)	(0.77)
Nicaragua	34.32	35.05	0.73
	(0.66)	(0.48)	(0.68)
Costa Rica	40.37	40.59	0.22
	(0.65)	(0.59)	(0.95)
Panama	39.68	39.56	-0.12
	(0.54)	(0.67)	(0.45)
Colombia	40.70	39.34	-1.36
	(0.46)	(0.89)	(0.83)
Ecuador	38.26	37.92	-0.34
	(0.55)	(0.74)	(0.52)
Bolivia	39.67	39.14	-0.52
	(0.46)	(0.82)	(0.55)
Peru	38.56	39.46	0.90*
	(0.47)	(0.78)	(0.50)
Paraguay	40.02	39.88	-0.15
	(0.53)	(0.69)	(0.77)
Chile	42.07	43.41	1.34
	(0.42)	(2.51)	(1.72)
Uruguay	45.81	46.93	1.12
	(0.55)	(0.88)	(0.69)
Brazil	39.24	38.83	-0.89
	(0.51)	(0.77)	(0.55)
Argentina	41.97	41.25	-0.72
	(0.54)	(0.85)	(0.68)
Dom. Republic	39.94	40.35	0.41
	(0.47)	(1.35)	(1.11)
Jamaica	40.40	39.77	-0.63
	(0.53)	(0.86)	(0.74)

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Table A3: Balance in Years of Education by Sex of the Interviewer

Country	Female interviewer	Male interviewer	Difference
Mexico	9.79	9.92	0.13
	(0.15)	(0.17)	(0.24)
Guatemala	8.15	8.08	-0.07
	(0.20)	(0.14)	(0.31)
El Salvador	9.11	8.78	-0.33
	(0.14)	(0.26)	(0.29)
Honduras	7.38	8.15	0.77***
	(0.12)	(0.24)	(0.24)
Nicaragua	8.65	8.76	0.11
	(0.20)	(0.14)	(0.32)
Costa Rica	9.61	9.46	-0.15
	(0.16)	(0.15)	(0.29)
Panama	10.81	10.46	-0.35
	(0.12)	(0.17)	(0.32)
Colombia	9.82	10.21	0.40
	(0.12)	(0.23)	(0.35)
Ecuador	11.41	11.65	0.24
	(0.12)	(0.18)	(0.24)
Bolivia	11.14	11.54	0.40
	(0.13)	(0.23)	(0.34)
Peru	11.46	12.25	0.79**
	(0.12)	(0.18)	(0.33)
Paraguay	9.59	9.73	0.14
	(0.15)	(0.19)	(0.24)
Chile	11.51	12.63	1.12***
	(0.09)	(0.43)	(0.40)
Uruguay	9.46	10.11	0.65
	(0.13)	(0.20)	(0.42)
Brazil	8.89	8.96	0.09
	(0.12)	(0.20)	(0.37)
Argentina	11.00	11.50	0.51
	(0.11)	(0.19)	(0.39)
Dom. Republic	9.70	9.61	-0.09
	(0.12)	(0.31)	(0.29)
Jamaica	10.34	10.21	-0.12
	(0.09)	(0.13)	(0.20)

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Table A4: Balance in Wealth Index by Sex of the Interviewer

Country	Female interviewer	Male interviewer	Difference
Mexico	-0.06	0.07	0.13**
	(0.03)	(0.04)	(0.06)
Guatemala	0.05	-0.02	-0.07
	(0.04)	(0.03)	(0.06)
El Salvador	0.00	-0.01	-0.01
	(0.03)	(0.05)	(0.06)
Honduras	-0.03	0.08	0.11*
	(0.03)	(0.05)	(0.07)
Nicaragua	-0.09	0.03	0.12
	(0.04)	(0.03)	(0.08)
Costa Rica	0.01	-0.00	-0.01
	(0.03)	(0.03)	(0.07)
Panama	-0.05	0.07	0.13
	(0.03)	(0.04)	(0.12)
Colombia	-0.02	0.09	0.10
	(0.03)	(0.05)	(0.08)
Ecuador	0.00	-0.02	-0.02
	(0.03)	(0.04)	(0.08)
Bolivia	-0.01	0.02	0.03
	(0.03)	(0.05)	(0.07)
Peru	-0.04	0.11	0.15*
	(0.03)	(0.05)	(0.09)
Paraguay	-0.00	-0.00	0.00
	(0.03)	(0.04)	(0.05)
Chile	-0.00	0.07	0.08
	(0.02)	(0.12)	(0.13)
Uruguay	-0.07	0.13	0.20**
	(0.03)	(0.04)	(0.08)
Brazil	0.01	-0.01	0.01
	(0.03)	(0.05)	(0.12)
Argentina	-0.02	0.05	0.07
	(0.03)	(0.04)	(0.09)
Dom. Republic	-0.01	0.06	0.07
	(0.03)	(0.08)	(0.08)
Jamaica	0.06	-0.17	-0.23*
	(0.03)	(0.05)	(0.12)

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Table A5: Balance in Share of Rural Respondents by Sex of the Interviewer

Country	Female interviewer	Male interviewer	Difference
Mexico	0.22	0.18	-0.04
	(0.01)	(0.01)	(0.04)
Guatemala	0.45	0.53	0.07**
	(0.02)	(0.02)	(0.03)
El Salvador	0.37	0.39	0.02
	(0.01)	(0.02)	(0.03)
Honduras	0.49	0.41	-0.08*
	(0.01)	(0.03)	(0.04)
Nicaragua	0.34	0.39	0.05
	(0.02)	(0.02)	(0.05)
Costa Rica	0.33	0.40	0.07
	(0.02)	(0.02)	(0.09)
Panama	0.11	0.60	0.49***
	(0.01)	(0.02)	(0.07)
Colombia	0.20	0.22	0.02
	(0.01)	(0.02)	(0.05)
Ecuador	0.31	0.39	0.08
	(0.01)	(0.02)	(0.07)
Bolivia	0.28	0.39	0.11
	(0.01)	(0.02)	(0.09)
Peru	0.28	0.13	-0.15**
	(0.01)	(0.02)	(0.07)
Paraguay	0.44	0.43	-0.01
	(0.02)	(0.02)	(0.03)
Chile	0.13	0.13	-0.00
	(0.01)	(0.05)	(0.08)
Uruguay	0.06	0.08	0.02
	(0.01)	(0.01)	(0.05)
Brazil	0.14	0.14	-0.01
	(0.01)	(0.02)	(0.07)
Argentina	0.11	0.17	0.06
	(0.01)	(0.02)	(0.06)
Dom. Republic	0.26	0.18	-0.08*
	(0.01)	(0.03)	(0.05)
Jamaica	0.42	0.35	-0.07
	(0.02)	(0.02)	(0.08)

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Table A6: Balance in Importance of Religion by Sex of the Interviewer

Country	Female interviewer	Male interviewer	Difference
Mexico	3.25	3.22	-0.03
	(0.03)	(0.04)	(0.05)
Guatemala	3.73	3.74	0.00
	(0.03)	(0.02)	(0.04)
El Salvador	3.71	3.72	0.01
	(0.02)	(0.03)	(0.04)
Honduras	3.75	3.82	0.08**
	(0.02)	(0.03)	(0.03)
Nicaragua	3.68	3.74	0.06
	(0.03)	(0.02)	(0.04)
Costa Rica	3.52	3.47	-0.05
	(0.03)	(0.03)	(0.04)
Panama	3.69	3.66	-0.03
	(0.02)	(0.03)	(0.04)
Colombia	3.53	3.52	-0.02
	(0.02)	(0.04)	(0.05)
Ecuador	3.49	3.57	0.08
	(0.03)	(0.03)	(0.05)
Bolivia	3.62	3.47	-0.15**
	(0.02)	(0.04)	(0.07)
Peru	3.49	3.41	-0.08
	(0.02)	(0.04)	(0.05)
Paraguay	3.65	3.59	-0.06
	(0.02)	(0.03)	(0.04)
Chile	2.85	2.80	-0.04
	(0.03)	(0.17)	(0.18)
Uruguay	2.46	2.38	-0.08
	(0.04)	(0.06)	(0.09)
Brazil	3.67	3.60	-0.10
	(0.02)	(0.04)	(0.07)
Argentina	2.90	2.80	-0.10
	(0.04)	(0.06)	(0.10)
Dom. Republic	3.70	3.74	0.04
	(0.02)	(0.05)	(0.06)
Jamaica	3.61	3.73	0.12**
	(0.03)	(0.03)	(0.04)

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

Table A7: Balance in Demographic Variables by Interviewer Gender in Mexico 2021 Survey

Variable	Female interviewer	Male interviewer	Difference
Female	0.47	0.51	0.01
	(0.01)	(0.02)	(0.03)
Age	41.84	40.01	-1.58
	(0.32)	(0.62)	(0.90)
Wealth index	-0.01	0.05	0.07
	(0.02)	(0.04)	(0.06)
Educational attainment			
None	0.02	0.02	-0.00
	(0.00)	(0.01)	(0.01)
Primary	0.14	0.14	-0.01
	(0.01)	(0.01)	(0.03)
Secondary	0.48	0.45	-0.02
	(0.01)	(0.02)	(0.03)
Tertiary or University	0.36	0.40	0.03
	(0.01)	(0.02)	(0.02)
Rural or urban area			
City	0.54	0.51	-0.02
	(0.01)	(0.02)	(0.03)
Surroundings of a city	0.12	0.11	-0.01
	(0.01)	(0.01)	(0.02)
Town near rural area	0.20	0.24	0.05
	(0.01)	(0.02)	(0.02)
Rural area	0.14	0.14	-0.03
	(0.01)	(0.01)	(0.02)

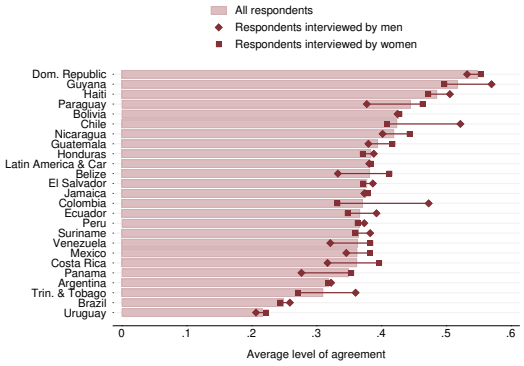
Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimates are produced using probability weights and correct standard errors following the survey sample design.

A.3 Differences in Gender Attitudes by Sex of IInterviewer

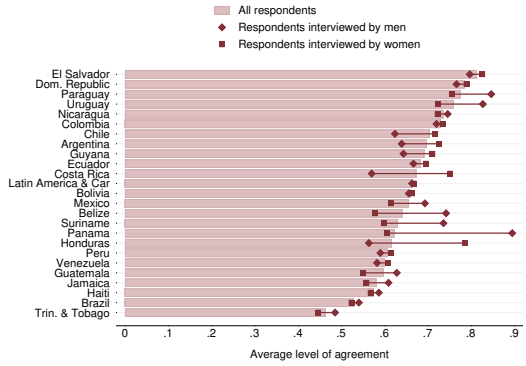
Figures A2 and A3 present the average level of agreement with each gender-related statement by country and by sex of the interviewer. Although not all differences are significant, these figures depict how different the country rankings could be if their pools of interviewers were composed of different shares of male and female interviewers. In some cases, the unbalanced composition of the pool of interviewers may distort the real average attitudes towards gender roles and inequality. Figures A4 through A7 show how significant these differences are when controlling for demographic variables.

Figure A2: Difference in Agreement by Sex of the Interviewer, Question by Question, Part 1

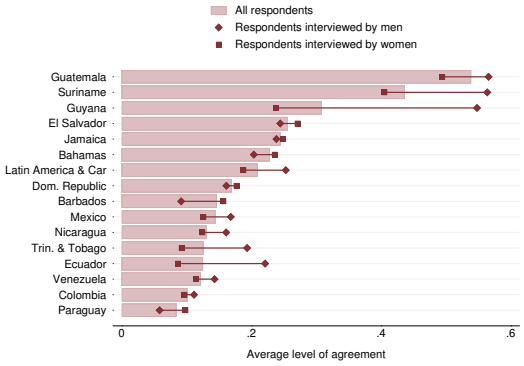
(a) Men should have more right to work



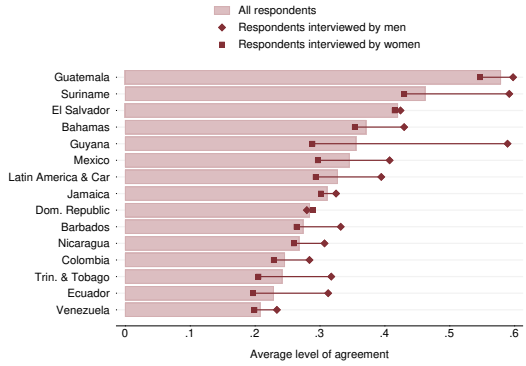
(b) Gender quotas in parliament



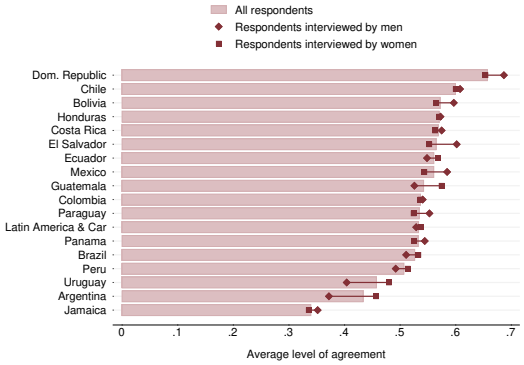
(c) Hit wife if she neglects the chores



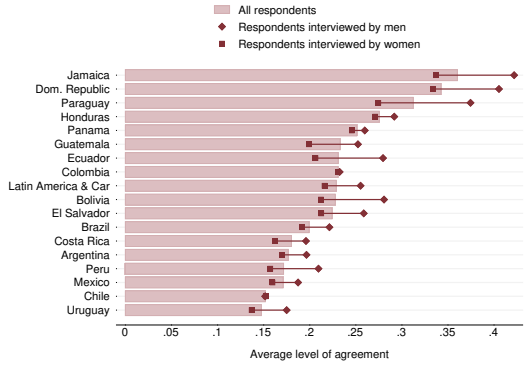
(d) Hit wife is she is unfaithful



(e) Children suffer when mother works



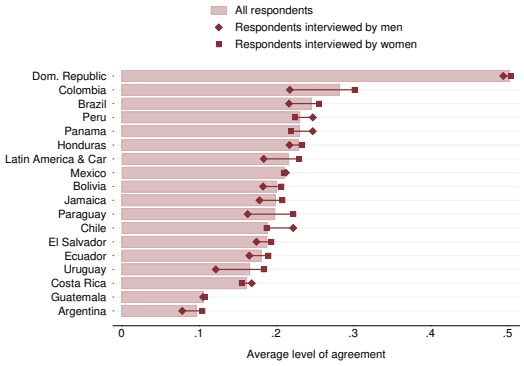
(f) Men are better political leaders



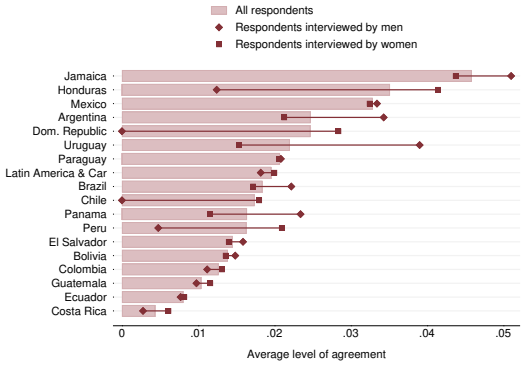
Note: Estimates are produced with probability weights following the survey sample design. See section 4 of the main text for a full description of the questions and Figures A4 and A5 for the significance of the differences when controlling for demographic variables and presence of spouse during the interview.

Figure A3: Difference in Agreement by Sex of the Interviewer, Question by Question, Part 2

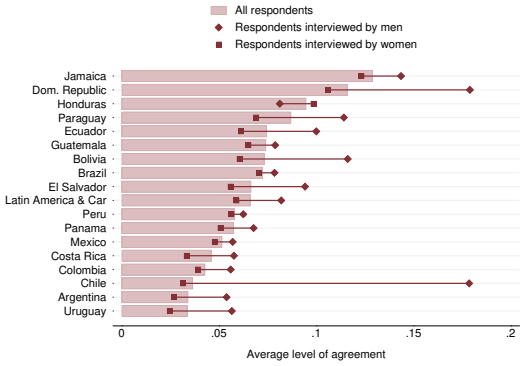
(a) A man would be more corrupt as a politician



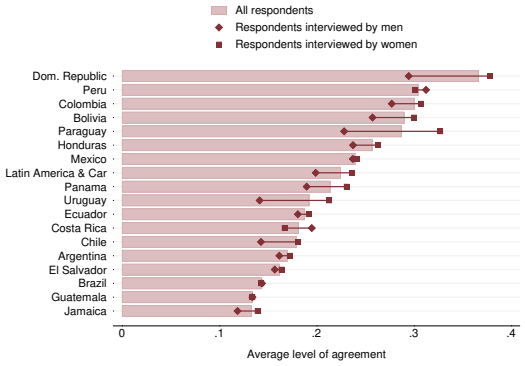
(b) A woman would be more corrupt as a politician



(c) Men manage better national economy



(d) Women manage better national economy



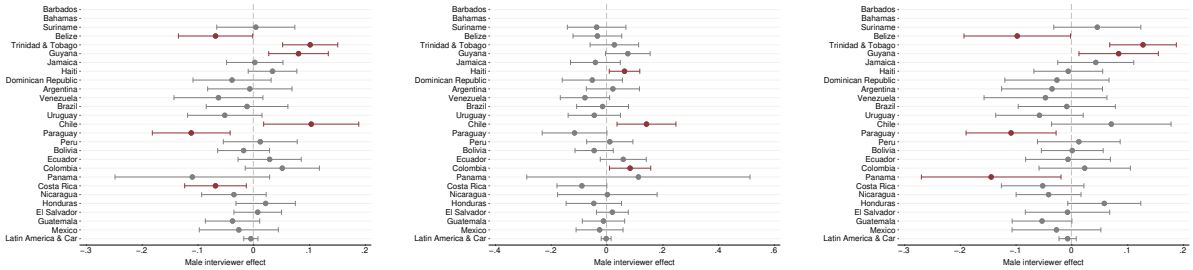
Note: Estimates are produced with probability weights following the survey sample design. See section 4 of the main text for a full description of the questions and Figures A6 and A7 for the significance of the differences when controlling for demographic variables and presence of spouse during the interview.

A.4 Full Set of Gender-of-Interviewer Effect Estimates by Country

Figure A4: Difference in Agreement by Sex of the Interviewer, Question by Question, Part 1

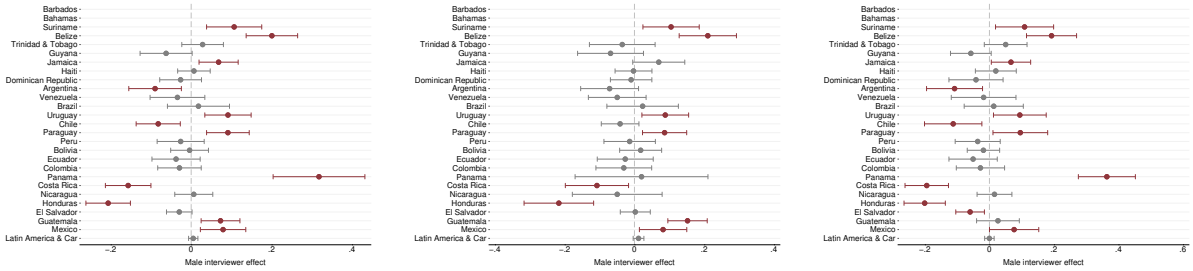
(I) Agreement: When there is not enough work, men should have a greater right to jobs

(a) All respondents (b) Female respondents (c) Male respondents



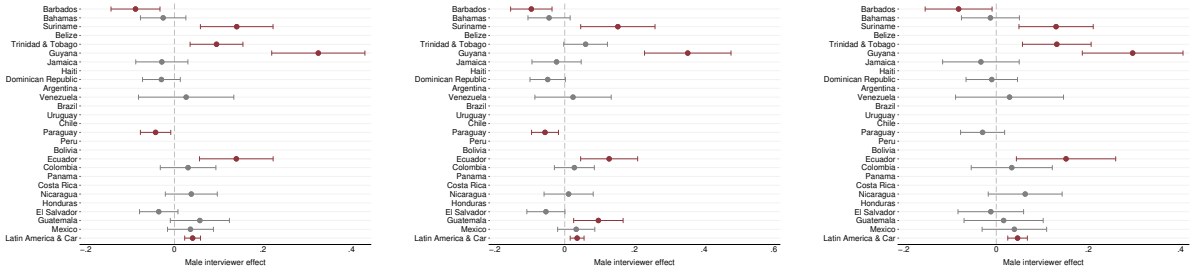
(II) Agreement with gender quotas in parliament

(d) All respondents (e) Female respondents (f) Male respondents



(III) It is justified that the husband hits his wife if she neglects the household chores

(g) All respondents (h) Female respondents (i) Male respondents

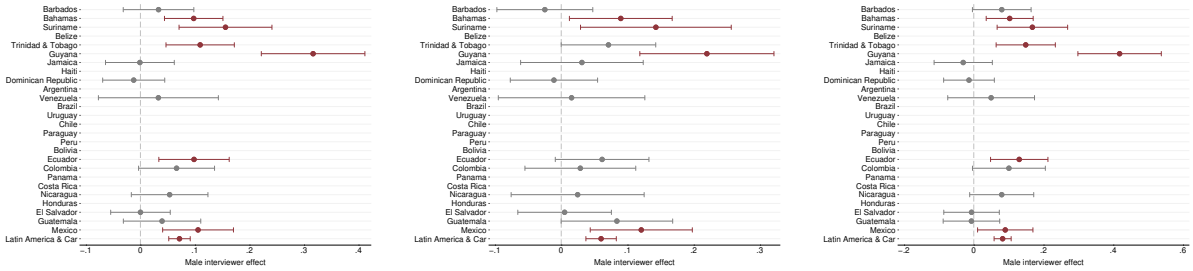


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents and presence of spouse during the interview. Estimates are produced using probability weights and correct standard errors following the survey sample design. See Section 4 of the main text for a full description of the questions.

Figure A5: Difference in Agreement by Sex of the Interviewer, Question by Question, Part 2

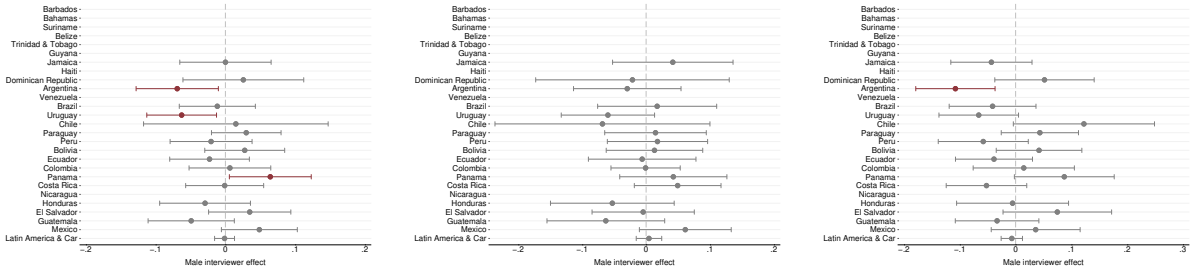
(IV) It is justified that the husband hits his wife if she is unfaithful

(a) All respondents (b) Female respondents (c) Male respondents



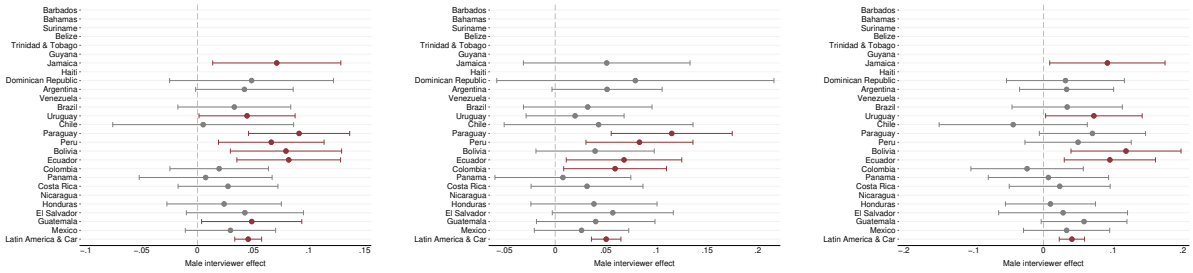
(V) Agreement: When a mother works outside the home, the children suffer

(d) All respondents (e) Female respondents (f) Male respondents



(VI) Agreement: In general, men are better political leaders than women

(g) All respondents (h) Female respondents (i) Male respondents

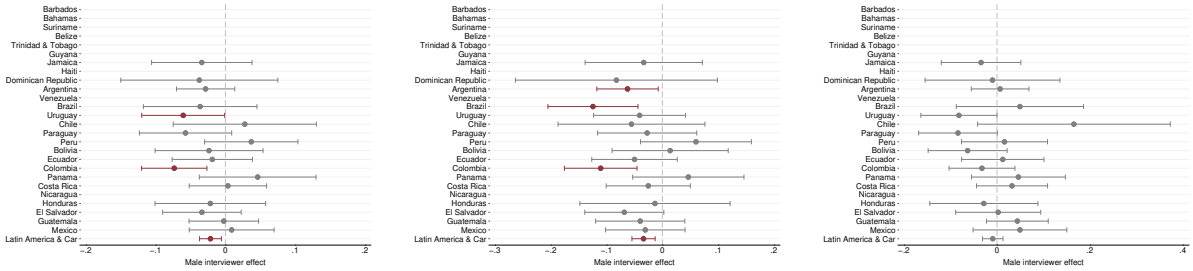


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents and presence of spouse during the interview. Estimates are produced using probability weights and correct standard errors following the survey sample design. See Section 4 of the main text for a full description of the questions.

Figure A6: Difference in Agreement by Sex of the Interviewer, Question by Question, Part 3

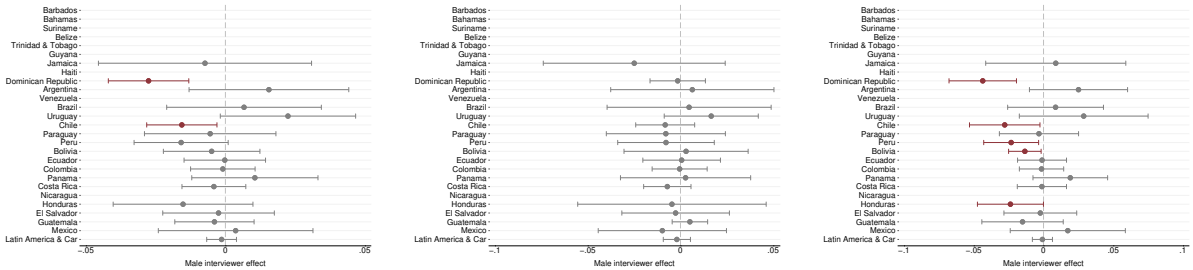
(VII) Thinks men would be more corrupt as politicians

(a) All respondents (b) Female respondents (c) Male respondents



(VIII) Thinks women would be more corrupt as politicians

(d) All respondents (e) Female respondents (f) Male respondents

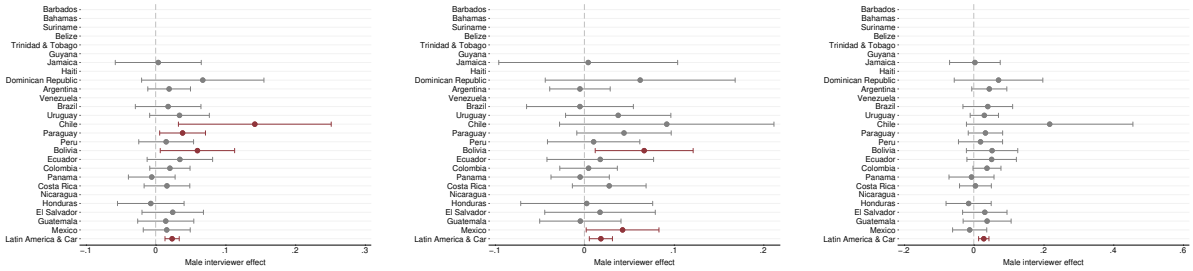


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents and presence of spouse during the interview. Estimates are produced using probability weights and correct standard errors following the survey sample design. See Section 4 of the main text for a full description of the questions.

Figure A7: Difference in Agreement by Sex of the Interviewer, Question by Question, Part 4

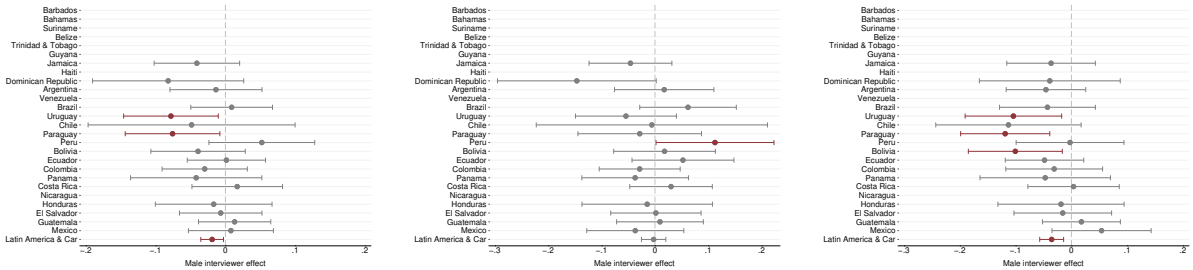
(IX) Thinks men would manage better the national economy

(a) All respondents (b) Female respondents (c) Male respondents



(X) Thinks women would manage better the national economy

(d) All respondents (e) Female respondents (f) Male respondents

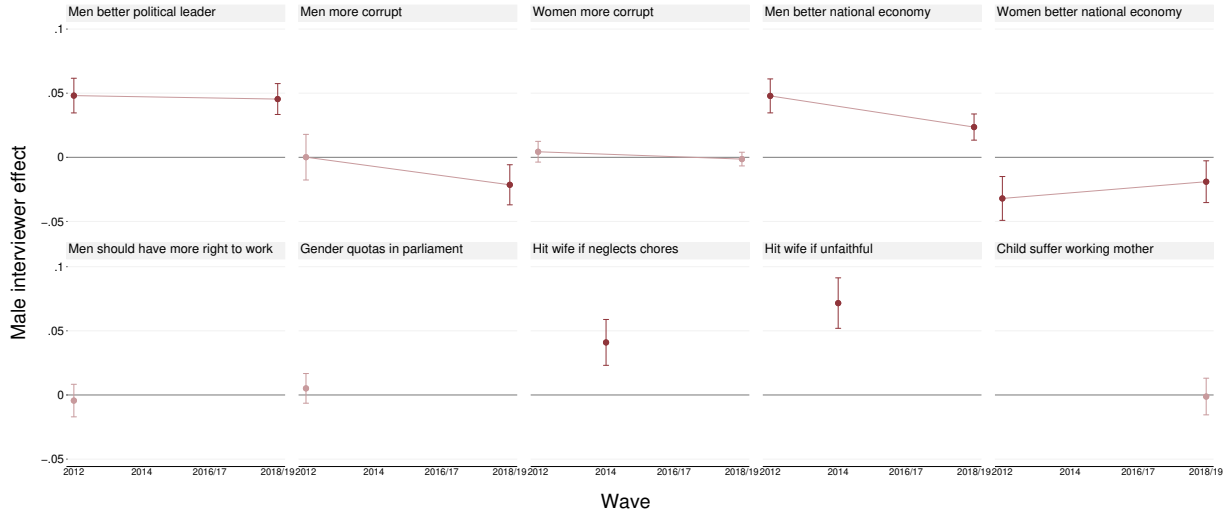


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondent and presence of spouse during the interviews. Estimates are produced using probability weights and correct standard errors following the survey sample design. See Section 4 of the main text for a full description of the questions.

A.5 Gender-of-Interviewer Effects in All Available Waves of the Survey

Figure A8 replicates Figure 1 for all the waves in which the questions of interest were included since 2010. The questions on the political abilities of men and women, presented in the upper row, were included in the 2012 and 2018/19 waves. All the questions in the lower row were included in only one wave of the survey.

Figure A8: Estimates for Gender-of-Interviewer Effects in All Available Waves



Note: 95% confidence intervals. The dark red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic characteristics of the respondents and presence of spouse during the interview. Estimates are produced using probability weights and correct standard errors following the survey sample design. See Section 4 of the main text for a full description of the questions.

A.6 Multiple Hypothesis Testing

Table A8 presents p-values corrected for multiple hypothesis testing. All p-values are calculated using the same specifications and observations that generate results presented in Figure 1.

Table A8: Multiple Hypothesis Testing with Complete Regional Sample

Question	Coefficient	SE	P-value	Westfall-Young	Bonferroni-Holm	Sidak-Holm
Men should have more right to work	-0.00	0.01	0.50	0.69	1.00	0.87
Gender quotas in parliament	0.01	0.01	0.38	0.53	1.00	0.86
Hit wife if neglects chores	0.04	0.01	0.00***	0.00***	0.00***	0.00***
Hit wife is unfaithful	0.07	0.01	0.00***	0.00***	0.00***	0.00***
Child suffer working mother	-0.00	0.01	0.87	0.86	1.00	0.87
Men better political leader	0.05	0.01	0.00***	0.00***	0.00***	0.00***
Men more corrupt	-0.02	0.01	0.01***	0.00***	0.04**	0.04**
Women more corrupt	-0.00	0.00	0.61	0.71	1.00	0.87
Men better national economy	0.02	0.01	0.00***	0.00***	0.00***	0.00***
Women better national economy	-0.02	0.01	0.02**	0.00***	0.11	0.11

Note: *** p < 0.01, ** p < 0.05, * p < 0.1. All p-values are calculated with the Stata command `wyoung`, with 100 bootstrap replications, and include the same controls as results presented in Figure 1.