ONLINE APPENDIX FOR: Explaining Women's Political Underrepresentation in Democracies with High Levels of Corruption

Aaron Erlich[1,2] and Edana Beauvais[3]

[1]Department of Political Science, McGill University

[2]Centre for the Study of Democratic Citizenship

[3]Department of Political Science, Simon Fraser University

V1.0

Contents

A In-Depth Theoretical Discussion of Scope Conditions

While many different types of regimes exist and the scale used will depend on the research question (Collier and Adcock 1999), to set out scope conditions for our paper, they can be broadly classified into democratic or authoritarian regimes (Alvarez et al. 1996).Again, while "corruption" varies by degree and the relevant typology will depend on the research question asked (Bussell 2015), for broad classification purposes, we can also dichotomize countries by their level of corruption (low levels of corruption/ high levels of corruption). These countries are those democracies with substantial violations of the rule of law and are often discussed as "electoral democracies" compared to "liberal democracies" where there are few violations of the rule of law (Klasnja and Pop-Eleches 2021).

In our present analysis, we limit our attention to electoral democracies and exclude non-democratic regimes because we are interested in the role of *voters' preferences* on *electoral outcomes*. While the preferences of citizens in non-democratic regimes may still influence rulers' political decisions (e.g., if rulers use policy responsiveness as a tool to maintain political order), authoritarian regimes (a) do not involve free and fair elections, and (b) citizens ultimately do not control collective decisions or agenda-setting. In authoritarian regimes, collective decision-making and agenda-setting—when and to what extent citizens' preferences are allowed to influence collectives—ultimately rests with the regime's power-holders (He and Warren 2011).[1] However, we will discuss the potential implications of our findings for non-democratic regimes in the conclusion.

While we have focused our analysis on the ways in which voter biases impact the electoral success of women candidates in countries where higher levels of corruption may create distinct gender-based opportunities and constraints, our findings may have implications for women's representation in authoritarian regimes. Voters' preferences are not the main force

[1] Another reason to limit the scope of our analysis to non-democratic regimes is our interest in the distinct relationship between corruption and democracy. While corruption harms both democracies and authoritarian regimes by creating inefficiencies (as rights or benefits become favors, to be paid for or repaid in kind), it is associated with additional harms that are specific to democracies (Warren 2004). As Warren (2004) explains, corruption in democracies breaks the link between decision-making and citizens' capacities to influence those decisions, reduces the domain of public agency, and undermines the culture of democracy (reducing trust, reciprocity, and increasing public cynicism—casting a shadow of doubt on both corrupt and non-corrupt politicians).

driving women's underrepresentation in highly corrupt, democratic contexts. Rather, it appears that women's underrepresentation in corrupt, democratic contexts is driven more by barriers that prevent women from winning party nominations and running for office in the first place, rather than overt discrimination at the polls. As such, women's underrepresentation is not a "democracy" problem (a problem that arises from empowering citizens to have a say over how they are governed). The same mechanisms that prevent women from running for office in democracies—such as the structural constraints imposed by gender inequality—may also prevent women from being appointed to office in non-democracies. Thus, in theory, solutions to achieving more women in political office, such as greater equality in terms of family-care responsibilities and women's empowerment in the economy, should be similar in democratic and non-democratic regimes. In practice, within democratic societies in which women have a say over their actions and the conditions of their actions, women can be expected to achieve greater eventual gains in social, economic, and political realms, than in the context of authoritarian regimes. This is an interesting area of future research.

B Supplementary Information on Ukraine

B.1 Ukraine's Quotas

Ukraine has no quotas for women in Parliament on the national level, and the idea remains controversial — though it does has significant support within the population (Martsenyuk 2015). Legislation for quotas at all levels was introduced in 2013 (see Haffert 2014) but never passed. There was also a recently implemented local election law that the requires that at least 30% of candidates participating in the elections must be women. However, as the International Foundation for Electoral Systems (IFES) (IFES 2015) documents, not having 30% women on the list has been found **not** to be grounds to deny registering a party's slate of candidates.

B.2 Women's Representation in Ukraine

Ukraine has changed electoral systems several times, so it is impossible to examine an unbroken time series. Birch (2003) has suggested that women politicians have fared better when Ukraine had a PR system and Thames (2018) has tested this using a Bayesian model and found evidence for a relationship between institutional design and women's representation. The theory that PR is causally linked with higher women's representation has found some support from the cross-national literature and within country case studies (see Paxton and Hughes 2016; Birch 2003).

The proportion of women representatives elected in single member districts (SMDs) in Ukraine's mixed-member proportional (MMP) system is much lower than the overall proportion of women's representation in the Parliament. Only four out of 198 (2%) of Ukraine's SMD representatives serving at the beginning 2017 were women. By comparison 9.5% of deputies elected in SMD districts in 1994 were women. This means that the women elected to Parliament in the 2014 elections, aside from the above-mentioned four, were elected from national lists.

B.3 Time Series of Ukrainian Women's Representation

The IPU, the source for the cross-national data in this paper, often records data on a monthly basis, though their reporting does not always occur at equal intervals. The IPU data is the source for Esarey and Schwindt-Bayer (2017), though this data is collapsed into Country-Year data. Figure B.1 shows all of the IPU data. The data show that, in the context of women's underrepresentation, a few bi-elections can change the percent women representation. While this may not make a large substantive difference, aggregating to yearly data obscures within legislative session variation.

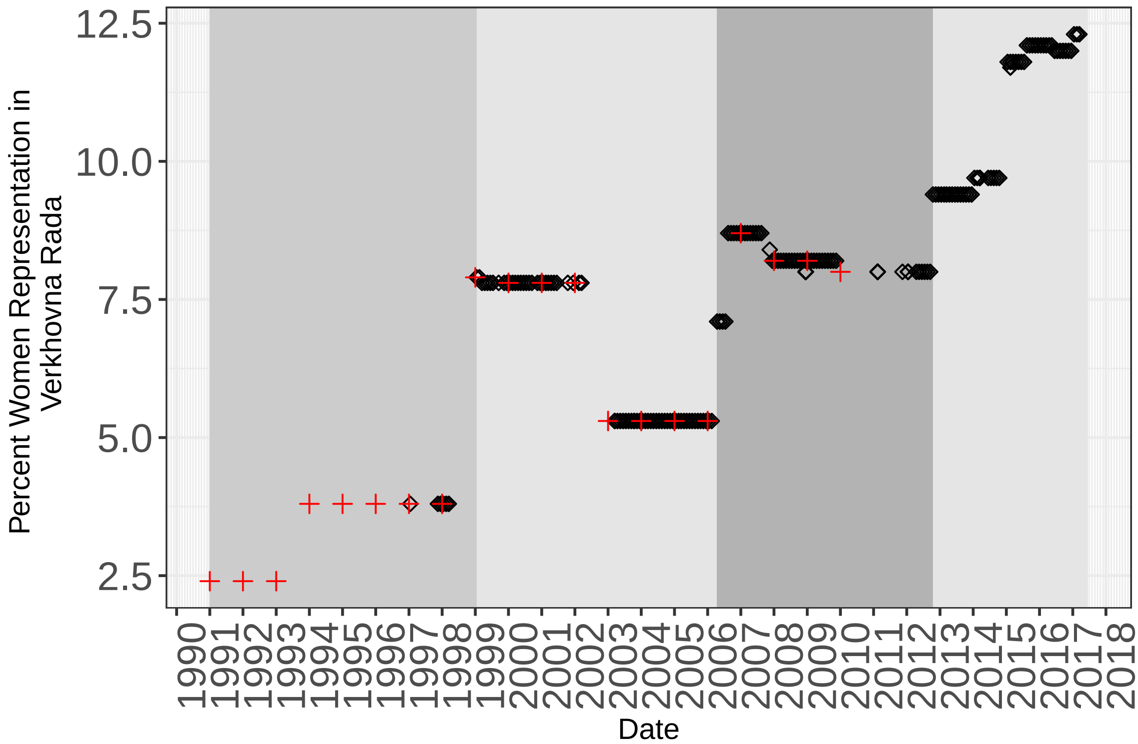

Figure B.1: Data show the collapsed (to yearly, with red plus symbols from Esarey and Schwindt-Bayer (2017)) and raw (generally monthly, with black diamonds) IPU data for Ukraine. Ukraine as changed its electoral system several times. The color in the background shows the type of electoral regime. Light grey represents an MMP system, dark grey a proportional system, and medium grey a SMP system.

C Study 1: Further Details of Survey Implementation and Design

Kyiv International Institute of Sociology (KIIS) carried out both waves of the face-to-face survey as part of a regular political survey, which collected independent (non-panel) samples. Both waves were carried out after the 2014 parliamentary elections (parliamentary elections occur every 5 years), and the Russian annexation of Crimea. Given the ongoing conflict in Eastern Ukraine, it is impossible to succinctly summarize all the events that occurred between the two survey waves, but key events can be viewed on *Ukraine Crisis Timeline* (http://ukraine.csis.org/).

Wave one was carried out between November 28, 2015 and January 5, 2016 and wave two was carried out between May 21 and June 15, 2016. The response rate for Kyiv only according to AAPOR standards was 49.8% in wave one of the experiment and 49% in wave two of the experiment. The entire study's response rate in Wave 2 was 50.8%.

Both waves of the Kyiv experiment were part of larger national survey. However, the gender experiment was only carried out in Kyiv and some pilot regions outside of Kyiv. For the national survey, the target population consisted of all adults aged 18 years and older, who were residents of one of the 24 Ukrainian oblasts or the city of Kyiv. Individuals permanently institutionalized in medical facilities, military quarters, and prisons were not included in the sample. The occupied territories (Autonomous Republic of Crimea, city of Sevastopol, and occupied parts of Donets'ka and Luhans'ka oblasts) and the contaminated territories around the Chornobyl Exclusion Zone were excluded from the sampling frame. Before drawing the sample population proportional to size (PPS), the populations were corrected based on the best extant estimates to account for internal and external migration as a result of the ongoing conflict in Donets'ka and Luhans'ka oblasts. Those who were internally displaced were added to the populations of recipient oblasts.

The survey employed a multi-stage cluster sample technique, where each oblast was divided into two or three strata: a rural stratum and an urban stratum, with any large cities over half a million potential respondents composing a third stratum (Kyiv was its own stratum). Oversamples were also conducted in some strata, so the sample was nationally not self-weighting nationally, but was self-weighting in Kyiv. All results are presented without survey weights for Kyiv, but we use survey weights in the robustness check in the pilot of villages. Within voting precincts, households were selected using a random route method, and individuals within a household were selected via a Kish Grid. Individuals for the experiment were randomized on the individual level.

D Validity of Images and Image Comparisons

D.1 Machine-learning Validation of Images

There are a range of ways researchers can cue candidate gender, such as by using feminine- versus masculine-sounding candidate names (Sapiro 1981), by using gendered pronouns (Barnes and Beaulieu 2014), or by asking respondents to judge images of actual politicians (Lawson et al. 2010; Carpinella and Johnson 2013; Johns and Shephard 2007; Carpinella et al. 2016). We choose to use images because seeing pictures should make the process of evaluating candidates more salient to survey respondents.

We used our fieldwork team to find models that were most similar in terms of attractiveness, ethnicity, age and hair color, given the model pool available. For the older couple, we chose dyed hair for the woman and baldness for the man based on assistants' judgments.

To validate our own judgment, we ran all of the images through the face++ API (www.faceplusplus.com) to judge the attractiveness of the images along with characteristics given to us by the algorithm: age, gender, and race. We find, in line with our own judgments, that each pair of images is similar and that the older pair is significantly less attractive. We do note that there somewhat of a difference in the API's rating of the "female beauty" of the brown-haired pair of candidates. We also note that the API misjudged the age of the bald older man candidate, likely because he is bald. The algorithm successfully predicted the gender and race of all the candidates.

Pair	Gender	Ethnicity	Male Beauty	Female Beauty	Age
brown	Man	White	84.5	83.4	34
brown	Woman	White	71.1	77.4	44
brunette	Man	White	81.8	84.8	34
brunette	Woman	White	78.1	82.0	41
older	Man	White	49.8	52.8	25
older	Woman	White	52.7	57.9	50

Table D.1: Results from the face++ app API, which uses machine learning to judge characteristics of images. We show the ethnicity, gender, attractiveness and, age of images

The face++ company documents their attractiveness measures in the following way:

Result of beauty analysis, including the following fields. The value of each field is a floating-point number with 3 decimal places between [0,100]. Higher beauty score indicates the detected face is more beautiful.

- male_score: male beauty score of the detected face
- female_score: female beauty score of the detected face

D.2 Verkhovna Rada Members in Similar Attire

These images are all taken from candidates official picture on an achived version of the Verkhovna Rada website: https://web.archive.org/web/20170714072050/http://gapp.rada.gov.ua/radatransl/Home/deps/en.

(a) Bakhteieva, Tetiana Dmytrivna; Opposition Bloc

(b) Katser-Buchkovska, Nataliia Volodymyrivna; People's Front

(c) Zalishchuk, Svitlana Petrivna, Petro Poroshenko Bloc

(d) Denysenko, Andrii Serhiiovych; non-affiliated

(e) Barvinenko Vitalii Dmytrovych, Revival

(f) Burbak Maksym Yuriiovych; People's Front

Figure D.2: Politicians across the Ukrainian political spectrum wear attire similar to the attire in our hypothetical images. Moreover, this clothing does not signify one particular party or position.

E Balance Tables

Table E.2: Orthogonality Table

	(1) 0	(2) 1	(3) Overall	(4) (1) vs. (2), p-value	(5) p-value from joint orthogonality test of treatment arms	(6) N from orthogonality test
♀ Respondent	0.63	0.59	0.61	0.37	0.37	466
	(0.03)	(0.03)	(0.02)			
♀ Interviewer	0.87	0.84	0.86	0.36	0.36	466
	(0.02)	(0.02)	(0.02)			
Age	49.78	48.89	49.32	0.62	0.62	466
	(1.27)	(1.28)	(0.90)			
> H.S. Ed.	0.92	0.90	0.91	0.64	0.64	464
	(0.02)	(0.02)	(0.01)			
Miss Cov.	0.00	0.01	0.00	0.17	0.17	466
	(0.00)	(0.01)	(0.00)			
Int. in Ukr.	0.31	0.32	0.32	0.88	0.88	466
	(0.03)	(0.03)	(0.02)			
Eth. Ukr.	0.96	0.93	0.94	0.15	0.15	466
	(0.01)	(0.02)	(0.01)			
Order 1	0.22	0.20	0.21	0.57	0.57	466
	(0.03)	(0.03)	(0.02)			
Order 2	0.07	0.16	0.12	0.00	0.00	466
	(0.02)	(0.02)	(0.01)			
Order 3	0.19	0.09	0.14	0.00	0.00	466
	(0.03)	(0.02)	(0.02)			
Order 4	0.19	0.26	0.23	0.08	0.08	466
	(0.03)	(0.03)	(0.02)			
Order 5	0.19	0.14	0.16	0.20	0.20	466
	(0.03)	(0.02)	(0.02)			
Order 6	0.15	0.15	0.15	0.90	0.90	466
	(0.02)	(0.02)	(0.02)			
N	226	240	466			

Wave 1 of survey data collection. Note that all the orders were not equally likely because of an issue with ODK, and two were out of balance.

Table E.3: Orthogonality Table

	(1) 0	(2) 1	(3) Overall	(4) (1) vs. (2), p-value	(5) p-value from joint orthogonality test of treatment arms	(6) N from orthogonality test
♀ Respondent	0.62	0.56	0.59	0.24	0.24	337
	(0.04)	(0.04)	(0.03)			
♀ Interviewer	0.88	0.87	0.88	0.92	0.92	337
	(0.02)	(0.03)	(0.02)			
Age	45.91	43.40	44.73	0.20	0.20	337
	(1.36)	(1.39)	(0.97)			
> H.S. Ed.	0.94	0.98	0.96	0.05	0.05	336
	(0.02)	(0.01)	(0.01)			
Miss Cov.	0.00	0.01	0.00	0.29	0.29	337
	(0.00)	(0.01)	(0.00)			
Int. in Ukr.	0.30	0.26	0.28	0.39	0.39	337
	(0.03)	(0.03)	(0.02)			
Eth. Ukr.	0.91	0.87	0.89	0.27	0.27	337
	(0.02)	(0.03)	(0.02)			
Order 1	0.15	0.18	0.16	0.51	0.51	337
	(0.03)	(0.03)	(0.02)			
Order 2	0.17	0.21	0.19	0.41	0.41	337
	(0.03)	(0.03)	(0.02)			
Order 3	0.17	0.22	0.19	0.33	0.33	337
	(0.03)	(0.03)	(0.02)			
Order 4	0.12	0.00	0.07	0.00	0.00	337
	(0.02)	(0.00)	(0.01)			
Order 5	0.21	0.15	0.18	0.11	0.11	337
	(0.03)	(0.03)	(0.02)			
Order 6	0.17	0.25	0.21	0.05	0.05	337
	(0.03)	(0.03)	(0.02)			
N	179	158	337			

Wave 2 of survey data collection. Note that all the orders were not equally likely because of an issue with ODK, and three were out of balance.

Table E.4: Orthogonality Table

	(1) 0	(2) 1	(3) Overall	(4) (1) vs. (2), p-value	(5) p-value from joint orthogonality test of treatment arms	(6) N from orthogonality test
♀ Respondent	0.61	0.66	0.63	0.17	0.17	594
	(0.03)	(0.03)	(0.02)			
♀ Interviewer	0.97	0.96	0.96	0.95	0.95	594
	(0.01)	(0.01)	(0.01)			
Age	48.80	48.16	48.53	0.66	0.66	594
	(0.93)	(1.14)	(0.72)			
> H.S. Ed.	0.56	0.60	0.57	0.33	0.33	592
	(0.03)	(0.03)	(0.02)			
Miss Cov.	0.00	0.00	0.00	0.82	0.82	594
	(0.00)	(0.00)	(0.00)			
Int. in Ukr.	0.18	0.14	0.16	0.18	0.18	594
	(0.02)	(0.02)	(0.02)			
Eth. Ukr.	0.66	0.69	0.68	0.46	0.46	594
	(0.03)	(0.03)	(0.02)			
Order 1	0.15	0.24	0.19	0.01	0.01	594
	(0.02)	(0.03)	(0.02)			
Order 2	0.18	0.20	0.19	0.44	0.44	594
	(0.02)	(0.03)	(0.02)			
Order 3	0.20	0.24	0.22	0.22	0.22	594
	(0.02)	(0.03)	(0.02)			
Order 4	0.11	0.00	0.06	0.00	0.00	594
	(0.02)	(0.00)	(0.01)			
Order 5	0.17	0.12	0.15	0.09	0.09	594
	(0.02)	(0.02)	(0.01)			
Order 6	0.19	0.20	0.19	0.79	0.79	594
	(0.02)	(0.03)	(0.02)			
N	343	251	594			

Wave 2 of survey data collection for the non-Kyiv rural sample. Note that all the orders were not equally likely because of an issue with ODK, but two were out of balance.

F Discussion of Differences in Means

Considering each of the paired photos separately, voters do seem to indicate a lower willingness to vote for the woman candidate in the third set of photos. However, this effect is not replicated in the other pairs of photos nor is it evident when considering the average of all three pairs of candidates, as shown in a simplified form in Figure F.3. Further statistical analysis reveals that any small average negative effect of the women candidate photos, if it exists, is less than one unit on the ten-point scale (see Appendix G.4).

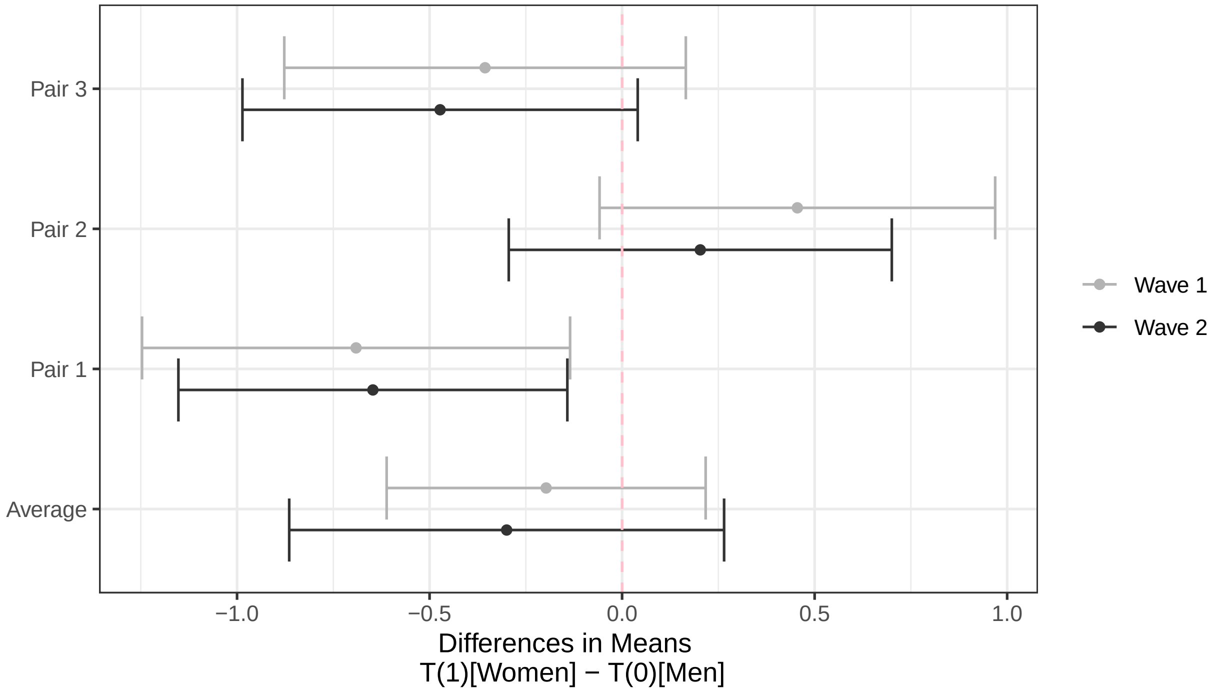

Figure F.3: The difference in means between men candidates and women candidates, where a positive number represents the women candidate being rated higher. Confidence intervals are associated with respective t-tests.

G Regression Tables

G.1 Robustness Checks Average Pic Rating

	Wave 1	Wave 2	Wave 1	Wave 2	Wave 1	Wave 2
Constant	4.29^*	4.17^*	2.78^*	5.16^*	2.84^*	5.15^*
	(0.15)	(0.20)	(0.99)	(1.22)	(0.99)	(1.23)
Women Photo = 1	−0.20	−0.30	−0.07	−0.17	−0.28	−0.13
	(0.21)	(0.29)	(0.22)	(0.30)	(0.35)	(0.44)
Woman Resp.			−0.03	1.02^*	−0.20	1.06^*
			(0.22)	(0.29)	(0.33)	(0.39)
Women Photo × Woman Resp.					0.34	−0.07
					(0.43)	(0.58)
Order effects	No	No	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes
R^2	0.00	0.00	0.05	0.11	0.05	0.11
Adj. R^2	−0.00	0.00	0.03	0.07	0.03	0.07
Num. obs.	466	337	464	336	464	336
RMSE	2.27	2.63	2.24	2.52	2.24	2.53

$^*p < 0.05$

Table G.5: Coefficients from least squares models where the outcome variable is the average rating of three pictures. Robust standard errors in parentheses.

G.2 Robustness Checks Average Pic Rating in Rural Areas

	Model 1	Model 2	Model 3
Constant	3.55^{***}	5.36^{***}	3.32^{*}
	(0.63)	(1.50)	(1.50)
Women Photo = 1	-0.42	-0.48	-0.51
	(0.43)	(0.41)	(0.41)
Woman Resp.	-0.37	-0.26	-0.20
	(0.35)	(0.33)	(0.33)
Order effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Sample Region FE	No	No	Yes
Num. obs.	594	592	592
RMSE	77.81	76.59	75.72
N Clusters	137	137	137

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table G.6: Coefficients from least squares models where the outcome variable is the average rating of three pictures. 95% confidence intervals based on robust standard errors in parentheses. Models are run only on Wave 2 where the experiment was piloted in villages. Fixed effects for macroregion are included in analysis in some models but omitted from output. Regressions account for survey design.

G.3 Robustness Checks Individual Pic Rating

	Pair 1-W1	Pair 1-W2	Pair2-W1	Pair2-W2	Pair 3-W1	Pair 3-W2
Constant	3.32^{*}	5.80^{***}	3.38^{**}	2.22^{*}	1.88	5.17^{***}
	(1.36)	(1.09)	(1.25)	(0.99)	(1.19)	(1.18)
Women Photo = 1	-0.56	-0.44	0.60^{*}	0.29	-0.19	-0.34
	(0.30)	(0.25)	(0.27)	(0.27)	(0.27)	(0.26)
Woman Resp.	-0.52	0.90^{***}	0.51	0.80^{**}	0.02	0.77^{**}
	(0.29)	(0.25)	(0.27)	(0.25)	(0.27)	(0.26)
Order effects	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	491	466	489	421	483	437
RMSE	3.10	2.62	2.88	2.52	2.89	2.69

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table G.7: Coefficients from least squares models where the outcome variable is the individual rating of each picture. 95% confidence intervals in brackets are shown based on robust standard errors.

G.4 Negligible Effect

We can use 90% confidence intervals, which are equivalent to a Two One-Sided *t*-test (TOST) to put bounds on the size of effect we can reject on the basis of our study (Rainey 2014). In all of our models, the effect is negative but statistically not significant. We set an effect of size one to statistically rule out such a large effect size.

That is, while the scale of 0-10 has no inherent meaning, we specify that negligible effect will be less that one point on this scale. We choose one point because we are looking for effects that are large and would make gender an important consideration in the evaluation process.

Using 90% confidence intervals provides statistical evidence for ruling out the possibility that effect size of the woman candidate photas is greater than one for both waves.[2] We show this finding in Figure G.4, along the 90% confidence intervals for the three regression models discussed in the papeer (no covariates, a vector of demographic controls, and a vector of demographic controls and the interaction between gender of respondent and gender of candidate photos). We see in Figure G.4 that regardless of model specification, that while we cannot rule out a small negative effect size we can rule out that effect size is greater than one.

[2] An effect size of one is equivalent to a Glass's Δ of .44 in both rounds of the survey in Kyiv.

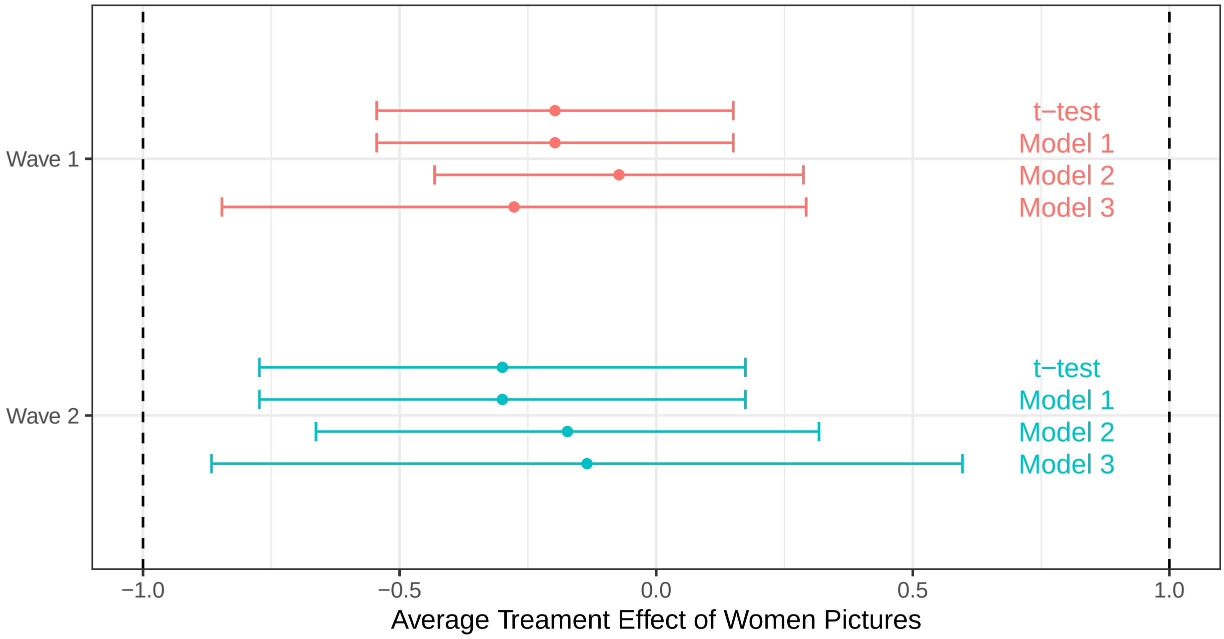

Figure G.4: The estimates are shown with 90% confidence intervals, which are equivalent to a TOST-test. Under no modeling assumptions do we see evidence to demonstrate that the treatment effect is greater than one.

H Differential Non-Response by Treatment Group

One assumption of any survey experiment with missing data is that any item non-response observed in the survey is independent of treatment status. This assumption is sometimes known as missingness independent of potential outcomes (MIPO). We examine how treatment status affects non-response to our questions about the randomly assigned gender of hypothetical images of candidates in both waves of the data collection. We classify item-non response as a respondent not giving a numerical answer to at least one of the three outcomes (their self-reported willingness to vote for each of the hypothetical candidates). As seen in Table H.8, we find no evidence that there is a difference in non-response between our men's and women's photo treatment in Wave 1 of the experiment. In Wave 1, the item non-response rate was lower for both treatment arms than in Wave 2. In Wave 2, we do find some evidence of item non-response being different between our men's and women's photo treatment.

Treatment	Wave 1	Wave 2
Men's photos	0.78	0.61
Women's photos	0.75	0.51

Table H.8: Item non-response rate for the ratings of the three experimental assigned pictures (Women's or Men's)

One approach to exploring this missingness is to examine whether there is a statistical relationship between observables that could theoretically differentially affect missingness and an interaction between observables and the treatment when predicting missingness. To formally test this, this we can compare a model that predicts missingness based on a vector of observables $\mathbf{B}_k$ ($Y_{missing} = \alpha + \sum B_k + \varepsilon$) with a model interacting an indicator of treatment status (D) and the same vector of observables (($Y_{missing} = \alpha + \sum B_k \times D + \varepsilon$)).

We choose five observable variables that could theoretically be related to treatment status. These include age, gender of respondent, whether respondents thought corruption was an important issue, education, and the gender of the interviewer. To examine whether missingness is differentially predicted by observables in our two treatment groups we run logistic regressions, as shown in Table H.9 and then conduct an F-test comparing the models with and without interactions.

As shown by the F-tests in both waves, we do not find that our theoretically relevant variables differentially affect missingness. What we do find in Wave 2 is that women interviewers have overall higher item response rates for the ratings of our three questions — though this is not true in Wave 1. It is possible that more of the women interviewers in Wave 2 were better trained so a greater number of respondents felt comfortable answering all of the questions from them.

Despite the fact there is no statistically significant difference between our models predicting missingness with and without interactions with treatment status, we conduct an additional robustness check by relaxing the MIPO assumption and instead making the assumption of MIPO$|\mathbf{X}$. As discussed in (Gerber and Green 2012), we now only assume missingness independent of potential outcomes conditional on a vector of observables. To do so we implement inverse probability weighting using the same assignment model in columns 4 of Table H.9 and then run a linear regression with the weights from this assignment model.

As shown in Table H.10, again we find no difference on average between the two treatment groups. The coefficient on our Women Candidate Treatment is very close to zero and not-statistically significant, giving us further assurance that our findings are not driven by missingness in our outcome variable, particularly in Wave 2.

	1 = Response to all three photo questions			
	Wave 1	Wave 1	Wave 2	Wave 2
Women Photo Treatment	-0.219	-0.879	-0.433^{*}	-0.809
	(0.199)	(1.420)	(0.172)	(1.052)
Age	-0.001	-0.001	0.005	0.006
	(0.005)	(0.008)	(0.005)	(0.007)
Woman Respondent	-0.179	0.061	-0.015	0.045
	(0.208)	(0.315)	(0.174)	(0.258)
Corruption Important	0.098	0.279	0.234	0.404
	(0.203)	(0.314)	(0.173)	(0.253)
Ed. > H.S.	0.705^{*}	0.793	0.647	0.366
	(0.276)	(0.413)	(0.378)	(0.473)
Interviewer Woman	-1.059^{*}	-1.999	1.178^{*}	1.206^{*}
	(0.414)	(1.032)	(0.219)	(0.312)
Treatment × Woman		-0.392		-0.110
		(0.421)		(0.351)
Treatment × Age		0.000		-0.002
		(0.011)		(0.010)
Treatment × Corruption Important		-0.311		-0.313
		(0.414)		(0.347)
Treatment × Education > High School		-0.175		0.798
		(0.557)		(0.837)
Treatment × Woman Interviewer		1.245		-0.083
		(1.133)		(0.439)
Constant	1.822^{*}	2.425^{*}	-1.404^{*}	-1.316^{*}
	(0.612)	(1.211)	(0.503)	(0.667)
AIC	642.665	649.431	793.882	802.095
BIC	673.467	702.234	824.649	854.838
Log Likelihood	-314.332	-312.715	-389.941	-389.048
Deviance	628.665	625.431	779.882	778.095
Num. obs.	602	602	599	599

$^{*}p < 0.05$

Table H.9: Logistic regression where the dependent variable is presence of responses for all three pictures. Models test for differences between treatment and control in response patterns. While the presence of women interviewers (compared to men) predict higher item response in Wave 2. F-tests from both waves ($Pr(> F) = .72$ and $Pr(> F) = .88$) fail to reject the null hypothesis that observable covariates in the two treatment arms do not differentially affect missingness.

	Model 1
Constant	3.96^{*}
	(0.21)
Women Photo Treatment	0.04
	(0.31)
R^2	0.00
Adj. R^2	-0.00
Num. obs.	336
RMSE	3.57

$^{*}p < 0.05$

Table H.10: Regression using IPW to weight for missingness conditional on observables.

I Conjoint

I.1 Descriptive Statistics

Table I.11: Descriptive Statistics

	Unweighted				Weighted	
Variable	Mean	Std. Dev	Min	Max	Mean	Std. Dev
Age	37.65	9.91	18	98	33.37	9.86
Children in HH	0.57	0.50	0	1	0.54	0.50
Woman	0.62	0.49	0	1	0.58	0.49
Income Scale	1.54	0.83	0	3	1.54	0.85
Married/Common Law	0.70	0.46	0	1	0.62	0.48
Single	0.21	0.40	0	1	0.31	0.46
Widowed	0.02	0.13	0	1	0.01	0.11
Divorced	0.07	0.25	0	1	0.05	0.23
Survey Russian	0.52	0.50	0	1	0.54	0.50

I.2 Names used in Conjoint

Table I.12: Attributes for Candidate Profiles in Conjoint Experiment (with all Names Included in the Experiment)

Attribute	Attribute Level	
Name	Ukrainian Names	
	Ukrainian Women's Names	*Ukrainian Men's Names*
	Lesia Shpak	Liubomyr Kulyk
	Ivanna Martseniuk	Mykhailo Humeniuk
	Mariia Protsiv	Vasyl Stefaniv
	Solomiia Didukh	Zynovii Shymanskyi
	Mariana Stanko	Ostap Stetsko
	Daryna Halytska	Maksym Zaporozhets
	Roksolana Kovalchuk	Nazar Boichuk
	Myroslava Kulish	Stepan Yarema
	Russian Names	
	Russian Women's Names	*Russian Men's Names*
	Larysa Kuznetsova	Viktor Lebedev
	Valentyna Iershova	Oleksii Fedorov

Halyna Nikitina	Hlib Smyrnov
Liudmyla Kalinina	Denys Krupenin
Aliona Tokareva	Andrii Ilyin
Liubov Kondratieva	Ivan Nikonov
Anastasia Volkova	Artem Popov
Lidiia Beliaieva	Leonid Ivanov
Neutral Names	
Neutral Women's Names	*Neutral Men's Names*
Olha Radchenko	Vladyslav Petrenko
Tetiana Shevchuk	Oleksandr Bondar
Iuliia Melnychenko	Hennadii Antonenko
Anna Vasylevska	Iurii Chyzhevskyi
Nataliia Rybak	Hryhorii Polishchuk
Anastasiia Khomych	Ihor Popovych
Kateryna Dub	Iaroslav Grushyn
Iryna Kolos	Oleh Ruban

I.3 Diagnostics

We conduct an F-test where we test for carryover effects on the AMCEs based on the order of the choice task. The F-test ($p = .13$) fails to reject the null hypothesis that the the order of the choice task does not matter.

We also conduct an F-test for profile order effects. This F-test also fails to reject the null hypothesis that the profile order had no effect ($p = .07$) on the AMCEs.

I.4 Marginal Means

In this plot, we show the overall weighted marginal means for the variables where there were no constraints (we exclude Age and Education). The plot clearly highlights how respondents more often say they will vote for candidates who campaign on "Fighting corruption." The plot also highlights the relative preference for married candidates over single or divorced candidates, and the preference for candidates with children.

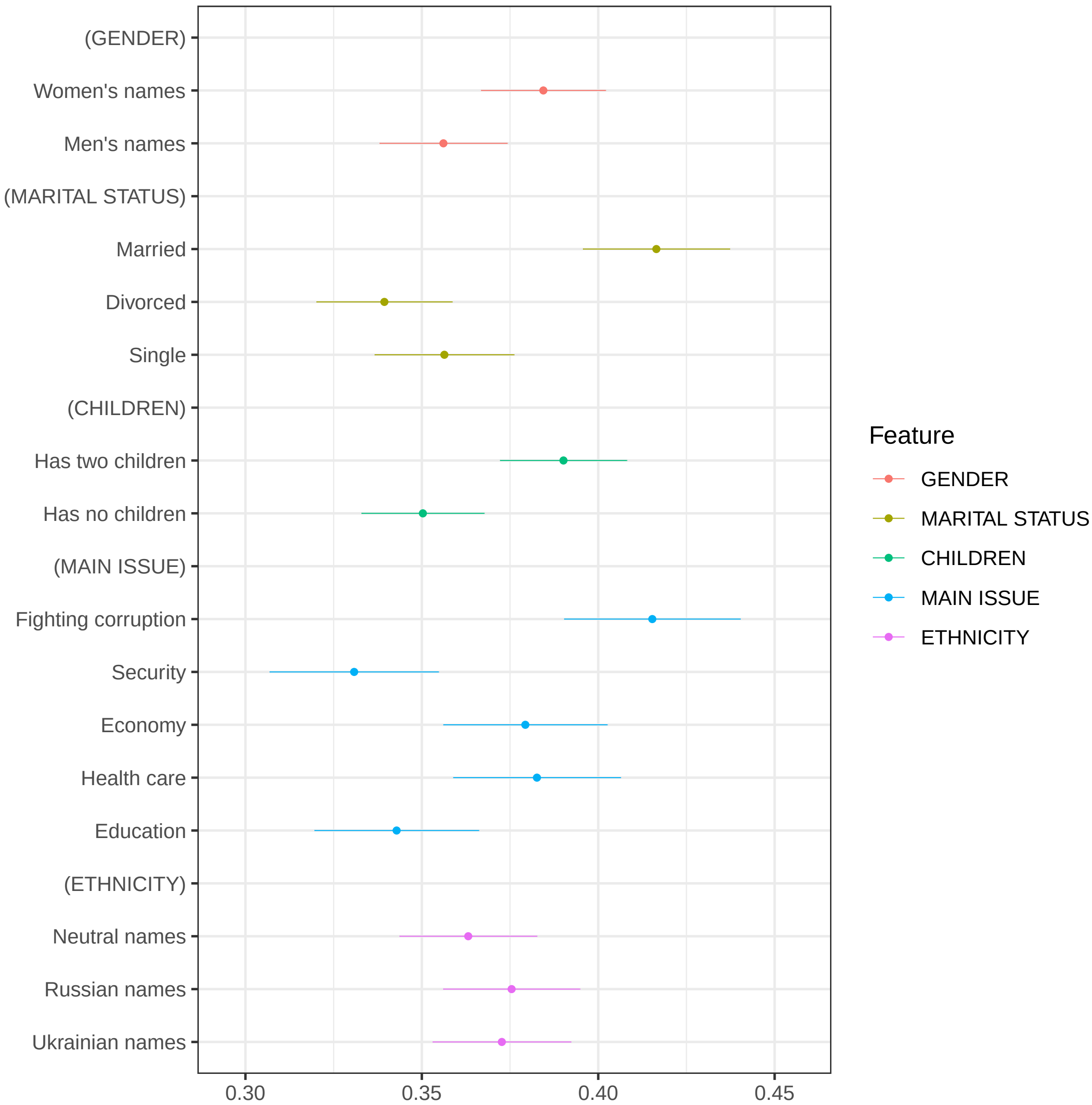

Figure I.5: Survey-weighted marginal means for vote variable

I.5 Weight construction

The weights were created in the data to make the data representative of the Ukrainian non-village population. The weights were created to take into consideration differential panel response and were then raked on five demographic variables based on census population estimates. They were created using the following five steps.

1. The survey company provided a dataset of invitations with panel social demographics which included a variable indicating response/non-response, and a unique respondent ID linking invitations to the survey data. Sampling invitations were one-stage random sample, so weight == 1 for everyone at this stage.

2. Fit a logistic model of invitation response as an outcome using invitations' demographic categories as predictors.

3. Construct five weight classes based on predicted probabilities of responding to the survey.

4. Assign each respondent a weight that is the mean predicted probability of its weight class. The weights were adjusted so that they sum to the total sample size.

5. Finally, the weights were raked with R's survey package (Lumley 2010) based on the following variables: gender, age, settlement size, region, education and income. The final raked weight is the weight used in the weighted analysis.

I.6 Unweighted analysis

In this appendix, we show the results without the survey weights applied. We show that we achieve nearly identical results to those shown in the main body of the paper.

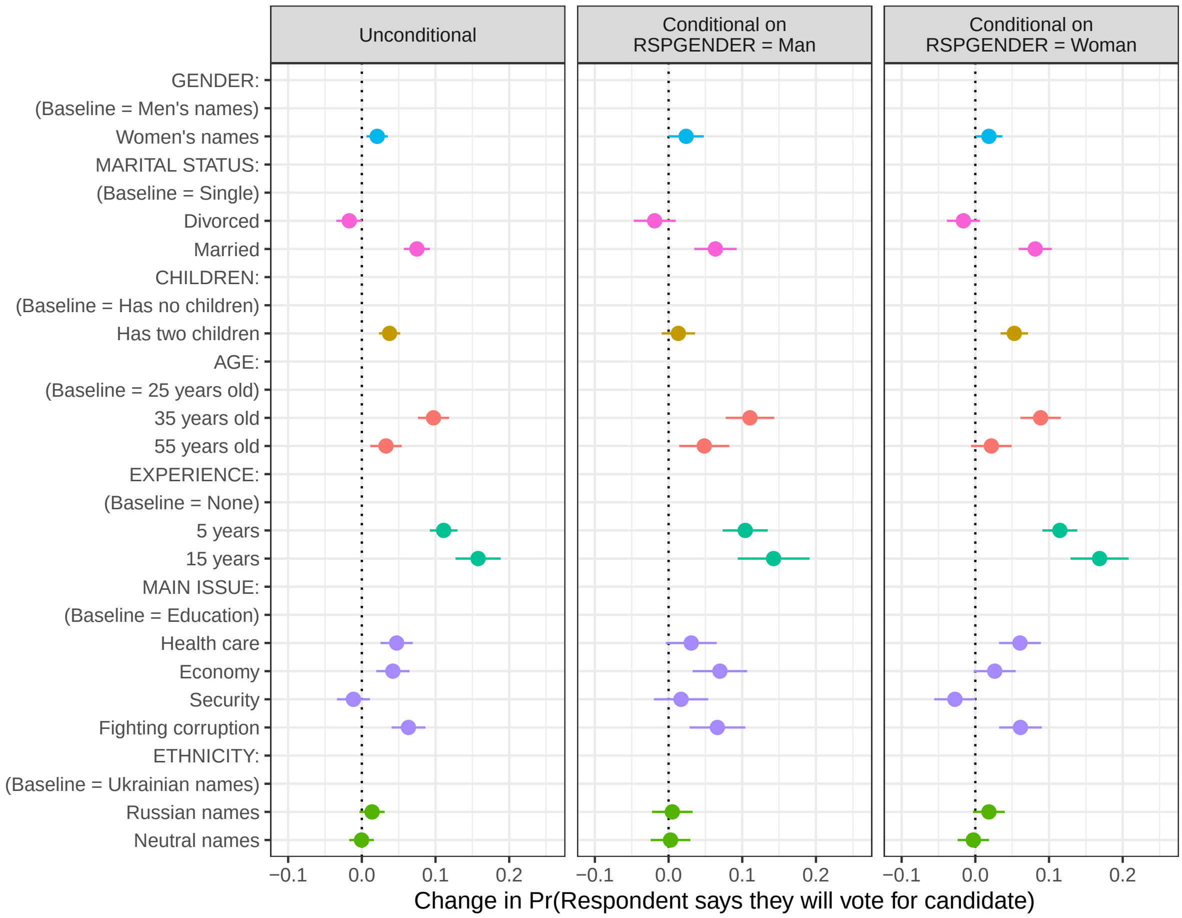

Figure I.6: AMCEs for Vote (Unconditional and Conditional on Respondent Gender with survey weights (unweighted))

I.7 Differences in Family Attributes by Candidate Gender

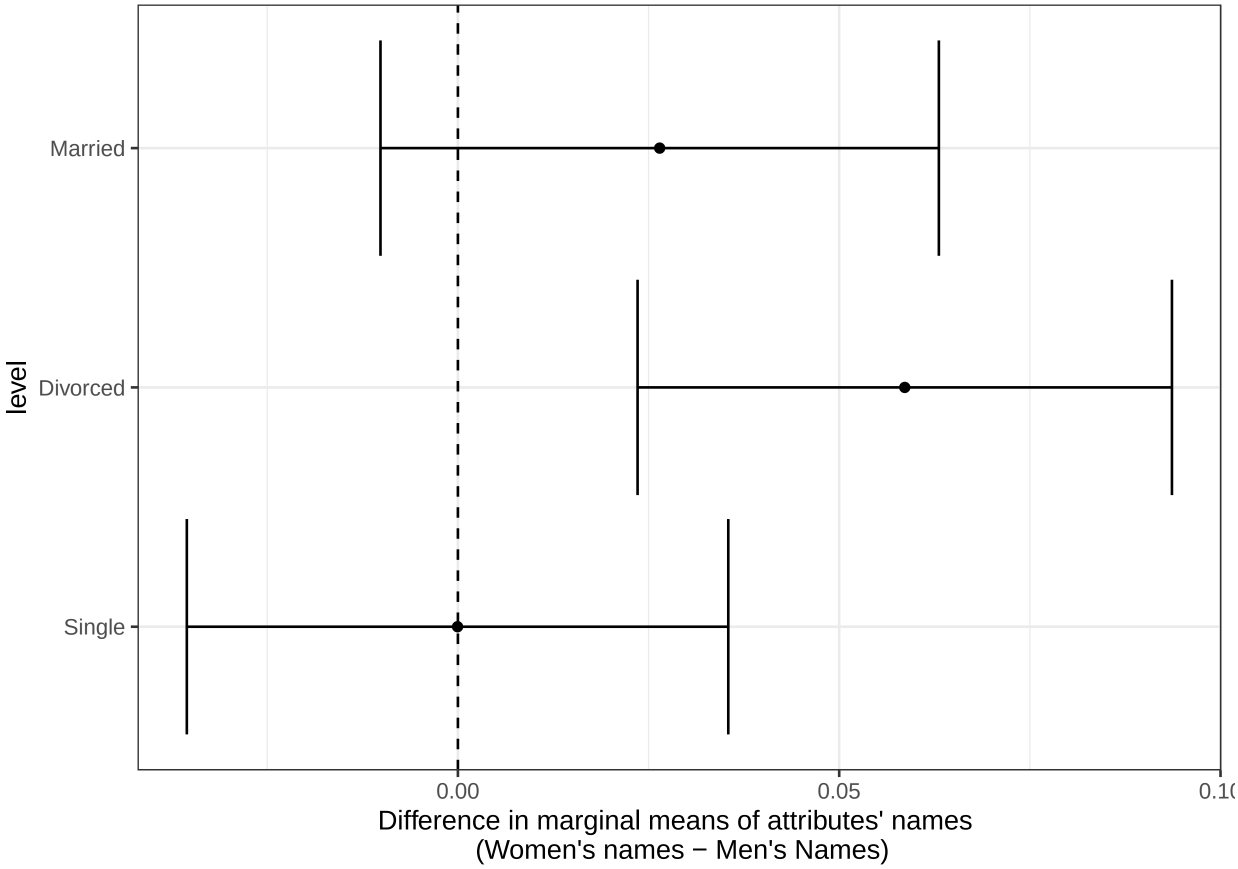

Figure I.7

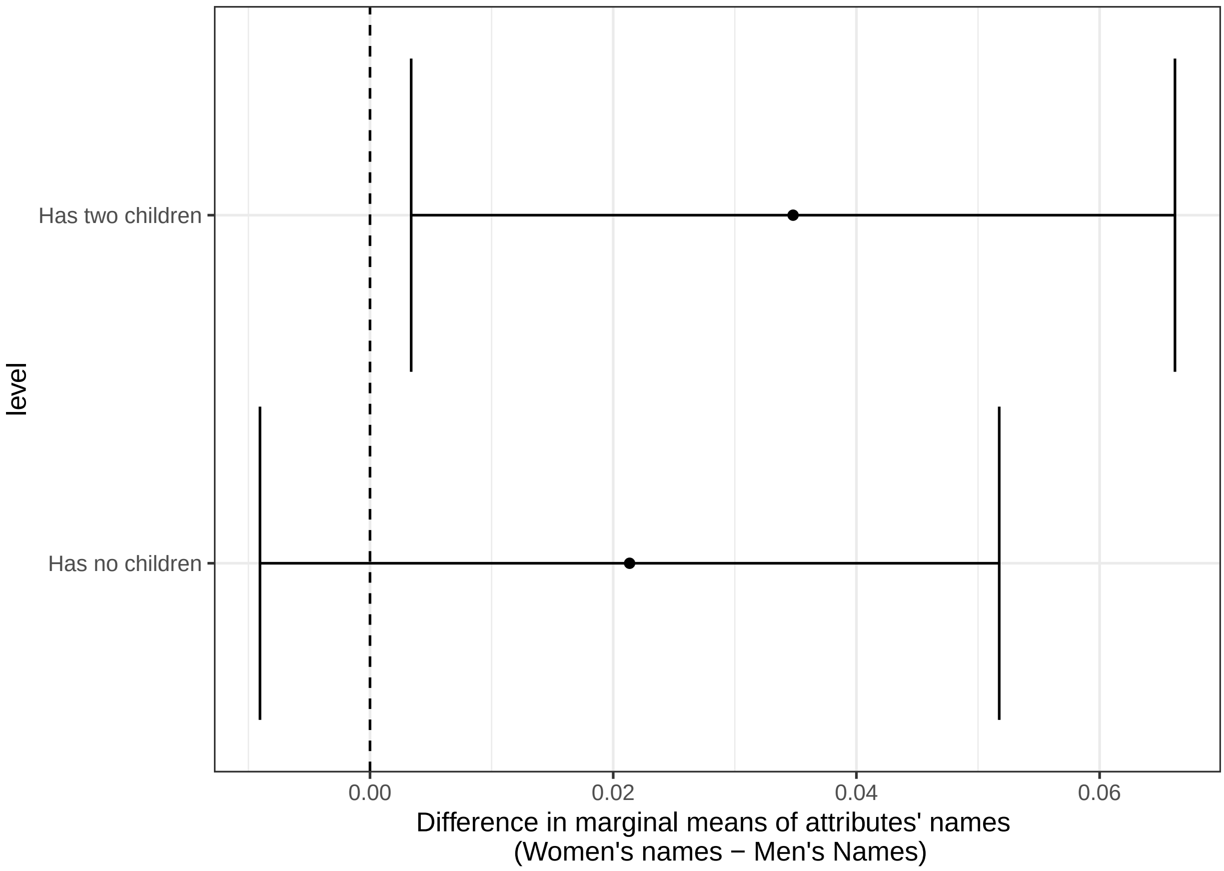
Has two children
Has no children
level
0.00
0.02
0.04
0.06
Difference in marginal means of attributes' names
(Women's names − Men's Names)

Figure I.8

References

Alvarez, Mike, José Antonio Cheibub, Fernando Limongi, and Adam Przeworski. 1996. "Classifying Political Regimes." *Studies In Comparative International Development* 31 (2).

Barnes, Tiffany D., and Emily Beaulieu. 2014. "Gender Stereotypes and Corruption: How Candidates Affect Perceptions of Election Fraud." *Politics & Gender* 10 (3).

Birch, Sarah. 2003. "Women and political representation in contemporary Ukraine." In *Women's Access to Political Power in Post-Communist Europe,* edited by Richard E. Matland and Kathleen A. Montgomery. Oxford: Oxford University Press.

Bussell, Jennifer. 2015. "Typologies of corruption: a pragmatic approach." In *Greed, Corruption, and the Modern State.* Edward Elgar Publishing.

Carpinella, Colleen M., Eric Hehman, Jonathan B. Freeman, and Kerri L. Johnson. 2016. "The Gendered Face of Partisan Politics: Consequences of Facial Sex Typicality for Vote Choice." *Political Communication* 33 (1).

Carpinella, Colleen M., and Kerri L. Johnson. 2013. "Politics of the Face: The Role Of Sex-Typicality in Trait Assessments of Politicians." *Social Cognition* 31 (6).

Collier, David, and Robert Adcock. 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices About Concepts." *Annual Review of Political Science* 2 (1).

Esarey, Justin, and Leslie A. Schwindt-Bayer. 2017. "Women's Representation, Accountability and Corruption in Democracies." *British Journal of Political Science.*

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* New York: W. W. Norton.

Haffert, Kristina. 2014. *Ukraine Gender Assessment 2014.* Technical report. Washington, D.C.: International Foundation for Electoral Systems.

He, Baogang, and Mark E. Warren. 2011. "Authoritarian deliberation: The deliberative turn in Chinese political development." *Perspectives on politics* 9 (2).

IFES. 2015. *Elections in Ukraine: October 25 Local Elections.* Technical report.

Johns, Robert, and Mark Shephard. 2007. "Gender, Candidate Image and Electoral Preference." *The British Journal of Politics and International Relations* 9 (3).

Klasnja, Marko, and Grigore Pop-Eleches. 2021. "Anti-Corruption Efforts and Electoral Manipulation in Democracies." *The Journal of Politics.*

Lawson, Chappell, Gabriel S. Lenz, Andy Baker, and Michael Myers. 2010. "Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies." *World Politics* 62 (4).

Lumley, Thomas. 2010. *Complex surveys: a guide to analysis using R.* Wiley series in survey methodology. Hoboken, N.J: John Wiley.

Martsenyuk, Tamara. 2015. "Women's Top-Level Political Participation" Failures and Hope of Ukrainian Gender Politics." In *New Imaginaries: Youthful Reinvention of Ukraine's Cultural Paradigm,* edited by Marian J. Rubchak. Berghahn Books.

Paxton, Pamela, and Melanie M. Hughes. 2016. *Women, Politics, and Power: A Global Perspective.* CQ Press.

Rainey, Carlisle. 2014. "Arguing for a Negligible Effect." *American Journal of Political Science* 58 (4).

Sapiro, Virginia. 1981. "If U.S. Senator Baker Were A Woman: An Experimental Study of Candidate Images." *Political Psychology* 3 (1/2).

Thames, Frank C. 2018. "The Electoral System and Women's Legislative Underrepresentation in Post-Communist Ukraine." *Comparative Politics* 50 (2).

Warren, Mark E. 2004. "What does corruption mean in a democracy?" *American Journal of Political Science* 48 (2).