Effects of Messengers' Gender on Government Learning: An Experimental Analysis of Policy-Makers in Peru, Tanzania, and India

Kaylee Hodgson, James Hodgson, Emily Ostler

Abstract *The purpose of this project is to identify how the gender characteristics of a messenger can affect policy makers' acceptance of policy information. Specifically, This study seeks to better understand the effects of gender on securing meetings with government officials. Enumerators in Tanzania, India, and Peru contacted 14,800 government officials, leading to 331 actual meetings with officials. In these meetings enumerators demonstrated a website database that summarizes over 400 policy evaluations. We have helped develop the website tool as the most comprehensive database of rigorous program evaluation summary results. We utilize a randomized control trial that occurred in these three countries to make a causal argument about officials' responses and willingness to meet through randomization of officials to male and female enumerators. We use randomization inference for difference-of-means testing and logistic regression to analyze the experimental results. We find that the evidence suggests that government officials did not prefer one gender when agreeing to meet with these enumerators. We found statistically significant effects in Peru, where officials were more likely to give positive responses to women (p-value=.002), a result that is in the opposite direction than we predicted. Such results are hopeful and indicate a lack of gender biases that we hope will continue to be studied.*

Introduction

Politicians are involved in crafting and implementing policy that can greatly impact the lives of their constituents. Academic researchers analyze these decisions and their outcomes, producing valuable information regarding the effectiveness of various policies. However, politicians often ignore the knowledge provided by these studies or lack the skills necessary to effectively use this information. These shortcomings in creating scientifically based policy decisions are partially because politicians may gain office due to birthright, wealth, corruption, charisma, or social biases, rather than a clear acquisition of skills. This can be a strength in representing citizens, but a weakness in finding and utilizing science-based research.

The purpose of this project is to better understand how to effectively present rigorous evaluation data on development policy to policy-makers in developing countries, specifically regarding effects of gender on securing meetings with government officials. To research these effects, we offered to present a database of more than 400 academic studies relating to various development topics to government officials in Tanzania, Peru, and India. We wanted to better understand if government officials would be more likely to express interest in data presented by men or women. It is increasingly important for government officials to have access to and knowledge of how to use academic studies in their policy-making, and this project will enable academics to more effectively disseminate their research among government officials and promote statistically-backed policy decisions. Furthermore, it will help us better understand what type of gender biases remain among government officials regarding their willingness to engage in conversation with students and academics.

Background

We wanted to test the effects of gender on securing meetings with government officials in three different geographic regions. Tanzania, India, and Peru are all developing countries with poor gender inequality. We measured this using the Gender Inequality Index produced by the UNDP, which includes various indicators, such as female participation in government, female participation in the workforce, and female access to education. Using 2014 data, we found that each of our countries had GII scores around average for their respective regions. Peru scored a 0.41, compared to a 0.43 for the Latin American region; Tanzania scored a 0.55, compared to a 0.57 for the Sub-Saharan African region; and India scored a 0.54, compared to a 0.50 for the South Asian region. (UNDP 2014). The average for developing countries was 0.55, which is almost exactly where both Tanzania and India scored. Additionally, cooperation from the USAID branches in these countries made partnership possible.

Furthermore, each of these countries have high population representation for their respective regions. Tanzania and Peru both have the fifth highest populations in Sub-Saharan Africa and Latin America, respectively, and India has the highest population in South Asia. Latin America and Sub-Saharan Africa do have more countries than South Asia (World Bank 2014). These countries also have fairly representative GDP per capitas (PPP) for their respective regions (World Bank 2014). In order to most effectively test the effects of variables such as the enumerators' gender, nationality, and contact methods on evoking interest among government officials, we needed similar levels of language fluency among all enumerators. Therefore, as English is widely spoken among the educated population in both Tanzania and India, English became the language medium between both local and Western enumerators and officials. We had a sufficient number of foreign enumerators who spoke Spanish, enabling us to select Peru without cause for concern regarding language.

Literature Review

As enumerators sought to reach government officials, asking them for thirty minutes of time to demonstrate the database, their success could be—in part—affected by the genders of themselves and of the officials.

Similar studies have been conducted, testing the discrimination of government officials to various demographics of constituents. For example, Broockman and Butler (2011) found that emails sent from constituents with stereotypically black aliases received fewer replies to requests than those with white aliases. However, minority legislators reacted oppositely, giving more responses to black aliases. Our enumerators initial contact with the officials was not face-to-face, but through emails and phone calls. Thus, the officials' only indicators of the enumerators' demographics was their name and possibly confirmed by their voice. These names were traditionally male or female and the gender of the enumerator would have been easily recognizable by the official. Thus, our study echoes the study mentioned above, but it includes real people who actually intend to meet with the official, if they agree; it includes the voice of the person seeking to contact the government official, in some scenarios; and it focuses on the effects of gender, rather than ethnicity.

We hypothesize that males will be more successful in securing meetings with government officials, due to biases towards male authority and legitimacy. In the majority of countries around the globe, women hold fewer than forty percent of parliamentary seats or governmental positions (Woman Stats 2016). In 2008, women were the heads of only thirteen countries (UNIFEM 2008). Though the percentage of females in leadership positions has increased significantly since the establishment of quotas in over one hundred countries (Beaman et al. 2009), many biases towards female leaders still remain. In India, women constitute only 18% of legislators, managers, and senior officials. In Peru, this number is 22.3%, and in Tanzania, it is only 16.5% (UNDP 2016). Even though women may be able to vote and run for office, it is often difficult for them to hold onto positions once elected or to have much power in decision-making (Hazarika 2008). We hypothesize that in these countries—where men hold the majority of leadership positions—that government officials will respond more positively to male enumerators. Whether consciously or subconsciously, they will view male enumerators as more legitimate and be more likely to express interest in meeting with them. Especially because we are presenting a database of information, which may be risky for government officials to log onto in their offices, they may be even more hesitant to meet with subordinates. Many of these officials are quite busy, and sacrificing even a half an hour of time to a student would require them to believe that the information to be presented will be of benefit to them. Therefore, officials will more likely meet with males, as males are more likely to be taken seriously.

Women tend to be undervalued in society, contributing to this preference for male authority. In India, for example, for every two males with a secondary education, only one girl has achieved a secondary education (UNDP 2016). In Tanzania, that ratio is 10:6, and in Peru it is 10:8 (UNDP 2016). In India, practices such as female infanticide, bride burning, and child marriage still persist (Hazarika 2008). In all three countries, there is a preference for male children (Woman Stats 2015), and women face great discrimination in terms of property and other rights (Woman Stats 2017). When women break out of customary gender roles, support for them decreases. For example, leadership and power are typically scene as masculine, and there is greater prejudice against women who try to take on positions requiring these characteristics (Garcia-Retamero and López-Zafra 2006). Thus, female enumerators who are seeking to advise government officials may be seen as deviating from traditional gender roles and may receive less support.

We understand that there may be some theoretical reasoning that may suggest that women would receive more positive responses. For example, in many trust game experiments, women have been perceived to be more trustworthy and less selfish than men (Buchan, Croson, and Solnick 2008; Innocenti and Pazienza 2006). Some studies have suggested that women act more altruistically than men, and thus can be more easily trusted (Dreber and Johannesson 2008), but this is not true in many cases (Cox, 2002; Andreoni and Vesterlund, 2001). We also consider a potential bias in favor of female attractiveness, as attractive female solicitors have been shown to have the most success in raising funds and receive higher wages for similar work (Landry et al., 2006; Andreoni and Petrie, 2008). Such biases in favor of female attractiveness may increase the effectiveness of our female enumerators, though we expect these effects to be small, as females will be contacting officials primarily over phone and email, where they cannot be seen.

There is evidence of heavy in-group biases favoring women in many situations, which would increase the response rates between female officials and female enumerators. Rudman and Goodwin (2004) found a significantly stronger in-group bias among women than among men.

Prentice and Carranza (2002) found that, due to social stereotypes, men may be perceived as more intimidating and insensitive than women, encouraging female officials to refuse more meetings with male enumerators. Furthermore, as there is such a great disparity in political participation and economic opportunity between men and women in these countries, we suspect that women may be eager to support other women. Therefore, when female enumerators ask to present the database, female officials may be more willing to meet with them, regardless of the benefit they may gain from the meeting. Because of the strength of the argument in favor of an in-group female bias, we have adjusted our hypotheses to include two sub-hypotheses. In the case of a female official, we expect female enumerators to be more successful at eliciting positive responses and securing meetings. However, because there are more male officials than female officials, the overall hypotheses remain consistent with the expectation that males will be more successful.

Research Design

The subjects in this experiment were national-level policy-makers in Peru, India, and Tanzania. Web-scrapers at Brigham Young University collected the publically available contact information of as many of these policy-makers as possible in these three countries. Once compiled, these names were randomly assigned to the enumerators assigned to each country. The enumerators in all three countries were balanced by nationality (local and foreign), and the enumerators in Tanzania and Peru were balanced by gender. The enumerators in India were majority female but still randomly assigned. The names of the government officials were randomly assigned among these enumerators, clustered by country assignments.

The American enumerators were selected from Brigham Young University in the United States. The American enumerators sent to Peru all spoke Spanish. There were no special language requirements for Tanzania or India. The local enumerators were selected from local universities in the respective countries. In Tanzania and India, these local enumerators were required to be proficient in English. Gender was self-reported and roughly balanced in the selection of enumerators. Once the enumerators were chosen, they were randomly assigned a list of government officials they were assigned to contact.

In all three countries, local and Western enumerators first contacted each subject to schedule a time to meet in person. This was completed over the phone or by email, depending on the contact information available to the enumerators. Enumerators in the three countries used scripts for initial and follow-up phone calls and emails. There were in-country differences in contact methods. In Tanzania, phone calls were primarily used, because email addresses were unavailable. In Peru, email was the primary initial contact method, and phone calls were used for follow-up contact. In India, emails and phone calls were both used for initial and follow-up contact. Government officials first received an initial email or phone call, and—if contact was not successful— at least two follow-up attempts were made to contact the subject. When contact was made, enumerators would request a visit to introduce the website to the subject.

There were two scripts for initial emails and phone calls. One included a social proof treatment, which was sent to randomly assigned officials and mentioned interest among other officials in their respective government ministries. The other script did not mention any responses from their peers. Enumerators were instructed to follow the script for the phone call as closely as possible before answering the government officials' questions. Enumerators could also attempt to make unannounced visits, in which they would visit the ministry and request a meeting with an assigned subject. The scripts for emails, phone calls, and unannounced visits can be found in the Supplemental Material section.

If the subject agreed to a visit, the enumerator who contacted that official (randomly assigned to the subject) visited the subject in person and asked him or her to complete a short baseline survey. The subject was invited to ask their colleagues to join the meeting. The enumerator then demonstrated a selected version of the website—either with a tablet that had a preloaded presentation with the particular version of the website that was randomly assigned to that official, or by visiting the actual website. The enumerators left the meeting, and subjects received no further prompts to use the website and were not contacted by research staff until the end line survey.

Enumerators would report their success in terms of "positive responses," or expressed interest in continued communication or the establishment of a meeting time; and "negative responses," or expressed disinterest in continued communication. Enumerators also noted appointment confirmations, if and when meetings were held, and whether the official invited colleagues to join the meeting.

Hypotheses

The purpose of this study is to evaluate the impact of messengers' gender in securing meetings with government officials in Tanzania, Peru, and India in order to improve future efforts to disseminate policy knowledge. We have the following hypotheses, which predict the impact of gender on positive responses and appointments held. We will test the following hypotheses, and their associated subgrouped hypotheses:

1. Male enumerators will elicit more positive responses from officials than female enumerators.
    a. Female enumerators will elicit more positive responses than male enumerators when contacting female officials.
    b. Male enumerators will elicit more positive responses than female enumerators when contacting male officials.
2. Male enumerators will hold more appointments than female enumerators.
    a. Female enumerators will hold more appointments than male enumerators when contacting female officials.

b. Male enumerators will hold more appointments than female enumerators when contacting male officials.

We predict that men will have greater success in securing positive responses than women. Because women are often taught by society to not be assertive, we expect that they may be perceived by both men and women as weaker or less authoritative and therefore receive less positive responses (Burns, Schlozman, and Verba 2001, 341). Nonetheless, we hypothesize that these biases will reverse among female officials. We predict that there will be a degree of kinship among women, making female officials more likely to meet with and respond positively to female enumerators than male enumerators. They may also feel more comfortable meeting with women, thus increasing meetings and trainings between female enumerators and female officials, as compared to those between males and females.

Methods

We measure the take-up rates for face-to-face trainings based on the gender of the enumerator contacting the government official. The outcome measures are estimated and compared based on the gender of the enumerator. Our outcome measures for each of the hypotheses are the likelihood of a positive response from a government official and the likelihood of holding an appointment with an official. We use difference of means tests using randomization inference, as well as logit and two-stage regression analysis to determine whether there are significant differences in the outcome measures based on gender. Difference of means tests are used to estimate average treatment effects (ATE) to evaluate how the likelihood of positive responses and appointments held changes based on the gender of the enumerator, as well as both the gender of the enumerator and official. We also use logit and two-stage regression analysis to evaluate the impact of the enumerator gender as well as other covariates on our outcomes, in order to ensure that the differences are truly due to the treatment of gender and robust to alternative specifications. Two-stage regression analysis is used to estimate the complier average causal effect (CACE) or treatment on treated. In the first stage of the two-stage regression model, we create a model with the positive response as the response variable, and all of the covariates as the explanatory variables. Then, at the second stage, we use the appointments held variable as the response, with both the gender of the enumerator and fitted values of positive response from the first stage regression as the explanatory variables. These analyses are performed both for each individual country (India, Tanzania, and Peru), and for the total dataset collectively.

In order to account for the large number of tests and multiple comparisons, we update the significance level, using the family wise error rate (FWER) adjustment. Using this method, we calculate an updated significance level of α=0.017 (from a base significance level of α=0.05) or a

less stringent α=0.035 (from a base significance level of α=0.10). Only the tests with p-values less than this updated significance values are considered significant in our tests.

We use randomization inference for our difference of means tests, a more robust method to analyze treatment effects. Randomization inference is especially useful because its results do not depend on distributional assumptions, because the p-value is calculated directly from the data. The algorithm for randomization inference randomly pairs the treatment assignments with the results and calculates the difference of means at each iteration. This allows us to plot a distribution of difference of means, and to identify where the actual difference of means from our treatment assignment falls in that distribution. We plot the results and calculate the p-value.

The covariates used for our regression analysis originally included the status of the enumerator (local or foreign), gender of the government official, social proof variable (control vs. social proof treatment), modes of contact (email, unannounced visit, phone call), status of the government official, and the interaction term between gender of the enumerator and gender of the government official. These terms were included in the model in order to control for alternative experimental conditions, along with government official demographics. We also use a fixed effect for country in the overall analyses.

While we originally included all of these variables in the analyses, we ultimately decided to remove modes of contact and government official status from the analysis. The modes of contact variables were removed for two main reasons. First, we ran a regression model with gender of the messenger as the dependent variable, and all of the control variables as explanatory variables, in order to verify that the treatment was not significantly related to the other covariates. We found that all three of the modes of contact (email, phone call, and unannounced visits) were significantly related to gender of the enumerator. This was inaccurately inflating the significance of gender. Our second reason for excluding modes of contact was that we found that the country fixed effects and the modes of contact variables were essentially measuring the same things, by performing a chi-square test for each mode of contact against the country variable. The results can be found in Table 1. We were not surprised by this result because of the idiosyncrasies within each country regarding contact methods. We also exclude the government official status from the analysis because we determined that the methods used to determine the status level were arbitrary. Additionally, because this variable was not significant in the original model, excluding the variable did not significantly change any of the results of the analysis.

Table 1: Pearson's Chi-Square Test: Country vs. Mode of Contact

Test	**Chi-Squared Statistic**	**p-value**
Country vs. Phone Call	2366.5	<0.0000
Country vs. Email	1173.8	<0.0000

Country vs. Unannounced Visit	576.9	<0.0000

Our final logistic models are written as follows:

Log Odds Ratio of Probability of Positive Response$_i$ = Enumerator Gender + Enumerator Status + Social Proof Treatment + Official Gender + Enumerator Gender * Official Gender + ε_i

Log Odds Ratio of Probability of Appointments Held$_i$ = Enumerator Gender + Enumerator Status + Social Proof Treatment + Official Gender + Enumerator Gender * Official Gender + ε_i

Additionally, our final two-stage regression model is written as:

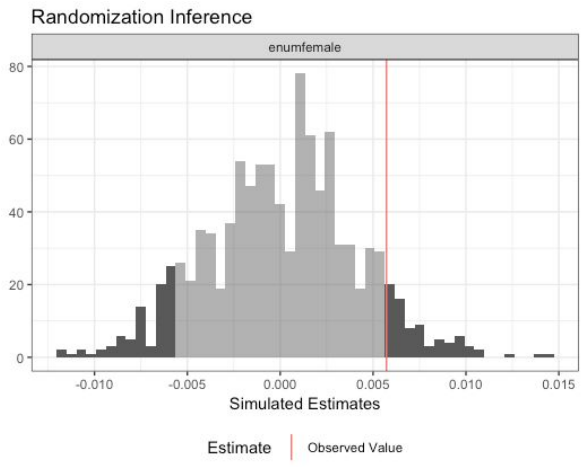
Appointments Held$_i$ = Enumerator Gender + Positive Response | (Enumerator Gender + Enumerator Status + Social Proof Treatment + Official Gender + Enumerator Gender * Official Gender + ε_i)

Results

We first report the results for the difference of means tests using randomization inference, which evaluated whether the gender of the enumerator affects the likelihood of a positive response and the likelihood of holding an appointment. Figures 1, 2, 3, and 4 plot the distribution of the difference of means values at each iteration for the randomization inference, with a red line indicating the difference of means we observed from our treatment assignment for gender. We find that gender is not significant for the overall analysis, or for India and Tanzania. However, the test for positive responses in Peru is significant (p-value=0.002). However, we find that the result is not in the predicted direction. We find that female enumerators are more likely to receive positive responses than male enumerators. While the others are not significant, we note that all of the positive response tests indicate that women were at least marginally more likely than men to obtain positive responses from the government officials.

Figure 1: Difference of Means Tests for Overall Data Using Randomization Inference

Positive Responses ~ Gender



Appointments Held ~ Gender

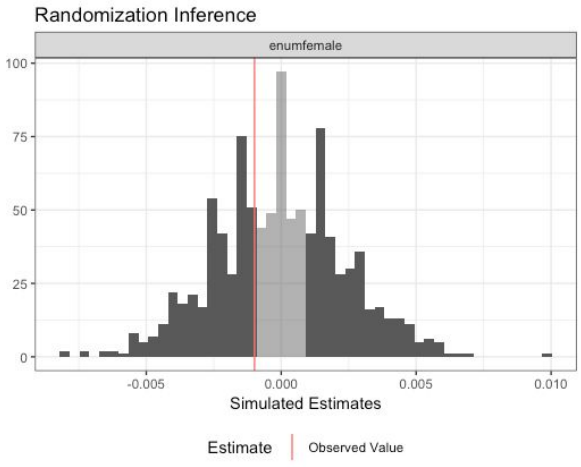
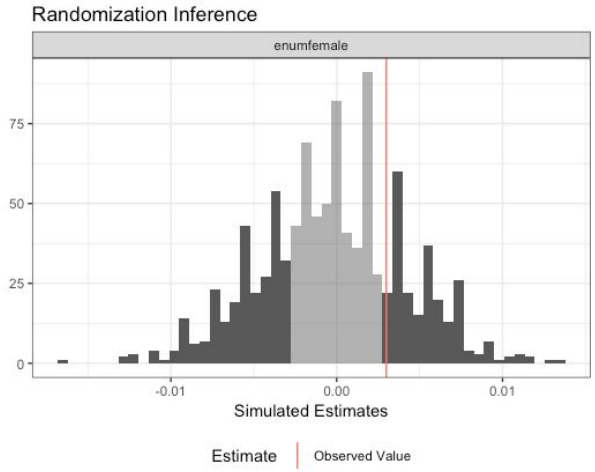


Figure 2: Difference of Means Tests for India Data Using Randomization Inference

Positive Responses ~ Gender



Appointments Held ~ Gender

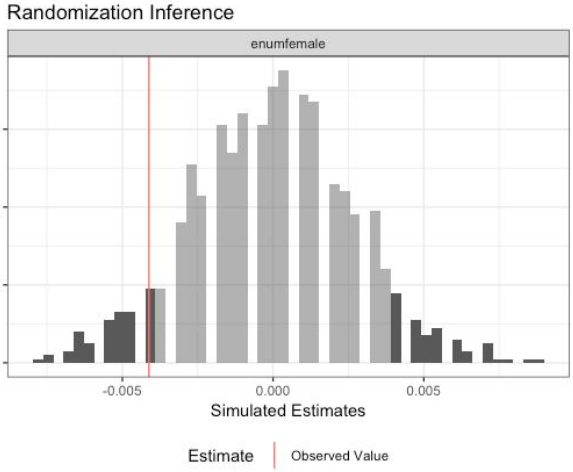
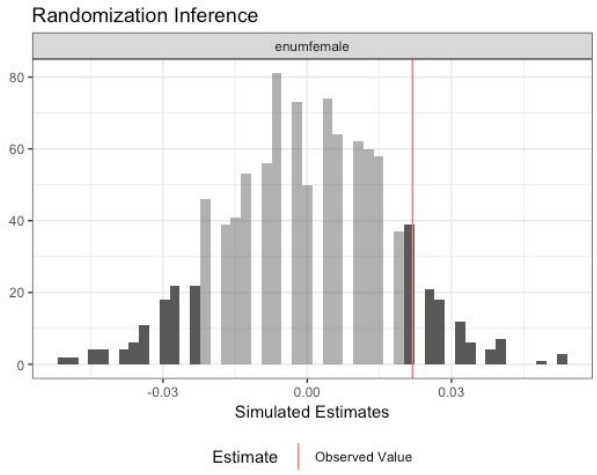


Figure 3: Difference of Means Tests for Tanzania Data Using Randomization Inference

Positive Responses ~ Gender



Appointments Held ~ Gender

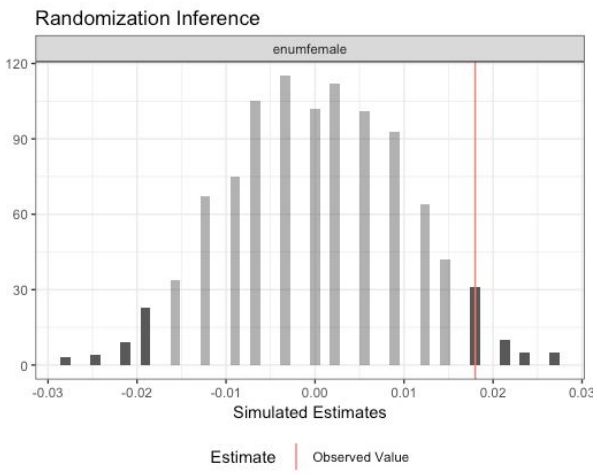
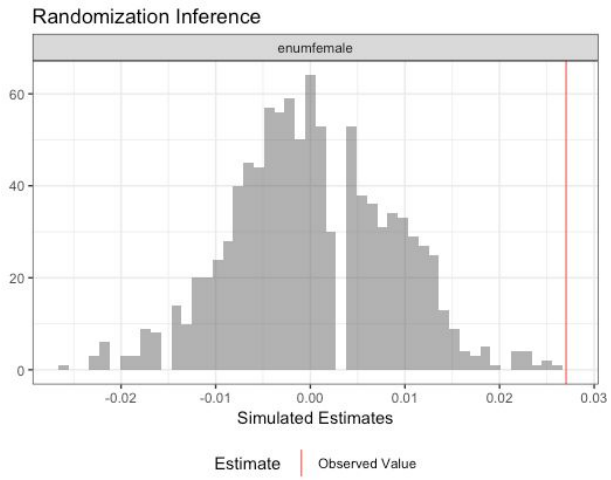
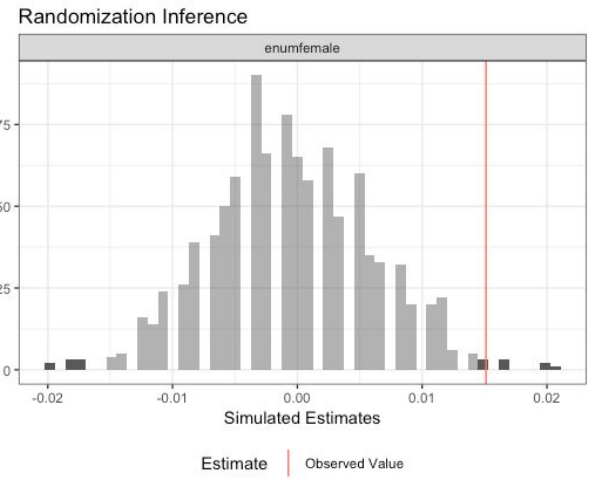


Figure 4: Difference of Means Tests for Peru Data Using Randomization Inference

Positive Responses ~ Gender



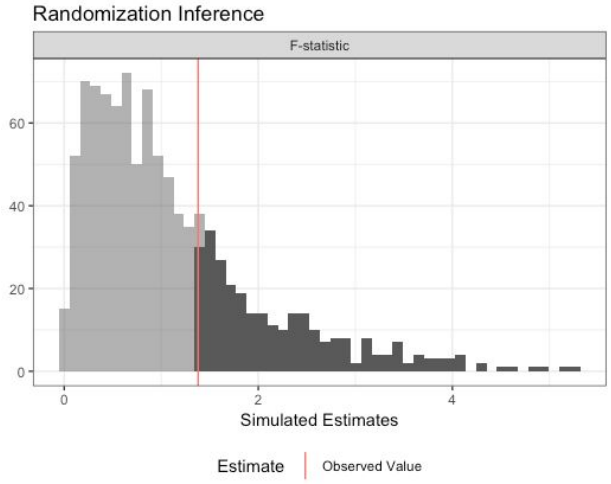
Appointments Held ~ Gender



We next performed the difference of means ANOVA tests using randomization inference to evaluate whether the interaction between the gender of the enumerator and the gender of the government official affects the likelihood of both a positive response and the likelihood of holding an appointment. The results are plotted in Figures 5, 6, 7, and 8. We find that the interaction between the gender of the enumerator and gender of the government official are not significant in any of the tests (at a=0.002). This indicates that the interaction of the genders of the official and enumerator are not significant factors in explaining our outcome measures.

Figure 5: ANOVA Difference of Means Tests for Overall Data Using Randomization Inference

Positive Responses ~ EnumGender*OfficialGender



Appointments Held ~ EnumGender*OfficialGender

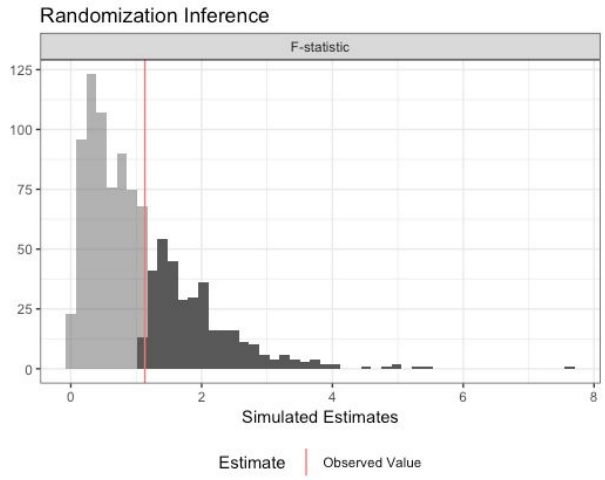
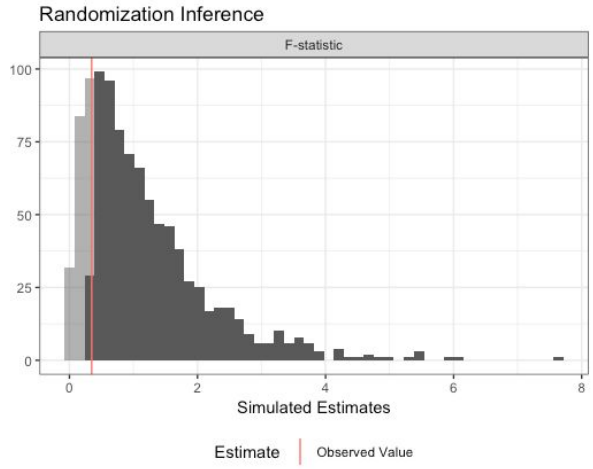


Figure 6: ANOVA Difference of Means Tests for India Data Using Randomization Inference

Positive Responses ~ EnumGender*OfficialGender



Appointments Held ~ EnumGender*OfficialGender

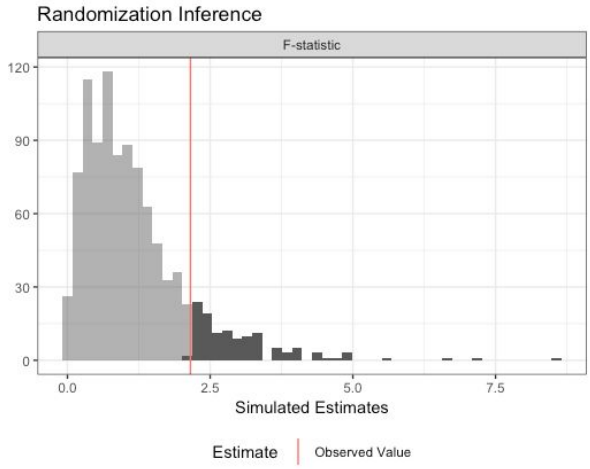
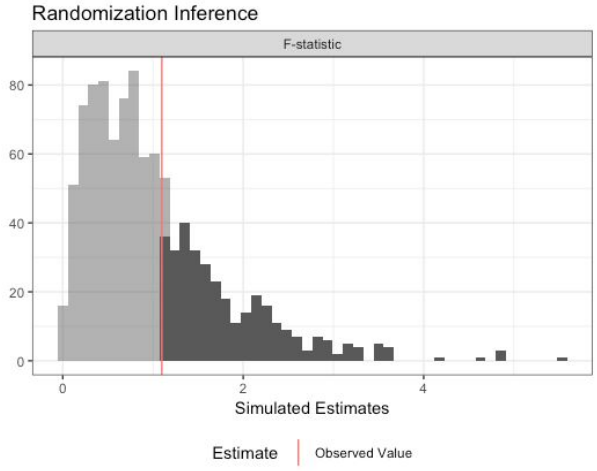


Figure 7: ANOVA Difference of Means Tests for Tanzania Data Using Randomization Inference

Positive Responses ~ EnumGender*OfficialGender



Appointments Held ~ EnumGender*OfficialGender

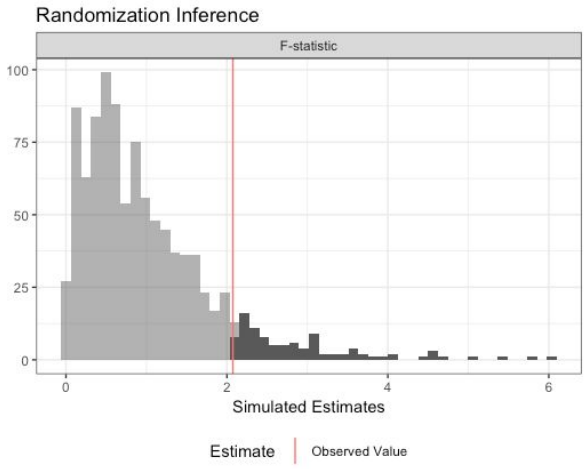
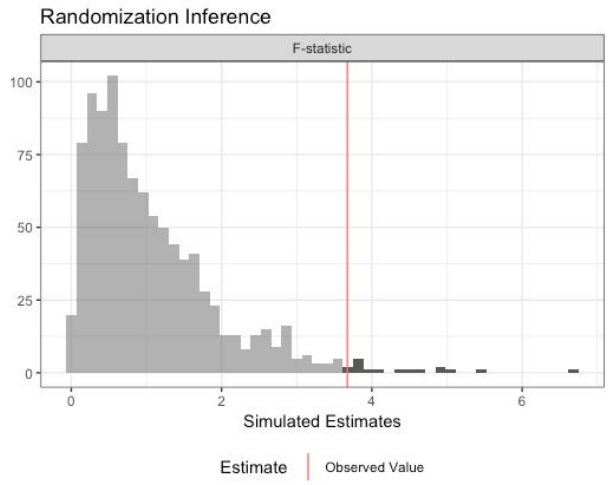
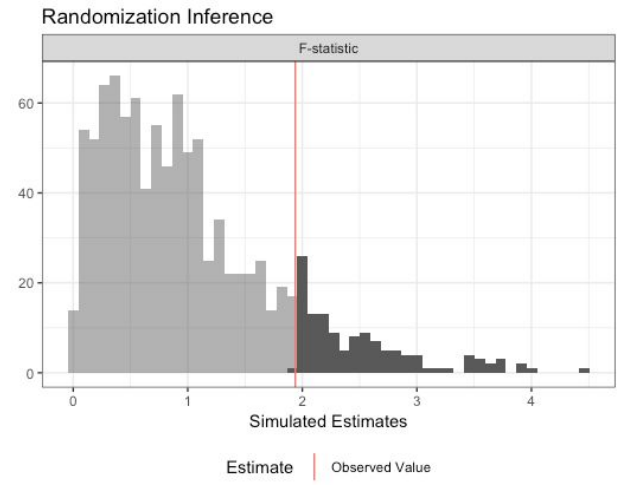


Figure 8: ANOVA Difference of Means Tests for Peru Data Using Randomization Inference

Positive Responses ~ EnumGender*OfficialGender



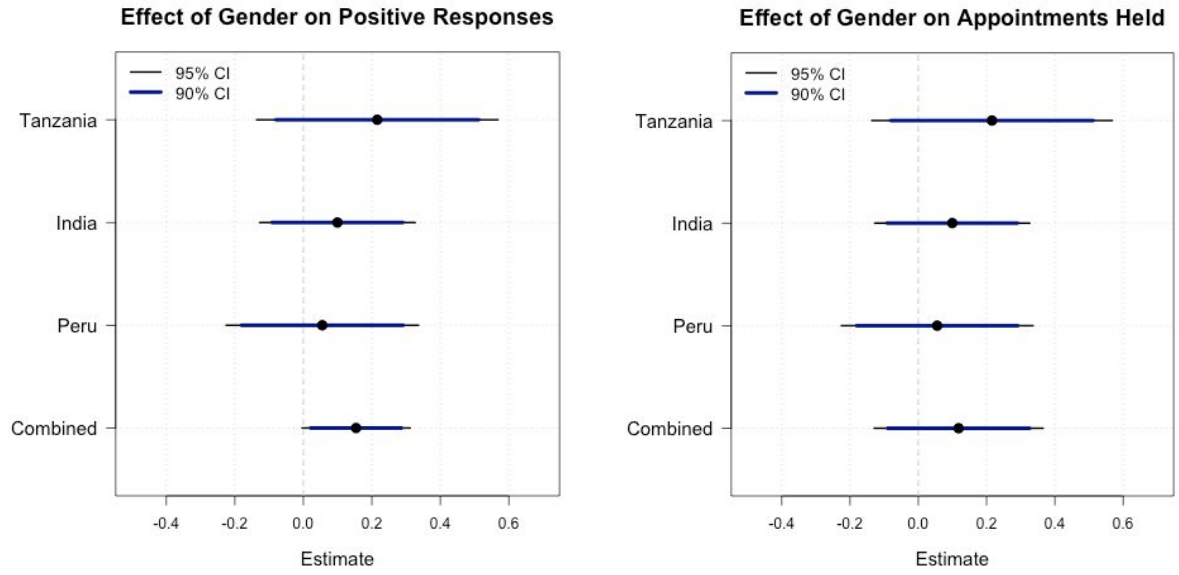
Appointments Held ~ EnumGender*OfficialGender



Additionally, we perform logistic regression analysis with the covariates specified and chosen in the previous section. The full results are displayed in Tables 2, 3, 4, and 5 in the Appendix. We provide plots in Figure 9 that show the estimates and uncertainty of the gender covariate for each of the regression analyses. We find that at the adjusted significance level, gender of the enumerator is not significant in any of the countries or overall once the other covariates are controlled for. However, we again note that all of the results are in the opposite direction from what we originally predicted. These results are informative and hopeful, given

that gender bias towards females does not appear to influence government officials at the level we predicted it would. In fact, we note that the coefficient for all of the logistic regression analyses are positive, indicating that women were actually more likely to both receive a positive response from a government official and to hold an appointment with an official.

Figure 9: Gender Covariate Estimates and Confidence Intervals from Regression Analyses



We lastly run our two-stage regression model to evaluate how the gender of the enumerator affects whether an appointment was held given that a positive response was already received. We report the regression coefficient values at the second stage of these models in Tables 10, 11, 12, and 13 in the Appendix. We again find that the enumerator gender is not significant in any of the individual countries or in the overall analysis. However, we are interested to find that coefficient for the gender of the enumerator is negative for both the overall second-stage analysis and the India second-stage analysis, which indicate that women were less likely to actually hold an appointment given that a positive response was received.

Conclusion

We were pleased to find that gender, in general, did not have a significant impact on the willingness of government officials to meet with enumerators. Additionally, that our efforts to reduce the barriers that officials face in using rigorous evidence in their decision making process was generally well received. The gender of the enumerator only significantly changed the likelihood of getting a positive response in Peru, and the coefficient was in the opposite direction from what we hypothesized. Thus, not only did we find a lack of evidence indicating gender biases among government officials in these three countries, but where we did see an effect of gender, it was in favor of females. Given the experimental method of this project, even null

results or significant results in the opposite direction from our original hypothesis teach us important information about the impact of gender on getting government officials willing to entertain new information.

We recognize that our study has many limitations, and we encourage more research to be done to better understand these effects. For example, we would like to see if there are connections between various demographics of government officials (such as their education level or status in the government) and their gender biases. We also recognize that this study may not fully explain or manifest the true gender biases existing in the government. For example, these biases may manifest themselves in other ways, such as a lack of willingness to respect some constituents' wishes (which may vary by gender) and by failing to address gender inequalities in society. Thus, we encourage future researchers to take these limitations and recommendations into account when seeking to better understand gender biases among government officials. Even so, our results are quite hopeful, indicating a lack of gender biases among government officials in these countries.

APPENDIX:

Table 2: Logistic Regression Results for Positive Responses: Overall Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-2.82	0.09	-30.62	0.00
Enumerator Gender	0.19	0.09	2.03	0.04
Enumerator Westerner	-0.44	0.08	-5.67	0.00
Official Gender	-0.05	0.13	-0.37	0.71
Social Proof	-0.39	0.08	-5.06	0.00
Country: Peru	0.49	0.09	5.67	0.00
Country: Tanzania	0.98	0.11	9.10	0.00
Enumerator Gender * Official Gender	0.08	0.17	0.46	0.65

Table 3: Logistic Regression Results for Appointments Held: Overall Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-4.10	0.15	-26.88	0.00
Enumerator Gender	0.24	0.14	1.69	0.09
Enumerator Westerner	-0.34	0.12	-2.83	0.00
Official Gender	0.20	0.18	1.09	0.27
Social Proof	-0.22	0.12	-1.84	0.07
Country: Peru	1.02	0.13	8.06	0.00
Country: Tanzania	0.78	0.19	4.17	0.00
Enumerator Gender * Official Gender	-0.36	0.25	-1.42	0.16

Table 4: Logistic Regression Results for Positive Responses: India Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-2.75	0.11	-24.65	0.00
Enumerator Gender	0.10	0.13	0.79	0.43
Enumerator Westerner	-0.21	0.11	-1.92	0.06
Official Gender	-0.09	0.21	-0.45	0.65
Social Proof	-0.66	0.11	-6.05	0.00
Enumerator Gender * Official Gender	0.08	0.26	0.33	0.74

Table 5: Logistic Regression Results for Appointments Held: India Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-3.89	0.19	-20.80	0.00
Enumerator Gender	0.01	0.21	0.05	0.96
Enumerator Westerner	-0.20	0.19	-1.07	0.28
Official Gender	0.31	0.29	1.04	0.30
Social Proof	-0.45	0.18	-2.44	0.01
Enumerator Gender * Official Gender	-0.89	0.44	-2.01	0.04

Table 6: Logistic Regression Results for Positive Responses: Tanzania Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-2.12	0.23	-9.33	0.00
Enumerator Gender	0.35	0.26	1.35	0.18
Enumerator Westerner	-0.47	0.19	-2.49	0.01
Official Gender	0.27	0.27	1.00	0.32
Social Proof	-0.05	0.18	-0.25	0.80
Enumerator Gender * Official Gender	-0.32	0.37	-0.85	0.39

Table 7: Logistic Regression Results for Appointments Held: Tanzania Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-3.96	0.47	-8.41	0.00
Enumerator Gender	0.88	0.50	1.78	0.07
Enumerator Westerner	-0.01	0.33	-0.04	0.97
Official Gender	0.61	0.53	1.14	0.25
Social Proof	-0.12	0.33	-0.36	0.72
Enumerator Gender * Official Gender	-0.80	0.68	-1.17	0.24

Table 8: Logistic Regression Results for Positive Responses: Peru Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-2.23	0.15	-14.66	0.00
Enumerator Gender	0.04	0.17	0.26	0.80
Enumerator Westerner	-0.82	0.15	-5.34	0.00
Official Gender	-0.18	0.21	-0.83	0.41
Social Proof	-0.13	0.14	-0.97	0.33
Enumerator Gender * Official Gender	0.29	0.29	1.01	0.31

Table 9: Logistic Regression Results for Appointments Held: Peru Data

	Estimate	Standard Error	z value	p-value
(Intercept)	-3.16	0.21	-15.07	0.00
Enumerator Gender	0.27	0.23	1.18	0.24
Enumerator Westerner	-0.43	0.19	-2.26	0.02
Official Gender	0.10	0.26	0.38	0.70
Social Proof	-0.03	0.18	-0.17	0.86
Enumerator Gender * Official Gender	-0.04	0.37	-0.10	0.92

Table 10: Two-Stage Regression - Second-Stage Results for Appointments Held: Overall Data

	Estimate	Standard Error	t value	p-value
(Intercept)	0.00	0.00	1.32	0.19
Enumerator Gender	-0.00	0.00	-1.28	0.20
Positive Response	0.36	0.04	8.64	0.00

Table 11: Two-Stage Regression - Second-Stage Results for Appointments Held: India Data

	Estimate	Standard Error	t value	p-value
(Intercept)	0.01	0.00	1.87	0.06
Enumerator Gender	-0.00	0.00	-1.86	0.06
Positive Response	0.23	0.08	2.86	0.00

Table 12: Two-Stage Regression - Second-Stage Results for Appointments Held: Tanzania Data

	Estimate	Standard Error	t value	p-value
(Intercept)	0.00	0.01	0.25	0.81
Enumerator Gender	0.01	0.01	1.64	0.10
Positive Response	0.13	0.14	0.91	0.36

Table 13: Two-Stage Regression - Second-Stage Results for Appointments Held: Peru Data

	Estimate	Standard Error	t value	p-value
(Intercept)	0.01	0.01	2.30	0.02
Enumerator Gender	0.01	0.01	1.27	0.20
Positive Response	0.31	0.10	3.12	0.00

References

Andreoni, James, and Lise Vesterlund. 2001. "Which is the Fair Sex? Gender Differences in Altruism." *The Quarterly Journal of Economics* 116 (1): 293-312. Accessed April 18, 2018. http://content.ebscohost.com/ContentServer.asp?EbscoContent=dGJyMNXb4kSep7Y4v%2BbwOLCmr1CeprFSrq%2B4TbaWxWXS&ContentCustomer=dGJyMPGuskyurK5IuePfgeyx43zx1%2BqE&T=P&P=AN&S=R&D=eoh&K=0563071.

Andreoni, James, and Ragan Petrie. 2008. "Beauty, Gender, and Stereotypes: Evidence from Laboratory Experiments." *Journal of Economic Psychology* 29 (1): 73-93. Accessed April 18, 2018. https://www.sciencedirect.com/science/article/pii/S0167487007000554.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124 (4): 1497-1540. Accessed April 17, 2018. https://academic.oup.com/qje/article/124/4/1497/1917190.

Broockman, David E., and David M. Butler. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55 (3): 463-477. Accessed April 18, 2018. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5907.2011.00515.x.

Buchan, Nancy R., Rachel T.A. Croson, and Sara Solnick. 2008. "Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game." *Journal of Economic Behavior and Organization* 68 (3-4): 466-476. Accessed April 18, 2018. https://www.sciencedirect.com/science/article/pii/S016726810800139X?_rdoc=1&_fmt=high&_origin=gateway&_docanchor=&md5=b8429449ccfc9c30159a5f9aeaa92ffb&ccp=y.

Burns, Nancy, Kay Lehman Schlozman, and Sidney Verba. 2001. *The Private Roots of Public Action.* Cambridge, Massachusetts: Harvard University Press. Accessed April 2018. https://books.google.com/books?hl=en&lr=&id=BM8YrTaavDwC&oi=fnd&pg=PR13&dq=Burns,+Schlozman,+and+Verba+&ots=9RzX8Ivw-b&sig=YlGOe69Jyer4IyUTcp5P5gC9fpI#v=onepage&q=Burns%2C%20Schlozman%2C%20and%20Verba&f=false.

Cox, James C. 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Groups Vs. Individuals and Males Vs. Females." *Experimental Business Research* 331-350. Accessed April 18, 2018. https://link.springer.com/chapter/10.1007/978-1-4757-5196-3_14.

Dreber, Anna, and Magnus Johannesson. 2008. "Gender Differences in Deception." *Economics Letters* 99 (1): 197-199. Accessed April 18, 2018. https://www.sciencedirect.com/science/article/pii/S0165176507002534.

Garcia-Retamero, Rocio, and Esther López-Zafra. 2006. "Prejudice Against Women in Male-Congenial Environments: Perceptions of Gender Role Congruity in Leadership." *Sex Roles* 55 (1/2): 51-61. Accessed April 18, 2018. http://content.ebscohost.com/ContentServer.asp?EbscoContent=dGJyMNXb4kSep7Y4v

%2BbwOLCmr1CeprBSsqi4TbCWxWXS&ContentCustomer=dGJyMPGuskyurK5IuePfgeyx43zx1%2BqE&T=P&P=AN&S=R&D=eft&K=507896998.

Hazarika, Sujata D. 2008. "Democracy and Leadership: The Gendered Voice in Politics." *Sociological Bulletin* 57 (3): 353-370. Accessed April 18, 2018. http://www.jstor.org/stable/23620806?seq=1#page_scan_tab_contents.

Innocenti, Alessandro, and Maria Grazia Pazienza. 2006. *Altruism and Gender in the Trust Game.* Working Paper, Department of Economics, University of Siena, Siena: Labsi: Experimental Economics Laboratory. Accessed April 18, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=884378.

Landry, Craig, Andreas Lange, John A. List, and Michael K., Rupp, Nicholas G. Price. 2005. *Toward and Understanding of the Economics of Charity: Evidence from a Field Experiment.* Working Paper, University of Chicago, Cambridge, MA: National Bureau of Economics Research, 1-52. Accessed April 18, 2018. http://www.nber.org/papers/w11611.pdf.

Prentice, Deborah A., and Erica Carranza. 2002. "What Women and Men Should Be, Shouldn't Be, Are Allowed to Be; and Don't Have to Be: The Contents of Prescriptive Gender Stereotypes." *Psychology of Women Quarterly* 26: 269-281. Accessed April 18, 2018. http://journals.sagepub.com/doi/pdf/10.1111/1471-6402.t01-1-00066.

Rudman, Laurie A., and Stephenie A. Goodwin. 2004. "Gender Differences in Automatic In-Group Bias: Why Do Women Like Women More Than Men Like Men?" *Journal of Personality & Social Psychology* (87): 494-509. Accessed April 18, 2018. http://sfx.lib.byu.edu/sfxlcl3?genre=article&isbn=&issn=00223514&doi=&title=Journal%20of%20Personality%20&%20Social%20Psychology&volume=87&issue=4&date=20041001&atitle=Gender%20Differences%20in%20Automatic%20In-Group%20Bias%3A%20Why%20Do%20Women%20Like%20.

UNDP. 2014. "Gender Inequality Index." *United Nations Development Programme.* Accessed April 18, 2018. http://hdr.undp.org/en/content/human-development-index-hdi.

—. 2016. "Human Development Report 2016: Human Development for Everyone." *The United Nations Development Program.* Accessed April 18, 2018. http://hdr.undp.org/sites/default/files/2016_human_development_report.pdf.

UNIFEM. 2008. "Who Answers to Women? Gender and Accountability." *United Nations Development Fund for Women.* Accessed April 17, 2018. http://www.unifem.org/progress/2008.

WomanStats Project. 2016. "Government Participation by Women, scaled 2016." *WomanStats Project.* Accessed April 18, 2018. http://www.womanstats.org/maps.html.

—. 2017. "Practice of Property Rights for Women: Scaled 2017." *WomanStats Project.* Accessed April 18, 2018. http://www.womanstats.org/maps.html.

—. 2015. "Son Preference and Sex Ratios: Scaled 2015." *WomanStats Project.* Accessed April 18, 2018. http://www.womanstats.org/maps.html.

World Bank. 2014. "GDP per Capita, PPP (Current International $)." *The World Bank: Data.* Accessed April 18, 2018. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD.

—. 2014. "Population, Total." *The World Bank: Data.* Accessed April 18, 2018. https://data.worldbank.org/indicator/SP.POP.TOTL.