Universidad de San Andrés

Departamento de Economía

Maestría en Economía

Gender-Based Research and Interviewer Effects. Evidence for Latin America and the Caribbean

Matías GÜIZZO ALTUBE

42017881

Mentor: Carlos SCARTASCINI

Victoria, Buenos Aires

28 de noviembre, 2023

Tesis de Maestría en Economía de

Matías GÜIZZO ALTUBE

"Investigación de Género y Efectos de Entrevistador. Evidencia para América Latina y el Caribe"

Resumen

A medida que la medición de opinión pública y de actitudes en cuestiones de género ha ganado cada vez más atención en las políticas y en la investigación de género, el diseño, la implementación y el análisis precisos de datos de encuesta se han vuelto primordiales. Este trabajo examina el impacto del género del entrevistador sobre respuestas de encuestas en América Latina y el Caribe, tanto a nivel individual como agregado, utilizando datos de LAPOP (2012-2018). Los hallazgos empíricos destacan la sustancial influencia que las normas sociales tienen sobre las opiniones de género, revelando que los encuestados son más propensos a sobreestimar las habilidades de los hombres en política en relación con las de las mujeres y a aceptar la violencia doméstica cuando son entrevistados por hombres en lugar de mujeres. Posteriormente, se presenta un modelo formal que ilustra cómo el género del entrevistador puede influir en las respuestas y se deriva un ajuste óptimo en el diseño y el análisis de la encuesta para corregir los sesgos introducidos. Como corolario, surgen dos acciones clave: las firmas encuestadoras deberían proporcionar siempre el sexo del entrevistador en sus datos, y los investigadores deberían ajustar este sesgo en sus análisis. Si no se aplican estas prácticas, es posible que los datos de encuestas no reflejen con exactitud lo que piensan los individuos de la región sobre las cuestiones de género, lo que podría sesgar las políticas públicas.

Palabras clave: Normas de género, Igualdad de género, Encuestas, Sesgos de comportamiento

"Gender-Based Research and Interviewer Effects. Evidence for Latin America and the Caribbean"

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. This paper examines 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). The empirical 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. The paper then presents a formal model that illustrates how the gender of the interviewer may influence responses, from which an optimal adjustment in survey design and analysis can derived to address the introduced biases. 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 public policy.

Keywords: Gender norms, Gender equality, Surveys, Behavioral biases

JEL Codes: J16, C83, D91

1 Introduction

Over the past several decades, academic research in the social sciences has demonstrated an escalating interest in unraveling the complexities of gender inequalities and their determinants. Early explanations for gender inequality often fell back on purported "biological differences" between men and women. However, the more contemporary scientific discourse has evolved to conceptualize gender as an intricate social construct, shaped by a multitude of attitudes, beliefs, and societal expectations that dictate human behavior.

This shift towards viewing gender through a sociocultural lens has not only catalyzed a surge in gender-focused research but has also underscored the importance of methodological precision in these studies. Gender-related articles culled from the JSTOR collection's social categories saw a leap from less than 1% in 1970 to more than 25% in 2016, as depicted in Figure A1 in the Appendix.[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 in recent years[2] (being published 1,053 articles on average every year from 2011 through 2016). A notable example of how modern science has been advancing in this direction is the work of Claudia Goldin, who was warded the Nobel Prize in 2023 for her pioneering research and contributions to the understanding of the changing roles of women in the economy.

The meteoric rise in survey-based, gender-related research underscores a critical need to scrutinize the processes through which such data is produced and interpreted. One crucial element often overlooked in this process is the potential impact of the gender of the interviewer on respondent answers, particularly when the survey questions touch upon gender issues and preferences. Existing studies have highlighted the potential for the 'gender-of-interviewer effect' to affect responses, but this phenomenon remains under-explored.

This paper aims to delve into the dynamics of the 'gender-of-interviewer effect' and its implications for survey design and data interpretation. I investigate how the gender of the interviewer can lead to biased inference from the data and potentially skew the understanding of gender issues, while also considering the effect that the gender of the respondent might have in their responses.

[1] This ratio surpassed 30% in 2022, but data is only reported until 2016 due to an abrupt decline in the number of articles accessible through Constellate since 2017.

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

To address this issue, after an empirical analysis using data from 26 countries of Latin America and the Caribbean, I introduce a formal model that quantifies the impact of the interviewer's gender on survey responses, providing a theoretical foundation for understanding and mitigating these biases. By doing so, I hope to not only shed light on an under-recognized source of bias in gender-focused research but also 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

In every face-to-face interview, similar to any social interaction, both the respondent and interviewer adhere to predefined rules and protocols. The interaction's purpose and the roles assigned to each participant are understood by both parties. For instance, the respondent expects the interviewer to ask questions for future analysis, while the interviewer expects the respondent to answer these questions as truthfully as possible. The interaction concludes after the questionnaire is completed.

However, this specific encounter is just one of many within the broader fabric of social interactions that make up society. 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 they are expected to behave by others. 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, economic status, among many other characteristics. However, their behaviors are also influenced by the characteristics of the other person and how those characteristics interact with their own. For instance, a woman may be expected to behave differently when she is surrounded by men versus when she is surrounded by 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", originate 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 biases, known as "role-independent effects", are 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 argument 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 are a result of social phenomena that transcend the interview and affect all other interpersonal interactions as well.

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).

Role-independent interviewer effects have been extensively studied over the last century, encompassing a growing variety of characteristics as the diversity of interviewers has increased. This research has also focused on characteristics deemed relevant for exploring specific issues. 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. Simi-

[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.

larly, 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, as exemplified by Benstead (2014a). First, the social-distance theory posits that respondents shape their answers to align with what they perceive to be the attitudes and beliefs of the interviewer. The second hypothesis concentrates on power dynamics, suggesting that respondents will mask their true attitudes and beliefs if they perceive the interviewer to be in a superior hierarchical position in terms of power, and will be truthful when they feel 'more or equally powerful' than the interviewer. Although these theories diverge on which respondents are 'telling the truth' and which are modulating their responses to please the interviewer, both align with a broader social desirability bias framework in which some respondents adjust their answers in the direction they expect the interviewer to favor.

These concerns about potential biases introduced by the selection of interviewers, and the challenge of distinguishing response effects from the 'true response,' have stimulated debate on interview design and conduct. Some early attempts by researchers to address this issue involved matching the characteristics of interviewers and respondents, or even allowing respondents to choose their interviewers. As these practices introduced other potential issues related to selection biases, subsequent efforts to mitigate role-independent interviewer effects (and social desirability bias in

general) sought to eliminate all human interaction from the data collection process, using technologies such as web surveys. This latter approach has become the most widely adopted due to its cost-effectiveness. However, it is not always the best option when seeking to gather data that is representative of large groups of people or targeted towards vulnerable populations that typically lack internet connectivity.

A common finding across the literature is that the bias is typically small in magnitude and not always statistically significant. This lack of robustness in the results could be due to the fact that the relevant interviewer characteristic must be sufficiently prominent in the question to influence the response. Alternatively, it may be that the level of analysis of the survey does not precisely align with the prevalence of certain social norms. For instance, in the case of gender-of-interviewer effects in Mexico, Flores-Macias and Lawson (2008) suggests that regional variance in the effects may have diluted their results. Consistent with their findings, the results presented in the subsequent section suggest that there is also variance in the magnitude, direction, and significance level of the effects at the country level in Latin America and the Caribbean.

3 Gender-of-interviewer effects in surveys in Latin America and the Caribbean

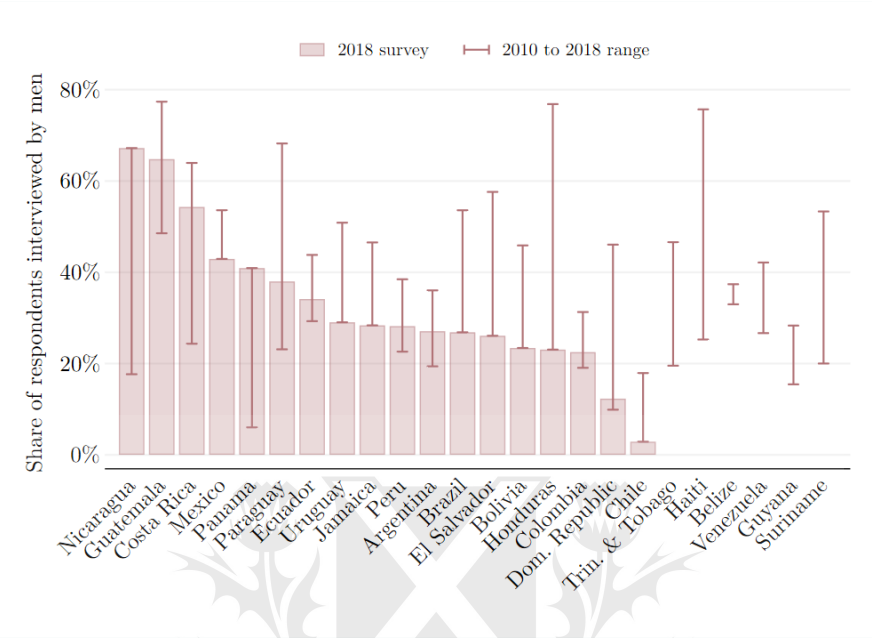
I employ data from the AmericasBarometer surveys conducted by the Latin American Public Opinion Project (LAPOP) to illustrate the contemporary effects of the interviewer's gender on responses in Latin America and the Caribbean.[5] The University of Vanderbilt has been conducting this survey on a biennial basis since 2004, encompassing up to 26 countries in the region.[6] The survey primarily utilizes face-to-face interviews for data collection, with the exception of the 2021 wave, that relied on telephone interviews (addressed separately towards the end of this section). The number of responses in each country and survey wave ranges from 1,000 to over 4,400, with the average and median number of responses being 1,677 and 1,507 respectively. LAPOP began documenting the gender of the interviewers in the 2010 wave of the survey. Figure 1 shows that the share of respondents interviewed by men vary significantly by country and over the years (as

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

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

identified by the range of the variable between 2010 and 2018).

Figure 1: 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 the University of Vanderbilt.
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.

The practice of recording the gender of the interviewer almost coincided with the start of the inclusion of a series of questions related to gender attitudes in the questionnaire (which happened in 2012). Having access to gender-focused questions as well as 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, I identified 11 gender-related questions. Nine 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"
- "It's justified to interrupt a pregnancy, that is, to have an abortion, when the mother's health is in danger"
- "When a mother works outside the home, the children suffer"
- "In general, men are better political leaders than women"
- "Same-sex couples having the right to marry"
- "Homosexuals being permitted to run for public office"

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

As it 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 I selected the most recent wave in which it was included[7] and estimated the male interviewer effect using the linear model in equation 1. Some of the questions offered two response options (agreement or disagreement), while others offered responses on agreement scales of 1 to 7 or 1 to 10, so all dependent variables were standardized to the interval [0,1], such that 1 represents the maximum level of agreement and 0, the maximum level of disagreement. 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,

[7] I 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.

whether they lives in a urban or rural area, how important religion is in their life,[8] and whether the spouse of the interviewee was present during the interview in interaction with the respondent's gender,[9] and ε_i captures the variance of y not explained by the model. I take the estimation of β through ordinary least squares as a measure of the male interviewer effect for each question.

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

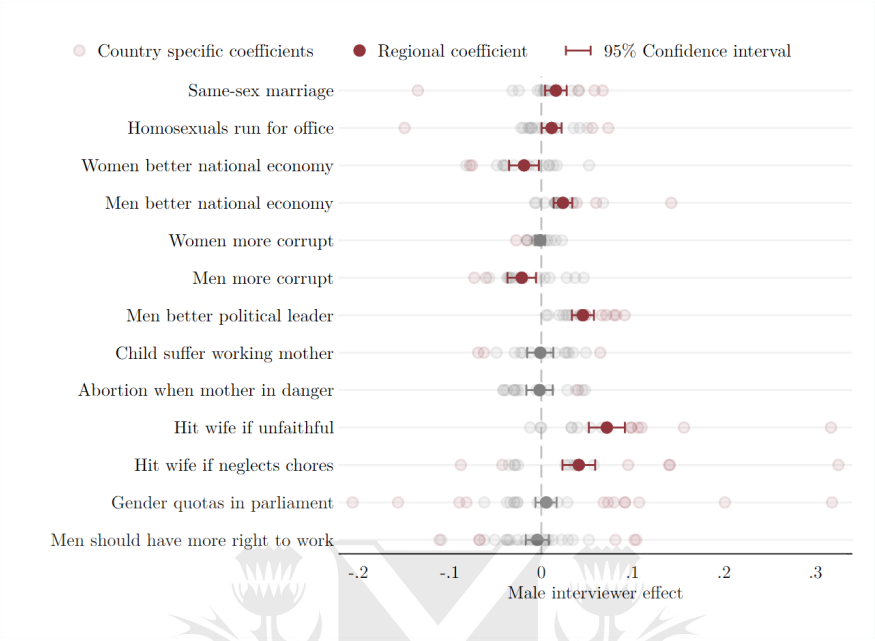
Figure 2 provides a summary of the estimates of the gender-of-the-interviewer effect, pooling both male and female respondents. Each line in the figure represents the results for a different question, plotting both the regional coefficient and the individual country coefficients. As all variables run 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 13 lines in place of 11 due to the structure of the questions concerning corruption and the management of the national economy, which provide the respondent the option to indicate a gender, as opposed to expressing agreement. Results are generally stable when using previous waves of the survey instead of the most recent one (see Figure A10 in the Appendix), and robust to corrections of p-values for multiple hypothesis testing (see Table A8 in the Appendix).

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

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

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

Figure 2: 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 the University of Vanderbilt.
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.

a man, compared to a woman.

Conversely, these significant effects are absent when the question does not directly compare men and women (such as questions about child suffering and abortion), or when discussing quotas and employment rights. There is a significant effect present in the question about same-sex marriage, although I lack a prior expectation for this result. However, these effects are not present in most previous waves (see Figure A10 in the Appendix) and are lose almost all significance in the multiple hypothesis tests (see Table A8 in the Appendix).

Secondly, these regional effects mask a considerable heterogeneity across countries. As we delve into the country-level effects, it can be noticed 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 abortion, 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 does 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 A5 through A9 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.

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.

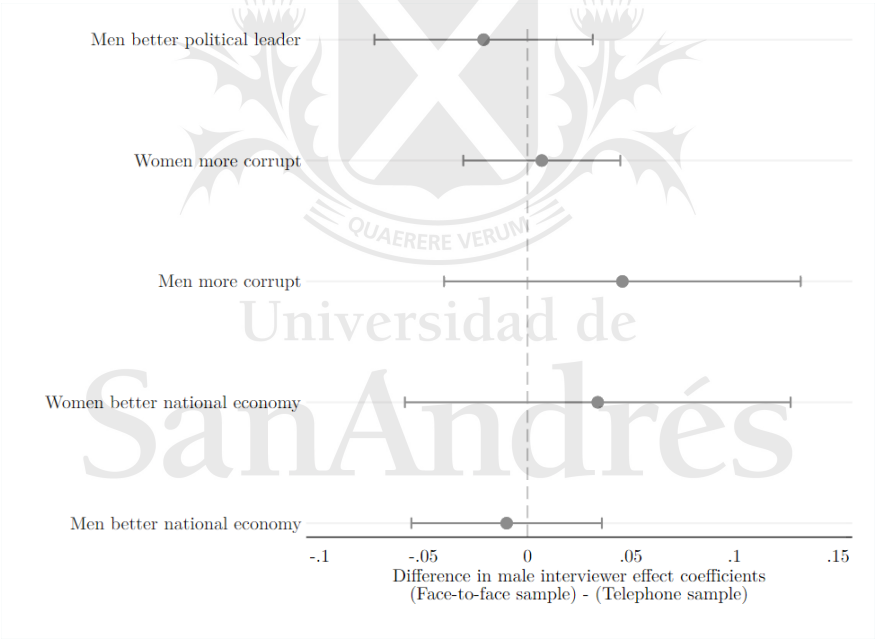
Anecdotal evidence suggests that there may exist some correlation between interviewer characteristics, for example due to recent changes in the interviewer labor market. As it has been a traditionally female-dominated occupation, male enumerators tend to be younger and less experienced than female ones, 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 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 there a female enumerator alone to these areas 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, I 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

Figure 3: 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.

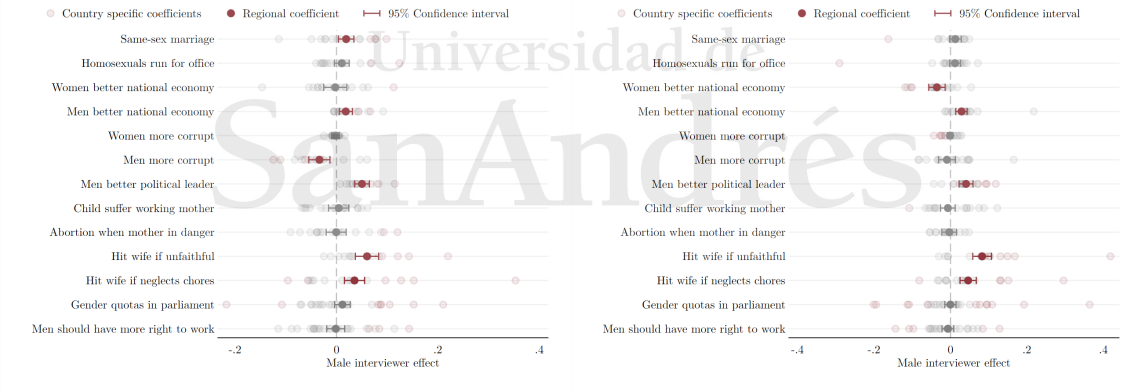
selection issues than face-to-face interviews[10], male interviewer effects should be less biased in the former. Figure 3 shows that the estimated effects of being interviewed by a male enumerator in the 2021 telephone survey are not statistically different from the same effects in the 2018 face-to-face survey.

Does the gender of the respondent matter too?

Figure 4 separates the effects presented in Figure 2 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 fine, 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 a men than when in front of another women. 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 men than a woman. The same does not happen for women.

Figure 4: Male interviewer effects by gender of the 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.

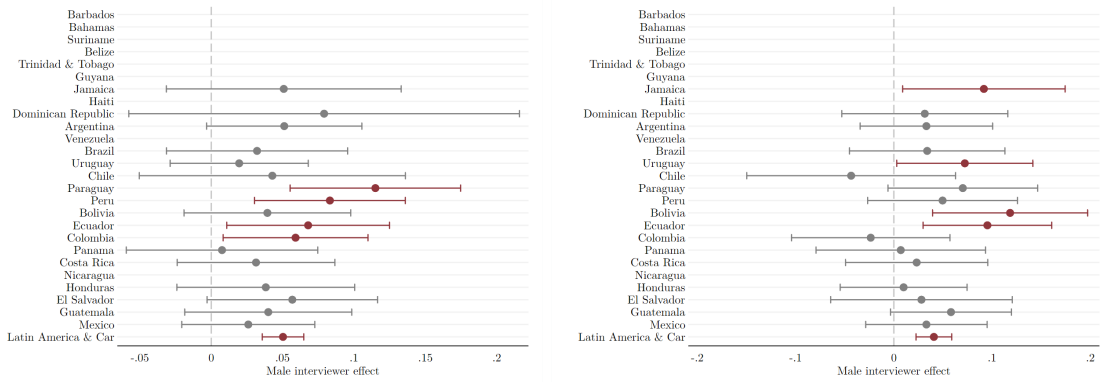
[10]Balance tests for all demographic variables included in the 2021 Mexico survey are presented in Table A7 in the Appendix.

Figure 5: Male interviewer effects by gender of the respondents - country by country

(I) Agreement: Men are better political leaders

(a) Female respondents

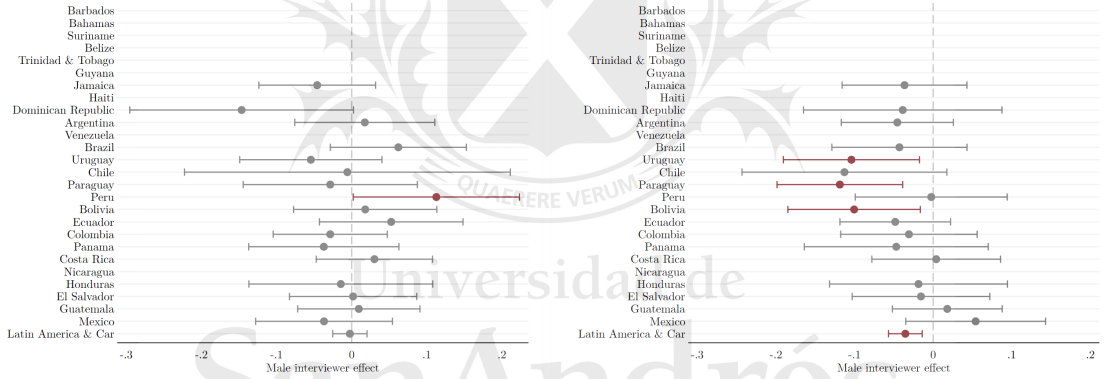
(b) Male respondents



(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 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.

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 expects 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 the interviewers on average.

Decomposing some of the results by country provides a more detailed perspective on what the regional averages reveal and obscure. Figure 5 depicts two such cases (the full set of questions is

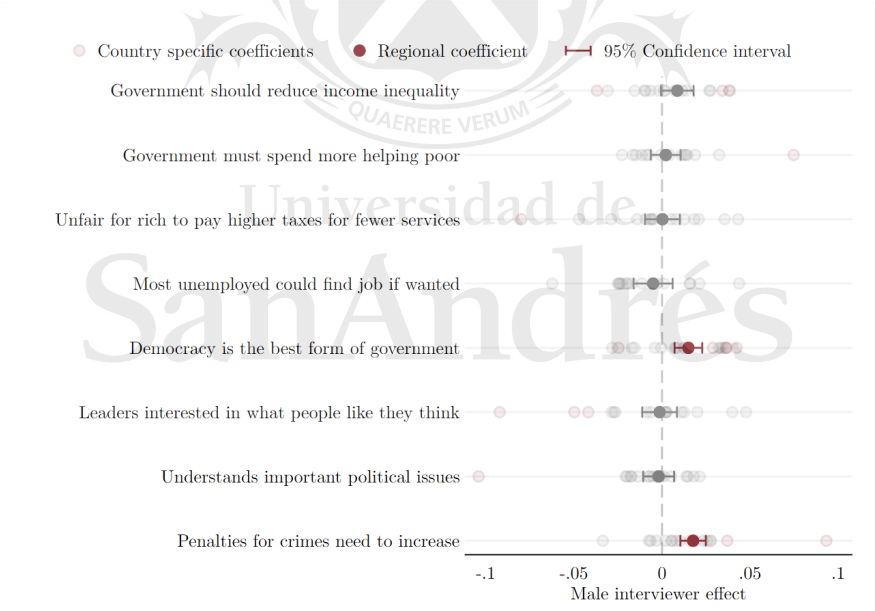
provided in section A.3 of the Appendix). The top row displays responses to the question of whether men are better political leaders. It can be observed that where the male interviewer effect exists, it is positive. Both men and women are more inclined to assert that men are superior political leaders 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.

Does the bias exist beyond gender-charged questions?

As mentioned in the introduction, gender-of-interviewer effect are not expected to exist for other questions that are unrelated to gender. I identified a module of eight questions about political

Figure 6: Male interviewer effects on political opinions unrelated to gender



Source: AmericasBarometer survey wave 2018/19. Latin American Public Opinion Project (LAPOP), by the University of Vanderbilt.
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 section A.7 in the Appendix for a full description of the questions.

opinions unrelated to gender that was included in the 2018/19 wave of the AmericasBarometer. These questions are also presented in terms of agreement, like the main gender questions in the analysis. The first half of the module is intended to measure preferences for redistribution and perceptions of inequality, while the second half broadens the focus to other perceptions and preferences in politics.[11] Figure 6 shows the results of this exercise with questions unrelated to gender. It can be noted that six questions do not show significant gender-of-interviewer effects, while the other two show a regional significant effect at the 5% level but no significant effects in almost all individual countries. Significance levels are robust to multiple hypothesis testing. Again, this exercise shows the complexity of the issue and the intricate relationship between social norms, gender, and responses. Results for the last question may indicate that respondents may believe that males are more prone to support higher penalties for crime. As such, they are more likely to respond that higher penalties are necessary when interviewed by a male than a female.

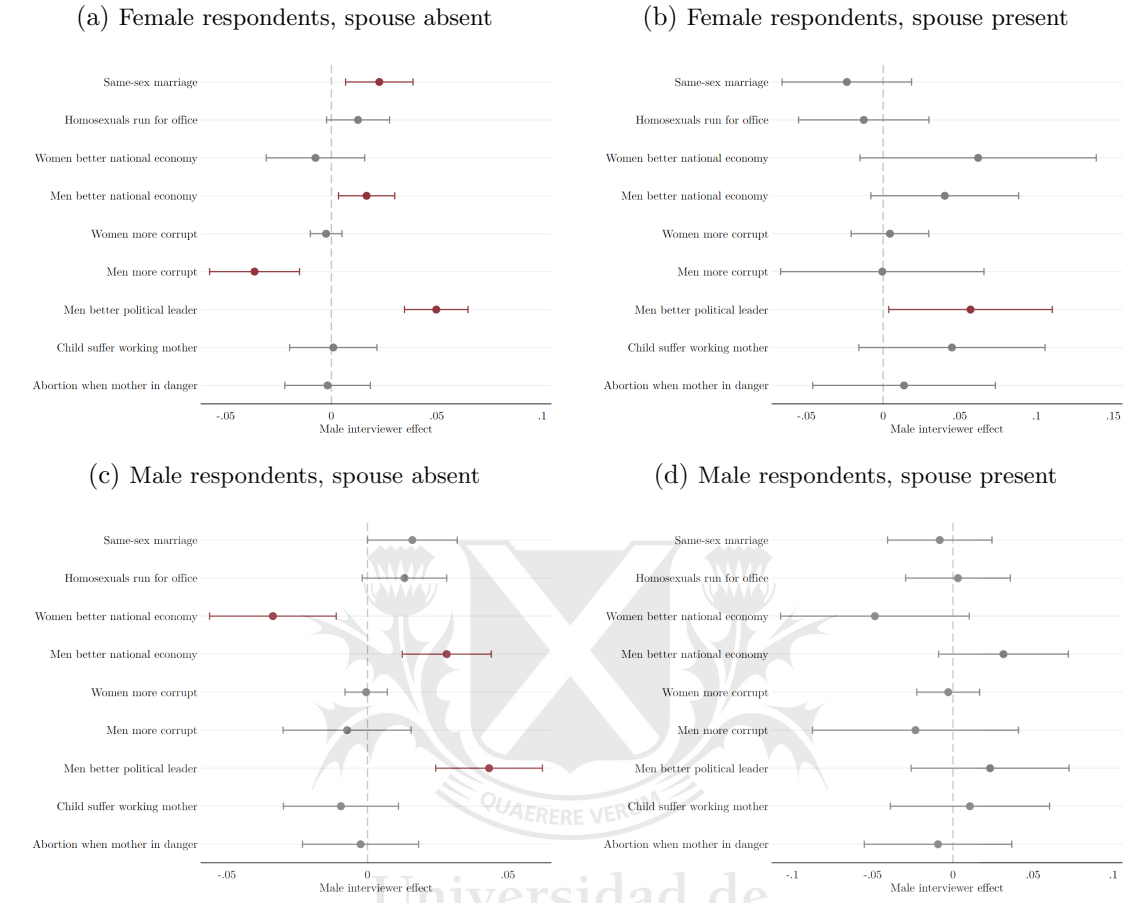
An additional complexity: having others in the room

As argued before, 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 who the other person is in society, and how they are expected to behave by others. The evidence so far 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. Furthermore, 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.

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

[11] See section A.7 in the Appendix for a full description of the questions used in this section.

Figure 7: 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 3 of this main text for a full description of the questions.

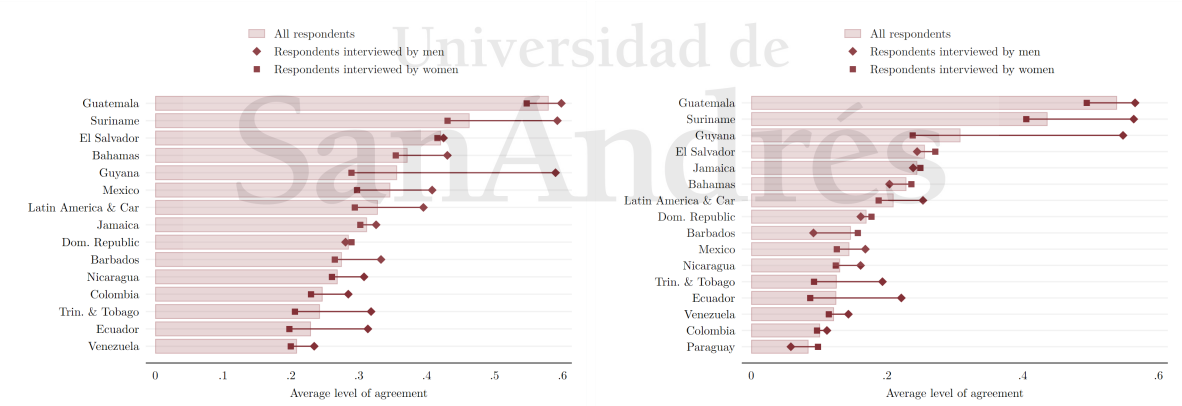
the sensitivity of the reported gender attitudes to the gender of the interviewer for both male and female respondents. The male interviewer effect dissipates when the respondent's spouse is present during the interview, as shown in Figure 7. If this means that the spouse presence dominates the gender-of-interviewer 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 Discussion: How well is gender opinion measured?

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. As an example, Figure 8 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 through A4 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.

Figure 8: 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.

Again, as mentioned before, it cannot be identified what is 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 next section presents a formal model that 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 also proposes an optimal adjustment to the survey responses derived from the model that can help correct for the biases generated.

5 Formalizing the gender of interviewer effect

Consider a respondent of gender (type) $\mu \in \{f, m\}$ being asked about their position regarding a specific issue by an interviewer of gender (type) $\theta \in \{f, m\}$. The respondent gives an answer $x \in \mathbb{R}$ trying to be as true to their real position, given by $v \in \mathbb{R}$, as possible, while trying to appeal to the interviewer at the same time, whose position is inferred by the respondent from the interviewer's gender as $\tilde{v}_\theta$. The problem faced by the 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 solution to this problem gives an optimal response x^* given by

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

with $\gamma \equiv \delta/(1+\delta) \in [0, 1]$ representing the weight of the interviewer position and $1 - \gamma$ the weight of the own real position in the reported response. A straightforward result from this expression is that, if the respondent has a minimum interest in appealing to the interviewer, i.e., $\gamma > 0$, and they think that the interviewer differs in his or her position towards the issue, given by $\tilde{v}_\theta \neq v$, then the reported opinion will differ from the true position of the respondent.

When scaling from individual to aggregate responses, however, researchers may try to cancel out individual distortions in the average response with a correct design (or reweighting). Assume that the distribution of v for each gender is centered in the average positions $\bar{v}_f$ and $\bar{v}_m$. Without loss of generality, preferences can be ordered such that v is increasing in a potentially "pro-female" direction, then$\Delta\, v \equiv \bar{v}_f - \bar{v}_m \geq 0$. Moreover, the fact that people may have a distorted perception of

the opinion distribution can be addressed with a distortion parameter $\varphi \geq 0$ such that $\tilde{v}_f = \bar{v}_f + \varphi$ and $\tilde{v}_m = \bar{v}_m - \varphi$.[12] With the additional 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

[12] 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.

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$\Delta\ v$ 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 no 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.

With slight modifications, the model also allows to address some of the different hypotheses discussed in the literature concerning the dynamics that lead to the existence of gender-of-interviewer effects. 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.

The ultimate goal of quantifying the male interviewer effect is 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:[13]

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

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 ad-

[13] 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 $n_{m}^{f} = n\left(\frac{\omega}{1+\omega}\right)\left(\frac{\gamma_{m}^{f}(\Delta v+\varphi_{f})}{\gamma_{m}^{m}\varphi_{m}+\gamma_{m}^{f}(\Delta v+\varphi_{f})}\right)$.

vocate 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.

6 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 towards gender equality. As highlighted in the literature and demonstrated in this paper, the interviewer's gender is especially pertinent. Intriguingly, the empirical 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 to contaminate 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 in 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's 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 to 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 it might be beneficial to plan the pool of interviewers in advance. In those cases, it may be helpful to add questions on normative expectations to the survey in order to perfect the weightings afterwards. 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 any scenario, researchers will always encounter trade-offs at every stage of research. When choosing a data collection technology, selecting interviewers, assigning respondents, and analyzing the data, there are always benefits and drawbacks to choosing the option that best contributes to the consistency of the targeted estimations.

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.

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.

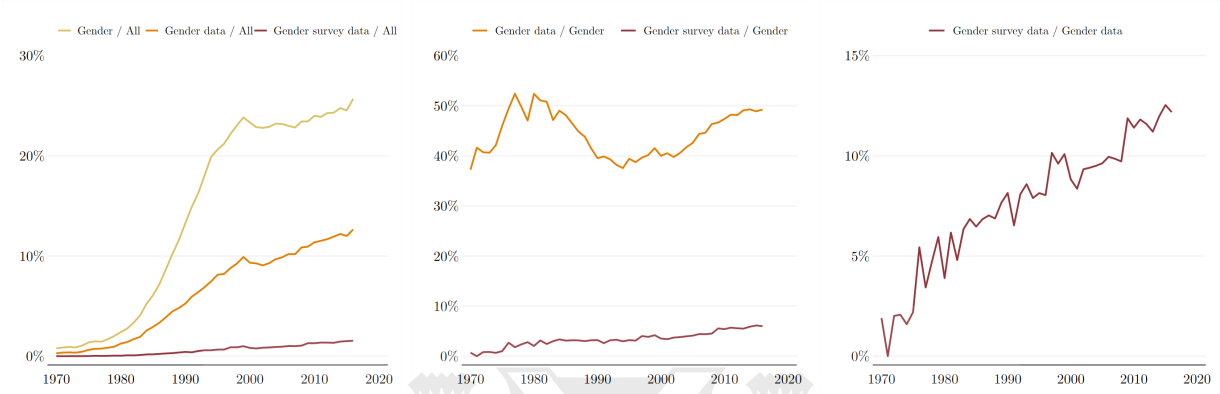
Weinreb, A. A. (2006). The limitations of stranger-interviewers in rural kenya. *American Sociological Review*, 71(6):1014–1039.

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 present 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 they 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)

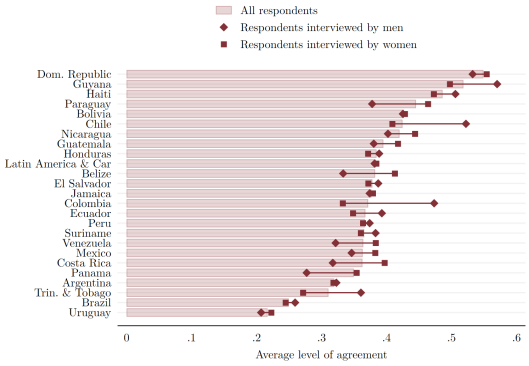
Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Estimates are produced using probability weights and correct standard errors following the survey sample design.

A.3 Differences in gender attitudes by sex of interviewer

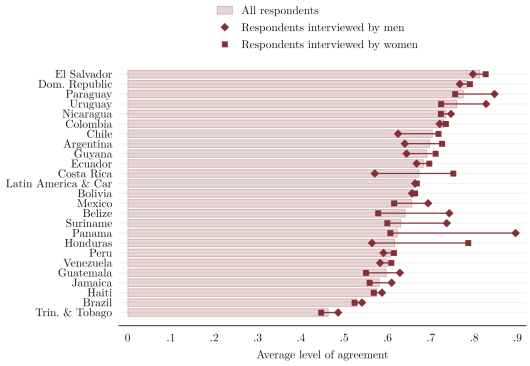
Figures A2 through A4 present the average level of agreement with each gender-related statement by country and by sex of the interviewer. Although not all differences are significant, these figures depict how different the country rankings could be if their pools of interviewers were composed by different 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 A5 through A9 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

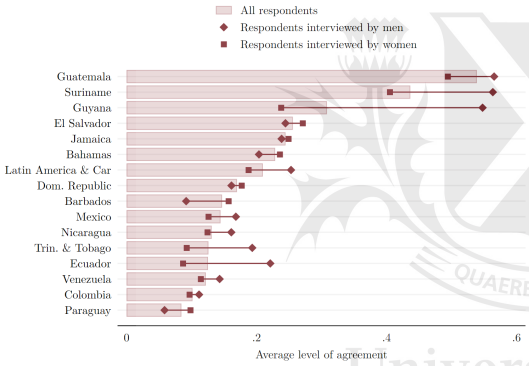
(a) Men should have more right to work



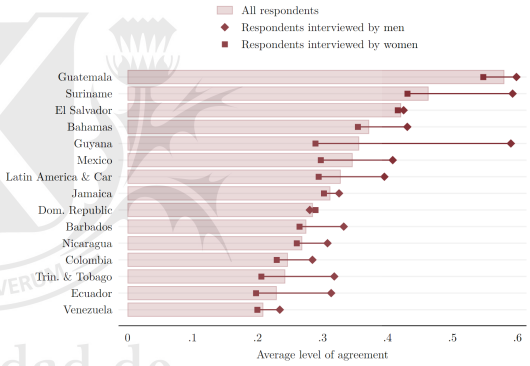
(b) Gender quotas in parliament



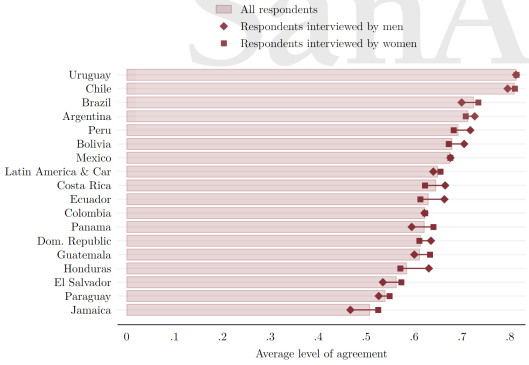
(c) Hit wife if she neglects the chores



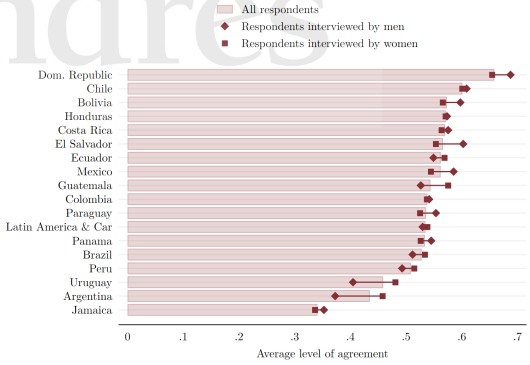
(d) Hit wife is she is unfaithful



(e) Abortion when mother is in danger



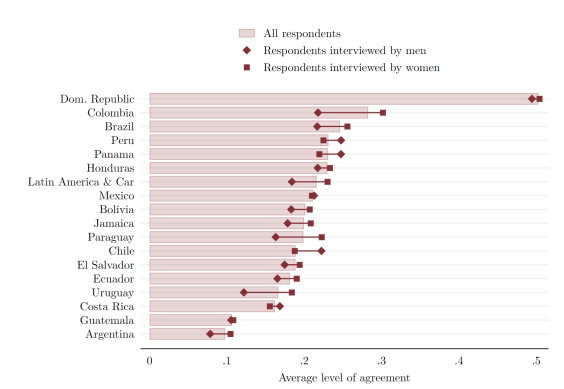
(f) Children suffer when mother works



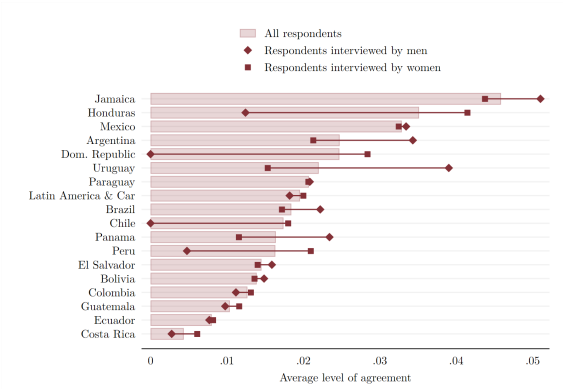
Note: Estimates are produced with probability weights following the survey sample design. See section 3 of the main text for a full description of the questions and Figures A5 and A6 for the significance of the differences when controlling for demographic variables and presence of spouse during the interview.

Figure A3: Difference in agreement by sex of the interviewer - question by question - Part 2

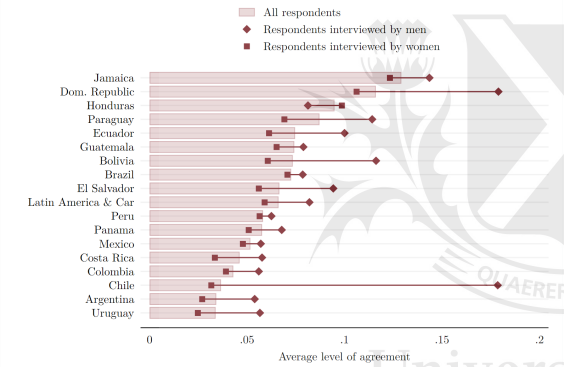
(a) A man would be more corrupt as a politician



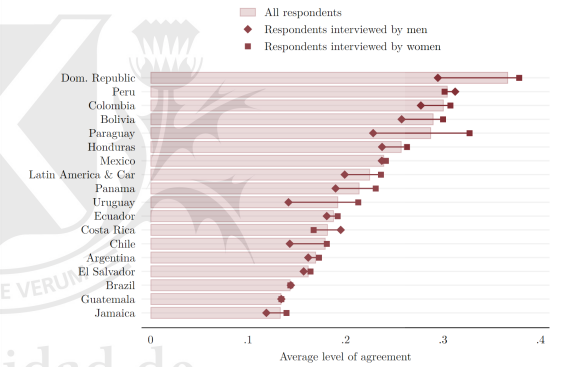
(b) A woman would be more corrupt as a politician



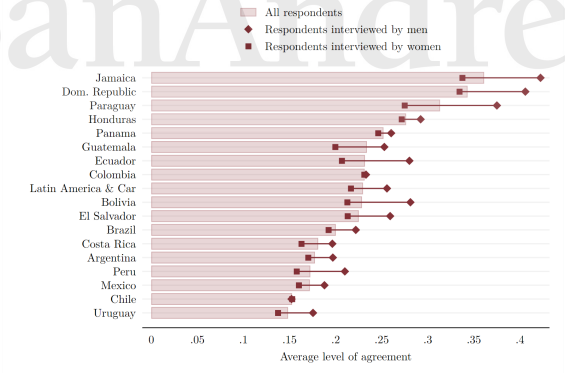
(c) Men manage better national economy



(d) Women manage better national economy



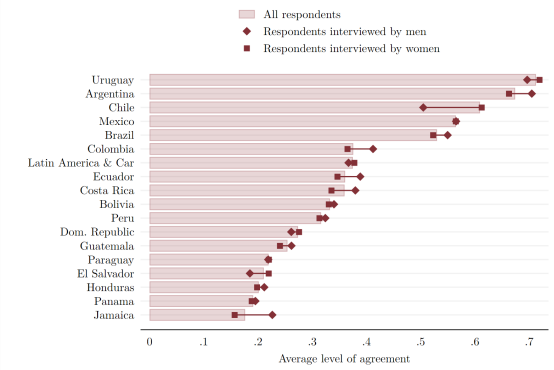
(e) Men are better political leaders



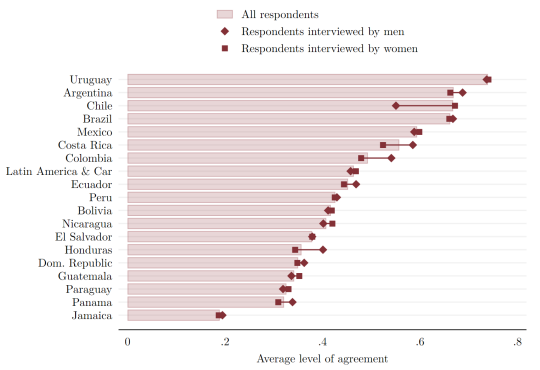
Note: Estimates are produced with probability weights following the survey sample design. See section 3 of the main text for a full description of the questions and Figures A7 and A8 for the significance of the differences when controlling for demographic variables and presence of spouse during the interview.

Figure A4: Difference in agreement by sex of the interviewer - question by question - Part 3

(a) Same-sex couples having the right to marry



(b) Homosexuals' right to run for public office



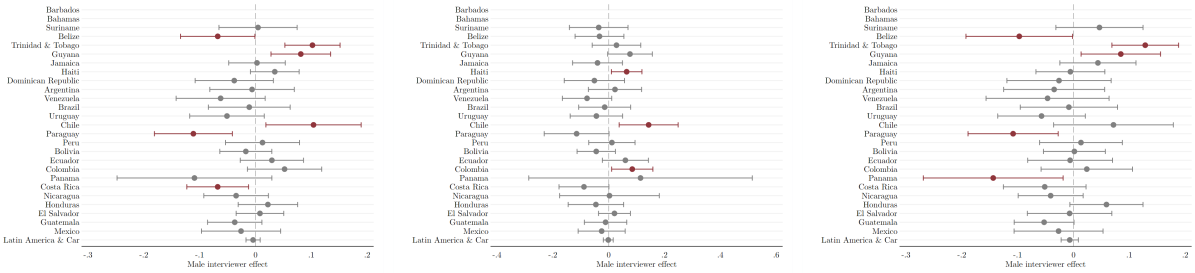
Note: Estimates are produced with probability weights following the survey sample design. See section 3 of the main text for a full description of the questions and Figure A9 for the significance of the differences when controlling for demographic variables and presence of spouse during the interview.

A.4 Full set of gender-of-interviewer effect estimates by country

Figure A5: Difference in agreement by sex of the interviewer - question by question - Part 1

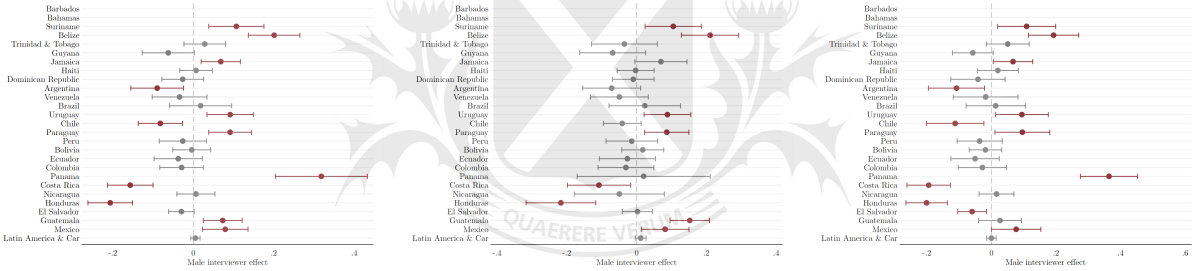
(I) Agreement: When there is not enough work, men should have a greater right to jobs

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



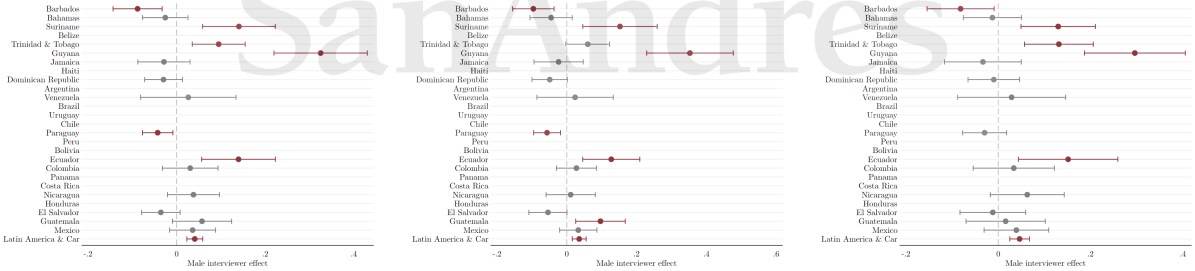
(II) Agreement with gender quotas in parliament

(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

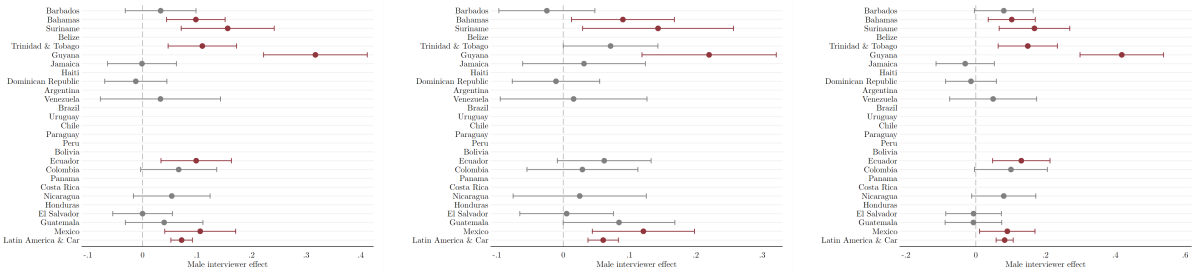


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic 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 3 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 2

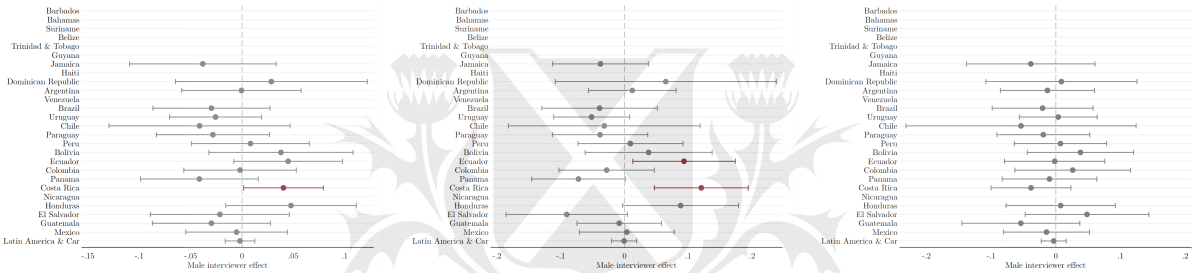
(IV) It is justified that the husband hits his wife if she is unfaithful

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



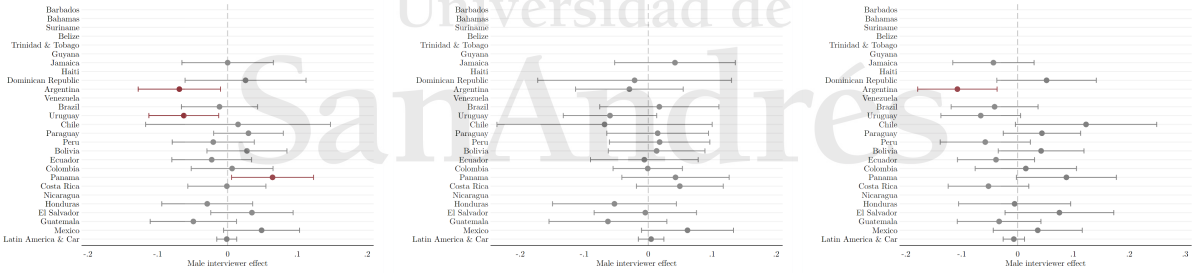
(V) It is justified to interrupt a pregnancy when the mother's health is in danger

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



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

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

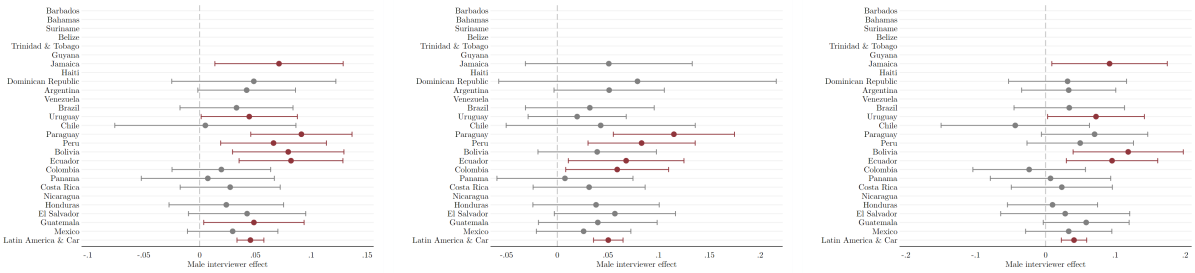


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic 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 3 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 3

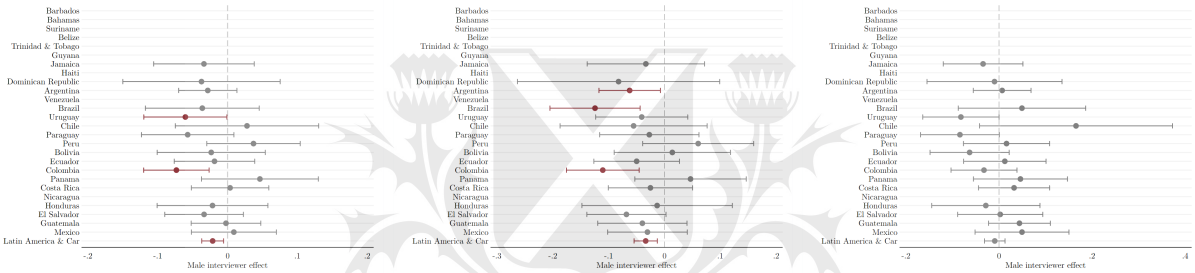
(VII) Agreement: In general, men are better political leaders than women

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



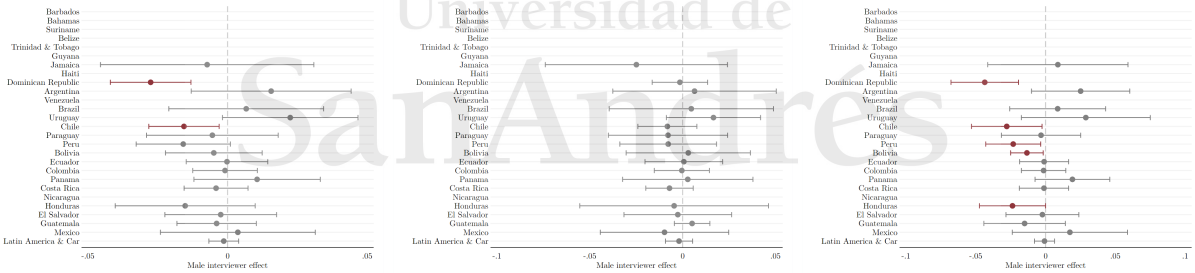
(VIII) Thinks men would be more corrupt as politicians

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



(IX) Thinks women would be more corrupt as politicians

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

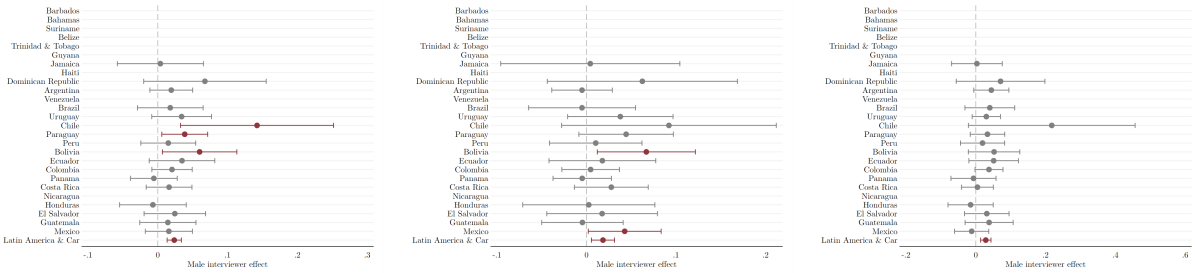


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic 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 3 of the main text for a full description of the questions.

Figure A8: Difference in agreement by sex of the interviewer - question by question - Part 4

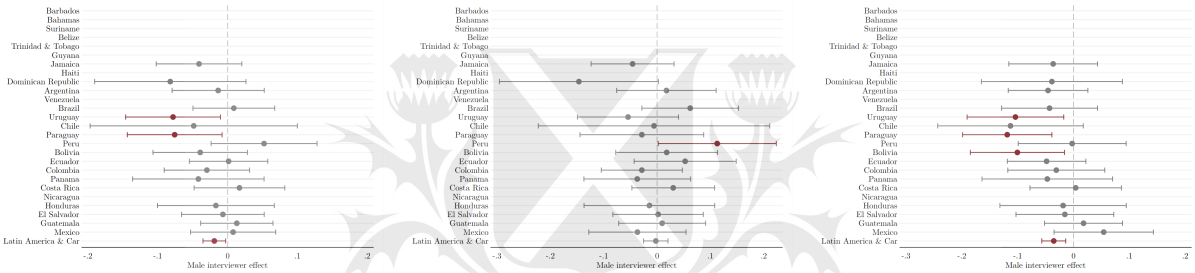
(X) Thinks men would manage better the national economy

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



(XI) Thinks women would manage better the national economy

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

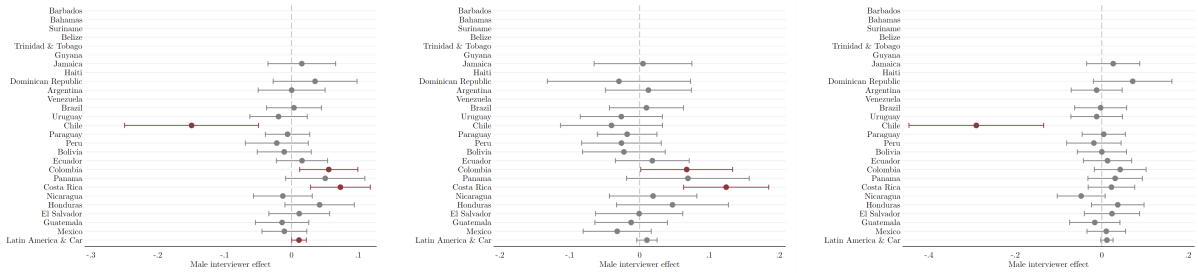


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic 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 3 of the main text for a full description of the questions.

Figure A9: Difference in agreement by sex of the interviewer - question by question - Part 5

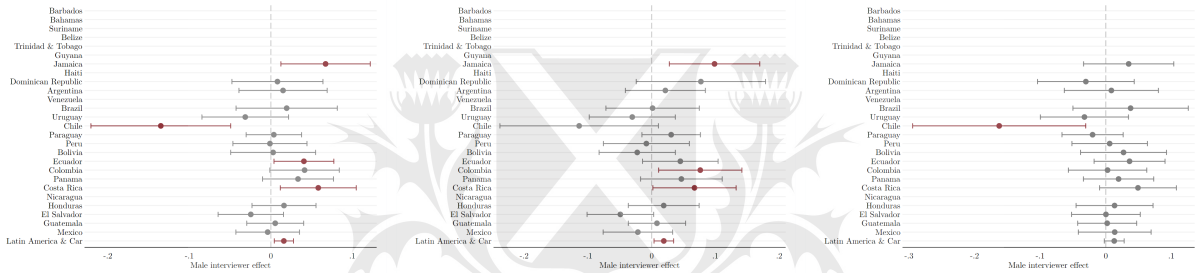
(XII) Approval: Homosexuals being permitted to run for public office

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



(XIII) Approval: Same-sex couples having the right to marry

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

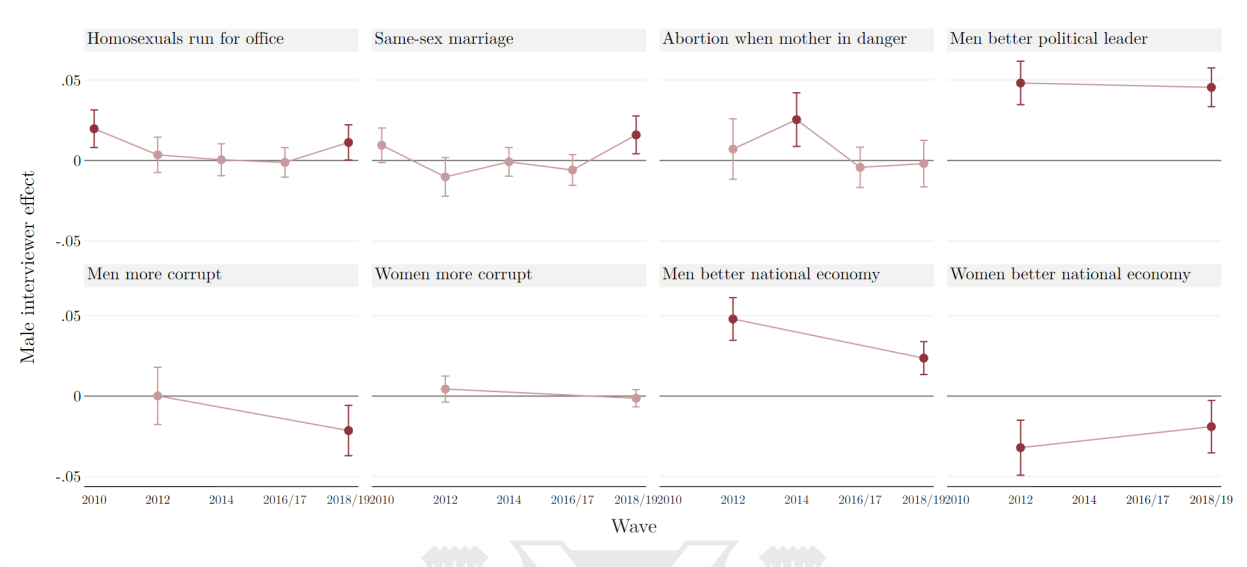


Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic 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 3 of the main text for a full description of the questions.

A.5 Gender-of-interviewer effects in previous waves of the survey

Figure A10 replicates Figure 2 for those questions that were included in at least two waves since 2010. Results are presented for each wave in which they were included. Those questions that were asked in only one wave are excluded of this figure.

Figure A10: Estimates for gender-of-interviewer effects in previous 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 3 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 2.

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	0.76	1.00	0.94
Gender quotas in parliament	0.01	0.01	0.38	0.62	1.00	0.91
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***
Abortion when mother in danger	-0.00	0.01	0.79	0.89	1.00	0.96
Child suffer working mother	-0.00	0.01	0.87	0.89	1.00	0.96
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***	0.07*	0.06*
Women more corrupt	-0.00	0.00	0.61	0.83	1.00	0.94
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***	0.16	0.15
Homosexuals run for office	0.01	0.01	0.04**	0.00***	0.27	0.24
Same-sex marriage	0.02	0.01	0.01***	0.00***	0.07*	0.06*

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 than results presented in Figure 2.

A.7 Questions unrelated to gender

We run an exercise with a module of eight questions included in the questionnaire of 2018/19 that measure levels of agreement in a similar way to the main attitudinal dependent variables of the analysis. The questions in the module ask for the level of agreement with the following statements, which are not specifically related to gender issues, in a scale of 1 (strongly disagree) to 7 (strongly agree):

- "The (*country*) government should implement strong policies to reduce income inequality between the rich and the poor."
- "The government should spend more on helping the poor."
- "It is unfair that the rich pay a lot in taxes but get very little back in government services."[14]
- "Most unemployed people could find a job if they wanted one."
- "Democracy may have problems, but it is better than any other form of government."
- "Those who govern this country are interested in what people like you think."
- "You feel that you understand the most important political issues of this country."
- "To reduce crime in a country like ours, punishment of criminals must be increased."[15]

All these questions were standardized to the interval [0,1] such that 0 represents the maximum level of disagreement and 1 represents the maximum level of agreement. Figures A11 and A12 show the results of regressing the agreement to each question against the sex of the interviewer and all the demographic control variables included in the main specifications, as well as the presence of the spouse during the interview, following the model in equation 1, country by country and for the complete regional sample. The estimations 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 they correct standard errors following the sample

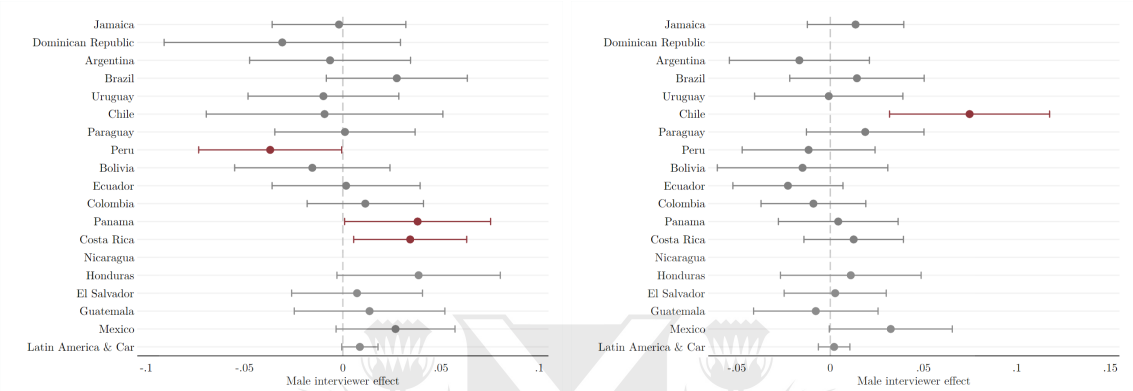
[14] In Colombia, Costa Rica, Honduras, and Panama, the statement was "It is okay that the rich pay a lot in taxes but get very little back in government services" instead. We reverted the scale and incorporated to the rest of the sample.

[15] This question was also included in the 2016/17 wave in a bigger set of countries than the 2018/19 wave. Therefore we use data from the 2016/17 wave instead.

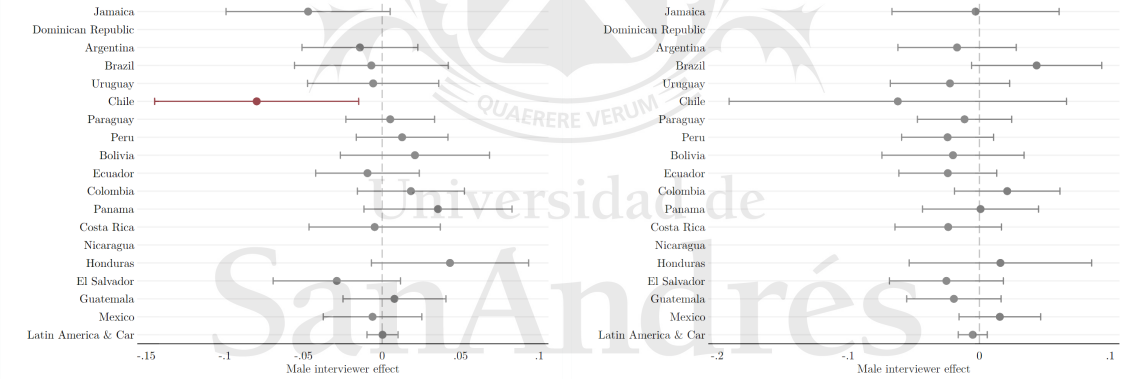
stratification structure. Barbados, Bahamas, Suriname, Belize, Trinidad and Tobago, Guyana, Haiti, and Venezuela were excluded from this exercise because the questions of interest were not asked in these surveys.

Figure A11: Male interviewer effects (unrelated to gender) by question and country - Part 1

(a) The government should reduce income inequality (b) The government must spend more helping poor



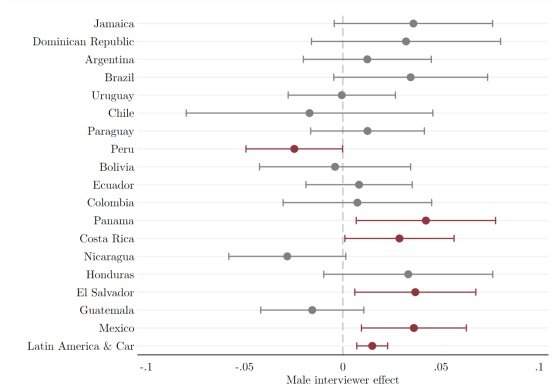
(c) Unfair for rich to pay higher taxes (d) Most unemployed could find a job if wanted



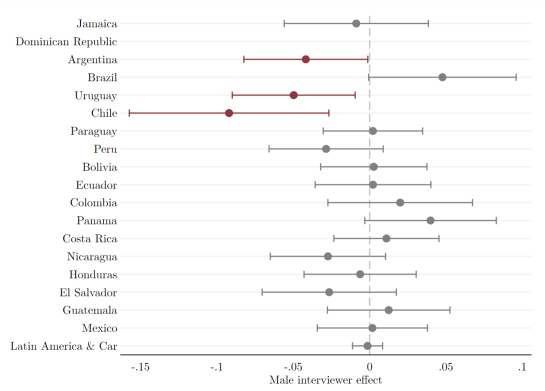
Note: 95% confidence intervals. The red color represents significance at the 5% level. The male interviewer effect is defined as the difference in agreement to each specific question between respondents interviewed by male enumerators and those interviewed by female enumerators, controlling for demographic 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 A.7 of this Appendix for a full description of the questions.

Figure A12: Male interviewer effects (unrelated to gender) by question and country - Part 2

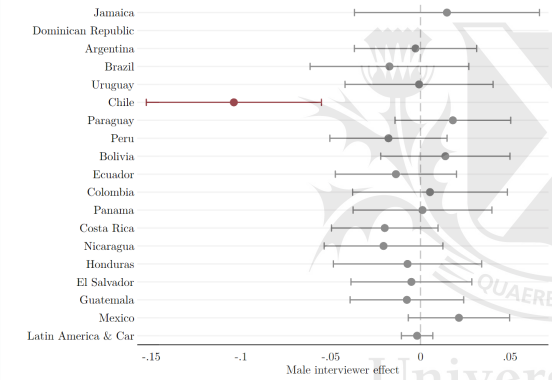
(a) Democracy is the best form of government



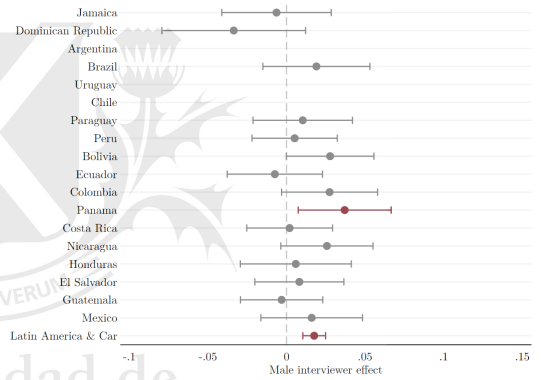
(b) Leaders interested in what people like they think



(c) Understands important political issues



(d) Penalties for crimes need to increase



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 A.7 of this Appendix for a full description of the questions.