Supplementary Information: Competent legislators or mere pawns? Experimental evidence of attitudes toward gender quota politicians

Carolyn Barnett[*] Alexandra Blackman[†] Marwa Shalaby[‡]

Contents

[*]University of Arizona, carolynbarnett@arizona.edu
[†]Cornell University, adb295@cornell.edu
[‡]University of Wisconsin-Madison, marwa.shalaby@wisc.edu

1 Data Collection and Sampling

1.1 Recruitment and Sampling

We implemented face-to-face survey experiments in respondents' households in Morocco. We relied on a nationally representative survey of 1,800 Moroccan citizens (voting age of 18 and older), who were randomly assigned one of two versions of the survey. The version containing the experiment discussed in this paper had 927 respondents. The survey was implemented by One to One Polling and Research. The local survey provider hired and trained enumerators. The research team members prepared training materials, maintained constant contact with survey team leaders, and monitored the incoming data. We also implemented a gender quota so that there were an even number of male and female enumerator-led interviews.

The survey employed stratified random sampling by region and urban versus rural areas within regions. There are 12 regions in Morocco and 1526 municipalities, each classified by the Moroccan government as either rural or urban. Three of the regions (Guelmim-Oued Noun, Laayoune-Sakia El Hamra, and Eddakhla-Oued Eddahab) were grouped into a single stratum due to their small population sizes, creating 10 de facto regions. Population sampling units (PSUs) equivalent to census districts (EAs, of which there are 36,581 in Morocco) were randomly selected within each stratum using the population proportional to size (PPS) method, based on the proportion of the population in the stratum in the last census. Each region was first allocated a portion of the sample according to its percentage of the population, and urban and rural areas within each region was allocated a portion of the sample according to the proportion of the population living in each type of area within each respective region. In each selected EA, 8 households were systematically drawn following a skip interval of households using starting points established by Morocco's High Commission for Planning (HCP) for the Moroccan census.

Detailed records were kept of hit rates, contacts, refusals, and unsuitable respondents. The selection of respondents within households was done using random selection with Kish tables. Interviewers alternated the selection of male and female respondents (ages 18+) from household to household. This sampling plan followed rigorous standards for selecting a nationally representative sample of adult Moroccans. Possible bias could emerge from the fact that the most recent census on which PSU sampling is based was conducted in 2014 (the next census is due in 2024). Systematic refusal to participate by certain demographic groups (e.g. less educated women) could also lead to an unbalanced sample. In Appendix 2, we compare our sample to other widely used national surveys of Morocco.

1.2 Ethics

We obtained ethical clearance from the Institutional Review Boards at the University of XX There was no deception involved and we included a consent form at the beginning of the survey to ensure the respondent understood what they agreed to and their rights regarding the storage and use of their data. Finally, we confirmed that the respondent was above the age of 18 (the voting age in Morocco) before continuing with the survey.

1.3 Data Quality

Following data collection, we ran several data quality checks. First, we compare the demographics of the sample to the demographics of other recent high-quality public opinion surveys conducted in Morocco, including the Arab Barometer, Afrobarometer, and World Values Survey (Table A.1). The only notable difference is that our survey has a smaller proportion of respondents with no formal education (7.5 percent).

We checked for evidence of near-duplicate responses (which would be evidence of enumerator fraud) using the Percentmatch software (Kuriakose 2015), following best practices outlined in Kuriakose and Robbins (2016). A high percentage of near-duplicate responses would indicate that some of the survey responses may be fraudulent. There were no exact duplicates in the data, and no two responses contained higher than a 0.78 match rate.

We pre-registered a set of robustness checks, including using a manipulation check, eliminating straightliners, and correcting for multiple hypotheses testing. We report these in Appendix 7.

2 Summary Statistics

Table A.1 summarizes the sample demographics and compares them to recent, high-quality surveys conducted in Morocco (Arab Barometer (2022), Afro Barometer (2022), and World Values Survey (2021)). As a result of blocking on gender, our sample includes 50 percent female respondents. The average age of respondents is 35. Of those surveyed, 67 percent are in the labor force, and 20 percent are unemployed (defined as in labor force and not employed). Over seven percent of the respondents report no formal education, 43 percent report that elementary or preparatory school are the highest levels attained, and 24.7 percent report secondary as the highest level. Looking at the Arab Barometer, Afrobarometer, and the World Values surveys from around the same period, our survey aligns with their survey demographics. The only notable difference is that our survey has a smaller proportion of respondents with no formal education (7.5 percent).

Table A.1: Survey Summary Statistics

Variable	N	Mean	St.Dev	Min	Max
Our Survey Sample (2023)					
Female	1800	0.500	0.500	0	1
Age	1800	35.202	13.230	18	77
In labor force	1800	0.672	0.470	0	1
Unemployment rate	1209	0.203	0.403	0	1
No formal education	1800	0.075	0.263	0	1
Elementary or preparatory	1800	0.429	0.495	0	1
Secondary	1800	0.247	0.431	0	1
Post-secondary	1800	0.241	0.428	0	1
Married or widowed	1800	0.500	0.500	0	1
Voted in 2021 parliamentary election	1770	0.457	0.498	0	1
Urban	1800	0.619	0.486	0	1
Arab Barometer Round 7 (2022), Morocco only					
Female	2404	0.498	0.500	0	1
Age	2312	37.980	14.394	18	99
In labor force	2404	0.546	0.498	0	1
Unemployment rate	1312	0.305	0.461	0	1
No formal education	2404	0.161	0.368	0	1
Elementary or preparatory	2404	0.332	0.471	0	1
Secondary	2404	0.220	0.414	0	1
Post-secondary	2404	0.285	0.451	0	1
Married or widowed	2404	0.538	0.499	0	1
Voted in 2021 parliamentary election	2404	0.416	0.493	0	1
Urban	2404	0.646	0.478	0	1
Afrobarometer Round 9 (2022), Morocco only					
Female	1200	0.500	0.500	0	1
Age	1199	39.437	14.315	18	85
In labor force	1200	0.632	0.482	0	1
Unemployment rate	759	0.163	0.370	0	1
No formal education	1200	0.150	0.357	0	1
Elementary or preparatory	1200	0.434	0.496	0	1
Secondary	1200	0.079	0.270	0	1
Post-secondary	1200	0.337	0.473	0	1
Married or widowed			Not asked		
Voted in 2021 parliamentary election	1200	0.475	0.500	0	1
Urban	1200	0.647	0.478	0	1
World Values Survey Round 7 (2021), Morocco only					
Female	1200	0.500	0.500	0	1
Age	1200	37.220	13.531	18	82
In labor force	1200	0.677	0.468	0	1
Unemployment rate	813	0.170	0.376	0	1
No formal education	1200	0.298	0.458	0	1
Elementary or preparatory	1200	0.217	0.412	0	1
Secondary	1200	0.153	0.360	0	1
Post-secondary	1200	0.332	0.471	0	1
Married or widowed	1200	0.572	0.495	0	1
Voted in 2021 parliamentary election			Not asked		
Urban	1200	0.683	0.465	0	1

3 Power Analyses

Pre-registered Hypotheses:
H1.1: Voters will view women elected via quota as less competent than either non-quota women or men.

H1.2: Voters will view women elected via quota as more pawn-like than either non-quota women or men.

H2.1: Among quota-elected women, voters will view those portrayed as working across party lines as more competent than those portrayed as working just with members of their own party.

H2.2: Among quota-elected women, voters will view those portrayed as working across party lines as less pawn-like than those portrayed as working just with members of their own party.

Power Analyses:
The quota vignette experiment was randomly presented to half the respondents. With N = 1800 respondents in total, we anticipated N = 900 for the experiment (the actual final sample size was N = 927). Respondents were asked three outcome questions regarding the vignette. All (competency, pawn-like, and cooperativeness) are measured on a 1-10 scale.

Hypotheses 1.1 and 1.2 are evaluated by comparing the outcome variable means among quota-elected women, non-quota women, and men.

Hypotheses 2.1 and 2.2 are evaluated by comparing the outcome variable means among politicians who work across party lines versus those who work with their co-partisans for quota-elected women.

Table A.2 summarizes the effect sizes at which the study has power = 0.8 to detect a true effect with alpha = 0.05 using the rating outcome (1-10). We note there that our preferred sample split of politician types is 50% quota women, 25% non-quota women, and 25% men.

Table A.2: Power Analysis Results Summary (Effect Size: Difference in Means of Rating Outcome (1-10)

Tests	Power=0.8	Power=0.95
H1: quota women vs. non-quota women (or vs. men, but not pooled)	0.3	0.32
H2: among quota women	0.34	0.36

We plot the power of this design at different effect sizes of interest for each of the hypotheses above. The figures below present the results of simulations relevant to the different hypotheses we have specified about how respondents evaluate the different politician vignettes. In each simulation, at any given effect size of interest, we run 1000 trials, each of which involves the following steps:

- Generate a dataset of randomized vignettes of the same size as our planned sample size. As in the real experiment, the different components of the vignette are randomized independently of one another. With respect to the politician type (quota woman, non-quota woman, or man), treatment assignments are generated with varying probabilities according to the sample split we are assessing the power of (see below). The other randomized elements of the vignette are assigned with equal probability.

- Generate random distributions of outcomes 1-10 with differences in means across subgroups of profiles that approximate the effect size of interest. For example, to test whether our analyses will correctly identify a difference of 0.5 in ratings across profiles describing quota women versus non-quota women, we generate two random distributions of integers 1-10, one with a mean of 5 for quota women, and one with a mean of approximately 5.5 (in expectation) for non-quota women.

- Analyze the randomly generated data by conducting the t-test or estimating the OLS regression described in the data analysis section. If the p-value for the relevant comparison or coefficient is significant in the result of a given trial, we save a value of 1 for that trial, or 0 otherwise.

- Return the proportion of trials that correctly identified a statistically significant effect. Given that we generated distributions of outcomes in which there was a true difference, this proportion is equal to the power of our design.

Figure A.1 shows the power of our design to detect different effect sizes in a comparison of ratings of quota-elected women compared to either of the other two groups. Figure A.2 shows the power of our design to detect different effect sizes in a comparison, among quota-elected women, of those described as working with other political parties vs. their co-partisans.

Figure A.1: Power with respect to H1 (Quota vs. Non-quota Women), 50-25-25 Sample Split

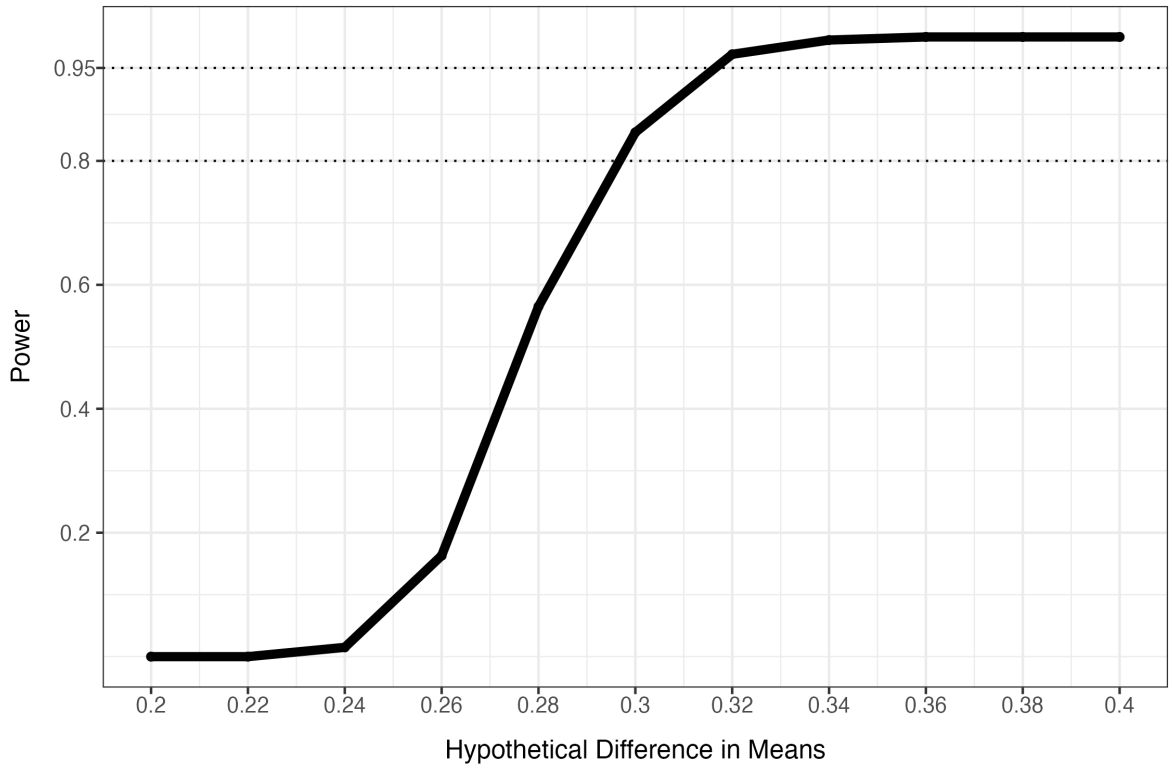
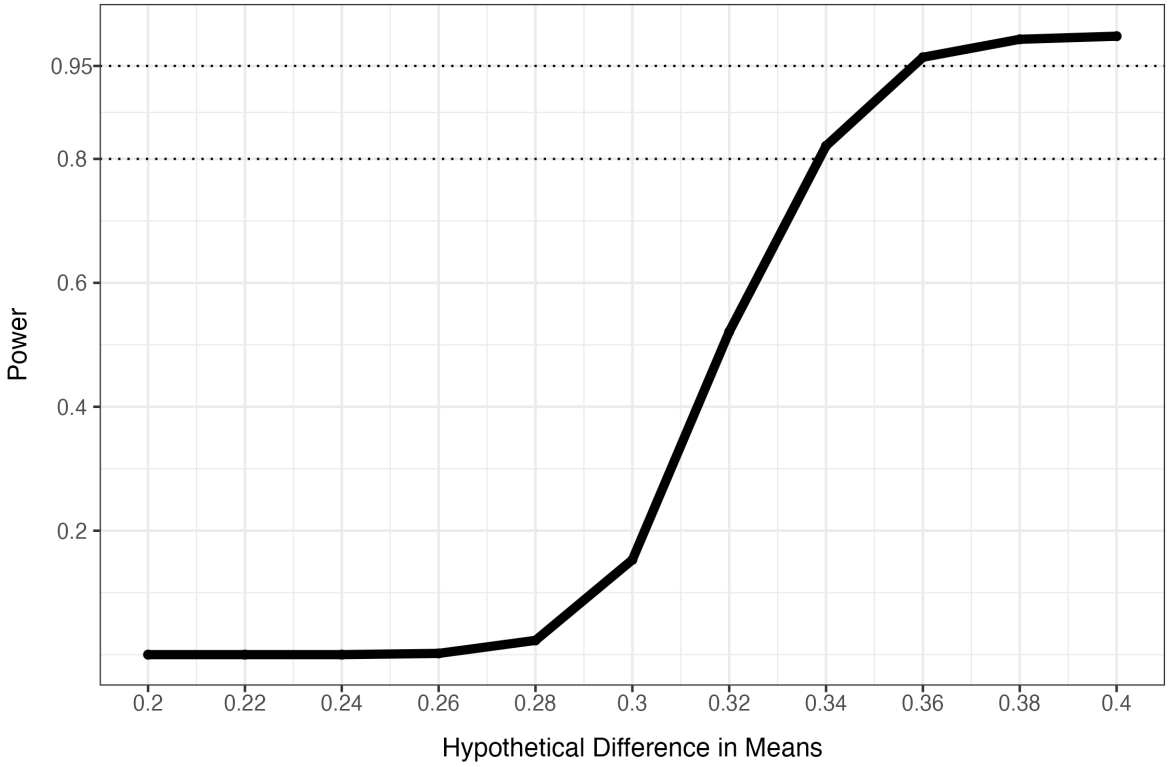


Figure A.2: Power with respect to H2 (Among Quota Women), 50-25-25 Sample Split



4 Additional Analyses: Effect of Collaboration Type

In addition to examining the effect of within versus cross-party collaboration for quota women, our pre-registered hypotheses included expectations that cross-party collaboration would improve competency ratings and reduce the perception of the politician as beholden to their party for non-quota male and female politicians as well. Tables A.3 and A.4 show the effect of cross-party collaboration for each type of politician (non-quota men, non-quota women, and quota women). Cross-party collaboration only reduces the perception of quota women as pawn-like, although the effect of cross-party collaboration on competence is in the expected direction (positive) for both quota and non-quota female politicians.

Table A.3: Effect of cross-party collaboration by politician gender and election type: Competence outcome

	Dependent variable:		
	Competence Outcome		
	(1)	(2)	(3)
	Man/general	Woman/general	Woman/quota
Work with another party	−0.039	0.378	0.286
	(0.276)	(0.257)	(0.197)
Constant	5.761***	5.475***	5.509***
	(0.196)	(0.177)	(0.138)
Observations	228	229	449
Adjusted R^2	−0.004	0.005	0.002
Note:			*p<0.1; **p<0.05; ***p<0.01

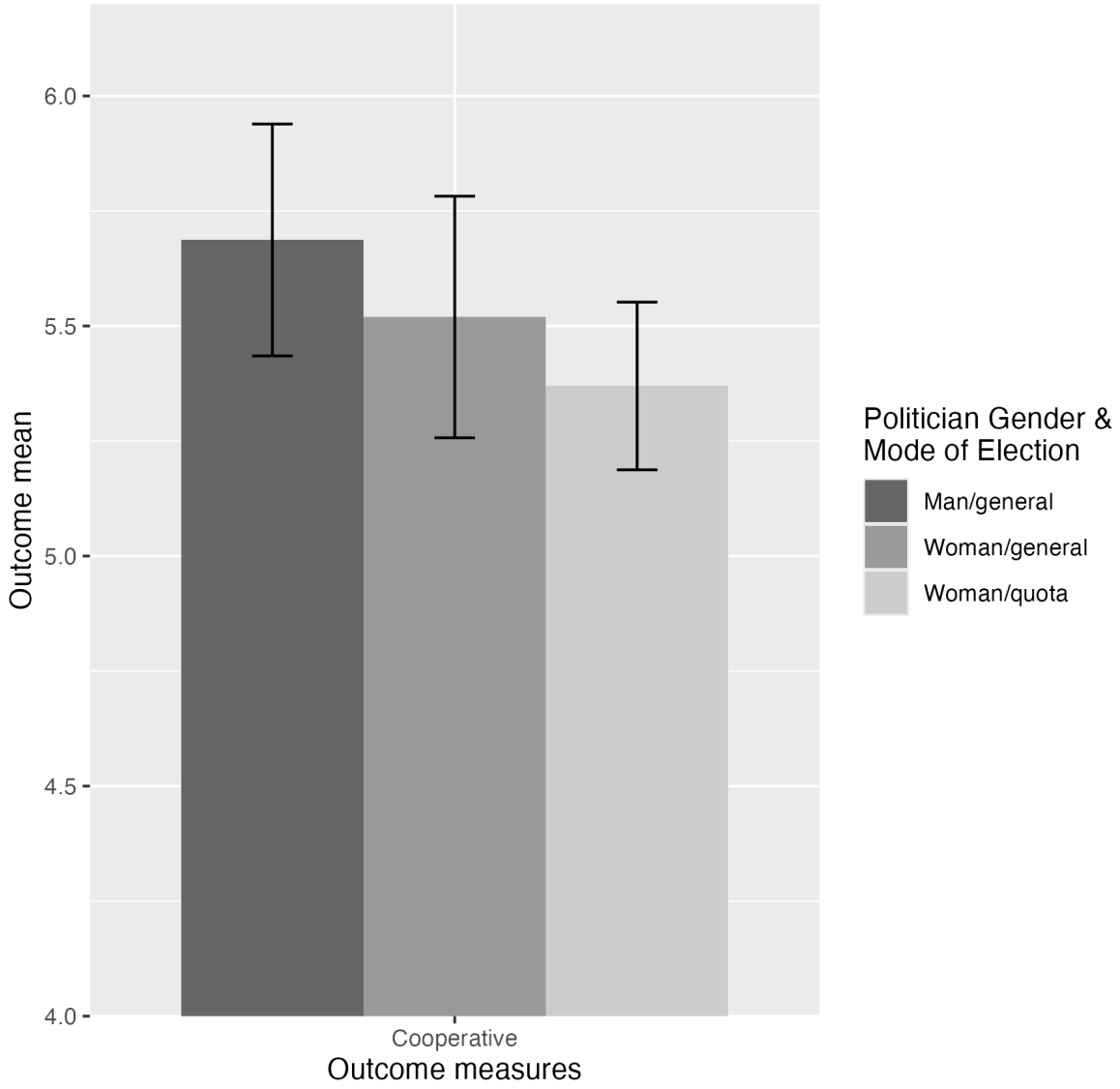
Table A.4: Effect of cross-party collaboration by politician gender and election type: Pawn-like outcome

	Dependent variable:		
	Pawn-like Outcome		
	(1)	(2)	(3)
	Man/general	Woman/general	Woman/quota
Work with another party	−0.100	0.149	−0.377*
	(0.296)	(0.294)	(0.205)
Constant	5.274***	4.824***	5.429***
	(0.210)	(0.204)	(0.143)
Observations	228	228	437
Adjusted R^2	−0.004	−0.003	0.005
Note:			*p<0.1; **p<0.05; ***p<0.01

5 Alternative Dependent Variable: Cooperativeness

In order to address concerns that respondents perceive being "pawn-like" as positive, we included a measure of the perceived cooperativeness of the politician. Figure A.3 shows the mean value of this outcome for each politician type. In the next section, we analyze the correlation of this outcome with our main outcome measures, showing that cooperativeness has has a clear positive valence, in contrast to the pawn-like outcome measure.

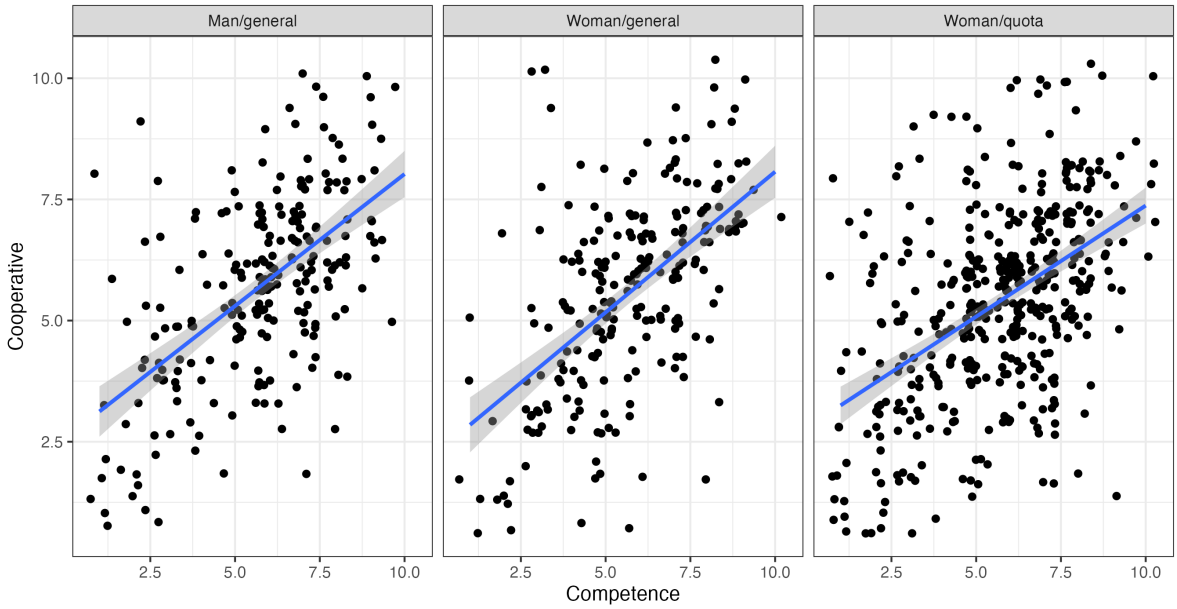
Figure A.3: Cooperativeness Outcome



6 Correlation of Outcome Measures

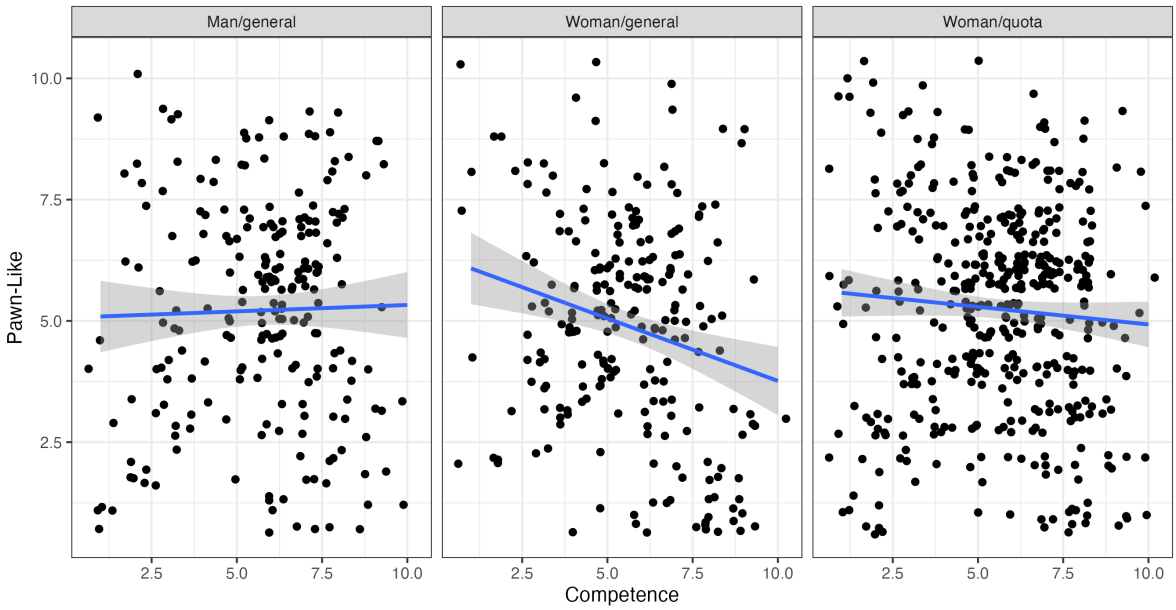
The figures in this section visualize the correlation of our different outcome measures: the competence and pawn-like ratings discussed in the main paper, and the cooperativeness outcome discussed in the prior section. In all treatment groups, the competence and cooperativeness ratings are strongly positively correlated ($p < 0.001$). This suggests that respondents do think of being "cooperative" as a positive politician trait. In contrast, the correlation between the pawn-like rating and each of the other two outcomes is either negative (for respondents in the non-quota woman treatment) or there is no correlation (for respondents in the quota woman and non-quota man treatments). In other words, respondents thinking about a woman elected via the competitive district lists who rated the politician as more competent were also less likely to describe that politician as pawn-like, and vice versa, suggesting that perceptions of such women as more autonomous are important to respondents' positive attitudes toward them. Other types of politicians—including quota women—are just as likely to be seen as pawn-like whether they are viewed as relatively competent or incompetent.

Figure A.4: Correlation of Competence and Cooperativeness Outcome Measures



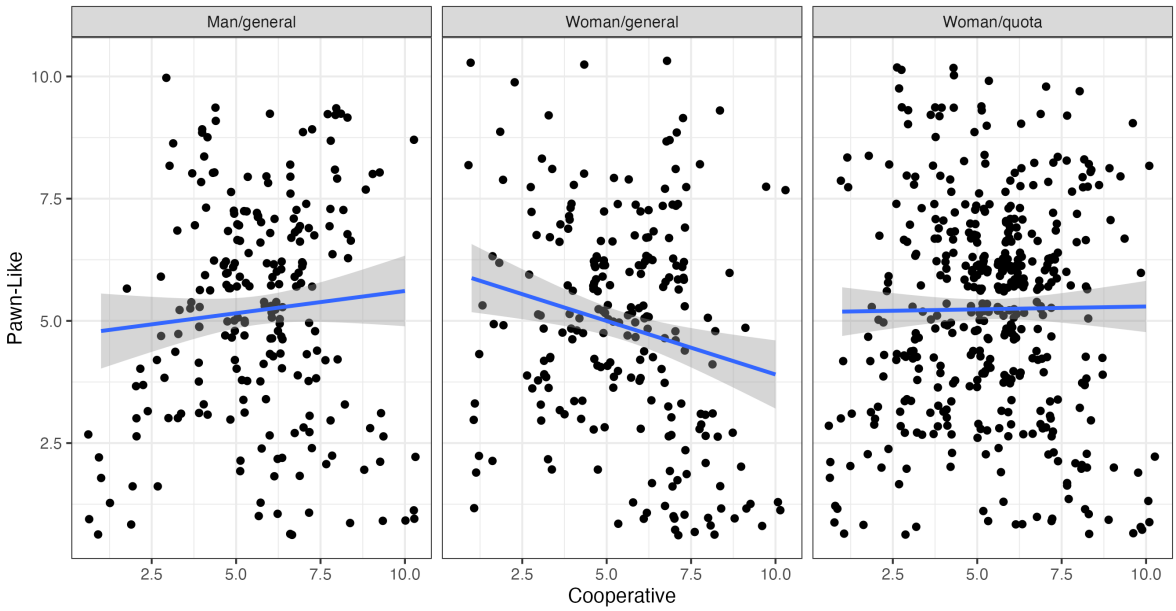
The figure displays the correlation of the competence and cooperativeness ratings of the politician within each treatment group, with a bivariate OLS regression estimated for each group.

Figure A.5: Correlation of Competence and Pawn-Like Outcome Measures



The figure displays the correlation of the competence and pawn-like ratings of the politician within each treatment group, with a bivariate OLS regression estimated for each group.

Figure A.6: Correlation of Cooperativeness and Pawn-Like Outcome Measures



The figure displays the correlation of the cooperativeness and pawn-like ratings of the politician within each treatment group, with a bivariate OLS regression estimated for each group.

7 Robustness Checks

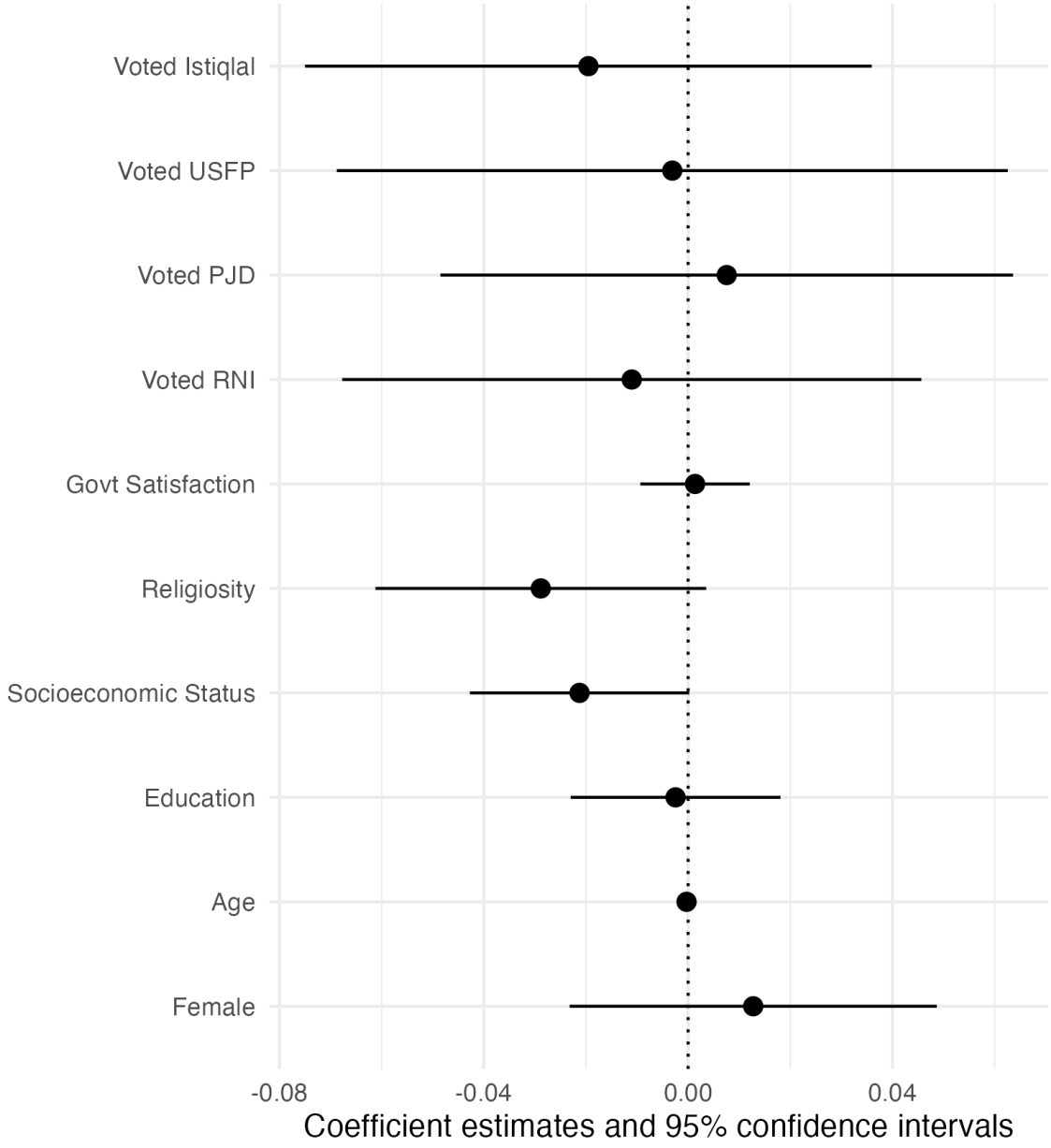
7.1 Missingness

Out of the 927 responses to the vignette experiment, enumerators recorded that respondents either answered ”don’t know” or refused to respond to the question with the following frequencies:

- Competence outcome: 21 responses (2.3 %)
- Pawn-like outcome: 34 responses (3.7 %)
- Cooperative outcome: 25 responses (2.7 %)

We do not impute missing data. To assess the relationship between non-response and respondent characteristics, we create a respondent-level indicator variable that takes a value of 1 if the responded ever provided a non-response (”don’t know” or refused) to any of the vignette outcome questions and 0 otherwise. 47 unique respondents out of the 927 (5.1%) ever gave a non-response. Figure A.7 displays the results of an OLS regression estimated with this indicator as the dependent variable and respondents’ demographic characteristics as predictors. Only subjective socioeconomic status and religiosity have a modest relationship with the propensity to non-respond. Both better-off respondents and more religious respondents were slightly more likely to respond, but these correlations are substantively small. We do not have a strong reason to believe that respondents vary along these dimensions in how they assess differences between quota and non-quota politicians such that this missing data would substantially bias our results. The low total number of non-response observations also suggests this is not a major concern.

Figure A.7: Correlates of Non-Response in Vignette Outcomes



The figure displays estimates from an OLS regression taking the demographic covariates as predictors and the indicator for ever providing a non-response as the outcome. The variables related to voting for specific political parties and the indicator for "female" are dummy variables. Government satisfaction is a 1-10 rating, where 10 indicates the greatest satisfaction. Religiosity is a numeric variable ranging from 1-3, where 1 indicates not religious and 3 indicates very religious. Socioeconomic status is a numeric variable ranging from 1-4, where 1 is the least well-off and 4 is the most well-off (based on a question about how well household income meets needs and whether the family can save money). Education is a numeric variable ranging from 1-4, where 1 indicates no formal education and 4 indicates a BA or higher. Age is a numeric variable ranging from 18 to 77.

7.2 Manipulation Check

Following the vignette experiment, respondents were asked whether the politician about whom they had just answered questions had been elected via a quota or not. Note that for respondents who were in one of the non-quota politician treatments, the correct answer to this question was "No," while for those in the female quota politician treatment, the correct answer was "Yes." Table A.5 shows the number and related percentage (in parentheses) of respondents in each treatment arm who correctly answered this question. Just under 60 percent of respondents in each of the non-quota treatments correctly answered "No," while 84 percent of respondents in the quota-woman treatment correctly answered "Yes." Respondents in the non-quota-man treatment were slightly more likely to state that they did not know the answer than respondents in other groups.

Table A.5: Manipulation Check Responses By Treatment Group

	Correct	Incorrect	Don't Know or Refused
Non-Quota Woman	138	83	15
	(58%)	(35%)	(6%)
Non-Quota Man	140	71	25
	(59%)	(30%)	(11%)
Quota Woman	384	52	19
	(84%)	(11%)	(4%)

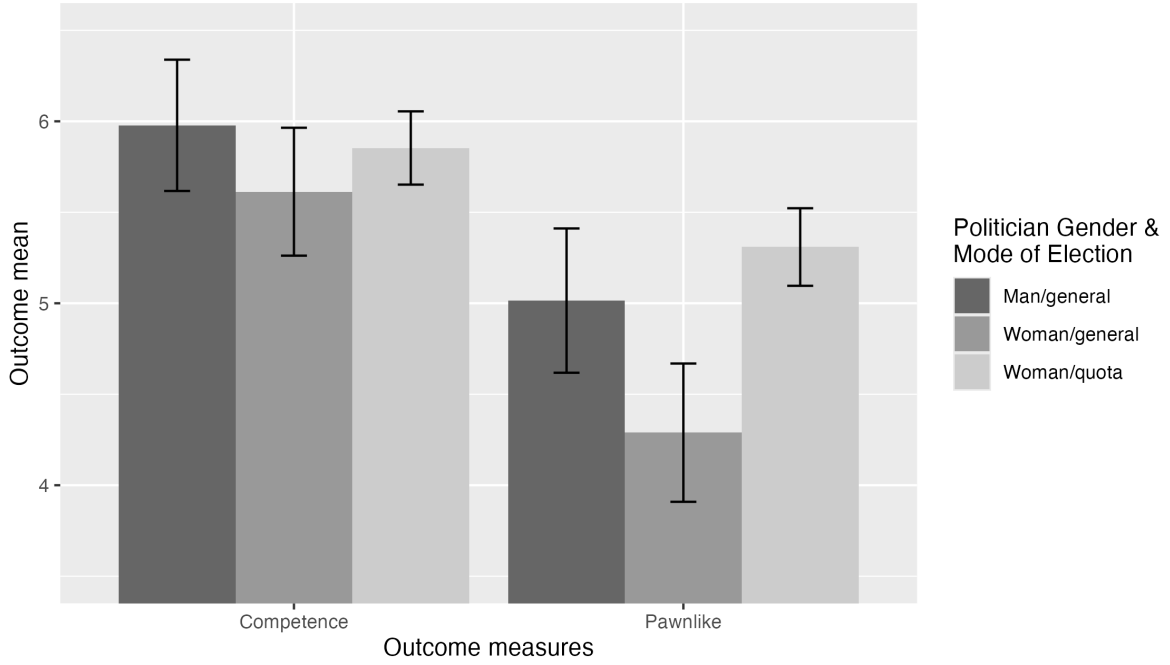
The remaining figures and tables in this section reproduce the main results discussed in the paper and this appendix using a sample that excludes respondents who did not correctly answer the manipulation check question. 662 respondents (71 percent of the original sample) correctly answered and remain in the sample.

Figure A.8 reproduces Figure 1 in the main manuscript with this reduced sample. The figure shows that while a similar pattern holds, the differences across groups with respect to the pawn-like outcome are more stark. Notably, in this sample, quota women are rated as more pawn-like than either women or men elected on the general district lists.

Table A.6 displays the difference-in-means for each outcome variable between quota politicians who worked with their own party vs. with another party. In this sample, quota women who work with another party are more likely to be rated as competent (6.1 vs. 5.6, p = 0.033) and less likely to be seen as pawnlike (5.1 vs. 5.5, p = 0.062). Figure A.9 reproduces Figure 2 in the main manuscript with the reduced sample, visualizing this result. Female quota politicians who work with members of other parties are viewed as more competent and less pawn-like than female quota politicians who work with their own parties.

Tables A.7 and A.8 show the effect of working with another party (compared to working with one's own party) on the competence and pawn-like outcomes, respectively, among respondents in each of the three treatment arms, estimated via OLS regressions of the respective outcomes on collaboration type separately for each type of politician. For quota women only, working with another party improves competency ratings ($p < 0.05$) and reduces the perception that the politician is pawn-like ($p < 0.1$).

Figure A.8: Effect of politician gender and mode of election: manipulation check sample



Recreates Figure 1 in the main manuscript but excludes respondents who failed the post-experiment manipulation check question (leaving N = 662).

Table A.6: Difference in Means for Type of Collaboration: Manipulation Check Sample (Quota women only)

Outcome	Mean (With her party)	Mean (With another party)	Difference	p-value
Competence	5.639	6.074	0.435**	0.033
Pawnlike	5.508	5.104	-0.404*	0.062

Note: *p<0.1; **p<0.05; ***p<0.01

Figure A.9: Mean Responses by Collaboration Treatment Among Quota Women: Manipulation Check Sample

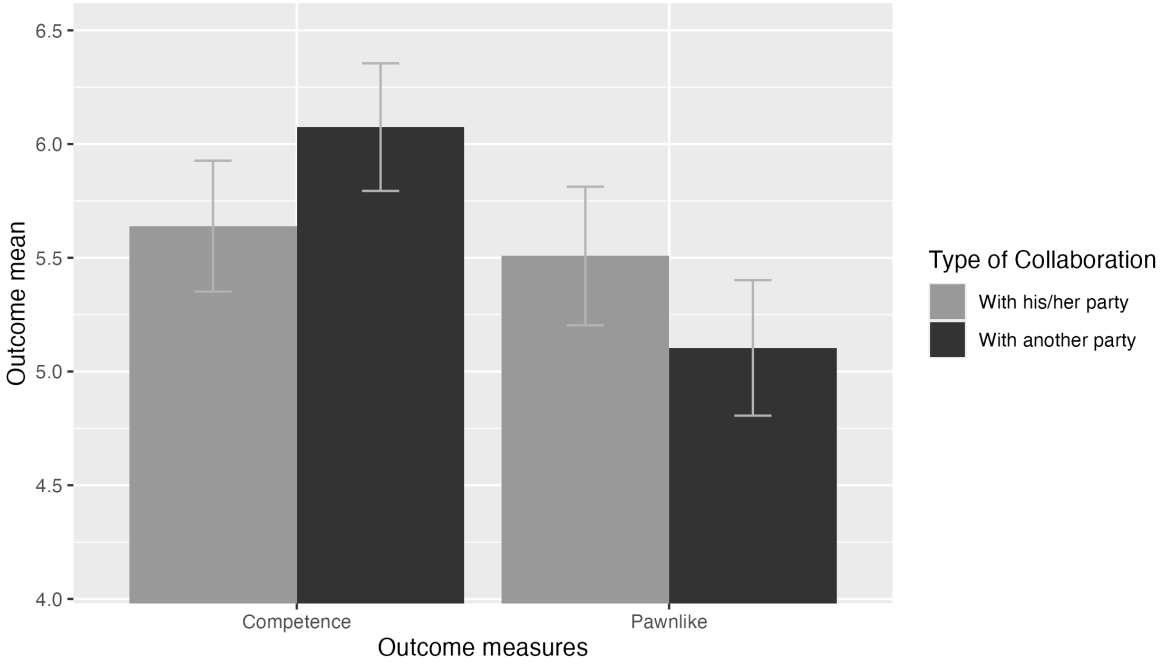


Table A.7: Effect of Working With Another Party on Competence Rating: Manipulation Check Sample, All Treatment Arms

	Dependent variable:		
	Competence Outcome		
	Man/general	Woman/general	Woman/quota
Work with another party	−0.074	0.441	0.435**
	(0.366)	(0.355)	(0.204)
Constant	6.014***	5.388***	5.639***
	(0.257)	(0.254)	(0.143)
Observations	136	137	382
Adjusted R^2	−0.007	0.004	0.009
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.8: Effect of Working With Another Party on Pawn-Like Rating: Manipulation Check Sample, All Treatment Arms

	Dependent variable:		
	Pawn-Like		
	Man/general	Woman/general	Woman/quota
Work with another party	0.676*	0.099	−0.404*
	(0.398)	(0.386)	(0.216)
Constant	4.676***	4.239***	5.508***
	(0.282)	(0.274)	(0.152)
Observations	136	135	372
Adjusted R^2	0.014	−0.007	0.007
Note:			*p<0.1; **p<0.05; ***p<0.01

7.3 Straightliners

Given the large number of items related to gender and sexism on the survey, we conducted a check to assess whether participants "straightlined" through responses to the items in the benevolent and hostile sexism, modern sexism, and egalitarianism batteries (always agreeing with either the first or last response option), although this is typically less of a concern in the face-to-face mode in which this survey was conducted than it is in online surveys. Straightlining through these items is unlikely to represent a genuine set of attitudes as the items are intentionally phrased from an egalitarian viewpoint on some items and from a conservative viewpoint on other items.

We find that among these survey items, a single respondent selected the first response ("strongly agree") more than two-thirds of the time, and no respondents selected the last response option ("strongly disagree") more than two-thirds of the time. An additional 43 respondents selected "agree" more than two-thirds of the time, and 4 respondents selected "disagree" more than two-thirds of the time. Although we did not pre-specify that such respondents would be treated as straightliners, for transparency, we identify "straightliners" according to these broader criteria.

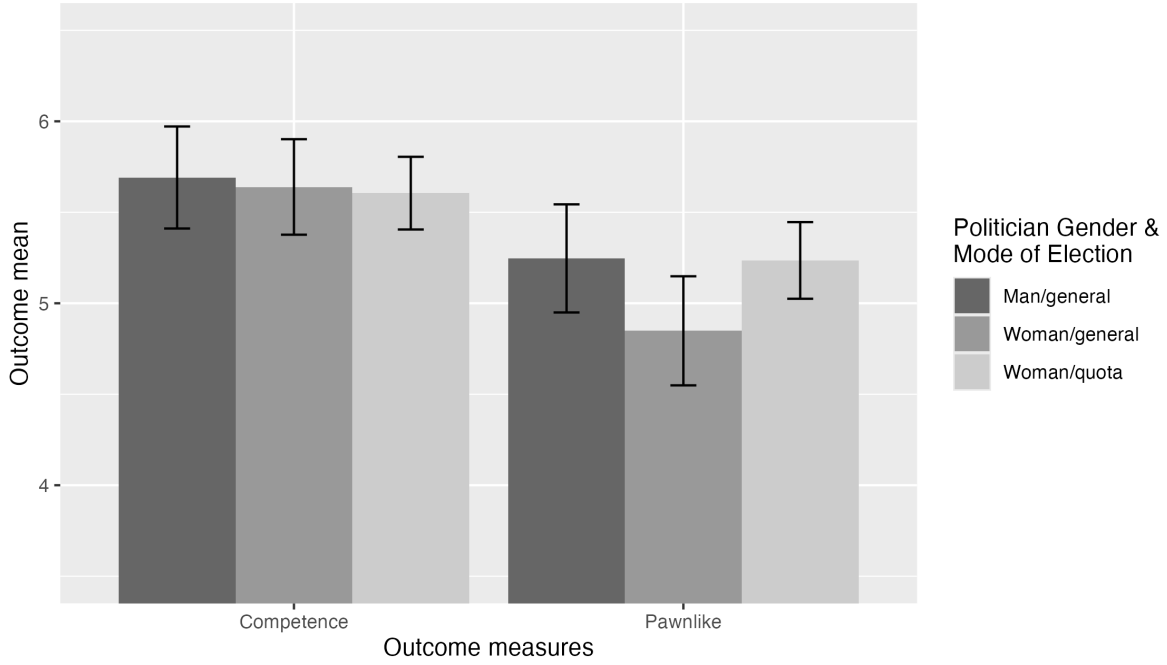
The remaining figures and tables in this section reproduce the main results discussed in the paper and this appendix using a sample that excludes the 48 straightliners, leaving an effective sample of 879 respondents.

Figure A.10 reproduces Figure 1 in the main manuscript with this reduced sample. The results do not differ substantially from those presented in the main text.

Table A.9 displays the difference-in-means for each outcome variable between quota women politicians who worked with their own party vs. with another party. In this sample, quota women who work with another party are less likely to be seen as pawnlike (5.1 vs. 5.5, p = 0.039). Politicians in the two groups do not differ in the extent to which they are viewed as competent, as in the main results. Figure A.11 reproduces Figure 2 in the main manuscript with the reduced sample, to visualize these results. Female quota politicians who work with members of other parties are viewed as less pawn-like than female quota politicians who work with their own parties.

Tables A.10 and A.11 show the effect of working with another party (compared to working with one's own party) on the competence and pawn-like outcomes, respectively, among respondents in each of the three treatment arms, estimated via OLS regressions of the respective outcomes on collaboration type separately for each type of politician. For quota women only, working with another party reduces the perception that the politician is pawn-like ($p < 0.05$).

Figure A.10: Effect of politician gender and mode of election: Non-Straightliner Sample



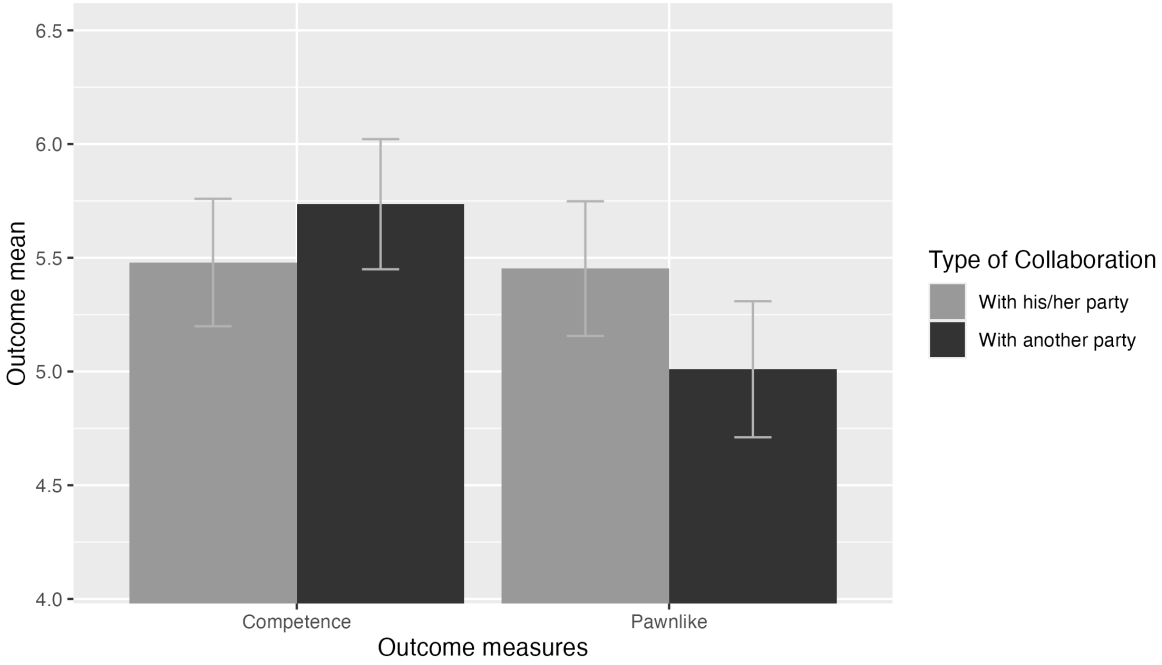
Recreates Figure 1 in the main manuscript but excludes respondents who straightlined (leaving N = 879).

Table A.9: Difference in Means for Type of Collaboration: Non-Straightliner Sample (Quota women only)

Outcome	Mean (With his/her party)	Mean (With another party)	Difference	p-value
Competence	5.479	5.736	0.257	0.207
Pawnlike	5.452	5.01	-0.442**	0.039

Note: *p<0.1; **p<0.05; ***p<0.01

Figure A.11: Mean Responses by Collaboration Treatment Among Quota Women: Non-Straightliner Sample



Recreates Figure 2 in the main manuscript but excludes respondents who straightlined (leaving N = 879).

Table A.10: Effect of Working With Another Party on Competence Rating: Non-Straightliner Sample, All Treatment Arms

	Dependent variable:		
	Competence Outcome		
	Man/general	Woman/general	Woman/quota
Work with another party	−0.013	0.437	0.257
	(0.285)	(0.266)	(0.203)
Constant	5.698***	5.430***	5.479***
	(0.204)	(0.184)	(0.142)
Observations	217	219	423
Adjusted R^2	−0.005	0.008	0.001
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.11: Effect of Working With Another Party on Pawn-Like Rating: Non-Straightliner Sample, All Treatment Arms

	Dependent variable:		
	Pawn-Like Outcome		
	Man/general	Woman/general	Woman/quota
Work with another party	−0.121	0.163	−0.442**
	(0.302)	(0.305)	(0.213)
Constant	5.308***	4.770***	5.452***
	(0.216)	(0.212)	(0.149)
Observations	219	218	412
Adjusted R^2	−0.004	−0.003	0.008
Note:			*p<0.1; **p<0.05; ***p<0.01

7.4 Multiple Hypotheses Testing

To account for testing multiple hypotheses, we report the adjusted p-values of our tests derived from the Bonferroni, Holm's, and Benjamini-Hochberg procedures for correcting for false discovery. Table A.12 reports the original and FDR-corrected p-values of our main hypothesis tests.

Table A.12: False Discovery Rate Adjustment Results

	Original	Bonferroni	Holm	Benjamini & Hochberg
H1.1 Quota Women vs. Non-Quota Women (Competence)	0.97	1	1	1
H1.1 Quota Women vs. Non-Quota Men (Competence)	0.58	1	1	0.87
H1.2 Quota Women vs. Non-Quota Women (Pawn-Like)	0.05	0.31	0.31	0.20
H1.2 Quota Women vs. Non-Quota Men (Pawn-Like)	0.91	1	1	0.97
H2.1 Co-Partisan vs. Cross-Party Among Quota Women (Competence)	0.15	0.89	0.59	0.30
H2.2 Co-Partisan vs. Cross-Party Among Quota Women (Pawn-Like)	0.07	0.40	0.33	0.20

8 Secondary Hypotheses: Sub-Group Analyses

We test heterogeneous effects along the following respondent characteristics: gender, hostile sexism, and patriarchal attitudes. The operationalization of each variable is described below. The hypotheses tested here were pre-registered. All results in this section are with respect to the competency and pawn-like rating outcomes. We did not pre-register sub-group hypotheses for the cooperativeness outcome. For each sub-group, we first hypothesize about the relationship contrasting quota and non-quota politicians (pooling across non-quota women and non-quota men). Then we hypothesize about the relationship comparing male and female politicians (pooling across quota women and non-quota women). All heterogeneous treatment effects by treatment and respondent sub-group (difference-in-difference estimates) are available in the replication code.

8.1 Respondent Gender

Pre-registered Hypotheses:

H1a: Male respondents are less likely to provide favorable competency evaluations for quota women (compared to non-quota politicians) than female respondents.

H1b: Male respondents are more likely to see quota women as "pawns" (compared to non-quota politicians) than female respondents.

H1c: Male respondents are less likely than female respondents to provide favorable competency evaluations for female politicians regardless of female politicians' quota status.

H1d: Male respondents are more likely than female respondents to see female politicians as "pawns" regardless of female politicians' quota status.

Results:

We find that male respondents are less likely than female respondents to provide favorable competency evaluations for quota women and more likely than female respondents to see quota women as "pawns." However, the difference-in-differences (i.e., the comparison to non-quota politicians) is not significant. Thus, we do not find support for H1a and H1b.

We find support for H1c (Table A.13). When we look just at female politicians, we find that male respondents are less likely than female respondents to provide favorable competency evaluations. We find no support for H1d. Male respondents are more likely than female respondents to see female politicians as "pawns" but the difference is not statistically significant.

Figure A.12: Effect of politician gender and mode of election on competency and pawn-like outcomes (By respondent gender)

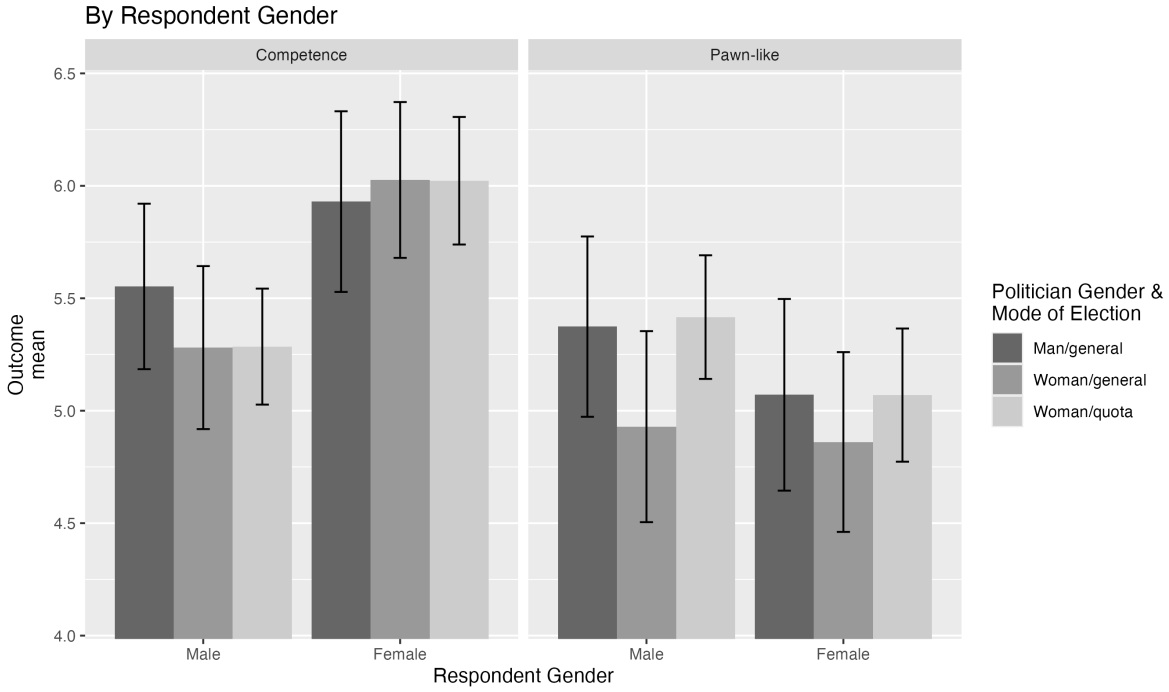


Table A.13: Differences between Male and Female Respondents (Women Politicians Sample Only)

Outcome	Mean (Male Respondents)	Mean (Female Respondents)	Difference	p-value
Competence	5.284	6.024	0.740***	0.000
Pawn-like	5.251	4.997	-0.255	0.131

Note: *p<0.1; **p<0.05; ***p<0.01

8.2 Hostile Sexism

To measure hostile sexism, we use the extent of respondents' agreement with three hostile sexism items on the survey instrument, which were all measured after the experiment: "women are too easily offended"; "women exaggerate the problems they have at work"; and "women seek to gain power by getting control over men." Coding responses to each of these items such that 1 = the least hostile sexist response and 4 = the most hostile sexist response, we then take the average response across the three survey items (the *hostile sexism score*). For the analysis of H2, we group respondents into those with below-mean scores (*low hostile sexism*) and those with scores at or above the mean (*high hostile sexism*).

Pre-registered Hypotheses:

H2a: Respondents who hold hostile sexist views are less likely to provide favorable competency evaluations for quota women (compared to non-quota politicians) than respondents who do not hold those views.

H2b: Respondents who hold hostile sexist views are more likely to see quota women as "pawns" (compared to non-quota politicians) than respondents who do not hold those views.

H2c: Respondents who hold hostile sexist views are less likely to provide favorable competency evaluations for women regardless of female politicians' quota status than respondents who do not hold those views.

H2d: Respondents who hold hostile sexist views are more likely to see female politicians as "pawns" regardless of female politicians' quota status than respondents who do not hold those views.

Results:

We find that respondents who hold more hostile sexist views are less likely than less hostile sexist respondents to provide favorable competency evaluations for quota women but are not more likely than less hostile sexist respondents to see quota women as "pawns." The difference-in-differences (i.e., the comparison to non-quota politicians) is significant (respondents who hold more hostile sexist views rate quota women -.580 lower than non-quota politicians (pooled across non-quota women and men), p-value = 0.033) for the competency outcome. Thus, we find support for H2a but not H2b.

We find support for H2c (Table A.14). When we look just at female politicians, we find that respondents who hold more hostile sexist views are less likely than less hostile sexist respondents to provide favorable competency evaluations. We do not find support for H2d.

Figure A.13: Effect of politician gender and mode of election on competency and pawn-like measures (By hostile sexism attitudes)

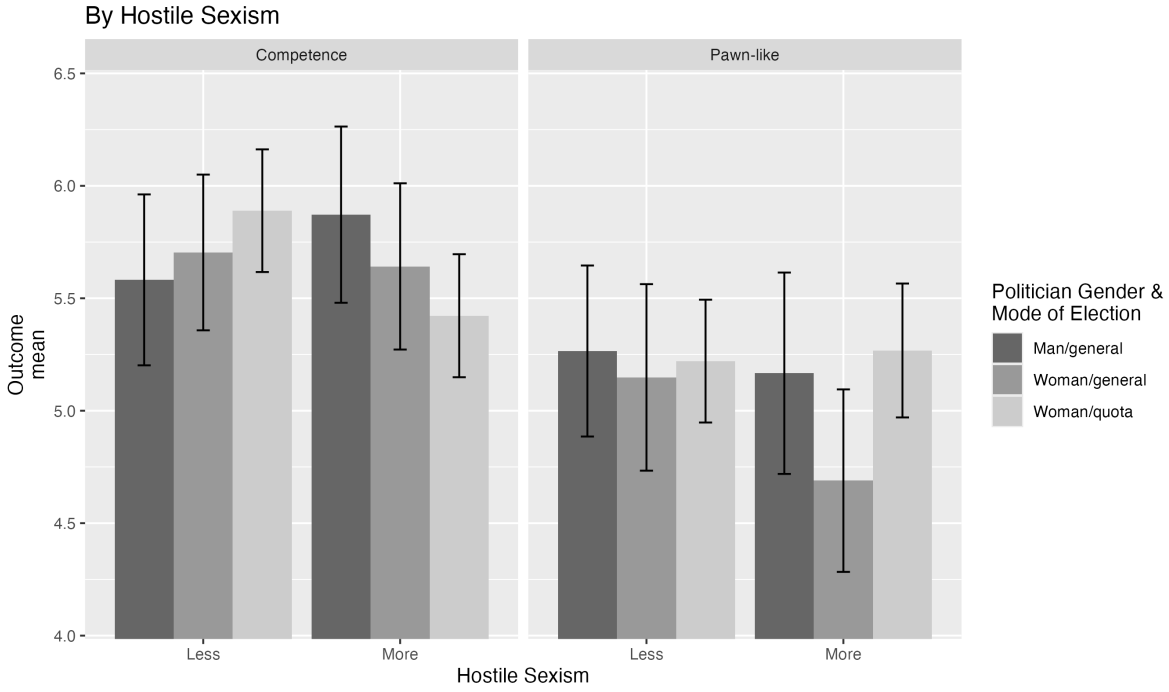


Table A.14: Differences between Low and High Hostile Sexism Respondents (Women Politicians Sample Only)

Outcome	Mean (Low Hostile Sexism)	Mean (High Hostile Sexism)	Difference	p-value
Competence	5.828	5.497	-0.331**	0.034
Pawnlike	5.196	5.067	-0.129	0.443

Note: *p<0.1; **p<0.05; ***p<0.01

8.3 Patriarchal Attitudes

To measure patriarchal attitudes, we use the extent of respondents' agreement with three items on the survey instrument, which were all measured after the experiment: "in general, men are better political leaders than women"; "a man should have final say in all decisions concerning the family"; and "when jobs are scarce, men should have more right to a job than women." Coding responses to each of these items such that 1 = the least patriarchal response and 4 = the most patriarchal response, we then take the average response across the three survey items (the *patriarchal attitudes score.*). For the analysis of H3, we group respondents into those with below-mean scores (*low patriarchal attitudes*) and those with scores at or above the mean (*high patriarchal attitudes*). Then we will test whether the marginal mean for a female politician is higher for respondents with

more egalitarian attitudes.

Pre-registered Hypotheses:

H3a: Respondents who hold patriarchal views are less likely to provide favorable competency evaluations for quota women (compared to non-quota politicians) than respondents who do not hold those views.

H3b: Respondents who hold patriarchal views are more likely to see quota women as "pawns" (compared to non-quota politicians) than respondents who do not hold those views.

H3c: Respondents who hold patriarchal views are less likely to provide favorable competency evaluations for women regardless of female politicians' quota status than respondents who do not hold those views.

H3d: Respondents who hold patriarchal views are more likely to see female politicians as "pawns" regardless of female politicians' quota status than respondents who do not hold those views.

Results:

We find that more patriarchal respondents are less likely than less patriarchal respondents to provide favorable competency evaluations for quota women but are not more likely than less patriarchal respondents to see quota women as "pawns." However, the difference-in-differences (i.e., the comparison to non-quota politicians) is not significant. Thus, we do not find support for H3a and H3b.

We find weak support for H3c (Table A.15). When we look just at female politicians, we find that more patriarchal respondents are less likely than less patriarchal respondents to provide favorable competency evaluations. We do not find support for H3d.

Figure A.14: Effect of politician gender and mode of election on competency and pawn-like measures (By patriarchal attitudes)

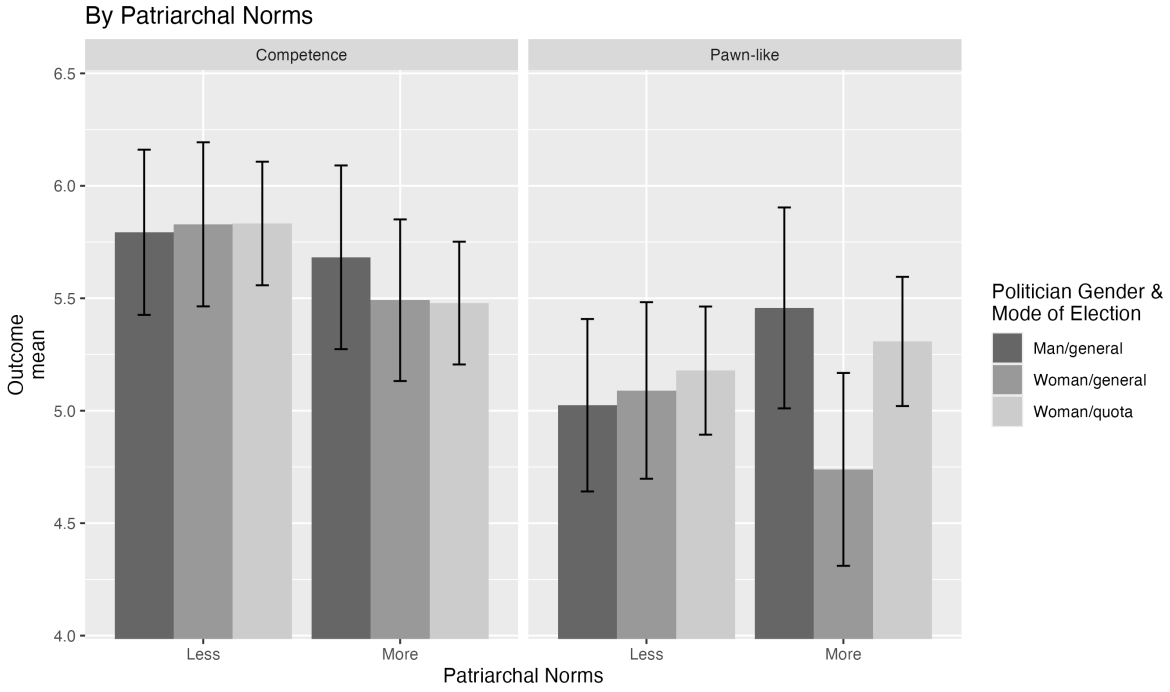


Table A.15: Differences between Low and High Patriarchal Attitudes Respondents (Women Politicians Sample Only)

Outcome	Mean (Low Patriarchal Attitudes)	Mean (High Patriarchal Attitudes)	Difference	p-value
Competence	5.831	5.483	-0.348**	0.026
Pawnlike	5.148	5.115	-0.033	0.844
Note:				*p<0.1; **p<0.05; ***p<0.01

9 Exploratory Analyses: Sub-Group Analyses

We test heterogeneous effects along the following respondent characteristics: voting in the previous election, trust in elections, and satisfaction with the government. The hypotheses tested here were not pre-registered. Please note that here we are looking how these respondent characteristics affect views of quota women specifically, as well as comparing quota women to non-quota politicians (pooled across non-quota men and women). This differs from the previous section where we also compared male and female politicians. All heterogeneous treatment effects by treatment and respondent sub-group (difference-in-difference estimates) are available in the replication code.

9.1 Trust in Elections

To measure trust in elections, we ask respondents: In your opinion, do you think elections in Morocco are fair and free? Those who "Agree" or "Strongly Agree" are coded as having **High Trust in Elections**, while those who "Disagree" or "Strongly Disagree" are coded as having **Low Trust in Elections**.

Results:

Table A.16 displays the difference-in-means for those with Low versus High Trust in Elections in the *Quota Women* treatment group only. We find that respondents who have high trust in the elections are more likely than low trust respondents to provide favorable competency evaluations for quota women *and* are more likely than low trust respondents to see quota women as "pawns." The difference-in-differences (i.e., the comparison to non-quota politicians) is not significant. Overall, respondents who have high trust in the elections provide higher competency ratings *and* higher pawn-like ratings.

Figure A.15: Effect of politician gender and mode of election on competency and pawn-like measures (By trust in elections)

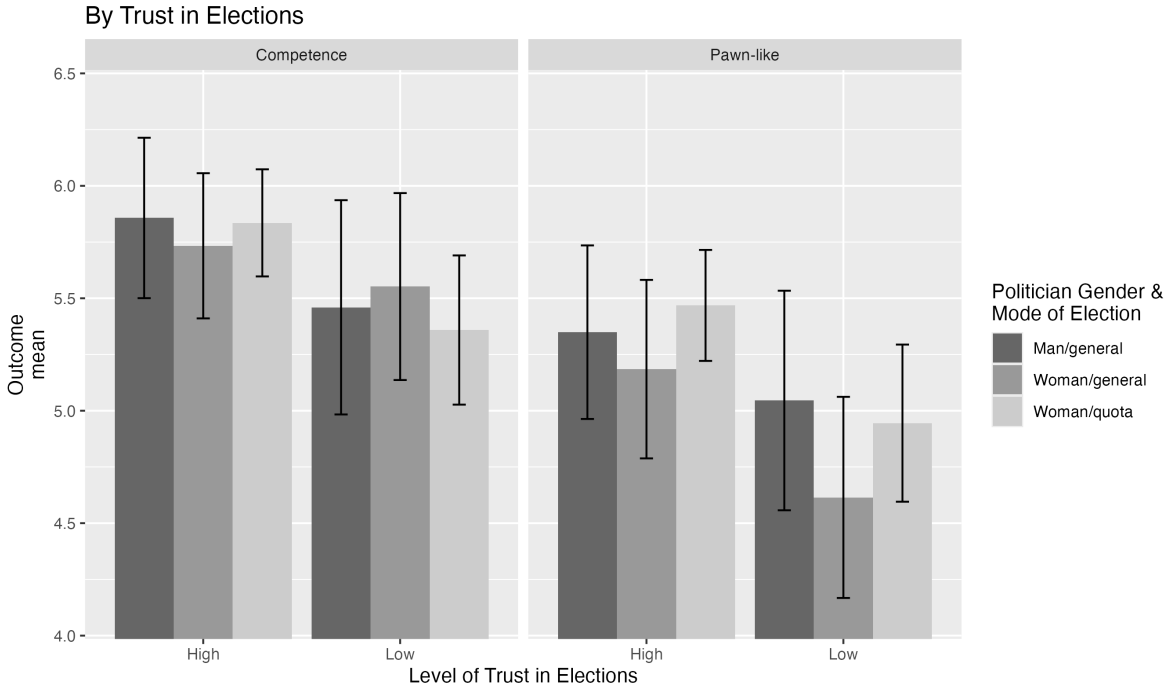
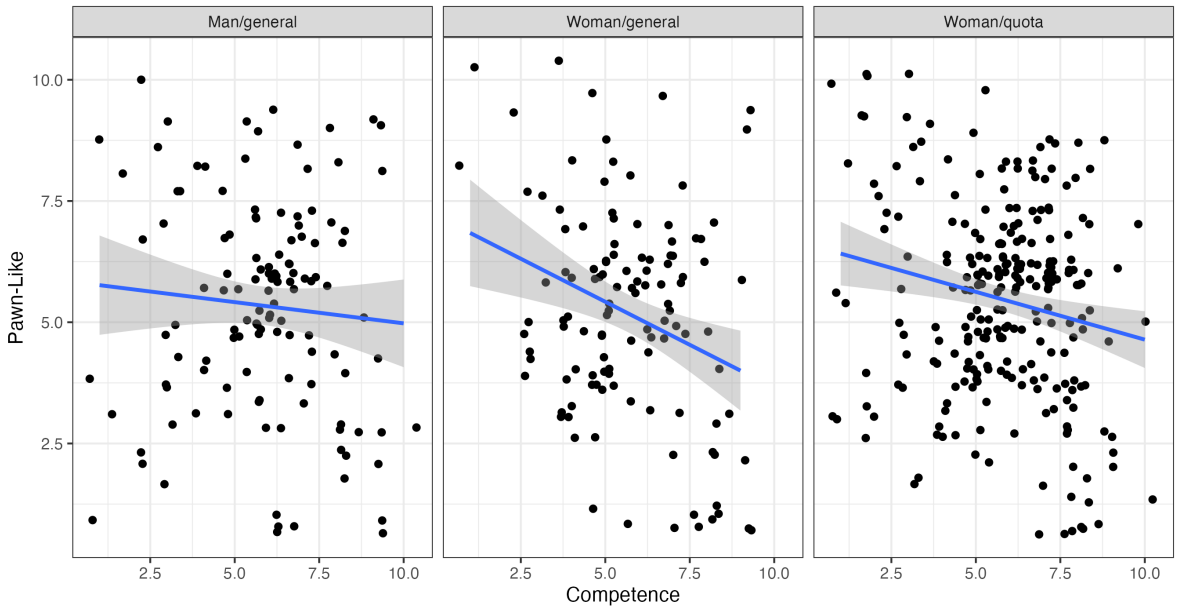


Table A.16: Differences between High and Low Trust in Elections Respondents (Quota Women Politicians Sample Only)

Outcome	Mean (High Trust)	Mean (Low Trust)	Difference	p-value
Competence	5.835	5.359	-0.476**	0.022
Pawnlike	5.468	4.945	-0.523**	0.016
Note:				*p<0.1; **p<0.05; ***p<0.01

We also check the correlation of our outcome measures among the subset of respondents with higher trust in elections. Figure A.16 shows the correlation of the competence and pawn-like measures in each treatment group subset to respondents who had higher trust in elections. As in the full sample, there is no correlation between these outcomes among those in the non-quota man treatments. There are negative correlations ($p = 0.002$) between these outcomes among those in both the quota and the non-quota woman treatments.

Figure A.16: Correlation of Outcomes Among Respondents with High Trust in Elections



9.2 Government Satisfaction

To measure government satisfaction, we ask respondents: On a scale from 1-10, where 1 is not satisfied at all and 10 is very satisfied, how satisfied are you with the current government? Those who express above mean government satisfaction (Mean = 4.913) are coded as having **High Government Satisfaction**.

Results:

Table A.17 displays the difference-in-means for those with Low versus High Government Satisfaction in the *Quota Women* treatment group only. We find that respondents who have high government satisfaction are more likely than low government satisfaction respondents to see quota women as "pawns." The difference in competency ratings for quota women is not statistically significant between high and low government satisfaction respondents. The difference-in-differences (i.e., the comparison to non-quota politicians) is not significant. Overall, like trust in elections, respondents who have high government satisfaction provide higher competency ratings *and* higher pawn-like ratings.

Figure A.17: Effect of politician gender and mode of election on competency and pawn-like measures (By government satisfaction)

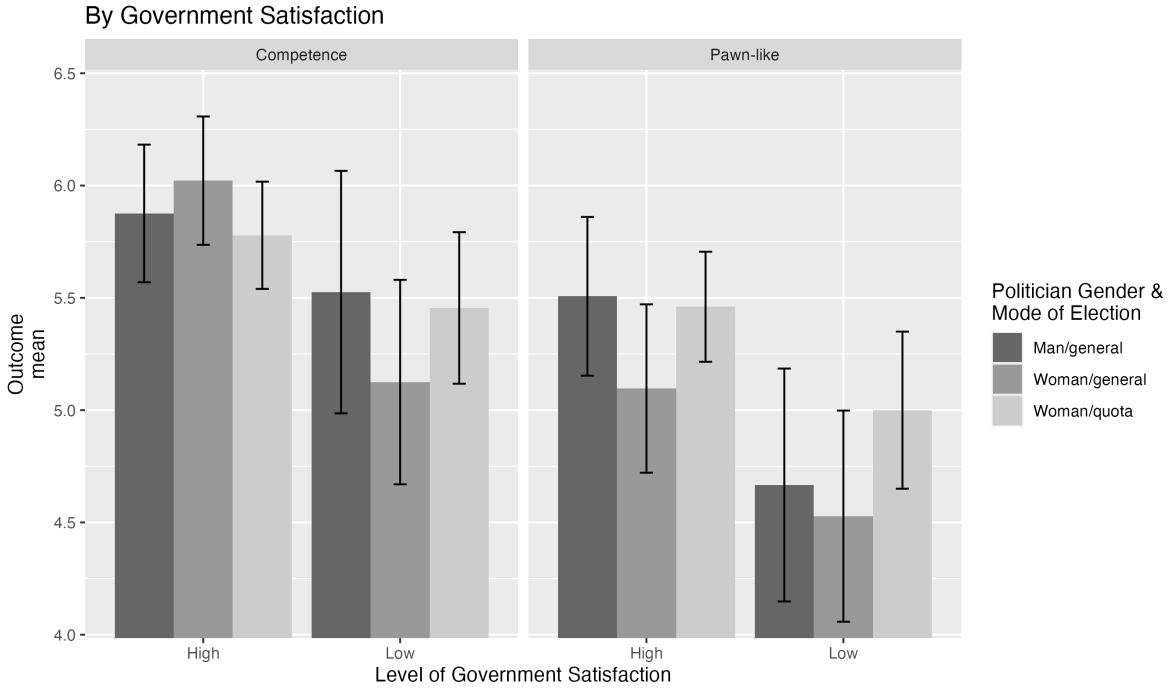
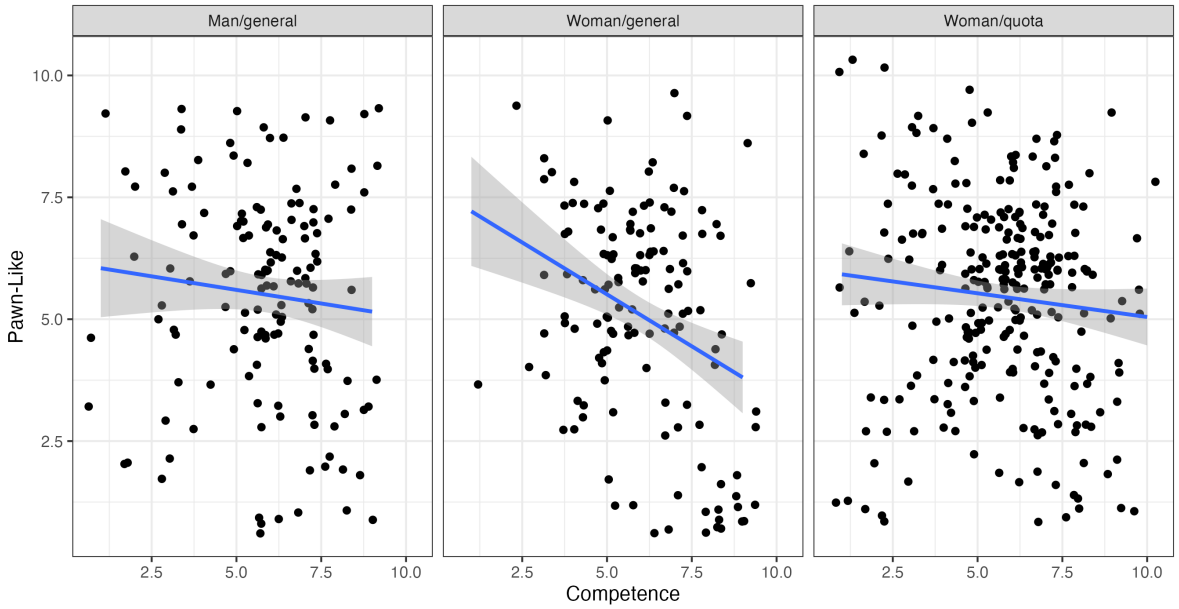


Table A.17: Differences between High and Low Government Satisfaction Respondents (Quota Women Politicians Sample Only)

Outcome	Mean (High Satis)	Mean (Low Satis)	Difference	p-value
Competence	5.779	5.455	-0.324	0.124
Pawnlike	5.460	5.000	-0.460**	0.034
Note:				*p<0.1; **p<0.05; ***p<0.01

We also check the correlation of our outcome measures among the subset of respondents who are more satisfied with the government. Figure A.18 shows the correlation of the competence and pawn-like measures in each treatment group subset to respondents who had higher satisfaction. As in the full sample, there is no correlation between these outcomes among those in the quota woman and non-quota man treatments, but a negative correlation ($p < 0.001$) between these outcomes among those in the non-quota woman treatment.

Figure A.18: Correlation of Outcomes Among Respondents with High Satisfaction with Government



9.3 Voted in Previous Election

To measure voting history, we ask respondents: Did you vote in the legislative elections in September 2021? Those who say "Yes" are coded as a **Voter**, and those who say "No" are coded as a **Non-Voter**.

Results:

Table A.18 displays the difference-in-means for Non-Voters and Voters in the *Quota Women* treatment group only. We find that respondents who report voting in the previous election (N=413) are less likely than non-voters (N=498) to provide favorable competency evaluations for quota women and more likely than non-voters to see quota women as "pawns." The difference-in-differences (i.e., the comparison to non-quota politicians) is significant for the competency outcome. Respondents who report voting in the previous election rate quota women's competence -0.652 lower than non-quota politicians (pooled across non-quota women and men, p-value=0.0182).

Figure A.19: Effect of politician gender and mode of election on competency and pawn-like measures (By previous voting)

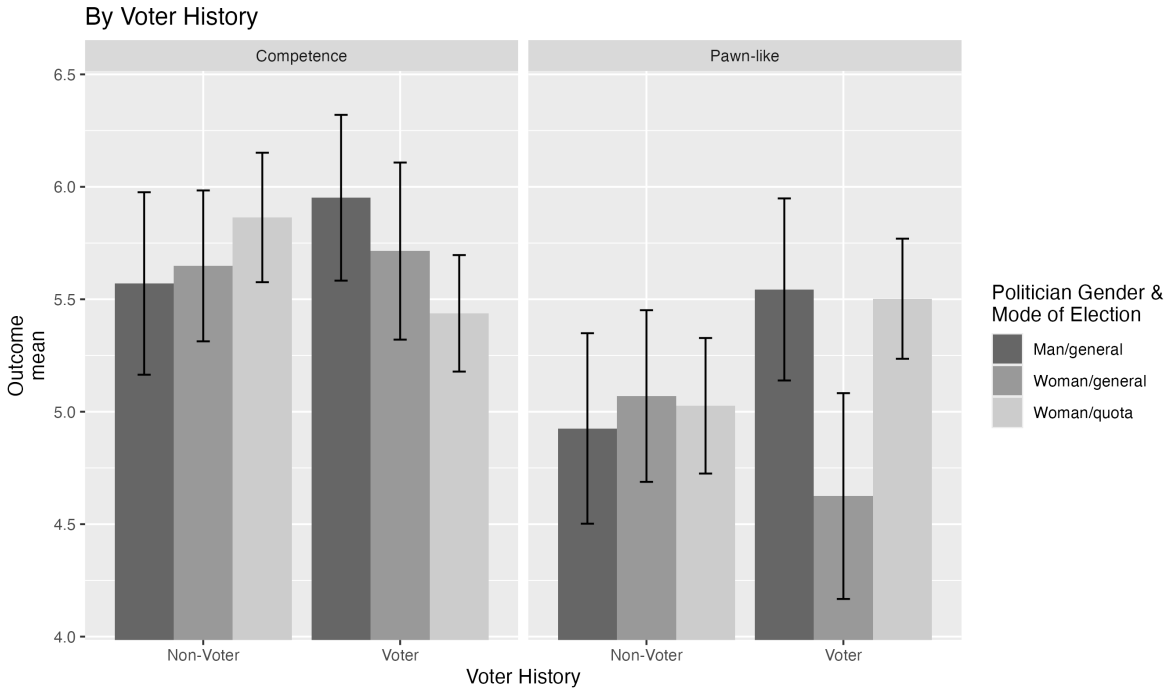


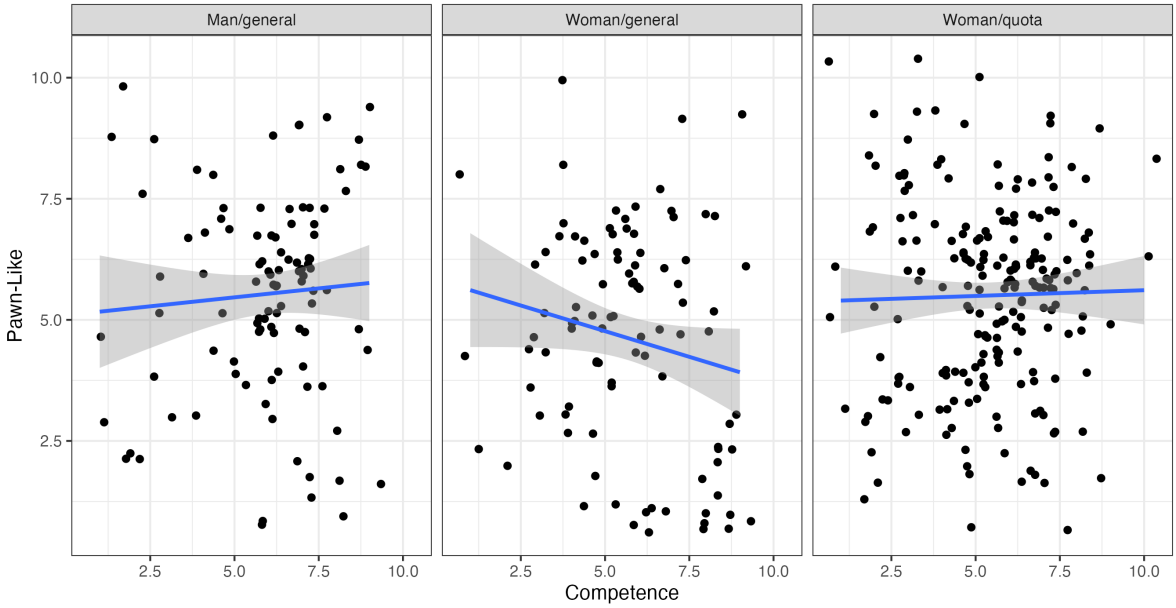
Table A.18: Differences between Non-Voters and Voters (Quota Women Politicians Sample Only)

Outcome	Mean (Non-Voters)	Mean (Voters)	Difference	p-value
Competence	5.864	5.438	-0.426**	0.031
Pawn-like	5.026	5.502	0.476**	0.020
Note:				*p<0.1; **p<0.05; ***p<0.01

We also check the correlation of our outcome measures among the subset of respondents who voted in the 2021 elections. Figure A.20 shows the correlation of the competence and pawn-like measures in each treatment group subset to respondents who were voters. As in the full sample,

there is no correlation between these outcomes among those in the quota woman and non-quota man treatments, but a negative correlation ($p < 0.001$) between these outcomes among those in the non-quota woman treatment.

Figure A.20: Correlation of Outcomes Among Voters



References

Kuriakose, Noble. 2015. "PERCENTMATCH: Stata Module to Calculate the Highest Percentage Match (near Duplicates) between Observations." Boston College Department of Economics: Statistical Softward Components. https://ideas.repec.org/c/boc/bocode/s457984.html.

Kuriakose, Noble, and Michael Robbins. 2016. "Don't Get Duped: Fraud through Duplication in Public Opinion Surveys." *Statistical Journal of the IAOS* 32 (3): 283–91.