Supplementary Appendix – Online Only

S1. Underrepresentation of women in parliamentary debate

I document the gender-gap in debate participation in the House of Commons by measuring the 'female speech ratio' (see equation S1 in section S3 below) for each parliamentary debate. The ratio is defined as the proportion of words in a debate spoken by women, divided by the proportion of women holding seats in parliament during the period of time in which the debate was held. When the ratio is equal to one the proportion of words spoken by women is equal to the proportion of seats held by women. When the ratio is less than one, women are underrepresented in parliamentary debate.

For the purposes of the analysis in this section, I focus only on backbench MPs (i.e. those who do not hold a government frontbench position, an opposition frontbench position, a position in the leadership of a political party, or a parliamentary committee chair position).[S1] Across the 14320 debates I study, the mean female speech ratio is 0.84 [0.82, 0.86] and the median ratio is 0.44 [0.42, 0.47]. This median figure implies that in the typical debate in parliament, women contribute a little more than half the words we would expect given their numerical strength. That is, although women occupied approximately 20% of the seats in the House of Commons during this period, in the typical debate they contributed only 10% of the words spoken by backbenchers.

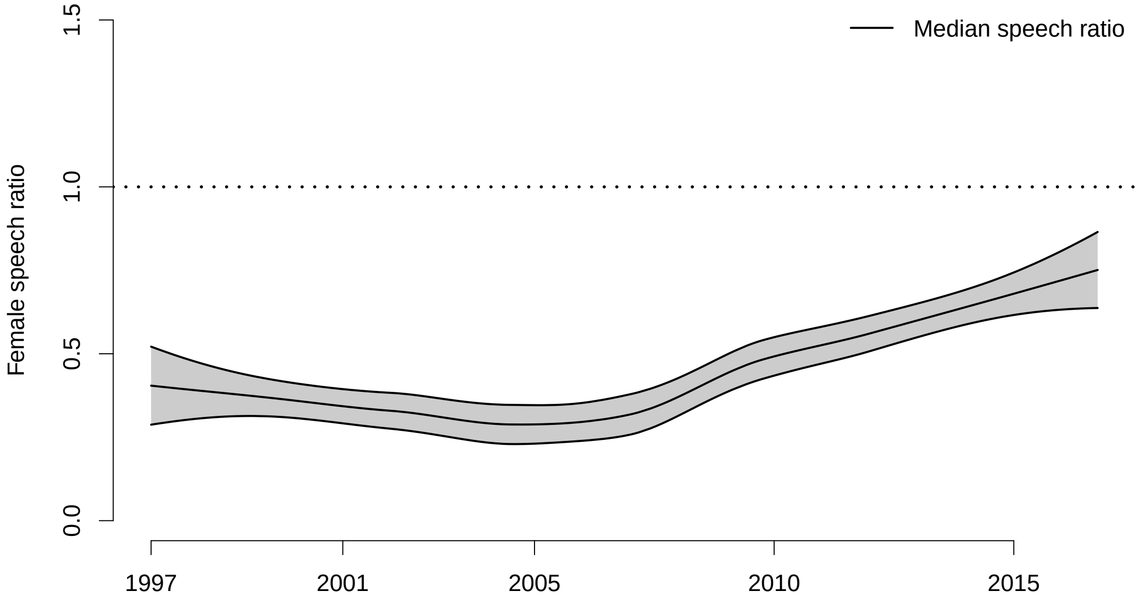
Figure S1 demonstrates how the median female speech ratio has changed over the 20 years of the study period. The solid black line is a loess curve of the median female speech ratio in each calendar month. The plot demonstrates that, even conditional on the fact that they are underrepresented numerically in parliament, women have also been underrepresented in parliamentary debates for the majority of this period, though there has been a somewhat more equitable balance of speaking time in debates in recent years.

Of course, there are numerous possible explanations for this underrepresentation. Note that these results are not driven by the fact that women are less likely to be appointed than men to cabinet positions, junior ministerial positions, committee chairs, or party leadership positions, as all MPs holding such positions are excluded from this analysis. However, Female MPs have, on average, lower levels of seniority in parliament than their male colleagues – particularly in this time period, where the 1997 Labour landslide brought many new women into parliament which may contribute to the proportionally low levels of speech documented here.

That the mean speech ratio is higher than the median speech ratio implies that there are certain sets of debates in which women participate much more than men. In figure S2 I plot the average female speech ratio in debates pertaining to each government department, pooling across years. As expected, there is significant variation across ministries. Women are somewhat overrepresented in debates pertaining to "Transport, Local Government and Regions", "Education", "Health", "Children, Schools and Families", and "Work and Pensions". By contrast, women are underrepresented in debates pertaining to a wide-variety of

[S1] In the analyses presented in the main body of the paper, I calculate equation S1 excluding only speeches made by the minister responsible for debate and the Speaker of the House of Commons. Sample restrictions are fully explained in the "Data and Methodology" section of the paper.

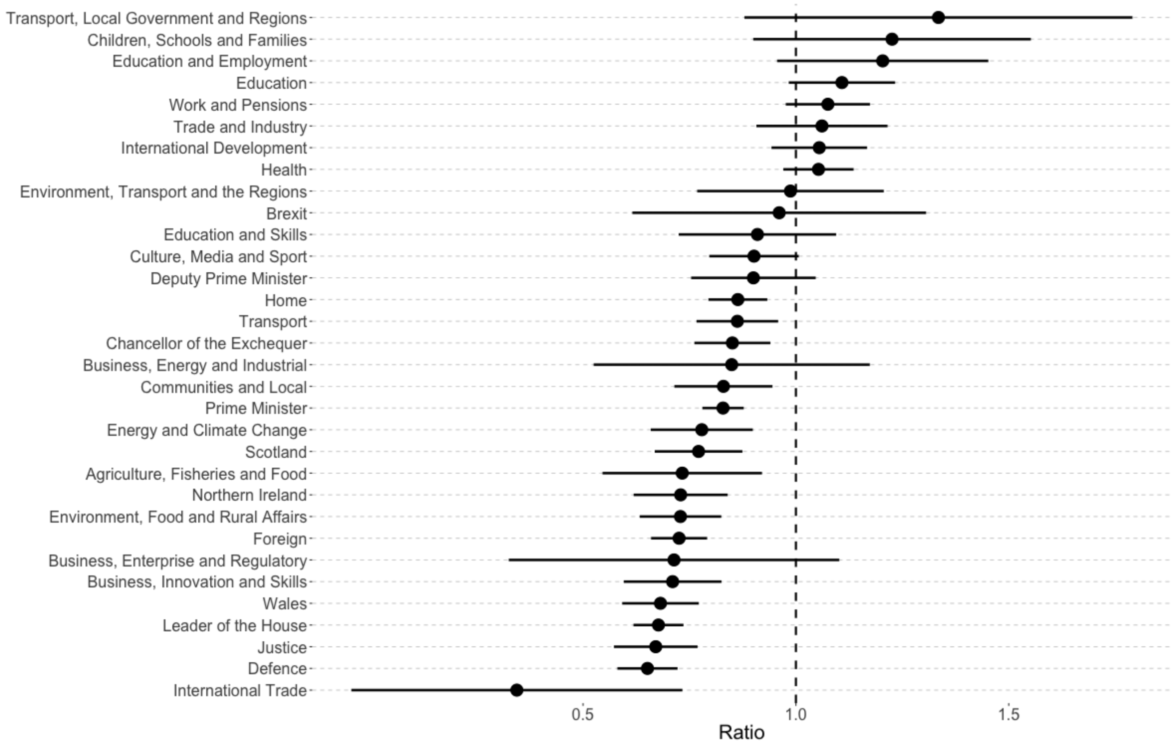
Figure S1: Female speech ratio, over time



NOTE: The figure shows smoothed loess curve of the median female speech ratio each month between 1997 to 2017 based on 14320 debates. The shaded section indicates the 95% confidence interval. The horizontal dotted black line indicates the expected level of speech when female MPs' contributions to plenary debate are equal to their representation in the House.

government departments, including, *inter alia*, the Defence ministry, the Foreign ministry, the Justice ministry, and debates presided over by the Prime Minister.

Figure S2: Female speech ratio, by ministry

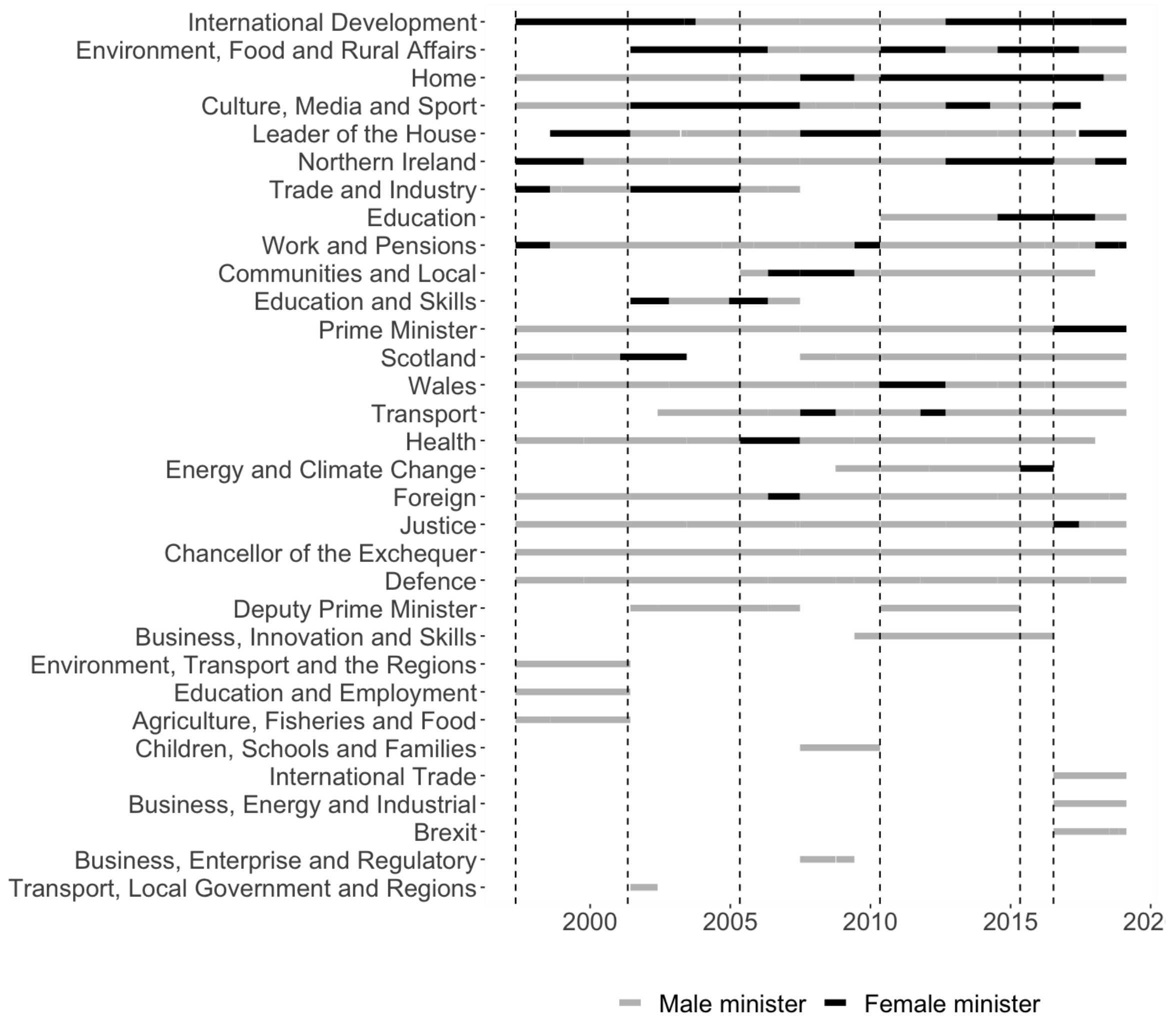


NOTE: The figure shows the average female speech ratio as defined in equation S1 for each ministry, pooled across all debates in the data. It is clear from the figure that some ministries are subject to greater levels of female participation than others.

S2. Independent variable

Figure S3 shows the variation in the independent variable over time for all 32 ministries included in the sample. Ministries are sorted by the proportion of the time period that the ministry is occupied by a female minister. Gray bars pertain to periods in which the minister responsible is male, and black bars represent female ministers. There are several ministries for which the responsible minister is never a woman, including the the Defence ministry and the position of Chancellor of the Exchequer.

Figure S3: Gender of ministers over time



Note: The figure shows the distribution of the independent variable over time. While some ministries are never held by a woman (those all in gray), the gender of the minister in several ministries varies over time.

S3. Alternative dependent variables

Equations S1, S2, and S3 provide alternative definitions of the dependent variable in equation 1. Results for the main fixed-effects models using these alternative operationalisations are presented in tables S1, S2, and S3 below. Regardless which of these measures is used, the main results hold: the appointment of a female minister leads to an increase in the level of participation in parliamentary debates by female MPs.

$$RatioWordsWomen_{d(mt)} = \frac{PropWordsWomen_{d(mt)}}{\text{Proportion of women in parliament}_t} \tag{S1}$$

$$PropSpeechesWomen_{d(mt)} = \frac{\text{\# speeches by women}_{d(mt)}}{\text{\# speeches by men and women}_{d(mt)}} \tag{S2}$$

$$RatioSpeechesWomen_{d(mt)} = \frac{PropSpeechesWomen_{d(mt)}}{\text{Proportion of women in parliament}_t} \tag{S3}$$

Table S1: Effect of appointing of a female minister on female speech (ratio of words)

	RatioWordsWomen						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.256***	0.253***	0.195***	0.188***	0.174***	0.174***	0.177***
	(0.072)	(0.064)	(0.070)	(0.052)	(0.054)	(0.048)	(0.036)
Constant	0.806***	0.541**	0.561***	0.227	0.264	5.517	0.811
	(0.045)	(0.259)	(0.183)	(0.274)	(10.249)	(569.392)	(2.381)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	32	31	24	23	22	22	22
95% CI	[14,49]	[16,47]	[7,41]	[11,36]	[8,35]	[10,33]	[13,31]
Observations	14,320	14,320	14,320	14,320	14,320	14,320	14,320
R^2	0.010	0.043	0.045	0.075	0.081	0.090	
Adjusted R^2	0.010	0.029	0.043	0.059	0.063	0.070	0.082

NOTE: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with bootstrapped robust standard errors (clustered by ministry) shown in parentheses. Models 5 and 6 include linear and quadratic ministry-specific time trends in addition to ministry fixed-effects and month fixed-effects. Model 7 is a GAM model including non-parametric, ministry-specific, flexible time trends. The "Effect Size" row indicates the percentage increase in female participation relative to the average female participation rate under male ministers. *p<0.1; **p<0.05; ***p<0.01

Table S2: Effect of the appointment of a female minister on female participation (proportion of speeches)

	PropSpeechesWomen						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.050***	0.043***	0.043***	0.034***	0.032***	0.032***	0.032***
	(0.015)	(0.011)	(0.015)	(0.008)	(0.009)	(0.008)	(0.007)
Constant	0.179***	0.092*	0.122*	0.052	0.048	−1.740	0.555
	(0.009)	(0.051)	(0.074)	(0.051)	(2.390)	(171.404)	(0.932)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	28	24	24	19	18	18	18
95% CI	[11,44]	[12,37]	[7,41]	[10,28]	[8,28]	[9,27]	[11,26]
Observations	14,320	14,320	14,320	14,320	14,320	14,320	14,320
R^2	0.008	0.062	0.040	0.085	0.089	0.097	
Adjusted R^2	0.008	0.048	0.038	0.069	0.072	0.077	0.082

Note: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with bootstrapped robust standard errors (clustered by ministry) shown in parentheses. Models 5 and 6 include linear and quadratic ministry-specific time trends in addition to ministry fixed-effects and month fixed-effects. Model 7 is a GAM model including non-parametric, ministry-specific, flexible time trends. The "Effect Size" row indicates the percentage increase in female participation relative to the average female participation rate under male ministers. *p<0.1; **p<0.05; ***p<0.01

Table S3: Effect of the appointment of a female minister on female participation (ratio of speeches)

	RatioSpeechesWomen						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.201***	0.198***	0.162***	0.156***	0.146***	0.144***	0.149***
	(0.053)	(0.053)	(0.044)	(0.038)	(0.045)	(0.042)	(0.032)
Constant	0.819***	0.525*	0.679***	0.299	0.275	−1.067	3.461
	(0.038)	(0.274)	(0.130)	(0.272)	(8.197)	(592.332)	(5.673)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	24	24	20	19	18	18	18
95% CI	[12,37]	[12,37]	[9,30]	[10,28]	[7,29]	[7,28]	[10,26]
Observations	14,320	14,320	14,320	14,320	14,320	14,320	14,320
R^2	0.006	0.033	0.033	0.056	0.060	0.068	
Adjusted R^2	0.006	0.018	0.031	0.040	0.042	0.048	0.053

Note: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with bootstrapped robust standard errors (clustered by ministry) shown in parentheses. Models 5 and 6 include linear and quadratic ministry-specific time trends in addition to ministry fixed-effects and month fixed-effects. Model 7 is a GAM model including non-parametric, ministry-specific, flexible time trends. The "Effect Size" row indicates the percentage increase in female participation relative to the average female participation rate under male ministers. *p<0.1; **p<0.05; ***p<0.01

S4. Subset analyses

In figure S4 I present the results from a series of subset analyses. For each of the subsets listed below, I estimate the main model (equation 3) from the body of the paper. The figure shows the main substantive effect of interest – the increase in female debate participation that results from the appointment of a female cabinet minister relative to the baseline under male ministers – for each subset. The x-axis in the figure measures the effect size, and each point represents the effect as measured for a particular subset. The gray shaded interval indicates the effect size in the full sample (as given in table 1). Confidence intervals are constructed by bootstrapping each model 500 times, blocking on ministry.

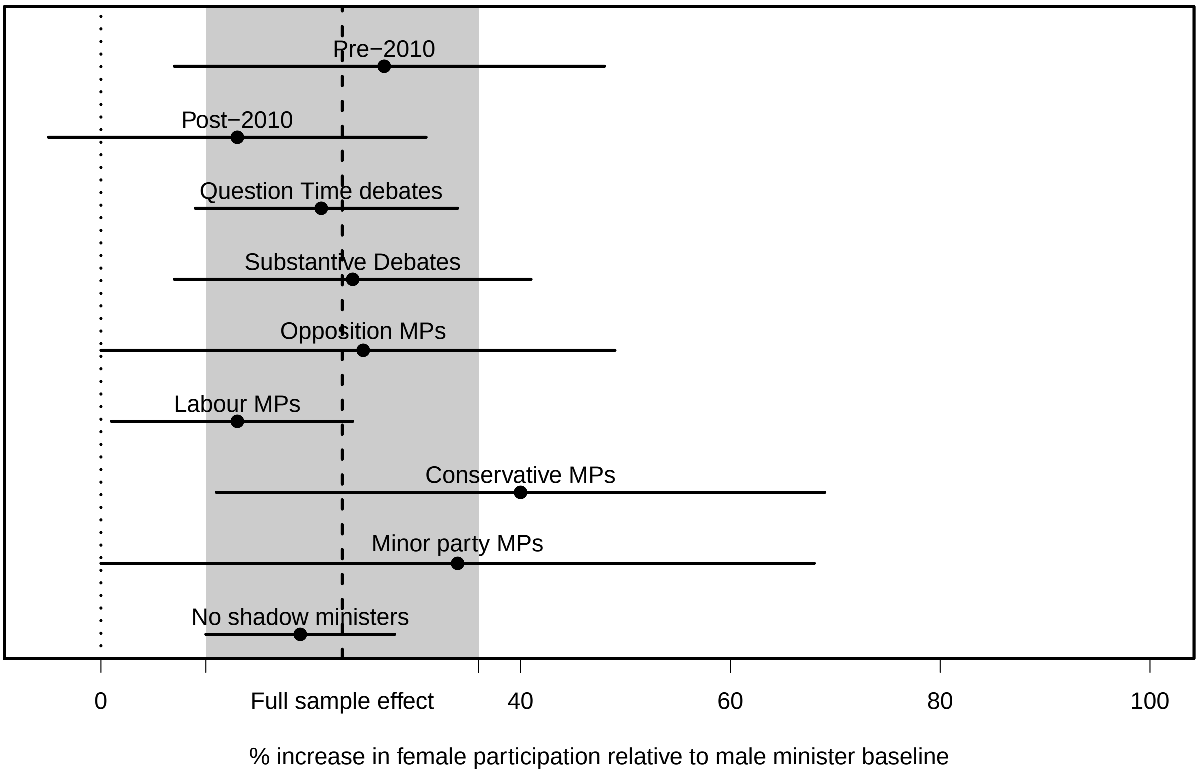
- *Before and After 2010* – From 1997 to 2010, under Labour-led governments, almost all of the female legislators talking in the presence of a female minister would have been directing their speeches to a copartisan. After 2010, under Conservative-led governments, the majority of female legislators talking in the presence of a female minister would have been be talking to a member of the opposition. It is possible that the leadership effect may manifest differently in these two scenarios and I therefore estimate the model separately for debates in the period before (8690 debates) and after (5630 debates) the general election on May 6th 2010.

- *'Question Time' vs 'Substantive' debates* – A key feature of parliamentary debate in the UK (and many other Westminster systems) is Question Time, where MPs have the opportunity to question government ministers about matters for which they are responsible. These weekly "debates" are not typically focused on in-depth discussion of a specific bill, but rather focus on a range of issues that concern a particular department. As these debates have a different structure and purpose to legislative bill debates, it is possible that the leadership effects would be different here than elsewhere. I therefore separate debates into "Question Time" and "Substantive" debates and estimate the model for each set of debates. Question Time debates are clearly marked in the debate titles with the phrase "Answers to Oral Questions." There are 9975 Question Time debates in the data, and 4345 substantive debates.

- *Opposition MPs* – From a theoretical perspective we might expect that the effects documented in the main body of the paper are dependent on shared partisanship. That is, it might be the case that female cabinet ministers change the behaviour of female backbench MPs from the governing party, but not for those from the opposition. In order to account for this possibility, I estimate the model while subsetting to focus only on the effects for members of the Opposition party. I drop all MPs from the governing party from the data, recalculate the proportion of words spoken by women amongst opposition MPs, and then estimate model. The resulting analysis is based on 13549 debates.

- *Labour, Conservative and Minor Party MPs* – In the UK, the Labour party has historically had a much higher level of descriptive representation of women than other parties, while women's presence is considerably lower among Conservatives and Liberal Democrats. It may therefore be the case that female leaders are more important in parties where women are less well represented. I therefore estimate the model while subsetting to MPs from different parties. Specifically, for Labour MPs, Conservative MPs, and then for MPs from other minor parties (including the Liberal Democrats and the Scottish National Party), I first drop all other MPs from the data, I then recalculate the proportion of words spoken by women MPs from the party of interest, and then finally I estimate model. The resulting analysis is based on 13131 debates for Labour MPs, 12852 debates for Conservative MPs and 9351 debates for MPs from minor parties.

Figure S4: Stability of participation effects across subsets



NOTE: The plot presents estimates of the main treatment effect for different subsets in the data. Points represent estimates from equation 3 for each subset, and bootstrapped confidence intervals are also presented. The gray-shaded region indicates the effect size estimated from the full data, as presented in table 1. Although there is some variability in the treatment effect across subsets, the effect is always positive and in most cases comparable in size to the main effect given in the body of the paper.

The message from the figure is clear: although there is a small amount of heterogeneity in treatment effects across these different subset analyses, the main findings stand. The effects are positive in all cases, are significantly different from zero in all but two cases (debates held before 2010 and amongst MPs from minor parties), and in general the subset effects are comparable to the effect size reported in the main body of the paper. In all cases the confidence intervals on the subset coefficients overlap with the point estimate from the main analysis.

The most interesting of these subset effects is the one associated with Opposition MPs. Although readers may suspect that female leadership effects are isolated to same-party MPs, the results here suggest otherwise. When focussing only on MPs who come from a separate party from the cabinet minister, it remains the case that the appointment of a female cabinet minister increases the participation of other female MPs in parliamentary debate.

Turning to the other effects, there is some evidence that the effects are somewhat stronger for substantive debates than for question time debates, though the small sample size for these subsets makes the estimation very imprecise. Similarly, there is some evidence that the effects are stronger for members of the Conservative party and minor parties than they are for the Labour party. However, again, the small sample sizes and the resultant estimation uncertainty prevents us from drawing strong conclusions.

Participation results by party

Tables S4 and S5 present the results of the main participation models described in equations 3, 4 and 5 in the main body of the paper for the Conservative and Labour party subsets defined above. Consistent with the findings presented in figure S4, across the different model specifications it seems that the female leadership point estimate is somewhat smaller for Labour MPs than Conservative MPs.

Table S4: Effect of the appointment of a female minister on female participation (Labour MPs)

	PropWordsWomen (Labour MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.059**	0.050***	0.047	0.033**	0.030***	0.024*	0.021
	(0.028)	(0.017)	(0.030)	(0.015)	(0.010)	(0.014)	(0.013)
Constant	0.263***	0.163***	0.198*	0.137*	0.094	10.657	1.012
	(0.011)	(0.057)	(0.106)	(0.072)	(3.610)	(441.233)	(1.608)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	22	19	18	13	11	9	8
95% CI	[1,43]	[6,32]	[-4,40]	[2,24]	[4,19]	[-1,19]	[-2,17]
Observations	13,131	13,131	13,131	13,131	13,131	13,131	13,131
R^2	0.006	0.060	0.033	0.083	0.093	0.098	
Adjusted R^2	0.005	0.045	0.031	0.065	0.074	0.077	0.095

Note: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with bootstrapped robust standard errors (clustered by ministry) shown in parentheses. Models 5 and 6 include linear and quadratic ministry-specific time trends in addition to ministry fixed-effects and month fixed-effects. Model 7 is a GAM model including non-parametric, ministry-specific, flexible time trends. The "Effect Size" row indicates the percentage increase in female participation relative to the average female participation rate under male ministers. *p<0.1; **p<0.05; ***p<0.01

Table S5: Effect of the appointment of a female minister on female participation (Conservative MPs)

	PropWordsWomen (Conservative MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.052***	0.051***	0.046***	0.044***	0.045**	0.048***	0.028***
	(0.014)	(0.016)	(0.012)	(0.016)	(0.018)	(0.018)	(0.010)
Constant	0.112***	0.091	0.063*	0.029	0.112	−3.305	−0.076
	(0.009)	(0.078)	(0.038)	(0.063)	(4.693)	(194.170)	(0.190)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	47	45	41	40	40	43	25
95% CI	[22,72]	[18,73]	[20,62]	[12,67]	[8,72]	[12,74]	[8,42]
Observations	12,851	12,851	12,851	12,851	12,851	12,851	12,851
R^2	0.007	0.037	0.040	0.066	0.071	0.089	
Adjusted R^2	0.007	0.021	0.038	0.048	0.051	0.067	0.094

Note: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with bootstrapped robust standard errors (clustered by ministry) shown in parentheses. Models 5 and 6 include linear and quadratic ministry-specific time trends in addition to ministry fixed-effects and month fixed-effects. Model 7 is a GAM model including non-parametric, ministry-specific, flexible time trends. The "Effect Size" row indicates the percentage increase in female participation relative to the average female participation rate under male ministers. *p<0.1; **p<0.05; ***p<0.01

S5. Spillover effects

One concern readers might have is that the effects presented in table 1 do not account for potential spillovers that might occur over time within ministries. In particular, it is possible that, through a process of habit formation, women who participate at higher rates under female ministers may "get used" to this participation, and continue to participate at high rates even after a female minister leaves office. If this were the case, we should expect to see higher than average levels of female MP debate participation in debates pertaining to ministries where there has been a female minister, but in the periods immediately after a female minister leaves office.

As discussed in footnote 8 in the main body of the paper, spillovers of this type would create higher levels of female participation in some periods under male ministers, and would therefore cause the estimates presented in table 1 to be downwardly biased. Nevertheless, in this section I address this point by reestimating the main models from the participation analysis, but here I include additional indicators for debates held in ministries in the periods *after* a female minister has held office. I include 6 dummy variables, which indicate if a debate was held in the first month after a female minister left office, the second month after a female minister left office, and so on.

These variables therefore capture whether the months after a female minister leaves office are marked by higher than average levels of female participation in debate. Importantly, I also include the female minister variable in the analysis. This means that the baseline to which these dummy variables refer is debates held in periods where there is a male minister, and when the debate occurs either at least more than *6 months after* a female minister has left office, or *before* a female minister takes office. If the spillover argument made above is correct, we should expect some or all of the coefficients associated with these dummy variables to be positive.

The results of this analysis, presented in table S6, suggest very limited evidence for these types of spillover effects. First, the estimated coefficients on the various dummy variables are generally indistinguishable from zero, and vary considerably in both sign and magnitude from month to month. There is therefore no evidence that the debates held in ministries that were previously occupied by female ministers are systematically higher after the female minister leaves office than they are at other points in time. Further, the estimated effect sizes associated with the 'Female minister' variable (the main treatment variable in the paper) are essentially the same here as they are in the main analysis presented in table 1.

One implication of this analysis is that the motivating effects of female leadership do seem to be largely confined to the periods in which the female minister holds office. There is little evidence – in the first six months after a female cabinet minister steps down – that the increased participation of female MPs persists in future debates led by male ministers.

Table S6: Temporal spillover effects

	PropWordsWomen						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.061***	0.055***	0.050**	0.041***	0.038***	0.039***	0.044***
	(0.019)	(0.014)	(0.021)	(0.012)	(0.012)	(0.011)	(0.008)
0-1 month after	0.018	0.007	0.022	0.006	−0.006	0.003	0.016
	(0.019)	(0.023)	(0.019)	(0.021)	(0.022)	(0.020)	(0.021)
1-2 months after	0.015	0.036	0.020	0.028	0.024	0.024	0.040
	(0.051)	(0.061)	(0.048)	(0.053)	(0.055)	(0.052)	(0.039)
2-3 months after	0.099**	0.094*	0.084**	0.069	0.067	0.065	0.085**
	(0.042)	(0.054)	(0.037)	(0.050)	(0.055)	(0.054)	(0.035)
3-4 months after	0.077**	0.049	0.065*	0.033	0.024	0.028	0.042
	(0.036)	(0.035)	(0.034)	(0.036)	(0.036)	(0.032)	(0.027)
4-5 months after	0.059	0.020	0.037	0.003	−0.006	−0.002	0.016
	(0.036)	(0.034)	(0.035)	(0.028)	(0.030)	(0.029)	(0.027)
5-6 months after	0.175***	0.094*	0.100	0.039	−2.214	1.173	0.056
	(0.011)	(0.054)	(0.082)	(0.051)	(310.789)	(166.629)	(0.166)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	34	31	28	23	22	22	25
95% CI	[13,56]	[16,46]	[5,51]	[11,36]	[8,35]	[10,34]	[16,34]
Observations	14,320	14,320	14,320	14,320	14,320	14,320	14,320
R^2	0.013	0.076	0.054	0.106	0.114	0.123	
Adjusted R^2	0.013	0.062	0.051	0.091	0.096	0.103	0.116

NOTE: Regression coefficients are shown with bootstrapped standard errors (clustered by ministry) shown in parentheses. The "Effect Size" row indicates the percentage change in female participation relative to the average under male ministers in debates that occur either at least more than 6 months after a female minister has left office, or before a female minister takes office. *p<0.1; **p<0.05; ***p<0.01

S6. Female shadow ministers and female Speakers of the House

Shadow ministers

To investigate the hypothesis that opposition parties may respond strategically to the appointment of a female government minister by appointing a female shadow minister to the opposition cabinet, I analyse the relationship between the sex of a *newly appointed shadow minister* and the sex of the *current government minister.* I estimate this relationship using logit models of the following form:

$$\text{logit}(\mathbb{E}[ShadowFemaleMinister_{mt}]) = \alpha + \beta_1 * FemaleMinister_{mt} + \lambda_m + \epsilon_{mt}$$

where $ShadowFemaleMinister_{mt}$ is equal to one when the shadow minister appointed to ministry m at time t is a woman, and zero otherwise. $FemaleGovernmentMinister$ is equal to one when the government minister responsible for a given ministry m in time t is a woman, and zero otherwise. λ_m is a ministry fixed effect. If opposition parties are responding strategically to the sex of the government minister in a given ministry, then the β_1 coefficient will be positive, indicating that the probability of appointing a female shadow minister is associated with the sex of the current cabinet minister for that ministry. The results of these regressions are given in table S7.

While the coefficient on the 'female government minister' variable are positive in both models, these effects are imprecisely estimated, and statistically indistinguishable from zero. This suggests that it is unlikely that the effects documented in the main analysis are driven by the strategic appointment of female shadow ministers by opposition parties.

Table S7

	Female Shadow Minister	
	(1)	(2)
Female cabinet minister	0.128	0.093
	(0.073)	(0.088)
Constant	0.257	0.000
	(0.037)	(0.316)
Ministry FEs	No	Yes
Observations	204	204
R^2	0.015	0.147
Note:	Standard errors in parentheses	

Table S8 replicates the main participation analysis, but here excludes any speeches made by shadow cabinet ministers from the calculation of equation 1. As the results show, there is essentially no difference in the results, suggesting that the effects in the main analysis cannot be attributed to the strategic appoint of female shadow ministers by opposition parties.

Table S8: Effect of appointing a female minister on female debate participation – shadow ministers excluded

	PropWordsWomen (Opposition ministers excluded)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.042***	0.036***	0.041***	0.033***	0.029***	0.028***	0.024***
	(0.016)	(0.012)	(0.015)	(0.009)	(0.010)	(0.008)	(0.007)
Constant	0.174***	0.099**	0.113	0.050	0.010	0.693	0.253
	(0.009)	(0.048)	(0.078)	(0.050)	(3.151)	(141.578)	(0.433)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	24	20	23	19	17	16	14
95% CI	[6,42]	[7,34]	[6,41]	[9,28]	[5,28]	[6,25]	[6,21]
Observations	14,307	14,307	14,307	14,307	14,307	14,307	14,307
R^2	0.006	0.057	0.039	0.081	0.086	0.092	
Adjusted R^2	0.006	0.043	0.037	0.066	0.068	0.072	0.077

NOTE: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017 with shadow ministers excluded from the data. Regression coefficients are shown with bootstrapped cluster-robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table S9 presents the results of the main participation models described in equations 3, 4 and 5 in the main body of the paper, though here the treatment variable is an indicator for whether the *shadow* cabinet minister in debate is a woman. To avoid issues of confounding related to the gender of the relevant cabinet minister, I focus on the 3422 debates which include an opposition shadow minister but where no cabinet minister is present. The results show that, consistent with the idea of female leadership effects, the presence of a female shadow cabinet minister is also associated with higher levels of participation in debate by other female MPs, though the effect is smaller on average and less precisely estimated than the effects for female cabinet ministers presented in the main body of the paper. Based on the estimates in model 6, women contribute approximately 18% more under female shadow cabinet ministers than under male shadow cabinet ministers. This compares to an effect size of 21% from the equivalent cabinet minister model in table 1.

Table S9: Effect of appointing a female shadow cabinet minister on female debate participation

	PropWordsWomen						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female shadow minister	0.046***	0.037*	0.048***	0.042**	0.040*	0.032	0.039***
	(0.017)	(0.021)	(0.017)	(0.020)	(0.021)	(0.022)	(0.015)
Constant	0.175***	0.390***	0.168***	0.417***	0.338*	−1.695	−0.339
	(0.015)	(0.135)	(0.013)	(0.158)	(0.178)	(3.361)	(0.765)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	26	21	27	24	23	18	22
95% CI	[8,45]	[-2,45]	[9,46]	[2,46]	[-1,46]	[-6,43]	[5,39]
Observations	3,422	3,422	3,422	3,422	3,422	3,422	3,422
R^2	0.006	0.076	0.058	0.126	0.139	0.168	
Adjusted R^2	0.006	0.017	0.051	0.063	0.071	0.094	0.163

NOTE: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017 with shadow ministers excluded from the data. Regression coefficients are shown with bootstrapped cluster-robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

Speakers of the House

Each Commons' debate is presided over either by the Speaker of the House or one of his/her Deputy Speakers. In the time period I study, there is variation in the gender of the MPs holding these positions, and – given that these Speakers are entrusted with selecting the order of MPs' debate contributions – it is possible that the *selection* of MPs' speeches follows some gendered dynamic.

In table S10, I address this concern by presenting the results of the main participation models described in equations 3, 4 and 5 in the main body of the paper, though here I also control for the gender of the Speaker of the House or the Deputy Speaker of the House presiding over each debate.[S2]

The results in table S10 suggest that controlling for the gender of the Speaker/Deputy speaker in charge of debate does little to affect the main estimates of interest. Model 6 in table S10, for example, suggests that the effect of a female minister on participation when controlling for Speaker gender is to increase the proportion of words spoken by other female MPs by 21% – exactly the same estimate as in the main analysis (table 1). Similarly, table S10 suggests that the gender of the Speaker does not any significant effect on female speech participation: across model specifications, the coefficient on the Speaker-gender variable is small in magnitude, varying in sign, and indistinguishable from zero.

[S2]For those debates where the Speaker/Deputy Speaker is directly mentioned in Hansard, it is straightforward to record the gender of that Speaker/Deputy Speaker for that debate, however Hansard does not always note who is 'in the Chair' for a given debate. To expand the number of debates for which I have information on the gender of the Speaker/Deputy Speaker, I search the speeches in each debate for references to either "Madam Speaker" and "Madam Deputy Speaker" or to "Mr Speaker" and "Mr Deputy Speaker", and assign debates as being presided over by a female Speaker/Deputy if there are more occurrences of the former strings than of the latter strings. For the remaining debates where I am still missing data on the Speaker's gender, I assign the gender of the Speaker/Deputy to be female for the period prior to October 2000 (when Betty Boothroyd's speakership ended) and male for the period after that date (the two Speakers in this latter – Michael Martin and John Bercow – were men).

Table S10: Effect of appointing a female cabinet minister on female debate participation (controlling for Speaker gender)

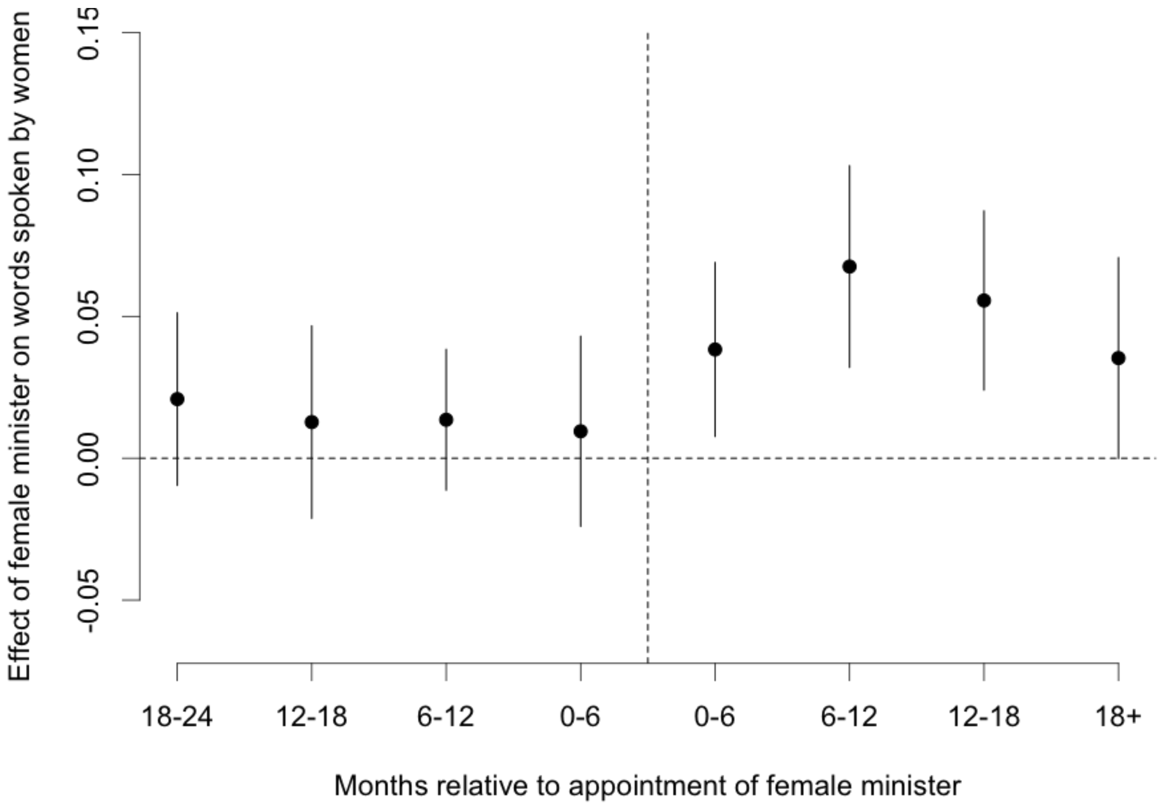
	PropWordsWomen						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.061***	0.055***	0.049**	0.041***	0.037***	0.037***	0.039***
	(0.018)	(0.013)	(0.020)	(0.012)	(0.011)	(0.010)	(0.008)
Female Speaker	−0.043***	0.0001	−0.037***	−0.001	−0.004	−0.005	−0.005
	(0.009)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)	(0.008)
Constant	0.184***	0.094*	0.125*	0.040	0.050	1.264	0.061
	(0.011)	(0.049)	(0.072)	(0.056)	(2.948)	(167.784)	(0.143)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	34	31	28	23	21	21	22
95% CI	[14,55]	[17,46]	[6,50]	[10,36]	[8,34]	[10,32]	[14,31]
Observations	14,320	14,320	14,320	14,320	14,320	14,320	14,320
R^2	0.017	0.075	0.056	0.106	0.113	0.122	
Adjusted R^2	0.017	0.061	0.054	0.091	0.096	0.103	0.115

NOTE: Models 1-6 represent OLS fixed-effect regressions for the period 1997-2017 with shadow ministers excluded from the data. Regression coefficients are shown with bootstrapped cluster-robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

S7. Dynamic panel model estimates

Figure S5 plots the estimates and 95% bootstrapped confidence intervals (clustered by ministry) from a dynamic panel model. Here I code a binary indicator for the first 6 months of the treatment period in a given ministry, and then add four leads and three lags of this indicator in addition to the full set of fixed-effects and linear and quadratic time-trends. The final lagged variable captures all treated periods from eighteen months until the end of the treatment period for a given ministry. The coefficients from this model therefore represent the estimated difference in the outcome between treated and untreated ministries in the periods before and after the treatment occurs.

Figure S5: Dynamic panel model estimates



Note: The plot presents estimates of switching from a male to a female minister before and after the actual change occurred. The vertical dashed line indicates the timing of the change, and the points indicate (at six month intervals) the difference between treated and untreated ministries at the given time point. Estimates are generated from a dynamic panel regression including ministry and time fixed-effects, ministry-specific linear and quadratic time trends, and indicator variables for four leads and three lags of the change in minister gender. 95% confidence intervals are constructed by bootstrapping the regression model, blocking on ministry.

S8. Details of the influence measurement procedure

I proceed in two steps: first, I construct similarity graphs for all speeches in each debate; second, I analyse the graphs using an iterative ranking algorithm to calculate a vector of centrality scores, P, which correspond to the influence of each speech in each debate.

Construction of a debate-specific similarity graph, S_d, begins with the selection of a metric which measures how linguistically similar two speeches are to one another. I represent each speech as an N-dimensional term-frequency-inverse-document-frequency (tf-idf) vector, where N is the number of unique words in the corpus. Each element in the vector is a count of the number of times a given word, w, appears in a given speech, s, multiplied by the logged *inverse document frequency* of that word, to create a weighted term-frequency score, v_{ws}, for each word in each speech. Where N is the total number of unique words in the corpus, n_w is the number times that word w appears in the corpus, and tf_{ws} is the number times that word w appears in speech s, the score for w in s is given by:

$$v_{ws} = tf_{ws} * log(\frac{N}{n_w}) \tag{S4}$$

A high value of v_{ws} occurs when a word is used frequently in a given speech, but infrequently in the corpus as a whole. The weights thus filter out very common words such as 'stopwords', and ensure that the vector representation of the speeches mostly reflects topically-salient features of the political debate.

Having calculated the tf-idf vectors for each speech in the corpus, I construct D similarity matrices (one for each debate), the typical element of which is:

$$S_d(i,j) = sim(v_i, v_j) = \frac{v_i \cdot v_j}{||v_i||||v_j||} = \frac{\sum_{w=1}^{W} v_{wi} \cdot v_{wj}}{\sqrt{\sum_{w=1}^{W} v_{wi}^2} \cdot \sqrt{\sum_{w=1}^{W} v_{wj}^2}} \tag{S5}$$

i.e. the cosine-similarity of the weighted word-count vectors of speeches i and j in debate d. Each graph (again, one for each debate) therefore consists of nodes that represent speeches in a debate, and edges which are placed between speeches for which $sim(v_i, v_j)$ is greater than some threshold value, S_{min}.[S3] The edges are then weighted by the similarity scores.

The cosine-similarity relation is symmetric (i.e. because $S_d(i,j) = S_d(j,i)$) and thus it is possible to construct either undirected (where edges between nodes run in both directions and receive the same weight) or directed (where edges between nodes run in only one direction) networks (Erkan and Radev, 2004). As I conceptualise influence as the degree to which language used in one speech is adopted in subsequent speeches, it is necessary to take the temporal ordering of debate into account when constructing the graphical network. Put simply, it does not make sense for speeches that occur later in the debate to 'influence'

[S3]Throughout the analysis I set S_{min} to 0.25, in line with Fader et al. (2007).

speeches that occur earlier in the debate. I therefore focus on only the upper triangle of the similarity matrices, S_d, while setting all elements in the lower triangle to zero. The consequence of this is that 'references' from one speech to another can only flow in one direction: later speeches can reference earlier ones, but not vice versa. Using a directed graph makes no difference to the computation of the influence scores (Mihalcea, 2004).

As described above, the influence of a speech is determined by the number of references it receives from other speeches within a debate (i.e. by the number of speeches which are linguistically similar to it), and by the influence of the referencing speeches. In the simplified case where all edges receive a weight of 1, an intuitive way of formulating this idea is to imagine that each speech has an influence value, and that this value to gets distributed to the speeches that it references:

$$p(i) = \sum_{j \in adj(i)} \frac{p(j)}{deg(j)} \tag{S6}$$

Where $p(i)$ is the influence of speech i, $adj(i)$ is the set of speeches that have edges with i, and $deg(j)$ is the degree of node j (the degree of a node is simply the number of edges that connects the node to other nodes). This formulation emphasises that a speech is more influential when it is referenced by many other speeches ($adj(i)$), when the influence of the referencing speeches ($p(j)$) increases, and when the referencing speeches reference relatively few other speeches ($deg(j)$). Weighting the edges of the network by $S_d(i,j)$ allows references to vary in strength (according to the similarity between speeches i and j) and we can reformulate equation S6 to include the weights in S_d via:

$$p(i) = \sum_{j \in adj(i)} \frac{S_d(i,j)}{\sum_{k \in adj(j)} S_d(k,j)} p(j) \tag{S7}$$

Equation S7 makes clear that the reference that speech i receives from speech j is determined by the linguistic similarity between i and j (the numerator), and the similarity between j and all of the speeches that j references (the denominator). Fader et al. (2007) and Erkan and Radev (2004) show that computation of the vector of speech-level influence scores, P, is achieved by calculating the left eigenvector of the row-normalised similarity matrix S_d via the *PageRank* algorithm, which was originally designed for computing webpage prestige in the Google search engine (Page et al., 1999).[S4] Finally, with these speech-influence scores in hand, as some MPs will speak multiple times in a debate, the influence score of an MP in a given debate is simply the sum of the influence scores p for the speeches given by that MP in that debate.

[S4]Mihalcea (2004) shows that either the Kleinberg (1999) HITS algorithm and PageRank can be used to calculate P and that both perform well in approximating human judgements. I present statistically and substantively similar results from the HITS algorithm in appendix section S9. I implement both algorithms using the iGraph package in R (Csardi and Nepusz, 2006)

S9. HITS algorithm influence results

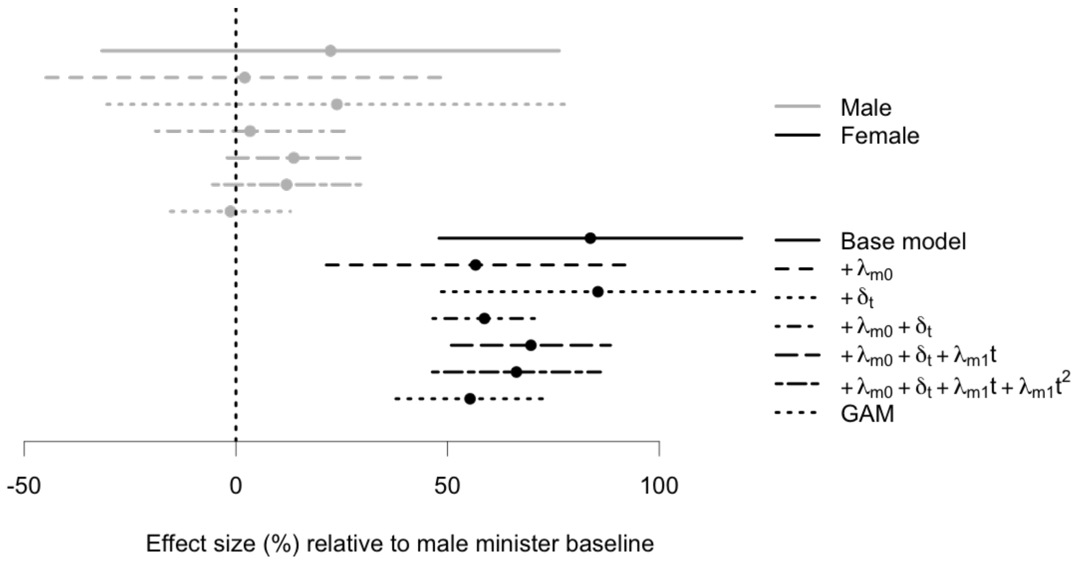
This section replicates the analysis in table 2 and figure 3 estimating P in equation S7 using Kleinberg's (1999) HITS algorithm. As table S11 and figure S6 show, the results are substantively and statistically very similar regardless of the estimation approach.

Table S11: Effect of appointing a female minister on MPs' debate influence – HITS

	Influence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.191***	0.211***	0.374***	0.290***	0.324***	−3.139*	3.662
	(0.030)	(0.011)	(0.041)	(0.024)	(0.034)	(1.815)	(6.335)
Female minister	0.024	0.002	0.025	0.004	0.014*	0.013	−0.001
	(0.029)	(0.025)	(0.029)	(0.012)	(0.008)	(0.009)	(0.008)
Female MP	−0.012	−0.010	0.002	−0.001	−0.001	−0.0001	−0.007*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.004)
Interaction	0.070***	0.061***	0.070***	0.062***	0.063***	0.061***	0.063***
	(0.018)	(0.012)	(0.018)	(0.013)	(0.013)	(0.013)	(0.008)
Party FEs	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	173,509	173,509	173,509	173,509	173,509	173,509	173,509
R^2	0.002	0.025	0.014	0.034	0.036	0.037	
Adjusted R^2	0.002	0.025	0.013	0.033	0.034	0.035	0.039

Note: Models 1-6 present OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

Figure S6: Marginal effect of female minister on male and female influence – HITS



NOTE: The plot shows the marginal effect of the appointment of a female cabinet minister on the debate influence of male (gray lines) and female (black lines) MPs, relative to the average level of influence when the minister is male.

S10. Split-sample results

In the main body of the paper I present results that evaluate the differential effects of female leadership on men and women's influence (table 2) and the responsiveness to MP's speeches (table 3) using models that include an interaction term between the gender of the cabinet minister and the gender of the MP. In this section, I provide an alternative illustration of these findings via a split-sample analysis, where I separately estimate the effects of female cabinet ministers on the influence of male and female MPs (tables S12 and S13) and on the responses received by male and female MPs (tables S14 and S15).

Consistent with the interaction models, the results in these models show that the appointment of a female minister has a large, positive, and significant effect of the influence of and responsiveness to female MPs (tables S13 and S15), but small and generally insignificant effects on the influence of and responsiveness to male MPs (tables S12 and S14).

Table S12: Effect of appointing a female minister on male MPs' debate influence

	Influence (male MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.003	0.001	0.003	0.001	0.001	0.001	-0.005^{***}
	(0.005)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.001)
Party FEs	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	140,604	140,604	140,604	140,604	140,604	140,604	140,604
R^2	0.001	0.045	0.015	0.056	0.063	0.067	
Adjusted R^2	0.001	0.044	0.014	0.055	0.061	0.065	0.077

Note: Models 1-6 present OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table S13: Effect of appointing a female minister on female MPs' debate influence

	Influence (female MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.014***	0.008***	0.016***	0.010***	0.008*	0.007**	0.005**
	(0.005)	(0.003)	(0.005)	(0.003)	(0.004)	(0.004)	(0.002)
Party FEs	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	32,905	32,905	32,905	32,905	32,905	32,905	32,905
R^2	0.006	0.053	0.025	0.067	0.075	0.079	
Adjusted R^2	0.006	0.052	0.018	0.060	0.067	0.070	0.085

NOTE: Models 1-6 present OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table S14: Effect of appointing a female minister on on the responsiveness to male MPs

	Responsiveness (male MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Minister	−0.651	−1.276	−0.778	−1.295***	−0.797	−0.686	−1.366**
	(1.034)	(1.725)	(0.760)	(0.433)	(0.507)	(0.532)	(0.533)
Constant	27.066***	27.105***	33.333***	29.309***	29.614***	297.810**	328.162
	(0.855)	(0.129)	(6.097)	(5.308)	(5.107)	(119.837)	(543.440)
Same Party Control	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	130,887	130,887	130,887	130,887	130,887	130,887	130,887
R^2	0.004	0.010	0.016	0.020	0.022	0.023	
Adjusted R^2	0.004	0.010	0.014	0.018	0.019	0.020	0.022

NOTE: Models 1-6 present OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with robust standard errors (clustered by ministry) shown in parentheses.
*p<0.1; **p<0.05; ***p<0.01

Table S15: Effect of appointing a female minister on the responsiveness to female MPs

	Responsiveness (female MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Minister	5.731***	3.580***	5.598***	3.629**	3.366***	4.709***	4.026***
	(1.206)	(0.871)	(1.761)	(1.616)	(1.057)	(1.202)	(1.158)
Constant	26.207***	22.428***	38.575***	30.386**	40.320***	−70.378	−6.208
	(2.504)	(1.650)	(12.888)	(13.645)	(12.909)	(153.288)	(31.267)
Same Party Control	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	28,579	28,579	28,579	28,579	28,579	28,579	28,579
R^2	0.005	0.017	0.021	0.030	0.035	0.037	
Adjusted R^2	0.005	0.016	0.014	0.022	0.025	0.026	0.036

NOTE: Models 1-6 present OLS fixed-effect regressions for the period 1997-2017. Regression coefficients are shown with robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

S11. Validating influence scores

I test two relatively unambiguous intuitions about which actors in the House of Commons we expect to be influential in parliamentary debate.[S5] First, government ministers should be on average *more* influential than other MPs when participating in plenary debate. Ministers play a crucial role in setting the agenda for parliamentary business, and their speeches are frequently used to outline policy that we would expect others to comment on extensively. Second, the Speaker of the House should be on average *less* influential than other members. The majority of the Speaker's contributions are procedural, having little to do with the substantive matters under discussion, and should not be referenced frequently by other members. I test these expectations by regressing the influence score on binary indicators for whether an MP is either the cabinet minister responsible for the current debate, or the Speaker. The results, presented in table S16 in the appendix, strongly support the expectations: cabinet ministers are on average 5 times more influential than backbench MPs, while the Speaker is less than half as influential as backbenchers.

Table S16: Ministers are more influential, and Speakers of the House are less influential

	Influence		
	(1)	(2)	(3)
Minister	0.956***		0.949***
	(0.006)		(0.006)
Speaker		−0.227***	−0.154***
		(0.008)	(0.007)
Constant	0.220***	0.299***	0.227***
	(0.002)	(0.002)	(0.002)
Observations	195,833	195,833	195,833
R^2	0.127	0.004	0.129

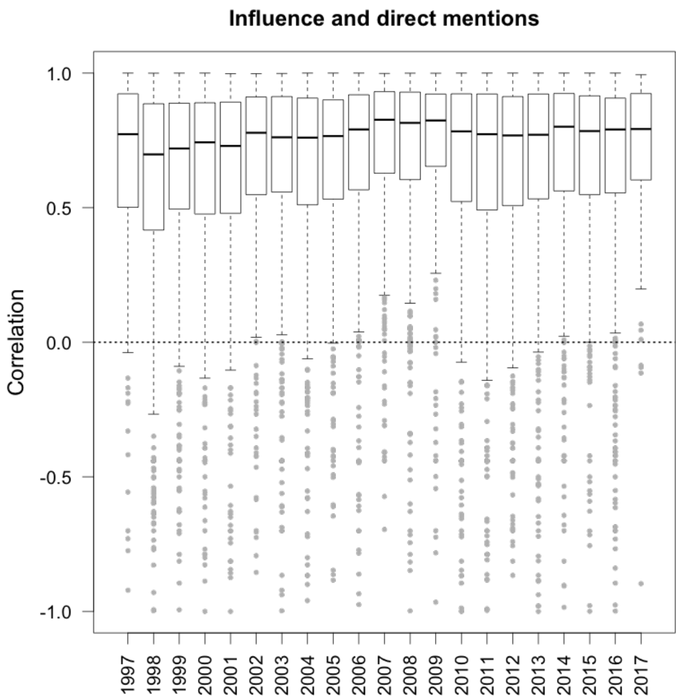
NOTE: OLS regressions where the outcome variable is *influence* as defined in equation S7, and the independent variables are "Minister" – an indicator that is equal to one when a speech is given by a government minister – and "Speaker" – an indicator that is equal to one when a speech is given by the Speaker of the House. The baseline corresponds to the average level of *influence* for speeches delivered by backbench MPs. *p<0.1; **p<0.05; ***p<0.01

In addition, we can also compare the influence score of an MP to how many direct references that MP receives during a debate. MPs follow strict conventions when directly referring to other members as they must not use the names of their colleagues, but instead refer to the "Member for Holborn and St Pancras" or the "Honourable Member for Taunton

[S5] Fader et al. (2007) have shown previously that these influence scores calculated for speeches made in the US Senate correlate strongly with membership and seniority in Senate legislative committees.

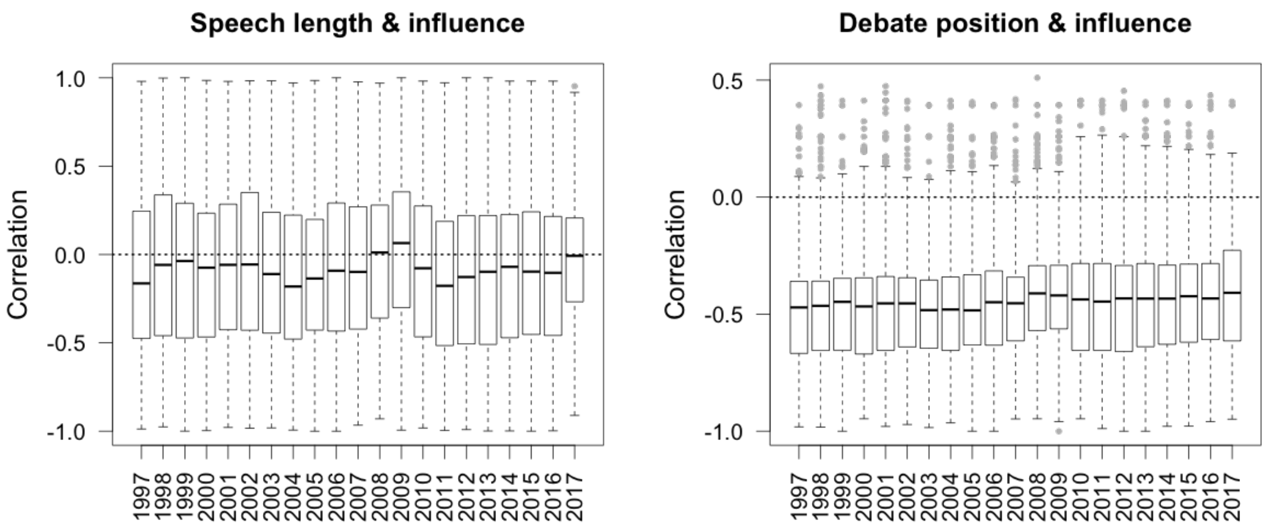
Deane" and so on. Constituency names are unique to each MP, and by searching for constituencies in the speech-texts it is possible to construct a count for the number of times any particular MP is directly mentioned by any other MP during the course of a debate. It seems clear that an MP who is directly mentioned by many other MPs in their speeches is playing an important role in the debate at hand and if the influence score defined in equation S7 is valid, it should correlate positively with the number of direct mentions that an MP receives in debate. In appendix figure S7 I show that this is the case: the average correlation across all debates in the sample is 0.66. Together, these comparisons provide reassuring evidence regarding the face validity of the measure of influence described above.

Figure S7: Correlation of speech 'influence' and direct mentions



Note: The plot shows the average correlation between the influence of an MP and the number of times that that MP was directly mentioned by other members measured at the debate level across all years in the sample. In line with expectations, influence is strongly positively correlated with direct mentions.

Figure S8: Correlation of speech 'influence' with speech length and debate position



NOTE: The left panel shows the average correlation between the length of a speech and influence measured at the debate level across all years in the sample. The right panel shows the equivalent correlation between speech position and influence.

S12. Validating responsiveness scores

First, within a debate, MPs might use similar words even when they are not responding to one another. Debates are normally focussed on a small number of topics, the discussion of which will lead MPs to use similar language regardless of whether they are talking directly to one another. However, if the measure defined in equation 8 captures responsiveness, and not merely topicality, then speeches that are adjacent to one another should demonstrate higher responsiveness scores than speeches that are not adjacent. Table S17 in the appendix tests this hypothesis. From each debate in the corpus, I randomly sample two speech pairs. One of the pairs is adjacent, and one is non-adjacent. I then regress the responsiveness score on a binary indicator which is equal to one for an adjacent pair of speeches, and zero otherwise. The coefficient on this indicator is statistically significant, and implies that adjacent speeches are approximately 30% more responsive than non-adjacent speeches. This provides strong evidence that equation 8 is capturing something distinct from topicality: comparing pairs of speeches *within the same debate*, those speeches that follow directly after each other are more responsive than speeches that are non-adjacent.

Table S17: Adjacent speeches are more responsive than non-adjacent speeches

	res
Adjacent	6.629***
	(1.277)
Constant	29.707***
	(0.903)
Observations	21,602
R^2	0.001

Note: OLS regression for adjacent and non-adjacent speeches (within a debate). Regression coefficients are shown with standard errors in parentheses. The outcome variable is *res* as defined in equation 8, and the independent variable is an indicator that is equal to one when a given pair of speeches occupy adjacent positions in the debate. The baseline corresponds to speeches that are non-adjacent. *p<0.1; **p<0.05; ***p<0.01

Second, in a subset of debates, government ministers go before the House to field questions from backbenchers, and are required to provide answers to these questions. In these 'Question Time' debates, questions by backbenchers need not address the same topic as the question just answered by the minister. For example, a first backbencher might ask the minister about schools, to which the minister will provide an answer, and then a second backbencher might ask about child care provision, to which the minister must also respond. In these debates, we should therefore expect that when a minister's speech follows a backbencher, that speech should be more responsive than when a backbencher's speech follows that of a minister. For each speech in each 'Question Time' debate, I code whether the speech is made by a minister responding to a backbencher, or a backbencher asking a new question.[S6] I then regress the responsiveness score on a binary indicator which is equal to one when the speech is made by a minister in response to a backbencher. The results in table S18 show that ministerial replies are more than twice as responsive than are questions posed by backbenchers to the minister. This indicates that the measure is accurately recovering intuitive properties of the concept of responsiveness.

Table S18: Minister and backbencher responsiveness

	res
Minister responding to backbencher	10.954***
	(0.180)
Constant	10.675***
	(0.131)
Observations	153,342
R^2	0.024

NOTE: OLS regressions for 9927 "question time" debates. Regression coefficients are shown with standard errors in parentheses. The outcome variable is *res* as defined in equation 8. The independent variable is an indicator that is equal to one when a speech is spoken by a minister, and comes immediately after a speech by a backbencher. The baseline corresponds to backbench speeches that follow directly after a speech by a minister. *p<0.1; **p<0.05; ***p<0.01

[S6] I exclude all instances where a backbencher follows from another backbencher.

S13. Debate level influence and responsiveness results

Tables 2 and 3 in the body of the paper present the results of the influence and responsiveness models with individual MP and individual speech data, respectively. Readers may be concerned that the identification strategy outlined in the earlier sections of the paper is designed for aggregate-level (debate) data, rather than individual-level data. However, (Angrist and Pischke, 2009, 235-237) show that the difference-in-difference model with individual level data and a group level treatment is equivalent to a group-level model with the regression weighted by cell size. To show this equivalence, I reproduce the results of the models described in equations 7 and 9, aggregating the data by debate and MP gender, weighting the regression models by the number of individuals (for the influence model) or the number of speeches (for the responsiveness model), and present the results in tables S19 and S20 below. Note that because I am averaging over MPs, it is not possible to include the individual partisanship controls that are included in the individual level models. Regardless of these modelling choices, the results are statistically and substantively very similar.

Table S19: Effect of appointing a female minister on MPs' debate influence (debate level)

	Influence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.047***	0.047***	0.055***	0.032	0.050	−0.717	−0.567
	(0.004)	(0.011)	(0.019)	(0.022)	(0.422)	(1.104)	(1.077)
Female minister	0.003	0.001	0.003	0.001	0.001	0.0004	−0.002
	(0.006)	(0.003)	(0.007)	(0.003)	(0.004)	(0.002)	(0.002)
Female MP	−0.001	−0.0004	0.001	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Interaction	0.011***	0.009***	0.012***	0.009***	0.009***	0.009***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Party FEs	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	23,494	23,494	23,494	23,494	23,494	23,494	23,494
R^2	0.002	0.061	0.021	0.077	0.086	0.091	
Adjusted R^2	0.001	0.060	0.012	0.067	0.075	0.079	0.065

NOTE: Models 1-6 present OLS fixed-effect regressions for the period 1997-2017, model 7 presents results from the GAM. Regression coefficients are shown with bootstrapped robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table S20: Effect of appointing a female minister on the responsiveness to MPs' speeches (debate-level)

	Responsiveness						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	−0.629	−0.368	0.387	0.370	0.422	0.475	0.526
	(1.636)	(1.630)	(1.588)	(1.677)	(1.565)	(1.626)	(0.361)
Female Minister	−0.772	−1.769	−0.873	−1.777***	−1.322**	−0.912	−1.557***
	(1.129)	(1.661)	(0.800)	(0.547)	(0.640)	(0.664)	(0.586)
Interaction	6.517***	6.445***	6.572***	6.452***	6.492***	6.458***	6.245***
	(1.875)	(1.871)	(1.896)	(1.918)	(1.833)	(1.805)	(0.808)
Constant	25.171***	25.055***	32.456***	28.292***	29.904	214.074	255.959
	(0.744)	(5.533)	(6.550)	(6.482)	(107.716)	(13,046.050)	(450.708)
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	22,757	22,757	22,757	22,757	22,757	22,757	22,757
R^2	0.003	0.030	0.055	0.072	0.081	0.086	
Adjusted R^2	0.003	0.029	0.046	0.062	0.070	0.074	0.085

NOTE: Models 1-6 present OLS regressions for ministerial responses, and model 7 presents the results of the GAM. Regression coefficients are shown with cluster-robust standard errors in parentheses (clustered on ministry). *p<0.1; **p<0.05; ***p<0.01

S14. Differential agenda-setting of male and female ministers

The main idea here is to measure the topical content of the issues under discussion in debate, and to evaluate whether topics which are associated with high levels of female participation (when the minister is male) increase when a female minister is appointed. In order to measure the topical content of the legislation under debate, I focus on the speeches made by *ministers* during each debate, rather than the speeches made by all members. In many cases, debates begin with a long opening statement by the minister, in which they put forward the purpose and detail of the legislation to be considered by the House. As the content of speeches made by other members may itself be a result of the appointment of a female minister, these speeches provide a useful resource for estimating the agenda proposed by the ministers.

I proceed in four steps. First, I estimate a series of topic models to produce debate-level topic proportions for all debates in the sample. These proportions indicate the topical content of each debate, and give a basis on which to find thematically similar debates under both male and female ministers. Second, I use the topic proportions for debates which are held under male ministers as explanatory variables in linear regressions, where the dependent variable is the ratio of words spoken by women as defined in equation S1. The coefficients from these regressions indicate the degree to which each latent topic is traditionally associated with female participation in debate. Third, in a second set of linear regressions, I estimate the relationship between the prevalence of a topic and the sex of a minister. The coefficients from these regressions indicate whether a topic increases or decreases when the minister is female. Finally, I compare the two sets of regression coefficients. If the agenda-setting hypothesis is correct, there should be a positive correlation between these two sets of coefficients: topics that are traditionally associated with female participation will increase when the minister is female. Such a finding would suggest that female ministers are indeed focussing on topics that are more conducive to female participation in legislative debate.

I start by applying a series of unsupervised topic models to all speeches made by ministers in the entire sample. I use the Correlated Topic Model (Blei and Lafferty, 2006), which, as with all topic models, assumes that the frequency with which terms co-occur within different documents (here, debates) gives information about the topics that feature in those documents. The key quantity of interest recovered from the CTM is θ, which is a T x D matrix of topic proportions that describe the fraction of each ministerial statement $d \in \{1, 2, ..., D\}$ that is from each topic $t \in \{1, 2, ..., T\}$. Analysts must choose how many topics to estimate from the data, and because the 'correct' number of topics is unclear, *a priori*, I estimate K topic models for a range of topic counts from 30 to 70, at 5 topic increments. This results in 9 separate θ_k matrices, with typical elements θ_{ktd}: the proportion of ministerial-statement d in topic t from topic-model k.

I then use each θ_k matrix as the model matrix[S7] for a linear regression predicting Y_d, the female speech ratio in debate d. As the goal of this first-stage model is to establish a

[S7] The topic proportions for each statement (θ_{kd}) sum to one, and so I could exclude one of the topics or the intercept term. I choose to exclude the intercept term.

baseline level of female participation associated with each topic, I estimate this model only for those debates where the presiding minister is male. I repeat this exercise K times, once for each topic model.

$$Y_d = b_{k1}\theta_{k1d} + b_{k2}\theta_{k2d} + ... + b_{kT}\theta_{kTd} + \epsilon_d \tag{S8}$$

The estimated b coefficients represent the degree to which each topic (collection of words) is associated with female participation in debates, holding other topics constant. An example of the substantive information that these coefficients contain is clear from table S21 , which contains each topic from the 30 topic model, ordered by their respective b coefficients. Reassuringly, the topics with the largest b coefficients deal primarily with topics that match intuitive notions of female interests, including children, parents, and women's issues. Additionally, women appear relatively more likely to contribute to debates that focus on the NHS, teachers and schools, and energy issues.

Next, I estimate a series of regressions to establish which topics are more prevalent under female ministers. As we are concerned here with establishing the differences in agenda-setting *within* government ministries, I estimate models of the following form:

$$\theta_{ktd(m)} = \alpha + \gamma_{kt} * FemaleMinister_d + \lambda_m + \epsilon_{ktd} \tag{S9}$$

Where $\theta_{ktd(m)}$ is the proportion of debate-text d (in ministry m) devoted to topic t from topic model k. $FemaleMinister_d$ is a binary variable equal to one when debate d is presided over by a female minister, and λ_m is a ministry fixed effect. The model is estimated separately for each *topic*, and, as in the previous step, I repeat this exercise for each of the 5 *topic models*. The estimation therefore results in K vectors of γ_t coefficients – one coefficient for each topic, in each topic model. When γ_{kt} is positive, this implies that the use of the topic increases when a female minister is appointed, and when it is negative it suggests that the use of the topic decreases on the appointment of a female minister.

Equations S8 and S9 therefore result in two vectors of coefficients: b_k gives the relationship between each of the topics in topic model k and the level of female debate participation under male ministers, and γ_k indicates how much each of the same topics in topic model k increases (or decreases) when a female minister is appointed. Assessing the correlation between these coefficient vectors allows us to test whether female ministers introduce legislation that focusses on topics which are associated with high levels of female participation under male ministers. Thus, to test the agenda-setting hypothesis, I regress the estimated b coefficients from equation S8 on the γ coefficients from equation S9 according to:

$$b_{t(k)} = \alpha + \zeta_k * \gamma_{t(k)} + \epsilon_{t(k)} \tag{S10}$$

If the agenda-setting hypothesis is correct, then the $b_{t(k)}$ and $\gamma_{t(k)}$ coefficients should be positively correlated, indicating that high female-participation topics (under male ministers) play a more prominent role on the policy agenda when a female minister is appointed. That is, we expect the ζ_k coefficient from equation S10 to be positive. Such a finding would contradict the legislative leadership hypothesis, as it would suggest that the increased lev-

Table S21: Topics ordered by their association with female participation under male ministers

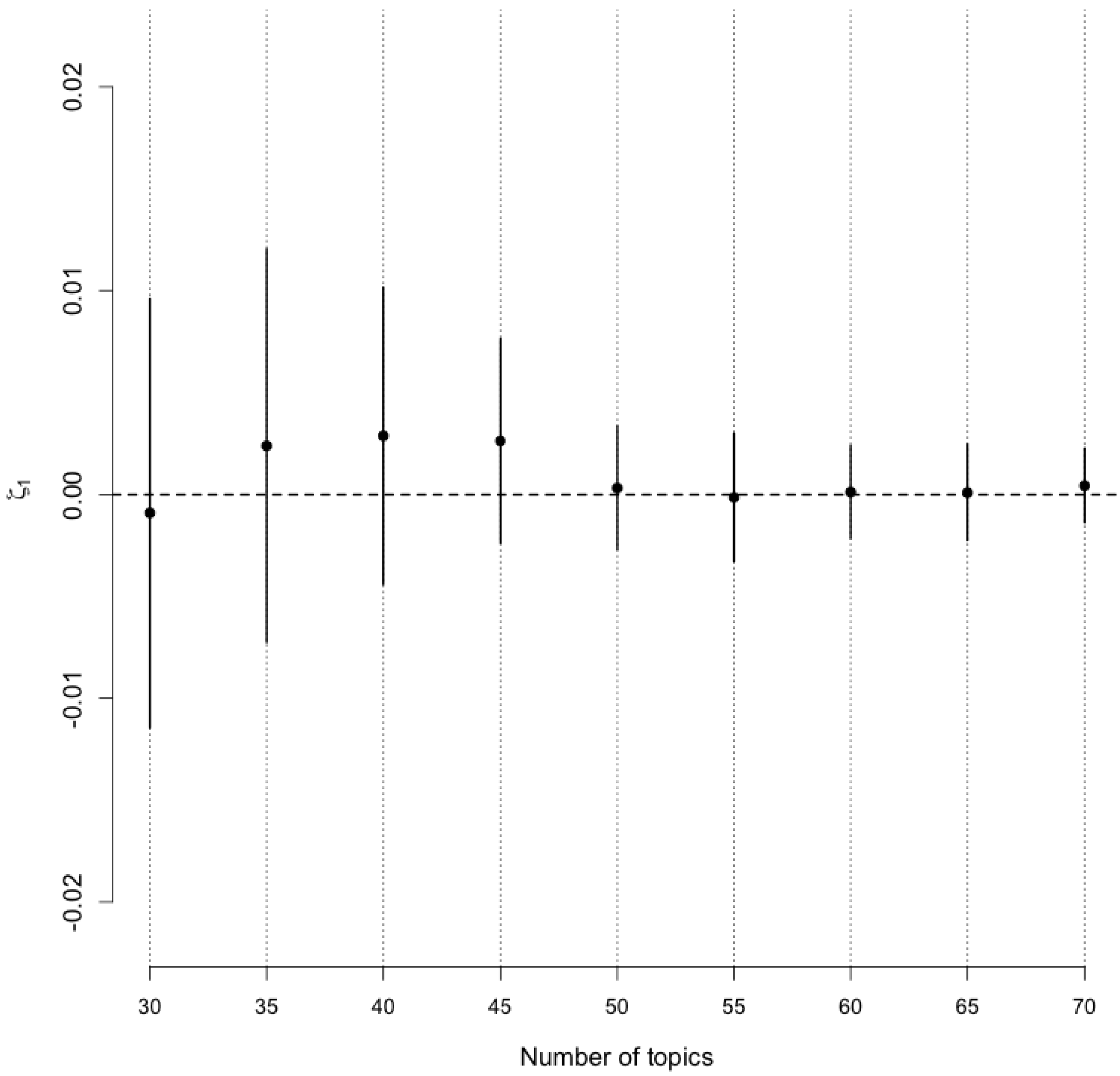
Topic	beta
labour_deficit_budget_left_coalit_shadow_let	1.116
doctor_hospit_trust_nhs_junior_patient_staff	0.880
school_teacher_pupil_educ_student_teach_children	0.723
pension_disabl_credit_child_incom_women_wage	0.678
eight_sport_olymp_game_lotteri_footbal_art	0.640
amend_bill_claus_legisl_provis_draft_order	0.633
energi_carbon_emiss_climat_gas_electr_renew	0.631
local_author_flood_council_fund_region_grant	0.620
prison_offend_crime_sentenc_polic_fine_victim	0.599
debat_motion_committe_monday_thursday_tuesday_hous	0.597
health_cancer_treatment_wait_patient_mental_nhs	0.556
railway_rail_transport_railtrack_airport_road_passeng	0.535
immigr_asylum_card_passport_ident_migrat_border	0.471
africa_aid_zimbabw_mine_african_intern_develop	0.458
busi_manufactur_industri_enterpris_job_skill_steel	0.426
risk_assess_advic_inform_test_identifi_emerg	0.363
paper_white_bbc_consult_propos_appoint_recommend	0.353
farmer_anim_farm_agricultur_food_diseas_outbreak	0.321
saddam_palestinian_syria_iran_israel_resolut_iraq	0.275
conserv_tori_liber_oppos_opposit_parti_polici	0.263
inquiri_investig_alleg_sir_report_evid_panel	0.250
bank_afternoon_post_lend_financi_custom_loan	0.245
scottish_scotland_hear_wale_elector_elect_assembl	0.162
afghan_afghanistan_troop_armi_defenc_deploy_arm	0.158
indic_terror_terrorist_intellig_threat_human_attack	0.113
treati_europ_european_union_negoti_honour_britain	0.112
sure_think_thing_look_absolut_import_get	0.076
cent_per_increas_figur_pay_compar_substanti	0.026
ireland_northern_beg_agreement_irish_paramilitari_decommiss	-0.0004
tax_rate_billion_inflat_cut_unemploy_budget	-0.008

els of legislative participation documented in the main results section could be attributed to the development of an increasingly ‘female-friendly’ agenda under female ministers. I present the estimated ζ_k coefficients – one for each of the topic models – in figure S9 along with their associated 95% confidence intervals.

Figure S9 provides no clear evidence that female ministers focus more attention on topics that are traditionally marked by high levels of female participation. As none of the slopes is statistically significant at traditional levels, the plot suggests that it is unlikely that changes to the legislative agenda are responsible for the changes in female participation documented in the main text.

Of course, the topics measured in this analysis are relatively coarse – even in the higher topic models – and so might not capture the more nuanced ways that female ministers might frame or structure debates which may be more conducive to female MPs’ participation. For instance, female committee chairs in the US seem to use a different tone when guiding deliberations than male committee chairs (Kathlene, 1994). Accordingly, this analysis does not rule out the possibility that female cabinet ministers are responsible for more subtle changes to the *way* in which the plenary agenda is discussed, but rather suggests that their appointment is not associated with a fundamental shift in that agenda overall.

Figure S9: There is no increase in the use of 'female friendly' topics on the parliamentary agenda when a female minister is appointed.



NOTE: The graph plots, on the x-axis, the number of topics, and on the y-axis, the estimated ζ coefficients from equation S10. There is no clear evidence that when female ministers are appointed, they focus more on topics that are traditionally popular with other female MPs.