

Corruption, accountability and gender: do female politicians face higher standards in public life?

Andrew C. Eggers[†]

Nick Vivyan^{*}

Markus Wagner[‡]

[†]Nuffield College and University of Oxford, 1 New Road, Oxford OX1 1NF, United Kingdom; andrew.eggers@nuffield.ox.ac.uk.

^{*}School of Government and International Affairs, Durham University, The Al-Qasimi Building, Elvet Hill Road, Durham DH1 3TU, United Kingdom; nick.vivyan@durham.ac.uk.

[‡]Department of Government, University of Vienna, Rathausstrasse 19/1/9, 1010 Vienna, Austria; markus.wagner@univie.ac.at.

Abstract

Previous research suggests that female politicians face higher standards in public life, perhaps in part because female voters expect more from female politicians than from male politicians. Most of this research is based on observational evidence. We assess the relationship between accountability and gender using a novel survey vignette experiment fielded in the UK in which voters choose between a hypothetical incumbent (who could be male or female, corrupt or uncorrupt) and another candidate. We do not find that female politicians face significantly greater punishment for misconduct. However, the effect of politician gender on punishment varies by voter gender, with female voters in particular more likely to punish female politicians for misconduct. Our findings have implications for research on how descriptive representation affects electoral accountability and on why corruption tends to correlate negatively with women's representation.

Keywords: accountability, conjoint analysis, corruption, gender

Supplementary material for this article is available in the appendix in the online edition.

Replication files are available in the JOP Data Archive on Dataverse (<http://thedata.harvard.edu/dvn/dv/jop>).

This research was supported by a grant from the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science (LSE). This research was conducted in line with the ethical standards contained in the 1964 Declaration of Helsinki and its later amendments.

Perceptions of corruption tend to be lower where female political representation is higher (Dollar, Fisman, and Gatti 2001; Swamy et al. 2001), but the mechanism behind this pattern remains uncertain. Female politicians may have fewer opportunities to participate in corruption as they are excluded from predominantly male networks (Bjarnegård 2013). Women in politics may also be more highly-qualified (Anzia and Berry 2011), more risk-averse (Esarey and Schwindt-Bayer forthcoming), or more opposed in principle to corruption (Dollar, Fisman, and Gatti 2001).

We focus on a further mechanism suggested by Esarey and Schwindt-Bayer (forthcoming): female politicians in high-accountability contexts may be less corrupt because they believe they are more likely to be held accountable by voters. We assess whether this belief is accurate: Is it true that female politicians are more heavily punished for comparable levels of misconduct? If so, is this due to differences in how men and women evaluate female politicians' records, as suggested by the work of Jones (2014)? We address these questions with a vignette experiment in the UK.

Our point of departure is gender stereotyping of politicians. One widely accepted view is that men are seen as more agentic, i.e. competent and assertive, and women as more communal, i.e. compassionate, warm and emotional (Dolan 2004). Women are also seen as more honest than men (Dolan 2014b; Fridkin, Kenney, and Woodall 2009; Alexander and Andersen 1993; Kahn 1992). Consistent with this, voters who value honesty are more likely to vote for women than for men (Dolan 2004; Frederick and Streb 2008).¹

These gender stereotypes imply that voters respond differently to misconduct by male and female politicians. In the absence of evidence of misconduct, voters who view women as more honest would support a female politician more than an otherwise-similar male politician. The flip-side of this stereotyping is that female politicians have “further to fall” if wrongdoing is revealed: to the extent that female politicians' support draws more on voters who are attracted to perceived integrity, their support stands to suffer more when

¹The persistence of gender stereotypes stands in contrast to the weak to nonexistent penalty for female candidates at the ballot box (Dolan 2014a). This discrepancy may exist because female politicians are generally of higher quality and because voters face multiple considerations in addition to candidate gender.

a lack of integrity is found (Funk 1996: p. 18).² Our first hypothesis is therefore: *female politicians are punished more severely than male politicians for equivalent misconduct* (H1).

Turning to the gender of voters, women have been found to on average be tougher on corruption (Alatas et al. 2009; Eckel and Grossman 1996), perhaps because they value honesty and integrity more. Our second hypothesis is thus: *women punish misconduct more severely than men* (H2).

Finally, the difference between male and female voters' behavior could depend on whether the *politician* is male or female. If female voters care more about corruption, and/or adhere more to gender stereotypes portraying female politicians as honest, then the stronger punishment for female politicians may come disproportionately from female voters. (For example, the “further to fall” hypothesis may apply only to female voters.) Alternatively, following Jones (2014), female voters might punish female politicians more because they are more engaged when evaluating female rather than male politicians. In either case, the higher aggregate punishment of female politicians might originate with female voters. Our third hypothesis is: *women punish female politicians more severely for misconduct than men* (H3).³

Experimental design

We use a population-based survey experiment similar to choice-based conjoint analysis (Hainmueller, Hopkins, and Yamamoto 2014). To our knowledge, our experiment is the first to consider how politician and voter gender affect punishment for corruption

²Formally, denote by $x_{g,c}$ the support level of a politician of gender $g \in \{m, f\}$ (where m is male, and f is female) given observed misconduct $c \in \{0, 1\}$. The claim is that $x_{f,0} - x_{f,1} > x_{m,0} - x_{m,1}$. In the simplest stereotype-based account, $x_{f,0} > x_{m,0}$, because the stereotype favors women in the absence of observed misconduct, but $x_{f,1} = x_{m,1}$, because once corruption is observed the stereotype does not matter.

³Accountability may also be lower when people evaluate politicians of the same gender as themselves (Jones 2014). Our experimental design allows us to test this.

(though Schwindt-Bayer, Verge, and Wiesehomeier (2016) have considered the *context* of corruption in conjunction with politician gender).

We surveyed 1,962 British voters on 2-3 June 2014, with respondents drawn from the YouGov online panel of over 360,000 people. The sample was designed to be representative of the British voting-age population in terms of age, gender, region, social grade and newspaper readership, though participants will be more experienced than the general population in taking (political) surveys. YouGov samples nevertheless accurately depict effect sizes of key predictors on vote choice (Sanders et al. 2007).

After an introductory screen, respondents were presented with five vignettes, each depicting a contest between an incumbent and a challenger (see Figure 1 for an example choice task). The incumbent could randomly exhibit good or bad conduct: ‘Last year, the current MP received a commendation for diligent and ethical service from a Westminster watchdog’ or ‘Last year, the current MP was found to have inappropriately claimed £10,000 in expenses’. By comparing bad to good behaviour (rather than to a neutral scenario), we provide the same amount of information in all scenarios and minimise the impact of respondents’ prior beliefs about MP behaviour. The gender, party, age and former job of both politicians also varied randomly. The incumbent could be Labour or Conservative, and the challenger Labour, Conservative, or Liberal Democrat. The possible ages were 45, 52 and 64 (incumbent) and 40, 52 and 64 (challenger). The previous jobs were GP, journalist, political advisor, teacher or business manager.

Respondents were asked: ‘If you were living in this constituency at the next general election, which party would you vote for?’ Depending on the parties of the politicians, the answer options were: ‘The current Labour/Conservative MP’, ‘the Labour/Conservative/Liberal Democrat challenger’, ‘the Labour/Conservative/Liberal Democrat candidate’, ‘a candidate from another party’ or ‘no one, I would not vote’.

In an observational study, it would be difficult to determine whether differences in how voters respond to misconduct by male and female MPs are due to hard-to-measure characteristics (such as the severity of misconduct or the nature of local political preferences) that may vary with MP gender. Different responses to comparable misconduct

Figure 1: Screenshot of example choice task

YouGov What the world thinks

Constituency A

This is a marginal constituency won narrowly by the Conservatives at the last election. Based on polls, the only other party with a chance of winning this seat are Labour. Here are the details of the current Conservative MP and the Labour challenger:

Current MP:  Conservative 64 years old Female Formerly a business manager	Main challenger:  Labour 62 years old Female Formerly a business manager
--	---

Last year, the current MP was found to have inappropriately claimed over £10,000 on expenses.

If you were living in this constituency at the next general election, who would you vote for?

- ☐ The current Conservative MP
- ☐ The Labour challenger
- ☐ The Liberal Democrat candidate
- ☐ A candidate from another party
- ☐ No one, I would not vote

by male and female MPs could also be caused by differences in how misconduct is reported: media coverage of misconduct depends on MP gender (Larcinese and Sircar 2017). In our experiment, the conduct and gender of the MP are, by design, unrelated to the political context and respondent characteristics, as is the information about MPs and their conduct. Thus, although differences in voter responses to similar behaviour by male and female MPs could in general be related to differences in the information voters receive about male and female MPs' behaviour or to voters' gender biases (or both), our experiment focuses only on the latter channel.

Our survey experiment cannot replicate real-world vote choice, but we try to maintain external validity in two ways. First, we present respondents with a multidimensional, reasonably realistic choice setting. This should also reduce social desirability bias, as respondents can justify their vote based on a number of considerations (Hainmueller, Hopkins, and Yamamoto 2014). Second, we primed respondents to think about partisan considerations rather than MP gender or conduct in office by displaying party logos in the choice tasks and including an introductory screen that characterized general elections as opportunities to select a national government (see Appendix A).

Results

Table 1 presents a series of linear probability models for incumbent vote. The raw descriptive results (Appendix B) exhibit the same patterns. Models with controls for respondent, incumbent and challenger characteristics (Appendix C); with interactions between treatments and all measured respondent characteristics (Appendix D); for the first choice task only (Appendix E); and taking into account challenger gender (Appendix F) also all yield substantively identical results.

Model 1 shows that corrupt MPs are penalized: the probability of choosing the incumbent MP is 24 percentage points lower for ‘bad’ than for ‘good’ MPs. Model 1 also shows that female incumbents in general are not less likely to be supported and that female voters are about as likely to vote for the incumbent as men.

Table 1: Probability of voting for incumbent, MP misconduct and gender

	(1)	(2)	(3)	(4)
Intercept	0.406*** (0.011)	0.401*** (0.012)	0.392*** (0.013)	0.400*** (0.015)
MP misconduct	-0.240*** (0.011)	-0.231*** (0.013)	-0.212*** (0.015)	-0.222*** (0.019)
MP female	0.014 (0.009)	0.023* (0.014)	0.014 (0.009)	-0.001 (0.020)
Respondent female	-0.008 (0.012)	-0.008 (0.012)	0.019 (0.017)	-0.005 (0.022)
MP misconduct × MP female		-0.018 (0.018)		0.020 (0.026)
MP misconduct × Resp. female			-0.055*** (0.021)	-0.019 (0.027)
MP female × Resp. female				0.048* (0.028)
MP misconduct × MP female × Resp. female				-0.072** (0.036)
Observations	9,810	9,810	9,810	9,810
R ²	0.071	0.071	0.072	0.072
Adjusted R ²	0.070	0.070	0.071	0.071

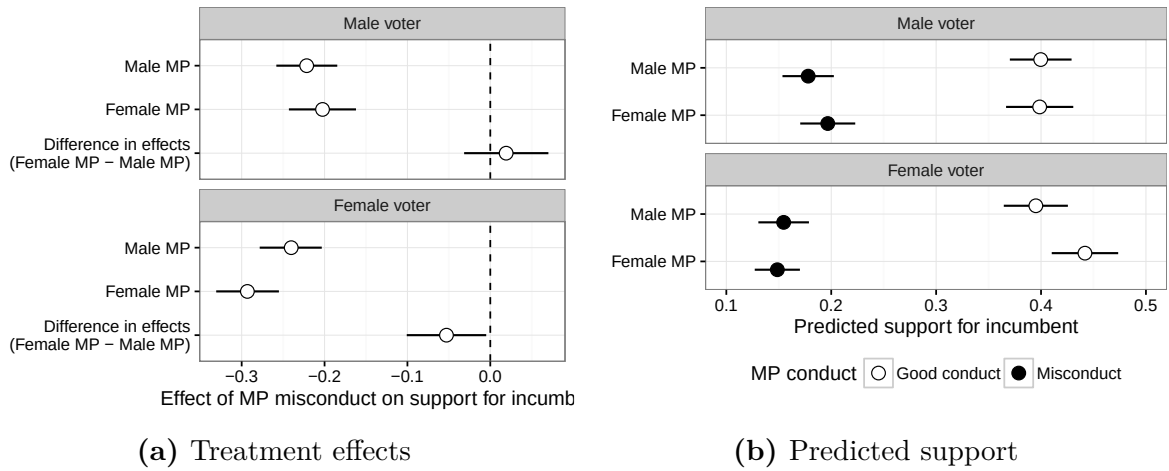
NOTE: OLS models. Dependent variable: Respondent votes for the incumbent MP (1) or not (0). Standard errors clustered by respondent. *p<0.1; **p<0.05; ***p<0.01.

Model 2 shows little support for H1: female politicians are *not* punished significantly more for wrongdoing than male politicians. The probability of voting for a male politician is 23 percentage points lower on average if the MP engaged in misconduct. This

punishment is only 2 percentage points larger for female politicians, and the difference in effects is not statistically significant.

Model 3 shows that, consistent with H2, female respondents punish MP misconduct more harshly (by over 5 percentage points) than male respondents. While this echoes a finding in Esarey and Chirillo (2013), the differential punishment by gender we detect cannot be attributed to differences in the extent to which male and female voters are aware of MP misconduct.

Figure 2: Effect of incumbent conduct by incumbent gender and respondent gender



Note: Panel (a) compares the treatment effects of MP misconduct as MP and voter gender varies. Panel (b) plots predicted probability of voting for the incumbent. 95% confidence intervals shown. Based on Model 4, Table 1.

Next, Model 4 tests H3 by including a three-way interaction that permits punishment by MP gender to vary by voter gender. This interaction term is significant: the conditioning effect of politician gender on punishment for misconduct is different for male and female voters. Figure 2a presents the estimated treatment effect of MP misconduct across different combinations of voter gender and MP gender. Among female voters, the negative effect of misconduct is about 5 percentage points greater for female incumbent MPs. In contrast, among male voters, the effect of misconduct does not differ significantly by MP gender, although the point estimate of punishment is slightly smaller (in absolute terms) for female MPs.

Figure 2b presents predicted probabilities based on Model 4. These show two reasons why the punishment is particularly large when the voter *and* the MP are female: the

highest incumbent support is when a female respondent faces a non-corrupt female incumbent, and the *lowest* incumbent support is when a female respondent faces a corrupt female incumbent. So, female respondents particularly like a female incumbent who is not corrupt and particularly dislike a female incumbent who is corrupt. The figure also helps to explain why we find no difference in punishment of female and male MPs (H1): female voters appear to punish female MPs more than male MPs, but male voters if anything punish them less.

Conclusion

Our results suggest that voters on average punish misconduct similarly among male and female politicians. If female politicians do face greater accountability, this is probably not because voters treat female politicians more harshly. Yet, our findings leave open the possibility that voters are more aware of misconduct by female politicians (e.g. Lar-cinese and Sircar 2017) or perceive similar behaviour by men and women differently. Our findings also suggest that female voters are more responsive to corruption among female than among male politicians, in particular because women react more to *good* behaviour by female politicians. This provides the first experimental evidence that men and women differ in how they hold male and female politicians accountable for misconduct.

Would similar effects be found in a real UK election? While we used a representative sample and included strong incumbency and party cues and a weak gender cue, our effects might be biased upwards if our participants reacted to key attributes more strongly than they would in a real election. Would similar effects would be found in other contexts? Electoral systems affect accountability patterns (Esarey and Schwindt-Bayer forthcoming). Gender effects might be greater under open-list than under closed-list proportional representation systems, as in the former it is easier to cast a personal vote. Future research in other institutional settings can usefully test this proposition.

Acknowledgements

We would like to thank Catherine de Vries, Miriam Golden, and three anonymous reviewers for useful suggestions.

References

- Alatas, Vivi, Lisa Cameron, Ananish Chaudhuri, Nisvan Erkal, and Lata Gangadharan. 2009. "Gender, culture, and corruption: Insights from an experimental analysis." *Southern Economic Journal* 75 (3): 663–680.
- Alexander, Deborah, and Kristi Andersen. 1993. "Gender as a Factor in the Attribution of Leadership Traits." *Political Research Quarterly* 46 (3): 527–545.
- Anzia, Sarah F, and Christopher R Berry. 2011. "The Jackie (and Jill) Robinson effect: Why do congresswomen outperform congressmen?" *American Journal of Political Science* 55 (3): 478–493.
- Bjarnegård, Elin. 2013. *Gender, Informal Institutions and Political Recruitment: Explaining Male Dominance in Parliamentary Representation*. Palgrave Macmillan.
- Dolan, Kathleen. 2014a. "Gender Stereotypes, Candidate Evaluations, and Voting for Women Candidates What Really Matters?" *Political Research Quarterly* 67 (1): 96–107.
- Dolan, Kathleen. 2014b. *When Does Gender Matter?: Women Candidates and Gender Stereotypes in American Elections*. Oxford University Press.
- Dolan, Kathleen A. 2004. *Voting for women: How the public evaluates women candidates*. Westview Press.
- Dollar, David, Raymond Fisman, and Roberta Gatti. 2001. "Are women really the "fairer" sex? Corruption and women in government." *Journal of Economic Behavior & Organization* 46 (4): 423–429.
- Eckel, Catherine C, and Philip J Grossman. 1996. "The relative price of fairness: Gender differences in a punishment game." *Journal of Economic Behavior & Organization* 30 (2): 143–158.
- Esarey, Justin, and Gina Chirillo. 2013. "'Fairer Sex' or Purity Myth? Corruption, Gender, and Institutional Context." *Politics & Gender* 9 (04): 361–389.
- Esarey, Justin, and Leslie Schwindt-Bayer. forthcoming. "Women's Representation, Accountability, and Corruption in Democracies." *British Journal of Political Science*.
- Frederick, Brian, and Matthew J Streb. 2008. "Women Running for Judge: The Impact

- of Sex on Candidate Success in State Intermediate Appellate Court Elections.” *Social Science Quarterly* 89 (4): 937–954.
- Fridkin, Kim L, Patrick J Kenney, and Gina Serignese Woodall. 2009. “Bad for men, better for women: The impact of stereotypes during negative campaigns.” *Political Behavior* 31 (1): 53–77.
- Funk, Carolyn L. 1996. “The impact of scandal on candidate evaluations: An experimental test of the role of candidate traits.” *Political Behavior* 18 (1): 1–24.
- Hainmueller, Jens, Daniel J Hopkins, and Teppei Yamamoto. 2014. “Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments.” *Political Analysis* 22 (1): 1–30.
- Jones, Philip Edward. 2014. “Does the descriptive representation of gender influence accountability for substantive representation?” *Politics & Gender* 10 (02): 175–199.
- Kahn, Kim Fridkin. 1992. “Does being male help? An investigation of the effects of candidate gender and campaign coverage on evaluations of US Senate candidates.” *The Journal of Politics* 54 (2): 497–517.
- Larcinese, Valentino, and Indraneel Sircar. 2017. “Crime and punishment the British Way: Accountability channels following the MPs’ expenses scandal.” *European Journal of Political Economy* 47: 75–99.
- Sanders, David, Harold D Clarke, Marianne C Stewart, and Paul Whiteley. 2007. “Does mode matter for modeling political choice? Evidence from the 2005 British Election Study.” *Political Analysis* 15 (3): 257–285.
- Schwindt-Bayer, Leslie, Tania Verge, and Nina Wiesehomeier. 2016. Corruption and gendered perception of political elites. Paper presented at the 2016 Annual Meeting of the American Political Science Association.
- Swamy, Anand, Stephen Knack, Young Lee, and Omar Azfar. 2001. “Gender and corruption.” *Journal of development economics* 64 (1): 25–55.

Biographical statement

Andrew C. Eggers is an Associate Professor in the Department of Politics and International Relations at the University of Oxford and a Professorial Fellow at Nuffield College.

Nick Vivyan is an Associate Professor in the School of Government and International Affairs at Durham University.

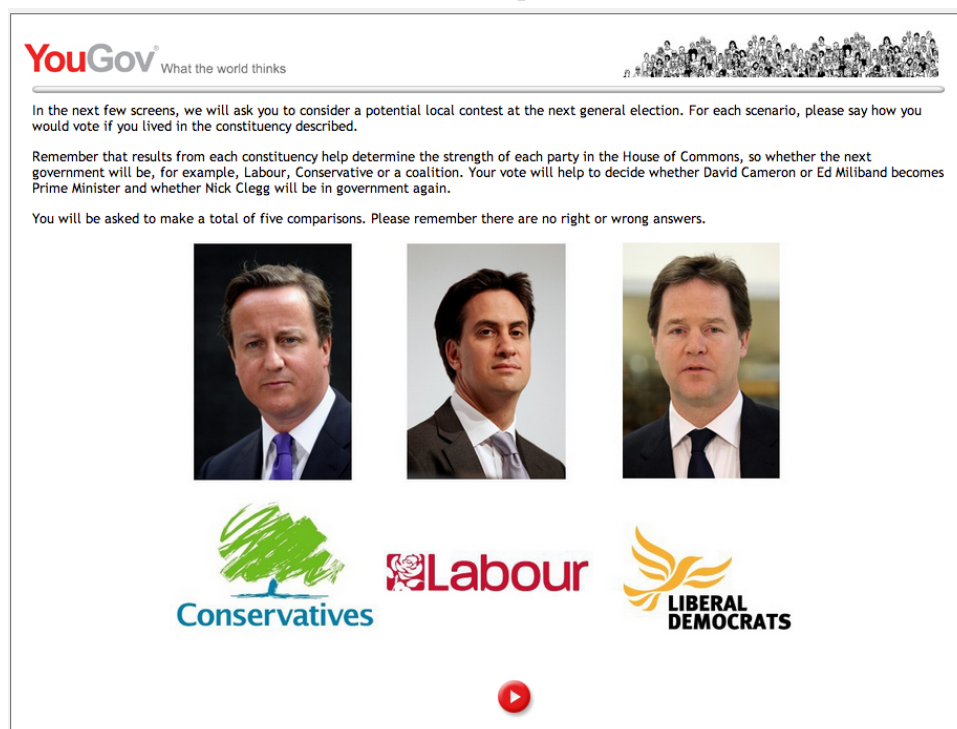
Markus Wagner is an Associate Professor in the Department of Government at the University of Vienna.

Appendices

A Introductory screen in online experiment

Figure A.1 shows the introductory screen that respondents received at the start of our experiment. Note that the screen primes respondents to think about a general election in terms of national government formation, political parties and party leaders. This emphasis was intended to enhance the external validity of respondent choices in the subsequent constituency contests. Partisan electoral considerations are often primed by national media in the run-up to general elections and such considerations may often overpower gender stereotypes (Dolan 2014a).

Figure A.1: Screenshot of introduction to experiment



B Raw incumbent voting rates

Table B.1: Voting rates by MP misconduct, MP gender and respondent gender

Resp. gender	MP gender	MP conduct	Incumbent	
			Share	N
Male	Male	Good	40.0	1246
		Bad	17.8	1191
	Female	Good	39.9	1224
		Bad	19.7	1139
Female	Male	Good	39.5	1210
		Bad	15.5	1261
	Female	Good	44.2	1247
		Bad	14.9	1292

NOTE: The table reports the percentage support for the incumbent MP among observations with each combination of MP conduct, MP gender and respondent gender, as well as the total number of each type of observation.

C Results with controls

The regression models reported in Table 1 of the main text do not include any control variables. In this Appendix, we show results when we control for respondent, MP and challenger characteristics, because including these controls substantially improves the precision of parameter estimates. The precise controls for the respondents are: gender, age group (25-39, 40-59, 60 and above), social grade (AB or C1 vs C2 or DE) and indicators for whether the respondent identifies with the incumbents' party, the challenger's party, or some other party (vs no reported party identification). The MP and challenger characteristics we control for are gender, age and previous occupation. Finally, we also control for the party match-up in vignette (Conservative vs Labour, Labour vs Conservative, Conservative vs Liberal Democrat, Labour vs Liberal Democrat). We re-estimate each of the models from Table 1 with these controls. Table C.1 shows the results for each model specification with and without controls. Comparing each pair of Models (e.g., Models 7 and 8), it is clear that adding or removing controls yields has little effect on the coefficient point estimates, though as expected standard errors are reduced in the models with controls.

Table C.1: Probability of voting for incumbent MP by misconduct and gender, including models without controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.406*** (0.011)	0.262*** (0.024)	0.401*** (0.012)	0.259*** (0.024)	0.392*** (0.013)	0.250*** (0.024)	0.400*** (0.015)	0.257*** (0.024)
MP misconduct	-0.240*** (0.011)	-0.243*** (0.009)	-0.231*** (0.013)	-0.236*** (0.011)	-0.212*** (0.015)	-0.216*** (0.012)	-0.222*** (0.019)	-0.230*** (0.015)
MP female	0.014 (0.009)	0.011 (0.007)	0.023* (0.014)	0.017* (0.010)	0.014 (0.009)	0.011 (0.007)	-0.001 (0.020)	-0.004 (0.014)
Respondent female	-0.008 (0.012)	-0.004 (0.009)	-0.008 (0.012)	-0.004 (0.009)	0.019 (0.017)	0.022* (0.013)	-0.005 (0.022)	0.001 (0.016)
MP misconduct × MP female			-0.018 (0.018)	-0.014 (0.014)			0.020 (0.026)	0.027 (0.019)
MP misconduct × Resp. female					-0.055*** (0.021)	-0.053*** (0.018)	-0.019 (0.027)	-0.013 (0.022)
MP female × Resp. female							0.048* (0.028)	0.043** (0.020)
MP misconduct × MP female × Resp. female							-0.072** (0.036)	-0.078*** (0.028)
Controls for voter, MP and chall. characteristics?	No	Yes	No	Yes	No	Yes	No	Yes
Observations	9,810	9,800	9,810	9,800	9,810	9,800	9,810	9,800
R ²	0.071	0.436	0.071	0.436	0.072	0.436	0.072	0.437
Adjusted R ²	0.070	0.434	0.070	0.434	0.071	0.435	0.071	0.435

NOTE: OLS models. Dependent variable is a binary measure of whether a respondent votes for the incumbent MP (1) or not (0). Controls are respondent gender, age group (25-39, 40-59, 60 and above), social grade (AB or C1 vs C2 or DE) and indicators for whether the respondent identifies with the incumbents' party, the challenger's party, or some other party (vs no reported party identification); incumbent and challenger gender, age and previous occupation and party match-up in vignette (Conservative vs Labour, Labour vs Conservative, Conservative vs Lib Dem, Labour vs Lib Dem). Standard errors clustered by respondent. *p<0.1; **p<0.05; ***p<0.01.

D Robustness to more extensive controls

In the main paper we find that female voters on average punish incumbent misconduct more than male voters, and that this difference is particularly pronounced when the incumbent is female rather than male. While MP misconduct and MP gender are randomly assigned treatments in our experiment, voter gender is of course not. Furthermore, even the models from Table C.1 of the Appendix C control only for the main effects of other voter characteristics. Therefore, one may wonder whether the observed differences in treatment effects among male and female voters could be attributable to some other respondent variable that co-varies with respondent gender in our sample.

To examine whether this might be the case, we re-estimated Models 6 and 8 from Table C.1 of Appendix C, which included an interaction between treatment variables and respondent gender, and added to these models equivalent interactions between treatment effects and each respondent control variable included in the original model. Thus, for the model which contains the MP misconduct $\times$ respondent gender interaction, we add an MP misconduct $\times$ Z interaction for every respondent control variable Z. For the model which contains the three-way interaction MP misconduct $\times$ MP gender $\times$ respondent gender, we add an MP misconduct $\times$ MP gender $\times$ Z interaction for every respondent control variable Z.

Table D.1 shows the results for each model specification with and without these additional interaction terms (Models 1 and 3 are equivalent to Models 6 and 8 in Table C.1 of Appendix C). Comparing each pair of Models, it is clear that adding or removing the interactions between treatments and respondent control variables has little effect on the magnitude or significance of the interactions between respondent gender and experimental treatments.

Table D.1: Probability of voting for incumbent MP by misconduct and gender, controlling for voter attribute $\times$ treatment interactions

	(1)	(2)	(3)	(4)
MP misconduct $\times$ Resp. female	−0.053*** (0.018)	−0.056*** (0.018)	−0.013 (0.022)	−0.078 (0.063)
MP misconduct $\times$ MP female $\times$ Resp. female			−0.078*** (0.028)	−0.070** (0.027)
Controls for main effects of voter, MP and chall. characteristics?	Yes	Yes	Yes	Yes
Controls for voter characteristic $\times$ MP misconduct interactions?	No	Yes	No	Yes
Controls for voter characteristic $\times$ MP misconduct $\times$ MP gender interactions and constituent terms?	No	No	No	Yes
Observations	9,800	9,800	9,800	9,800
R ²	0.436	0.464	0.437	0.466
Adjusted R ²	0.435	0.462	0.435	0.463

NOTE: OLS models. Dependent variable is a binary measure of whether a respondent votes for the incumbent MP (1) or not (0). Voter characteristic controls are gender, age group (25-39, 40-59, 60 and above), social grade (AB or C1 vs C2 or DE) and indicators for whether the respondent identifies with the incumbents' party, the challenger's party, or some other party (vs no reported party identification). Incumbent and challenger characteristic controls are incumbent and challenger gender, age and previous occupation and party match-up in vignette (Conservative vs Labour, Labour vs Conservative, Conservative vs Lib Dem, Labour vs Lib Dem). Standard errors clustered by respondent. *p<0.1; **p<0.05; ***p<0.01.

E First choice task versus all choice tasks

The regression models reported in Table 1 of the main text are estimated based on respondents' choices in all five choice tasks. To check that our results are not an artefact of respondent learning as they progress through choice tasks, we re-estimate each model on data from respondents' first choice task only. Tables E.1 and E.2 show the results for each model specification in the first-task subsample and the full sample, without and with controls, respectively. Comparing each pair of Models (e.g., Models 7 and 8), it is clear that subsetting to data from the first choice task generally yields little change in the direction and magnitude of coefficient point estimates, although standard errors are substantially increased due to the substantial reduction in sample size. The one exception is Model 8 in Table E.1, which shows different patterns than Model 7; note, however, that in the Models with controls in Table E.2, the patterns for the first task and all tasks remain very similar.

Table E.1: Robustness of main regression results in first choice task subsample

	All tasks (1)	First task (2)	All tasks (3)	First task (4)	All tasks (5)	First task (6)	All tasks (7)	First task (8)
Intercept	0.406*** (0.011)	0.398*** (0.019)	0.401*** (0.012)	0.405*** (0.022)	0.392*** (0.013)	0.372*** (0.022)	0.400*** (0.015)	0.371*** (0.027)
MP misconduct	-0.240*** (0.011)	-0.254*** (0.019)	-0.231*** (0.013)	-0.266*** (0.027)	-0.212*** (0.015)	-0.202*** (0.028)	-0.222*** (0.019)	-0.204*** (0.039)
MP female	0.014 (0.009)	0.027 (0.019)	0.023* (0.014)	0.013 (0.028)	0.014 (0.009)	0.028 (0.019)	-0.001 (0.020)	0.030 (0.039)
Respondent female	-0.008 (0.012)	-0.003 (0.019)	-0.008 (0.012)	-0.004 (0.019)	0.019 (0.017)	0.049* (0.028)	-0.005 (0.022)	0.065* (0.039)
MP misconduct $\times$ MP female			-0.018 (0.018)	0.026 (0.039)			0.020 (0.026)	0.003 (0.055)
MP misconduct $\times$ Resp. female					-0.055*** (0.021)	-0.101*** (0.039)	-0.019 (0.027)	-0.125** (0.055)
MP female $\times$ Resp. female							0.048* (0.028)	-0.033 (0.056)
MP misconduct $\times$ MP female $\times$ Resp. female							-0.072** (0.036)	0.048 (0.078)
Observations	9,810	1,962	9,810	1,962	9,810	1,962	9,810	1,962
R ²	0.071	0.081	0.071	0.081	0.072	0.084	0.072	0.084
Adjusted R ²	0.070	0.079	0.070	0.079	0.071	0.082	0.071	0.081

NOTE: OLS models run on observations from respondents' first choice task only (odd-numbered models) or observations from all choice tasks (even-numbered models). Dependent variable is a binary measure of whether a respondent votes for the incumbent MP (1) or not (0). For models run on data from all choice tasks, standard errors clustered by respondent. For models run on data from respondents' first choice task only, standard errors are heteroskedasticity robust (HC3). *p<0.1; **p<0.05; ***p<0.01.

Table E.2: Robustness of main regression results in first choice task subsample, incorporating controls

	All tasks (1)	First task (2)	All tasks (3)	First task (4)	All tasks (5)	First task (6)	All tasks (7)	First task (8)
Intercept	0.262*** (0.024)	0.278*** (0.044)	0.259*** (0.024)	0.279*** (0.045)	0.250*** (0.024)	0.259*** (0.045)	0.257*** (0.024)	0.279*** (0.047)
MP misconduct	-0.243*** (0.009)	-0.264*** (0.015)	-0.236*** (0.011)	-0.267*** (0.022)	-0.216*** (0.012)	-0.227*** (0.022)	-0.230*** (0.015)	-0.254*** (0.031)
MP female	0.011 (0.007)	0.035** (0.015)	0.017* (0.010)	0.032 (0.022)	0.011 (0.007)	0.035** (0.015)	-0.004 (0.014)	-0.003 (0.031)
Respondent female	-0.004 (0.009)	0.003 (0.016)	-0.004 (0.009)	0.003 (0.016)	0.022* (0.013)	0.041* (0.022)	0.001 (0.016)	0.006 (0.031)
MP misconduct $\times$ MP female			-0.014 (0.014)	0.006 (0.031)			0.027 (0.019)	0.054 (0.044)
MP misconduct $\times$ Resp. female					-0.053*** (0.018)	-0.073** (0.031)	-0.013 (0.022)	-0.026 (0.044)
MP female $\times$ Resp. female							0.043** (0.020)	0.072 (0.045)
MP misconduct $\times$ MP female $\times$ Resp. female							-0.078*** (0.028)	-0.095 (0.062)
Controls for voter, MP and chall. characteristics?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,800	1,960	9,800	1,960	9,800	1,960	9,800	1,960
R ²	0.436	0.429	0.436	0.429	0.436	0.430	0.437	0.431
Adjusted R ²	0.434	0.421	0.434	0.421	0.435	0.422	0.435	0.422

NOTE: OLS models run on observations from respondents' first choice task only (odd-numbered models) or observations from all choice tasks (even-numbered models). Dependent variable is a binary measure of whether a respondent votes for the incumbent MP (1) or not (0). Controls are respondent gender, age group (25-39, 40-59, 60 and above), social grade (AB or C1 vs C2 or DE) and indicators for whether the respondent identifies with the incumbents' party, the challenger's party, or some other party (vs no reported party identification); incumbent and challenger gender, age and previous occupation and party match-up in vignette (Conservative vs Labour, Labour vs Conservative, Conservative vs Lib Dem, Labour vs Lib Dem). For models run on data from all choice tasks, standard errors clustered by respondent. For models run on data from respondents' first choice task only, standard errors are heteroskedasticity robust (HC3). *p<0.1; **p<0.05; ***p<0.01.

F Conditioning effects of challenger gender

Here we present evidence on how challenger gender interacts with voter and incumbent gender in shaping reactions to misconduct. It could be that having two main candidates who are women increases the attention that women pay to the political situation, or that women respond to the female-only contests differently because they can hold the female incumbent accountable without sacrificing another potential goal, namely increasing substantive representation through having a female MP.

To examine these possibilities we divide the sample into male and female respondents, then for each sub-sample estimate a regression model of incumbent voting that includes a three-way interaction between MP conduct, MP gender and the gender of the main challenger. We thus allow for the interaction between MP misconduct, MP gender and challenger gender to vary for male and female respondents. Table F.1 shows the estimated models. The coefficient estimates are very similar whether we include controls (Models 3 and 4) or not (Models 1 and 2).

Based on the models with controls, Figure F.1 compares the marginal effects of MP misconduct across varying combinations of MP, challenger and respondent gender. There is some suggestive evidence in Figure F.1 that female voters (bottom panel) punish misbehaving female MPs more than misbehaving male MPs to a greater extent when the main challenger is also female. According to the p-value on the three-way interaction term in Model 4 of Table F.1 this difference is significant at the 0.1 level. In contrast, for male voters (top panel) there is less of a difference in punishment for misconduct regardless of MP or challenger gender.

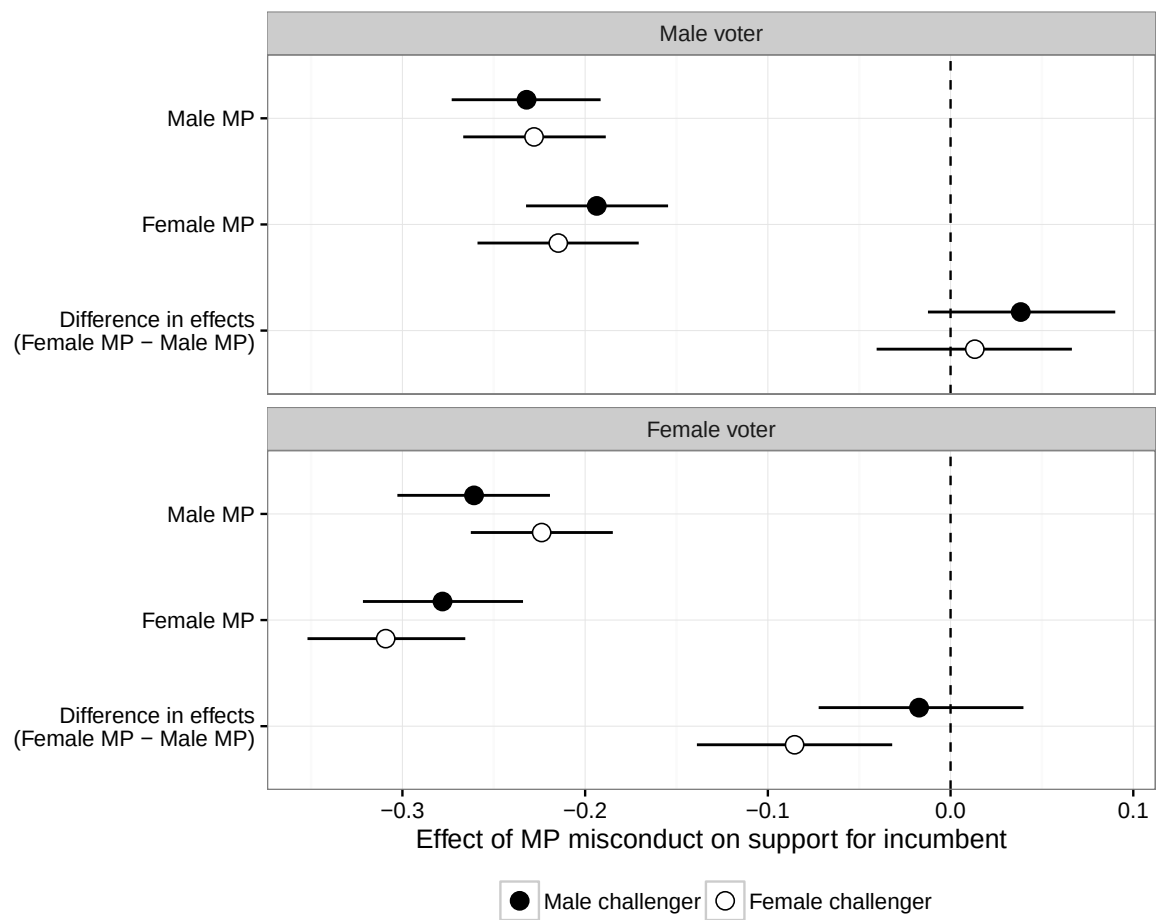
In sum, we find some tentative evidence that, among female voters, the difference in punishment of male and female politicians may be more pronounced when the main challenger is also female.

Table F.1: Probability of voting for incumbent, MP misconduct, MP gender and challenger gender

	Male resp. (1)	Female resp. (2)	Male resp. (3)	Female resp. (4)
Intercept	0.402*** (0.021)	0.396*** (0.022)	0.274*** (0.034)	0.257*** (0.034)
MP misconduct	-0.224*** (0.026)	-0.251*** (0.026)	-0.232*** (0.021)	-0.261*** (0.022)
MP female	0.012 (0.028)	0.033 (0.028)	-0.007 (0.020)	0.024 (0.021)
Challenger female	-0.004 (0.028)	-0.002 (0.028)	-0.012 (0.020)	-0.015 (0.020)
MP misconduct $\times$ MP female	0.023 (0.037)	-0.022 (0.034)	0.038 (0.026)	-0.017 (0.029)
MP misconduct $\times$ Chall. female	0.005 (0.035)	0.020 (0.034)	0.004 (0.027)	0.037 (0.026)
MP female $\times$ Chall. female	-0.026 (0.039)	0.028 (0.040)	0.005 (0.027)	0.030 (0.030)
MP misconduct $\times$ MP female $\times$ Chall. female	-0.009 (0.051)	-0.063 (0.048)	-0.024 (0.038)	-0.068* (0.039)
Controls for voter, MP and chall. characteristics?	No	No	Yes	Yes
Observations	4,800	5,010	4,795	5,005
R ²	0.055	0.090	0.463	0.417
Adjusted R ²	0.053	0.088	0.460	0.413

NOTE: OLS models. Dependent variable is a binary measure of whether a respondent votes for the incumbent MP (1) or not (0). Controls are respondent gender, age group (25-39, 40-59, 60 and above), social grade (AB or C1 vs C2 or DE) and indicators for whether the respondent identifies with the incumbents' party, the challenger's party, or some other party (vs no reported party identification); incumbent and challenger gender, age and previous occupation and party match-up in vignette (Conservative vs Labour, Labour vs Conservative, Conservative vs Lib Dem, Labour vs Lib Dem). Standard errors clustered by respondent. *p<0.1; **p<0.05; ***p<0.01.

Figure F.1: Effect of misconduct by voter, incumbent and challenger gender



Note: Effects calculated based on Models 3 and 4, Table F.1.