

THE EFFECTS OF FEMALE LEADERSHIP ON WOMEN'S VOICE IN POLITICAL DEBATE

Jack Blumenau

UNIVERSITY COLLEGE LONDON

Forthcoming in the British Journal of Political Science

ABSTRACT

Do female leaders amplify the voices of other women in politics? I address this question by examining parliamentary debates in the UK House of Commons. In the context of a difference-in-differences design which makes use of over-time variation in the gender of cabinet ministers, I demonstrate that female ministers substantially increase the participation of other female MPs in relevant debates, compared to when the minister is male. Further, using a measure of debate influence based on the degree to which words used by one legislator are adopted by other members, I show that female ministers also increase the influence of female backbenchers. To explore the mechanisms behind these results, I introduce a new metric of ministerial responsiveness and show that female ministers are significantly more responsive to the speeches of female backbenchers than are male ministers.

Word count: 10224

This version: May 2019.

I thank Laura Bronner, Alexandra Cirone, Andy Eggers, Guy Grossman, Simon Hix, Sara Hobolt, Benjamin Lauderdale, Nicola Mastrococco, Øyvind Skorge and Dawn Teele for helpful conversations. I also thank seminar participants at the European Bank for Reconstruction and Development, the University of Pennsylvania, the London School of Economics, EPSA, and MPSA. This work was supported by the Economic and Social Research Council (grant number: ES/N016297/1).

Does the promotion of women to high political office increase the voice of other women in politics? The factors that determine the appointment of women to political leadership roles have been the subject of increasing study in recent years (Heath, Schwindt-Bayer and Taylor-Robinson, 2005; Escobar-Lemmon and Taylor-Robinson, 2009; Krook and O'Brien, 2012; O'Brien, 2015; Barnes and O'Brien, 2018), but we know considerably less about the implications of these appointments for the behaviour of other politicians. This is surprising as the idea that female leaders can have important effects on the experiences of other women is well established in electoral politics (Wolbrecht and Campbell, 2007; Beaman et al., 2008; Gilardi, 2015), education (Nixon and Robinson, 1999; Bettinger and Long, 2005; Brajer and Gill, 2010; Beaman et al., 2012) and business (Wang and Kelan, 2013; Bertrand et al., 2014). In this paper, I evaluate the effects of female leadership in a legislative setting by studying whether the appointment of female cabinet ministers in the UK increases the participation and influence of other female members of parliament (MPs) in House of Commons' debates.

Political debate is an important setting for evaluating female leadership effects. The ways in which individuals interact in group discussions can provide important insights into relative distributions of power, particularly with regard to gender (Karakowsky, McBey and Miller, 2004), and when speaking is a mechanism for collective decision-making, inequalities in participation and influence may reflect deeper inequalities between groups (Karpowitz and Mendelberg, 2014, 115). Gender-based inequalities are present in British political debate: in the period I study, although women occupied approximately one fifth of the seats in parliament, they accounted for only one tenth of the speaking time in the typical parliamentary debate.¹ Such descriptive evidence is concerning because if “some kinds of people routinely speak more than others in deliberative settings. . . then participation isn't equal, and one democratic standard has fallen.” (Sanders, 1997, 365) Understanding the conditions under which female legislators participate and hold influence in political debates is therefore important for evaluating the representation of women's interests in politics more broadly.

¹See appendix section S1

There is growing evidence that increasing the *number of women* in a group has important consequences not only for gender participation gaps in political discussion (Karpowitz, Mendelberg and Shaker, 2012; Ban et al., 2018), but also for the set of issues that discussion addresses (Mendelberg, Karpowitz and Goedert, 2014), and how often female discussants are interrupted (Mendelberg, Karpowitz and Oliphant, 2014). However, no research to date has considered whether *female leadership* also affects the processes or outcomes of political debate. Cabinet ministers – the leaders I study – are key figures in the UK legislative process who play the central role in parliamentary business, including debate, that relates to their ministries. I argue that the appointment of a female cabinet minister has the potential to impact upon the behaviour of other women in two ways.

Firstly, the appointment of women to cabinet minister positions may help to break down historically constructed stereotypes about the policy domains to which women are well-suited to contribute. By undermining these stereotypes, it is possible that female cabinet ministers act as ‘role models’ to other female parliamentarians. Secondly, female ministers may also behave differently from their male colleagues, and in so doing may promote a debating culture that is more conducive to, and encouraging of, the participation and influence of other female MPs than that which has traditionally existed in the UK. Taken together, these arguments suggest that the appointment of a female minister will lead to higher levels of female participation and influence in debates that are presided over by the new minister.

However, causal identification of leadership effects presents an empirical challenge. Women are more likely to be appointed to lead traditionally “feminine” cabinet posts (Escobar-Lemmon and Taylor-Robinson, 2009; Krook and O’Brien, 2012), and women also disproportionately contribute to debates that deal with traditionally “feminine” policy areas (Bird, 2005; Catalano, 2009). Because of this, simple estimates of the relationship between cabinet minister gender and female debate participation will be upwardly biased. To make progress, I focus on within-ministry variation in the gender of the cabinet minister over time. I compare female debate participation in ministries before and after a

switch in the gender of the minister, and compare this difference to changes in female participation in other ministries where the gender of the minister remains constant. This approach allows me to rule out any omitted variable bias that could be attributed to any fixed tendencies of women to engage with the work of particular ministries and not others. Using this framework to analyse almost half a million speeches between 1997 and 2017, I demonstrate that female ministers increase the participation of female MPs in relevant debates by approximately 20% over the level of female participation under male ministers.

However, if the issues women raise in their speeches are ignored by other parliamentarians, then the substantive importance of this effect may be negligible. I address this issue by building on new techniques for identifying important speakers in political debate (Erkan and Radev, 2004; Fader et al., 2007), which I use to examine the relative influence of male and female MPs' speeches. The measurement strategy is based on the assumption that the more that an MP's language is adopted by other MPs in subsequent speeches, the more influential is the MP. Using this measure, I show that women are also more influential in debate when their female colleagues are elevated to high-office, but the influence of men in debate remains constant regardless of minister gender.

Turning to mechanisms, I provide evidence that female ministers do, in fact, behave systematically differently in debate towards female MPs than do male ministers. I introduce a new quantitative measure of ministerial responsiveness which assumes that a minister is more responsive when the language they use to reply to a backbencher is more similar to the words that the backbencher uses. I show that female ministers are substantially more responsive than their male counterparts to the speeches made by female MPs, but that there is no gendered difference in ministerial responses to the speeches made by male MPs. These findings do not rule out the possibility of purely symbolic 'role model' effects, but they do suggest that the ways in which these women behave towards their female colleagues is a potentially important mechanism through which they can increase the prominence of other women in parliament.

A central focus of legislative scholars studying gender issues has concerned the link

between descriptive representation – the number of women elected to parliament – and substantive representation – the incorporation of women’s interests into policy outcomes (Wängnerud, 2009). This paper contributes to a growing literature which argues that, in addition to their numerical strength, the heights to which female politicians rise also matters for the incorporation of women’s preferences in the policy process (Chattopadhyay and Duflo, 2004; Humphreys, Masters and Sandbu, 2006; Childs and Krook, 2009). The findings here also suggest that female political leaders matter for women’s representation, though through an under-appreciated mechanism: their presence amplifies the voice of other women in politics.

FEMALE LEADERSHIP AND VOICE IN PARLIAMENTARY DEBATE

Why might the appointment of women to positions of high office affect the participation and influence of other women in political debate? I consider two potential mechanisms through which gendered leadership effects might operate in the parliamentary setting.

First, female leaders are often thought to be “role models” who provide examples of success for other women, and in so doing help to undermine the stereotypic beliefs (Dasgupta and Asgari, 2004; Asgari, Dasgupta and Cote, 2010) that are often the source of negative evaluations of women’s capabilities (Eagly and Johnson, 1990). An extensive literature argues that female role models in politics can have inspirational effects, because seeing other people ‘like them’ being active and successful in political life encourages women to increase their rates of political participation. In electoral politics, for example, where countries have higher proportions of female representatives, women are more likely to discuss politics, and to participate in political activities (Campbell and Wolbrecht, 2006; Wolbrecht and Campbell, 2007). The election of female politicians also affects the career aspirations of adolescent girls (Beaman et al., 2012) and increases the propensity for other women to stand for elections (Beaman et al., 2008; Gilardi, 2015). Similarly, in education, assignment to same-sex teachers can significantly improve students’ educational achievements (Nixon and Robinson, 1999); influence course choices (Bettinger and Long, 2005); and improve communication between students and teachers (Brajer and Gill, 2010). Appointing women

to corporate boards can also increase the number of women occupying other leadership positions within business (Wang and Kelan, 2013; Bertrand et al., 2014).

In the parliamentary setting, the historical under-representation of women in leadership roles is one factor that makes the role-model mechanism plausible. Women are systematically under-represented in leadership positions cross-nationally (Heath, Schwindt-Bayer and Taylor-Robinson, 2005; Krook and O'Brien, 2012); are appointed to leadership roles in unfavourable circumstances (O'Brien, 2015); and tend to control low prestige and “feminine” portfolios (Studlar and Moncrief, 1999; Escobar-Lemmon and Taylor-Robinson, 2009). The historical marginalisation of women in high-power roles may therefore create perceptions that certain policy areas, and even politics in general, represent distinctly “male domains” (Sapiro, 1981, 712). Therefore, by breaking with historical patterns, the appointment of women to powerful cabinet positions may reverse the impression that women are unsuitable for participation in politics (Mansbridge, 1999). Observing the success of one woman in a policy area may therefore send a signal that women in general are qualified to contribute to that domain.

However, while there is anecdotal evidence that some female MPs in the UK do see women with cabinet appointments as role-models,² these effects are likely to be less powerful for women in parliament than elsewhere in politics. Female politicians may inspire other women to participate actively in electoral politics, but women in the legislature already have extensive political experience, and are less likely to be affected by seeing other people like them in positions of power. As Karpowitz and Mendelberg (2014, 334) argue, many well-known gendered behaviours may not apply to elite women, because elites are different: “it takes an unusual woman to seek...office...and an even more unusual woman to do what it takes to obtain it.” Because of this, we should not expect role-modelling to be as strong a determinant of behaviour for elite women as for women in the population at large.

Beyond simple role-model effects, however, female leaders may also *behave* differently

²For example, Betty Boothroyd, the former speaker of the Commons, has emphasised that working for Barbara Castle – the first woman to hold a series of important cabinet minister positions – had important motivational effects for her: “She was my role model because I felt, well, if Barbara can do it then I can do it.” (Boothroyd, 2013)

from their male counterparts, and do so in ways that is conducive to the participation and influence of other women. Theorists argue that the “deeply embedded culture of masculinity” (Lovenduski, 2005, 48) that pervades parliament is contra-indicated to female influence in political discussion. Legislatures are marked by highly gendered conversational dynamics in which male contributions to policymaking are “heard” more than female contributions (Kathlene, 1994; Hawkesworth, 2003). In the UK, the declamatory and adversarial style of Westminster debate (Childs, 2004, 10; Lovenduski, 2005, 54) is seen as particularly antithetical to the participation and influence of women in the policy-process. More broadly, women frequently face speaking environments that are less conducive to their participation than men, even in highly professionalised settings. For example, female Supreme Court justices and lawyers in the US are more likely to be interrupted and ‘talked over’ than men, and that these experiences are associated with them speaking less in deliberations (Dietrich, Enos and Sen, 2017). Similarly, experimental evidence suggests that in male-dominated groups women are likely to be interrupted more often by men (Mendelberg, Karpowitz and Oliphant, 2014); discussion will focus less on traditional “women’s issues” (Mendelberg, Karpowitz and Goedert, 2014); and each individual women will speak less (Karpowitz, Mendelberg and Shaker, 2012).

Female leadership has the potential to alter these dynamics. For example, women tend to be more democratic in their approach to leadership (Eagly and Johnson, 1990) and in the legislative setting, female committee chairs act more as facilitators, rather than directors, of committee discussions, speaking less and make fewer interruptions than their male counterparts (Kathlene, 1994). In the UK, female politicians employ a distinct form of language and debating style which is more cooperative, approachable, and practical than that of their male colleagues (Childs, 2000, 68). In general, female rhetorical styles are less aggressive, more inclusive, and more cooperative than male speech patterns (Karpowitz, Mendelberg and Shaker, 2012, 534). Women also tend to be characterised by facilitative styles of speech, marked by high levels of politeness and responsiveness, while male speech is seen as less facilitative (Hannah and Murachver, 1999, 2007) and these styles are strongly

predictive of the speaking time of conversational partners (Thomson, Murachver and Green, 2001). An important component of these styles is the degree to which an individuals' contributions to discussion are responded to and acknowledged by other participants (Hannah and Murachver, 2007). Accordingly, one way in which female leaders in parliament affect the behaviour of other female parliamentarians is if there is differential *responsiveness* of male and female high-office-holders. If female cabinet ministers provide higher quality responses to the speeches of female legislators than do their male counterparts, the status of women in legislative debate is likely to increase when a woman is appointed. Such an increase in status could increase the degree of influence that women enjoy, and thus their willingness to participate in plenary debate.

Female leaders may affect the behaviour of other women in debate both by acting as role models and by changing debate dynamics. Although I do not directly measure the relative importance of these two mechanisms, I do provide evidence that is consistent with the idea that female cabinet ministers behave differently from male cabinet ministers during parliamentary debates. Nevertheless, both mechanisms imply the same reduced-form relationship between the appointment of a female cabinet minister and the behaviour of other women. Specifically, when a female MP is appointed to lead a ministry previously held by a man, I expect other female MPs will be more likely to *participate* in debates that pertain to that ministry than they would have been previously. In addition, as participation and influence in debate are closely related concepts (Kathleen, 1994, 573; Karpowitz, Mendelberg and Shaker, 2012), I also expect that the *influence* of women in political discussion will increase when debates are presided over by a high-ranking women.

DATA AND METHODOLOGY

HOUSE OF COMMONS DEBATES

I study House of Commons debates between May 1997 and February 2017. The full sample contains 53398 debates and just over a million individual speeches.³ These debates cover a

³This information comes from theyworkforyou.com, a public website that catalogues all speeches made by UK MPs.

wide variety of parliamentary business, the most important (and most common) of which are those relating to substantive motions, where MPs express an opinion on some policy matter, and ministerial Question Time where MPs ask questions to government ministers on matters that pertain to their department’s jurisdiction. Cabinet ministers play a central role in both substantive debates, where they speak to propose legislation for consideration, and in Question Time, where they answer questions from backbenchers.

The analysis requires that each debate is mapped to an individual government ministry. To assign debates to ministries, I note whether a current cabinet minister speaks in a given debate, and assign the debate to the ministry for which that cabinet minister is responsible. Where more than one cabinet minister speaks, I assign the debate to the ministry of the most frequently appearing cabinet minister. As some debates do not contain speeches from any cabinet ministers, the final sample for analysis contains 14320 debates consisting of approximately 460,000 speeches.

The rules of debate participation in the Commons are permissive relative to other legislative settings. Party leaders have no control over which MPs participate as the non-partisan Speaker of the House allocates speaking time amongst members. MPs can register their interest to speak in a particular debate by writing to the Speaker, or can ‘catch the eye’ of the Speaker during debate by standing up in the Chamber. An MP can also ‘intervene’ on the speech of another MPs. Interventions are shorter impromptu remarks normally seeking to ask a question or make a point related to the current speech, and are granted only if the MP currently speaking ‘gives way’. Interventions are almost always granted by MPs both to co-partisans and to MPs from opposing parties. Overall, although the Speaker has formal powers to select speakers, this power is used impartially, and restrictions are placed on MPs’ speeches only in the most popular debates. In addition, interventions are outside of the control of the Speaker completely. As a consequence, observed patterns of participation in debate are highly likely to reflect MPs’ desires to participate in debate.⁴

⁴One potential concern is that there may be gendered-dynamics in the ways that male and female Speakers of the House – or Deputy Speakers, who also occasionally preside over debate – select MPs to participate in debate. In table S10 in appendix section S6 I control for the gender of the relevant Speaker/Deputy Speaker in charge of debate and show that the main effects of my treatment are unaffected.

METHODOLOGY

The key independent variable throughout the analyses, $FemaleMinister_{mt}$, is equal to one when the minister responsible for a given ministry m in time t is a woman, and zero otherwise. Figure S3 in the appendix shows the variation in this variable over time for all 32 ministries included in the sample.⁵ During the study period, there are several ministries for which the responsible minister is never a woman – including the Defence ministry and the position of Chancellor of the Exchequer – but no ministry is always controlled by a female minister. While I include all ministries in the analysis, identification relies only on those ministries which see a change in the gender of the cabinet minister over time.

The outcome variable in this section is the proportion of words spoken by female legislators in a debate, d , with each debate pertaining to a ministry m and year-month t :

$$PropWordsWomen_{d(mt)} = \frac{\# \text{ words spoken by women}_{d(mt)}}{\# \text{ words spoken by men and women}_{d(mt)}} \quad (1)$$

I exclude speeches made by ministers themselves (whether they are male or female), ensuring that the figures are not artificially inflated by female ministers speaking more after they are appointed. I also remove speeches made by the Speaker of the House, which are almost exclusively procedural. The measure defined in equation 1 is clearly sensitive to the number of women in parliament – a quantity that varies over time – and so in appendix section S3 I replicate the analysis with alternative specifications of the dependent variable.⁶

Simple comparisons between debates held under male and female ministers are likely to result in biased estimates of the effect of minister gender. There are two main identification concerns. First, systematic differences between ministries will likely affect the degree to which female legislators choose to participate in legislative debate, and also affect the likelihood that women will be appointed to lead particular ministries. For example, previous

⁵I define a cabinet minister as any individual who is paid a government salary and regularly attends cabinet meetings. I treat each change in the name of the ministry as a new ministry.

⁶Most importantly, I specify models for:

$$RatioWordsWomen_{d(mt)} = \frac{PropWordsWomen_{d(mt)}}{\text{Proportion of women in parliament}_{mt}} \quad (2)$$

which accounts for the changing gender composition of the parliament over time.

research shows that women are significantly more likely to participate in legislative debates that relate to areas of traditional concern to women, including health care and children and family issues (Catalano, 2009; Pearson and Dancey, 2011). Figure S2 in the appendix suggests that unobserved ministry characteristics such as these are clearly influential in the data here. Women speak significantly more in ministries such as ‘Education’, ‘Health’ and ‘Children, Schools and Families’, and significantly less in debates pertaining to the ‘Defence’, ‘Foreign’, and ‘Justice’ ministries. In addition, cross-nationally, women are more likely to be appointed to ministerial portfolios that address “feminine” issue areas (Escobar-Lemmon and Taylor-Robinson, 2009; Krook and O’Brien, 2012; Barnes and O’Brien, 2018). If the ministries to which women are appointed are those for which the rate of female participation is already high, then naive comparisons between debates presided over by female and male ministers are likely to be upwardly biased.

Second, the proportion of women holding seats in parliament changes over time, and in periods with more female legislators, women are clearly likely to contribute a greater proportion of words in parliamentary debates and will also be more likely to be selected for ministerial office. Again, if parliamentary turnover increases the number of women in parliament, then naive comparisons of female participation between male- and female-led ministries will be upwardly biased.⁷

To overcome these problems, I estimate linear fixed-effect regressions of the following form:

$$PropWordsWomen_{d(mt)} = \beta_1 * FemaleMinister_{mt} + \lambda_m + \delta_t + \epsilon_{d(mt)} \quad (3)$$

where λ_m is a ministry fixed-effect that washes out any omitted variable bias from unobserved ministry characteristics that are fixed over time (such as the degree to which a ministry deals with policy that is traditionally of greater concern to women), δ_t is a year-month fixed-effect to control for common shocks across ministries in a given month (such as the number of women in parliament), and $\epsilon_{d(mt)}$ is the error term. β_1 is the coefficient

⁷Note that the alternative specification of the dependent variable in equation 2 is not subject to this concern. Therefore, in addition to the fixed-effects specification, the results from the models (section S3 in the appendix) based on the *RatioWordsWomen* measure provide a reassuring robustness check here.

of interest, and captures the reduced-form causal effect of the appointment of a female minister on the participation of women in debates for those ministries that experienced a change in minister gender over time.

This fixed-effect design is equivalent to a multi-period ‘difference-in-differences’ in the style of Angrist and Pischke (2009, 234). β_1 identifies the effect of switching from a male to female minister based on the within-ministry variation of the outcome variable among those ministries that see changes in the gender of the minister over time. By accounting for fixed characteristics of ministries that might predict both female debate participation and the appointment of a female minister, the model compares changes in female debate participation in ministries that experience a switch in minister gender to ministries where the gender of the minister remains constant, while differencing out the general trends across ministries in a given month.

Identification of the causal effect relies on changes in minister gender being exogenous to the level of female debate participation, conditional on time and ministry fixed-effects. The key identifying assumption is that treated ministries would have followed the same trend as non-treated ministries in the absence of treatment. I relax this assumption by estimating further models which include ministry-specific linear (λ_{m1}) and quadratic (λ_{m2}) time trends:

$$\begin{aligned} PropWordsWomen_{d(mt)} = & \beta_1 * FemaleMinister_{mt} + \lambda_{m0} + \delta_t \\ & + \lambda_{m1}t + \lambda_{m2}t^2 + \epsilon_{d(mt)} \end{aligned} \tag{4}$$

where t is a time variable. Furthermore, in contrast to the typical multi-period ‘difference-in-difference’ model, in this setting the treatment (the presence of a female minister) switches on and off over time. That is, once appointed, a female minister might also leave office, and ministries often see multiple female ministers (appointed at different times) over the study period.⁸ To account for the possibility that differential *local* trends within min-

⁸Female MPs might increase their rate of participation in debates under female ministers, but then continue to participate at a similar rate when that female minister is replaced by a man. Spillovers of this type would cause a downward bias in the treatment estimates presented here. Nevertheless, in appendix section S5, I investigate whether rates of female participation in debate persist after a female minister

istries might confound the causal effect, I also estimate generalised additive models (GAM) which include non-parametric ministry-specific time trends:

$$\begin{aligned} PropWordsWomen_{d(mt)} = & \beta_1 * FemaleMinister_{mt} + \lambda_{m0} + \delta_t \\ & + \lambda_{m1}f(t) + \epsilon_{d(mt)} \end{aligned} \tag{5}$$

These models represent extremely conservative specifications, as the addition of the ministry-specific trends means that all unobserved and smoothly varying confounding differences are removed from the estimate of β_1 , and that only sharp changes to the trend in the outcome variable that occur at the same time as the change in minister gender contribute to this estimate. As none of the substantive or statistical results change noticeably when this crucial identifying assumption is relaxed, this lends significant support to the empirical design. I also provide further evidence for the validity of the identification assumption by estimating a dynamic panel model, estimates the treatment effect in the time periods before and after the actual change in minister gender and which I describe in more detail below. Finally, as there are only 32 ministries in the data, I construct bootstrapped standard errors, clustering at the ministry level (Cameron and Miller, 2015).⁹

FEMALE MINISTERS AND DEBATE PARTICIPATION

Before turning to the main results, I present a simple graphical analysis. Figure 1 shows, on the y-axis, the proportion of words spoken by female MPs in each month in each ministry that experienced a change in the gender of the minister, and the x-axis gives the date. Black line segments represent periods in which the presiding minister is female, and gray segments represent male ministers. The plot provides clear evidence of a female-leadership effect whilst also revealing heterogeneity across ministries. In many cases, the appointment of a female minister is accompanied by an increase in the proportion of words spoken by other female MPs. The effect appears to be particularly pronounced in the ‘Trade and

steps down from office. The findings suggest that the motivating effects of female leadership do seem to be largely confined to the periods in which the female minister holds office.

⁹I bootstrap 1000 times, resampling ministries from the full data with replacement. Because the GAM model is computationally very burdensome, I do not bootstrap this model. Nevertheless, the point estimates are very similar to the other models.

Industry’, ‘Home’, ‘Culture, Media and Sport’, and ‘International Development’ ministries. By contrast, there is less evidence of an effect in some other ministries, though in no cases does the appointment of a female minister appear to *decrease* the participation of other female MPs.¹⁰

Table 1 presents the results of the regression analyses. Model 1 presents the naive estimate of the effect of a female minister, without controlling for ministry or year-month fixed-effects. Models 2 and 3 introduce these fixed-effects separately, and model 4 presents the results of the ‘difference-in-differences’ model which includes both fixed-effects. The coefficient of the main variable of interest, *FemaleMinister*, is positive and significant in all four models, but it decreases noticeably when accounting for ministry. This suggests that female ministers are indeed appointed to lead ministries where the level of debate participation of other female MPs is already high. Nevertheless, the effect remains significant in model 4, implying that the appointment of a female minister leads to an increase in debate participation of other female MPs. The size of the effect is also substantial. Based on model 4, the appointment of a female minister increases the proportion of words used by women by 4.4 percentage points of total words. This corresponds to an increase of approximately 23% [10%, 36%] over the average speech rate of women in debates under male ministers.

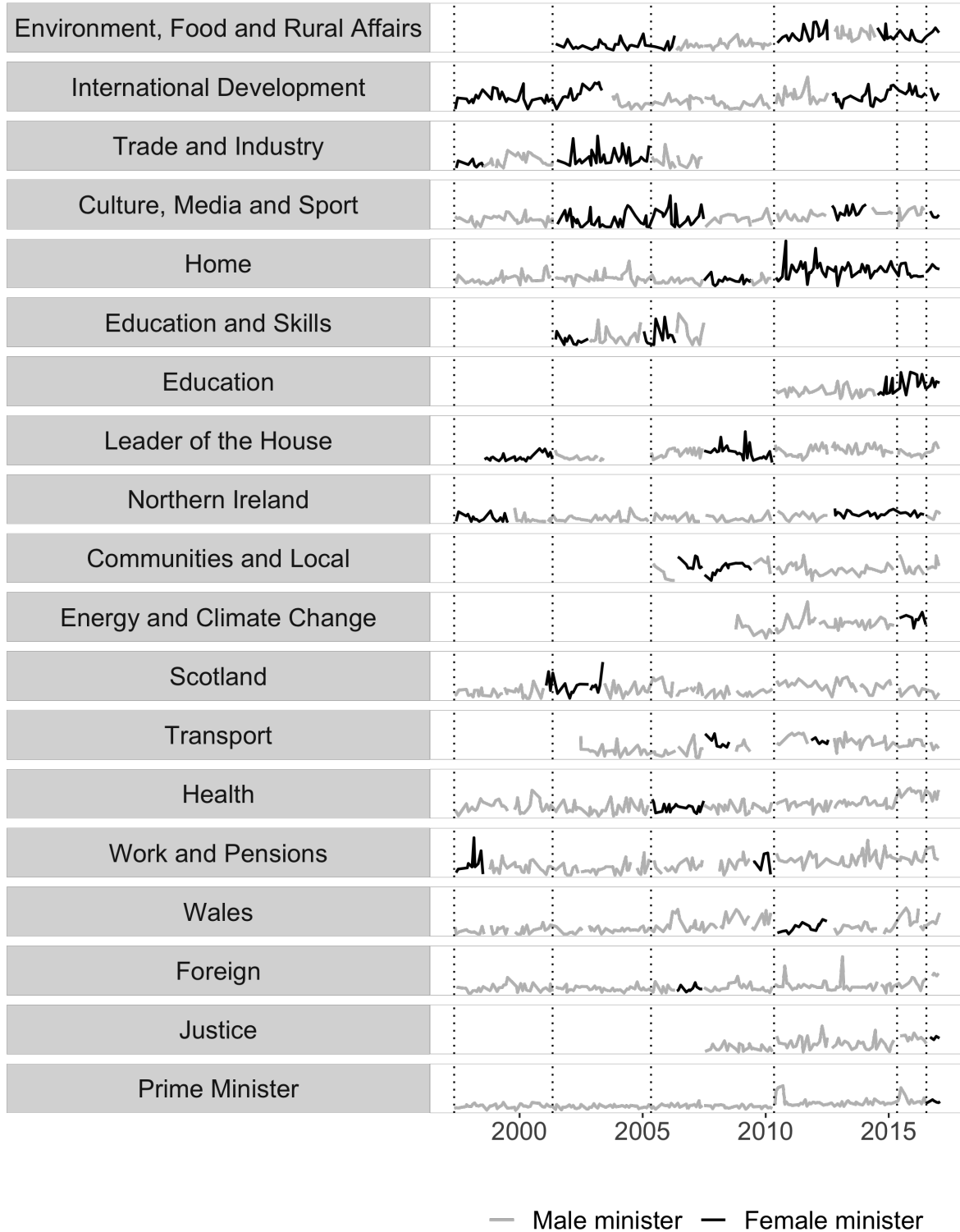
These results are robust to the introduction of linear, quadratic and non-parametric ministry-specific time trends in models 5, 6 and 7. Based on model 6, the appointment of a female minister increases the proportion of words spoken by other female MPs by between 10% and 33%. That the inclusion of ministry-specific time trends changes the estimates so little is encouraging, as it rules out the possibility that the effect is driven by either global or local trends in unobserved confounding variables.¹¹

An additional robustness check is presented in appendix figure S5, where I plot the

¹⁰In the regression analyses below, which also control for general (across ministries) and local (within ministries) trends in female participation over time, I average over this underlying heterogeneity.

¹¹In addition to being subject to bias, the naive models are also subject to significantly higher residual variance than the fully specified models. The uncertainty around the main effects reduces considerably in the more flexible model specifications.

Figure 1: Proportion of words spoken by women in treated ministries, over time



NOTE: The plot shows the proportion of words spoken by women in each calendar month in each ministry that experienced a change in the gender of the presiding minister.

Table 1: Effect of appointing a female minister on female debate participation

	<i>Prop Words Women</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.062*** (0.019)	0.055*** (0.014)	0.050** (0.021)	0.041*** (0.012)	0.037*** (0.011)	0.037*** (0.010)	0.040*** (0.008)
Constant	0.177*** (0.011)	0.094* (0.050)	0.100 (0.082)	0.039 (0.050)	0.046 (3.085)	1.243 (155.730)	0.059 (0.143)
Month FEs	×	✓	×	✓	✓	✓	✓
Ministry FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Effect Size %	35	31	28	23	21	21	22
95% CI	[13,56]	[15,47]	[5,51]	[10,36]	[8,34]	[10,33]	[14,31]
Observations	14,320	14,320	14,320	14,320	14,320	14,320	14,320
R ²	0.012	0.075	0.053	0.106	0.113	0.122	
Adjusted R ²	0.012	0.062	0.051	0.091	0.096	0.103	0.115

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. *p<0.1; **p<0.05; ***p<0.01

estimates and 95% bootstrapped confidence intervals (clustered by ministry) from a dynamic panel model. The coefficients from this model represent the estimated difference in the outcome between treated and untreated ministries in the periods before and after the treatment occurs. The results strongly support the identifying assumption, as I find no significant ‘placebo’ effects in the two years prior to the change in minister gender. This strengthens the plausibility of the design, as it suggests that there are no unobserved variables leading to differential trends in the outcome between the treatment and control ministries prior to the appointment of a female minister.

Finally, in appendix section S4 I report results from a series of analyses in which I reestimate the main model (equation 3) for different subsets of the data. First, leadership effects may be more likely to manifest for MPs that belong to the same party as the newly appointed cabinet minister, and may be less pronounced for MPs from opposition parties. I therefore subset to focus only on the effects for opposition MPs (all MPs whose party

is not in government). Second, in the UK, the Labour party has historically had more female MPs than other parties and it is possible that these pre-existing gender differences across parties also influence the strength of the effects. I therefore also rerun the main analysis focussing on the MPs of each party in turn. Third, I investigate whether the Question Time debate format drives the main results here, by subsetting to Question Time and non-Question Time debates. Figure S4 summarises the results of these analyses, and shows there is in fact very little heterogeneity in the effect across the different subsets.¹² In all cases, the main effect is positive and falls within the confidence intervals for model 4 in table 1, though in some cases the smaller sample size decreases the precision of these subset estimates.

Overall, the results presented in the graphical analyses, the main specifications, and in robustness checks provide strong support the main claim of this paper: when a female minister is appointed, other women speak approximately 20% more in debates pertaining to that ministry than when the responsible minister is male.

FEMALE MINISTERS AND INFLUENCE

Changes in participation tell us little about how debate contributions are received by others in the House. If women speak at an increasing rate, but the issues and concerns that they raise are ignored by other parliamentarians, then the substantive importance of the effects documented above may be limited. In this section, I use the texts of the parliamentary speeches to ask whether female backbenchers also play a more *influential* role in political debate under female ministers than under male ministers.

How might we identify ‘influential’ speakers? I consider an MP to be influential when the issues and concerns she raises in her speeches are adopted and discussed by other members in subsequent speeches. Having other people pick up on your framing of an issue is a way of controlling how the debate proceeds: it means other people are taking up your perspective, whether or not they agree with it. Influential MPs are therefore literally ‘shaping the

¹²Additional detail for the analysis subsetting by Labour and Conservative MPs can be found in tables S4 and S5 in appendix section S4.

debate’. The intuition behind the measurement strategy is to identify distinctive language that first appears in the statement of one MP, but then gets used subsequently by later MPs.

Building upon methods for detecting influence in text corpora (Erkan and Radev, 2004; Mihalcea, 2004; Fader et al., 2007), I assess the influence of a speech, i , by calculating how many ‘references’ i receives from other speeches within the debate. One speech, j , can be understood to ‘reference’ another, i , when it occurs after i in the debate and when it comprises language which is sufficiently similar to that used by i . By using similar language to i , j is implicitly indicating that i is relevant and important for the discussion at hand. There are myriad reasons why one speech may use similar language to another (direct quotation; expression of criticism; statement of support) but the goal here is not to assess the substantive meaning of each link. Rather, I assume that a speech that shares language with many other speeches is being *collectively referenced* and thus can be considered an important and influential speech within the debate. A simple way of assessing influence would therefore be to simply count the number of references each speech receives. I consider not only the number of references, but also incorporate information about the influence of the referencing speeches. Thus, the more references i receives, and the higher the influence of the referencing speeches, the higher is the influence of i within the debate.

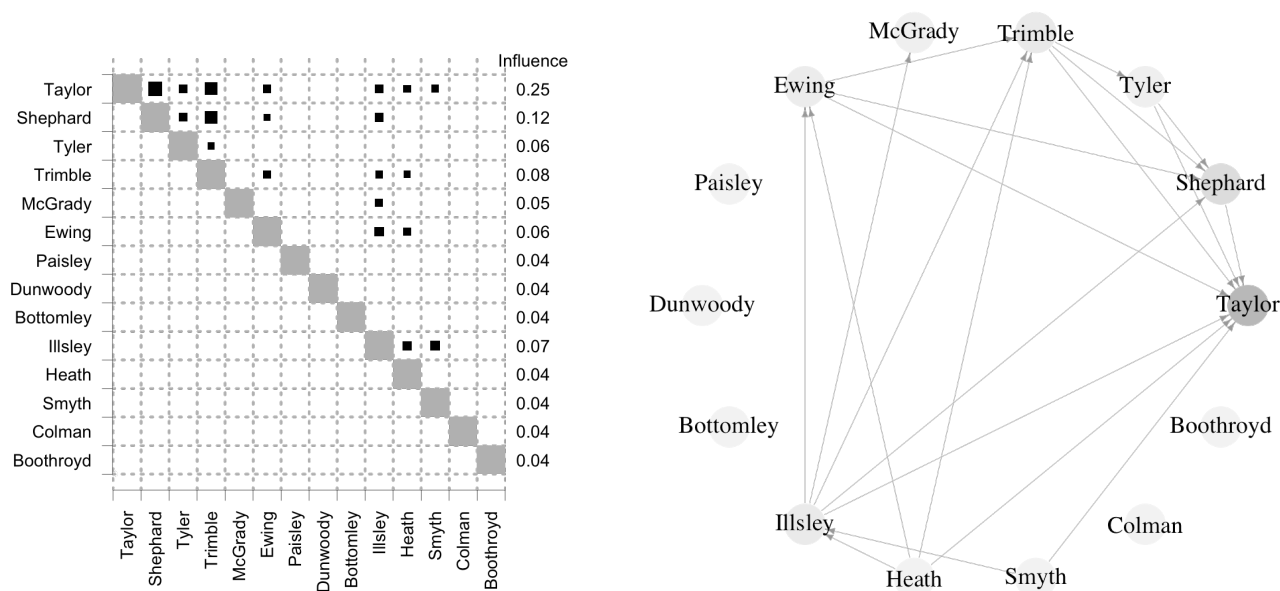
The basic steps of the measurement procedure are as follows.¹³ First, I construct debate-specific similarity matrices which measure the cosine-similarity between all speeches in each debate, where speeches are represented as term-frequency-inverse-document-frequency (tf-idf) vectors.¹⁴ Second, these matrices are converted into directed graphs (again, one for each debate) where the nodes represent speeches, and edges are placed between nodes for

¹³Full details can be found in appendix section S8. For manipulation of the text data I use the `quanteda` package (Benoit et al., 2018) in R. For the network models I use the `igraph` package (Csardi and Nepusz, 2006).

¹⁴Each element in the speech-vector v 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. 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.

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:

Figure 2: Example of ‘influence’ in a debate



NOTE: The left panel shows the ‘reference’ patterns for an example debate. The right panel depicts the similarity matrix as a network graph.

the speech-pairs whose cosine-similarity is greater than some threshold value.¹⁵ The edges are then weighted by the similarity scores. Third, I analyse these matrices using an iterative ranking algorithm (Page et al., 1999) to calculate a vector of centrality scores, P , which correspond to the influence of each *speech* in each debate.¹⁶ Finally, the influence score of an *MP* in a given debate is the sum of the influence scores for the speeches given by that *MP* in that debate.

In figure 2, the left-hand plot depicts the similarity matrix for an example debate with 14 separate speeches. Speakers are sorted according to the order in which they participated

$$S_d(i, j) = \text{sim}(v_i, v_j) = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|} = \frac{\sum_{w=1}^W v_{wi} \cdot v_{wj}}{\sqrt{\sum_{w=1}^W v_{wi}^2} \cdot \sqrt{\sum_{w=1}^W v_{wj}^2}} \quad (6)$$

i.e. the cosine-similarity of the weighted word-count vectors of speeches i and j in debate d .

¹⁵In line with Fader et al. (2007) I set this threshold to 0.25.

¹⁶Mihalcea (2004) shows that either the Kleinberg (1999) HITS algorithm or the Page et al. (1999) PageRank algorithm can be used to calculate P . Results from the HITS algorithm can be found in appendix section S9.

in the debate, such that Taylor is the first speaker and Boothroyd is the last speaker. As I only allow one speech to reference another when it occurs later in the debate than the speech it is referencing, the bottom triangle of the matrix is empty. I also exclude the possibility that a speaker can reference herself (gray shaded boxes). The black squares indicate the cosine similarity between two speeches, and are scaled such that when the similarity between a pair of speeches is 1 (i.e. when the tf-idf vectors are identical) the black square will fill the dashed box that contains it. The empty elements of the upper triangle correspond to speech pairs where the similarity between the speeches is lower than the minimum threshold. The left panel shows, for example, that Taylor’s speech is referenced by many subsequent speeches, while Tyler’s speech is referenced only by Trimble. The right margin of the plot gives the vector of influence scores for this debate. Taylor’s speech has an influence score of 0.25 and Tyler’s speech has an influence score of just 0.06.

The right-hand panel depicts the same similarity matrix as a directed network graph, with speeches as nodes (shaded proportionally to the influence scores) and edges as the ‘references’ flowing from one speech to another. Taylor’s speech is referenced by many other members, while Bottomley, Colman, Dunwoody, and Paisley’s speeches are not sufficiently similar to any subsequent speeches and which therefore receive no references.

Validation is essential for text-based measures of political concepts (Grimmer and Stewart, 2013) and closely related measurement strategies have been subjected to validity checks in previous work (Erkan and Radev, 2004; Fader et al., 2007). In the appendix, I test two relatively unambiguous intuitions about which actors in the House of Commons we expect to be influential in parliamentary debate. First, in appendix table S16 I show that cabinet ministers (who have agenda-setting privileges in debate) are on average 5 times more influential than backbench MPs, while the Speaker of the House (whose speeches are almost exclusively procedural) is less than half as influential as backbenchers. In addition, in appendix figure S7 I show that the influence scores estimated from the procedure above correlate strongly and positively with the number of times that an MP is directly mentioned by other MPs in debate. These comparisons provide reassuring evidence regarding

the face validity of the measure of influence described above.

One potential concern is that this measure is simply proxying either for speech *length* or, because references only flow from later speeches to earlier ones, the *position* that a speech occurs in a debate. In appendix figure S8 I show there is a very weak relationship between length and influence (the average correlation across all debates is -0.05), and although there is a stronger negative association between influence and debate position, the influence measure is clearly picking up information above and beyond simple debate ordering (the average correlation is -0.45).

With this measure in hand, I now analyse the effect of the appointment of a female cabinet minister on the influence of female MPs. In contrast to the previous analysis here I concentrate on effects at the individual – rather than debate – level.¹⁷ As before, I exclude all speeches made by cabinet ministers and by the Speaker of the House and estimate models of the form:

$$\begin{aligned}
 influence_{id(mt)} = & \beta_1 * FemaleMP_i + \beta_2 * FemaleMinister_{mt} + \\
 & \beta_3 * (FemaleMP_i * FemaleMinister_{mt}) + \\
 & \sum_{p=1}^P \beta_{party_p} * Party_i + \lambda_{m0} + \delta_t + \epsilon_{id(mt)}
 \end{aligned} \tag{7}$$

where $influence_{id(mt)}$ represents the influence of member i in debate d pertaining to ministry m at time t . β_1 captures the average difference in influence between male and female MPs when the minister is male. β_2 represents the marginal effect of a female minister on the influence of male MPs, and the equivalent effect for female MPs – and the main quantity of interest – is given by $\beta_2 + \beta_3$. If $\beta_2 + \beta_3 > 0$, this implies that female MPs’ influence increases after the appointment of a woman minister.¹⁸ As previously, in addition to ministry and time fixed-effects (λ_{m0} and δ_t , respectively), in some specifications I also relax

¹⁷Angrist and Pischke (2009, 235-237) show that the difference-in-difference model with individual level data and a group level treatment is equivalent to an appropriately weighted group-level (debate) model which I include in appendix section S13.

¹⁸In appendix section S10 I present equivalent results from a split-sample analysis, where I evaluate the effects of female leadership on the influence of male (table S12) and female (table S13) MPs separately. The results are substantively and statistically very similar.

Table 2: Effect of appointing a female minister on MPs' debate influence

	<i>Influence</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.038*** (0.006)	0.033*** (0.002)	0.048*** (0.014)	0.021 (0.019)	0.040*** (0.015)	-0.803*** (0.265)	32.145 (159.545)
Female minister	0.003 (0.005)	0.001 (0.003)	0.003 (0.005)	0.001 (0.003)	0.001 (0.003)	0.0003 (0.003)	-0.004*** (0.001)
Female MP	-0.001 (0.002)	0.0001 (0.002)	0.001 (0.001)	0.001 (0.002)	0.002 (0.002)	0.002 (0.002)	0.001 (0.001)
Interaction	0.011*** (0.003)	0.009*** (0.003)	0.012*** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.009*** (0.001)
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	0.045	0.016	0.056	0.063	0.066	
Adjusted R ²	0.002	0.045	0.015	0.055	0.061	0.065	0.076

NOTE: Regression coefficients are shown with cluster-robust standard errors (clustered by ministry) shown in parentheses. *p<0.1; **p<0.05; ***p<0.01

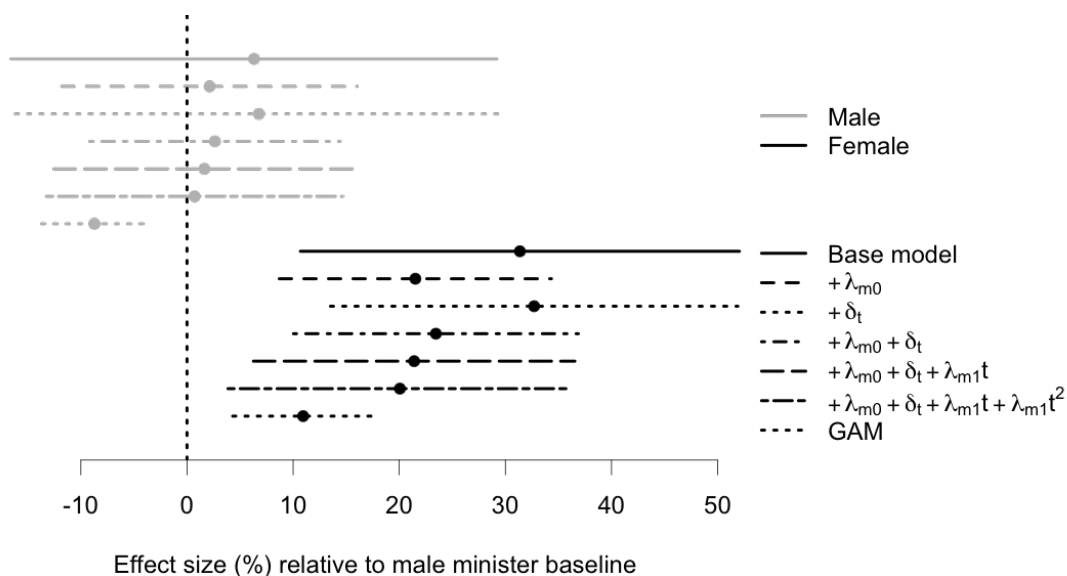
the common trend assumption with the addition of ministry-specific linear, quadratic, and non-parametric time trends. In these individual-level models, I also control for the party of each MP ($Party_i$). Errors are again clustered at the ministry level,¹⁹ and table 2 presents the results.

Model 1 gives the results of a naive specification without controlling for ministry or time fixed-effects, and indicates that while the appointment of a female minister has no effect on the influence of *male* MPs in parliamentary debates, *female* MPs' influence does increase when a female minister is appointed.²⁰ The introduction of ministry and time fixed-effects in models 2, 3 and 4, does not change the estimate dramatically: the appointment of a

¹⁹The individual-level models are computationally more burdensome than the debate-level models, and so I present traditional cluster-robust standard errors in table 2. For robustness I present bootstrapped clustered standard errors of the equivalent debate-level models in the tables in appendix section S13.

²⁰F-tests comparing the full models in table 2 to restricted models that do not include either of the interacted variables show clearly that the interacted variables are also jointly significant.

Figure 3: Marginal effect of female minister on influence



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.

female minister is significantly related to an increase in the influence of female MPs in parliamentary debate, but has no effect on the influence of male MPs. Models 5, 6 and 7 include ministry-specific linear, quadratic and non-parametric time trends. As before, that the effect does not disappear once controlling for these trends gives additional support to the design-based identification strategy.

Figure 3 shows the percentage change in influence for male (gray lines) and female (black lines) MPs after the appointment of a female minister, relative to a baseline where the minister is male. The marginal effect for male MPs is close to zero, varying in sign, and statistically insignificant for all models except for the GAM. For female MPs, the effect is always positive and significant, and the magnitude is non-trivial: based on model 6, female MPs are 20% [4%, 36%] more influential under a female minister than when the minister is male.²¹ In sum, the results indicate that the appointment of a female minister leads

²¹Although the confidence intervals in figure 3 overlap for some models, the effect of minister gender for female MPs is significantly greater than that for male MPs in all models, as evidenced by the significant interaction terms in table 2.

to an increase not only in the degree to which female MPs participate in plenary debate, but also in the level of influence that female MPs enjoy when debating with their fellow parliamentarians.

MINISTERIAL RESPONSIVENESS

The results above clearly indicate that female leadership has important effects on the experiences of female MPs in political debate. How might we account for these reduced-form effects? The processes underpinning these findings are likely to be many and varied, and isolating the mechanisms behind causal effects is notoriously difficult (Bullock, Green and Ha, 2010; Imai et al., 2011). In this section, I investigate the plausibility of one particular mechanism – that female leaders behave differently in debate towards female MPs than male ministers do – and in the subsequent section I report results that rule out some potential alternative mechanisms.

In particular, one explanation for these findings might be that female cabinet ministers behave in a systematically different manner towards female MPs than do male ministers. In particular, female ministers may be more *responsive* to the speeches of female MPs. Such a hypothesis is consistent with findings in the literature on social linguistics: “conversational partners who offer encouragement and are attentive and responsive are more likely to elicit frequent and active participation from speakers in the conversation.” (Hannah and Murachver, 1999, 157) Similarly, qualitative evidence from the UK suggests male MPs are often unresponsive to the speeches made by female MPs (Childs, 2004, 6). If female ministers give other female MPs responses to their speeches that are of higher quality, this is likely to signal that other women’s contributions are more influential in discussion, and may motivate higher levels of participation in future debates.

What are the important properties of responsiveness? I assume that a speech, j , responds to another speech, i , when it occurs directly after i and when it engages with the same thematic content as i . I also consider one speech to be more responsive to another when that speech is longer, on the assumption that longer responses give a greater impression of attentiveness and fullness of reply than shorter responses. I define a metric which

measures how similar two (consecutive) speeches are in terms of the words that they use. Making use of the same tfidf-representation of speeches as employed above (described in detail in equation S4), the responsiveness of speech j to speech i is given by:

$$res_{j \rightarrow i} = sim(v_i, v_j) * n_j \tag{8}$$

where the first term on the right-hand side of the equation is the cosine-similarity between the two tf-idf vectors, and n_j is the number of words in speech j . When all elements of v_i and v_j are positive, as they are here, the cosine-similarity of two documents is bounded between zero and one. An intuitive interpretation of $res_{j \rightarrow i}$ is therefore the (weighted) number of words in speech j that are responding to speech i .²²

I provide two types of validation for this measure in appendix section S12. First, I show that equation 8 captures something distinct from topicality, as 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. Second, I demonstrate that patterns of responsiveness conform with basic intuitions of parliamentary behaviour in the Commons by leveraging the particular structure of minister-backbencher interactions in Question Time debates.

I now turn to the main analysis. To reiterate, if female MPs speak more and become more influential because they receive higher quality responses from female ministers than male ministers, then ministerial speeches subsequent to female speeches should be marked by higher levels of res when the presiding minister is female. I therefore subset the data to those speeches made by backbench MPs which are immediately followed by speeches made by ministers, and estimate models of the following form:²³

²²Note that as i occurs prior to j , it therefore cannot be understood to ‘respond’ to j . For this reason, $res_{i \rightarrow j}$ is not meaningful in our context, and I calculate equation 8 only for sequentially adjacent speeches.

²³As before, an equivalent debate-level model, with very similar results, is presented in appendix section S13.

$$\begin{aligned}
res_{s(i)d(mt)} &= \beta_1 * FemaleMP_i + \beta_2 * FemaleMinister_{mt} + \\
&\beta_3 * (FemaleMP_i * FemaleMinister_{mt}) + \\
&\beta_4 * MinisterSameParty_s + \lambda_{m0} + \delta_t + \lambda_{m1}t + \lambda_{m2}t^2 + \epsilon_{s(i)d(mt)} \quad (9)
\end{aligned}$$

The unit of analysis in these models is a speech made by a backbencher, which is immediately followed by a speech made by a minister. Thus, $res_{s(i)d(mt)}$ is the response *received* by a speech s made by MP i in debate d pertaining to ministry m at month t . β_1 indicates the difference in responsiveness received by male and female MPs when the minister is male. β_2 captures the effect of the appointment of a female minister on the responses received by male MPs. β_3 therefore captures the interaction between the gender of the MP speaking, and the gender of the minister responding. A positive value for $\beta_2 + \beta_3$ would indicate that the appointment of a female minister leads to an increase in ministerial responsiveness to speeches by female MPs.²⁴ As before, I include ministry and time fixed-effects and the various ministry-specific time trends. Additionally, in order to account for the possibility that responsiveness might differ between ministers responding to speeches by members of their own party (rather than an opposition party), I also include a dummy for whether the minister is from the same party as the speaker of speech s . Errors are again clustered at the ministry level.

Table 3 presents the results. The interaction effect of interest, β_3 , is positive, significant, and sizeable in magnitude across all model specifications. I plot the substantive magnitude of these effects in figure 4, where the baseline is the average responsiveness of male ministers to male and female speeches. Based on the estimates in model 6, the appointment of a female minister increases the responsiveness to female speeches by 22% [11%, 33%]. By contrast, the appointment of a female minister has no consistent effect on the responsiveness to male speeches: across all models in table 3, β_2 is small in magnitude and in many cases

²⁴In appendix section S10 I again present equivalent results from a split-sample analysis, where I evaluate the effects of female leadership on the responsiveness to male (table S14) and female (table S15) MPs' speeches separately, rather than via the interaction model described here. The results are again very similar.

Table 3: Effect of appointing a female minister on the responsiveness to MPs' speeches

	<i>Responsiveness</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.212 (1.512)	0.026 (1.517)	0.683 (1.467)	0.656 (1.487)	0.713 (1.486)	0.757 (1.483)	0.832*** (0.288)
Female Minister	-0.493 (0.937)	-1.460 (1.355)	-0.581 (0.727)	-1.482*** (0.464)	-1.024** (0.446)	-0.721 (0.469)	-1.154** (0.474)
Interaction	5.738*** (1.505)	5.665*** (1.416)	5.772*** (1.518)	5.660*** (1.456)	5.678*** (1.477)	5.633*** (1.385)	5.390*** (0.645)
Constant	24.985*** (0.755)	24.629*** (0.146)	31.388*** (5.407)	27.300*** (4.899)	28.794*** (4.524)	186.167* (100.319)	264.770 (450.669)
Same Party Control	✓	✓	✓	✓	✓	✓	✓
Ministry FEs	×	✓	×	✓	✓	✓	✓
Month FEs	×	×	✓	✓	✓	✓	✓
Linear time trends	×	×	×	×	✓	✓	×
Quadratic time trends	×	×	×	×	×	✓	×
Flexible time trends	×	×	×	×	×	×	✓
Observations	158,500	158,500	158,500	158,500	158,500	158,500	158,500
R ²	0.004	0.010	0.014	0.018	0.020	0.021	
Adjusted R ²	0.004	0.010	0.013	0.016	0.018	0.019	0.021

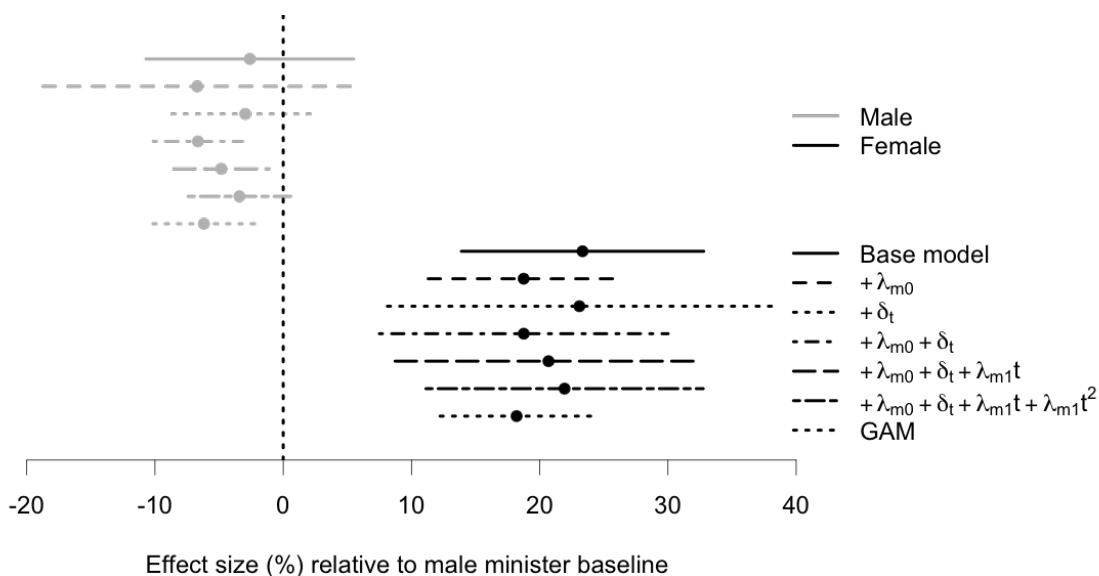
NOTE: Models 1-6 present OLS regressions for ministerial responses, and model 7 presents the results of the GAM. Regression coefficients are shown with cluster-robust standard errors in parentheses (clustered on ministry). *p<0.1; **p<0.05; ***p<0.01

statistically indistinguishable from zero. However, when the first speaker is a woman, then the gender of the responding minister matters.²⁵

That female MPs receive systematically different responses from male and female ministers helps to explain the increase in influence of female MPs detailed in the previous section. Female MPs become more influential in parliamentary debate (the language that they use in debate is adopted more often in subsequent speeches) after the appointment of a female minister, and this effect is at least partially driven by higher levels of responsiveness of the female minister herself. This may also explain the increase in the participation of female MPs in debate, as higher levels of ministerial responsiveness indicate that the concerns of female MPs are receiving more attention from powerful government figures, and send a signal that the issues that female MPs raise are worthy of governmental concern.

²⁵As before, F-tests comparing the full models in table 3 to restricted models that do not include either of the interacted variables show clearly that the interacted variables are also jointly significant.

Figure 4: Marginal effect of a female minister on responsiveness



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

ALTERNATIVE EXPLANATIONS

Differential responsiveness is not the only mechanism through which leadership effects might operate. Female ministers may have purely symbolic ‘role-model’ effects which impact the behaviour of other women. Such effects are very difficult to study empirically as they rely on essentially unobservable signals that result from the promotion of a woman to high office. Definitively adjudicating between the role-model and responsiveness mechanisms is therefore difficult. Nevertheless, in appendix section S6, I show that the effects of female leadership on the participation of other female MPs seem to be confined to the periods in which the female minister holds office. In particular, there is little evidence that – at least in the first six months after a female cabinet minister leaves office – the increased participation of female MPs persists in future debates led by male ministers.

These findings seem more consistent with the responsiveness mechanism that I articulate above, and less so with a role-model effect. If female leaders weaken historically constructed stereotypes about women in politics, then such effects should not be expected to disappear

once the female leader leaves office. On the other hand, if increased participation is due to female ministers behaving differently from their male colleagues, then the effects of female leadership are more likely to be closely tied to the exact time periods in which the female leaders hold office. The findings in section S6 therefore provide some additional support to the argument that female ministers' responsiveness may be key to affecting the participation and influence of female MPs in debate. However, these mechanisms are not mutually exclusive, and whether the results above stem purely from the "responsiveness" mechanism I outline, or are in part attributable to female cabinet ministers acting as role-models to other female MPs, they remain consistent with the idea that female leadership amplifies the voices of other women within policymaking.

More concerning is the possibility that the reasons for the increase in female participation and influence are completely distinct from the hypothesised leadership effects. I consider two alternative explanations here.

First, if appointing women to visible positions confers a political advantage to the governing party, opposition parties may respond by strategically appointing a woman to lead the competing *shadow* ministry. If this is the case, the documented effects may be due to the fact that institutional rules give both ministers and shadow ministers more time to speak on the House floor than other MPs. I investigate this hypothesis in appendix section S6 by analysing whether the appointment of a female shadow minister is positively associated with the presence of a female cabinet minister. I find little empirical support for such an argument. In addition, table S8 in the appendix re-runs the main analysis when excluding speeches made by shadow ministers. The results are very similar to those from the full sample. In combination, these tests suggest that it is unlikely that the effects above are driven by the strategic appointment of female shadow ministers.²⁶

Second, ministerial positions come with significant agenda-setting powers, and ministers determine the substance of legislation deriving from their ministries. One possible

²⁶Do female *shadow* ministers also have motivational effects for other women? In appendix section S6, I present results that are consistent with the main argument I make here: female *shadow* cabinet ministers also increase the participation of other female MPs.

explanation for the increase in female participation and influence is that female ministers may propose legislation that focuses on topics which are traditionally of greater interest to women. In section S14 of the appendix, I examine whether topics that are typically associated with high levels of female participation become more prevalent when a female minister is appointed. I use statistical topic models to estimate which legislative topics are associated with high levels of female participation under male ministers, and then assess the degree to which these topics increase when a female minister takes office. I find no evidence that female ministers are disproportionately introducing legislation that is traditionally associated with high levels of female participation. While the analysis in section S14 does not rule out the possibility that female cabinet ministers frame certain topics differently to men, it does indicate that the appointment of a female minister is not associated with an overall shift in the policy focus of a given ministry.

CONCLUSION

Legislatures are hierarchical institutions in which some actors have access to positions which confer important powers to the office holder. While the general consequences of these institutional powers are well-studied, less has been written about the implications of female occupation of such roles for the representation of women. Cabinet posts, committee chairs, and other high-profile legislative offices are normally marked by high levels of visibility and prestige, and make the politicians that hold these posts natural focal points for the public, but also for other members of the legislature. The results here suggest that when women hold high-profile offices, they have significant effects on legislative behaviour, and, crucially, that their appointment can increase the voice of other women in the policy process.

In particular, I show that in the context of the UK House of Commons, when a female cabinet minister is appointed, other female legislators become more active and central participants in parliamentary debate than is the case under male ministers. One interpretation of the findings presented here is that they provide evidence for a female ‘role-model’ effect in a legislative setting. It is plausible that, given that women have been underrepresented in cabinet minister positions both in the UK and cross-nationally, the women who are pro-

moted to these positions are seen as exemplars of success, and that their presence acts as a motivation and inspiration to their junior female colleagues. However, I also demonstrate that, beyond their mere presence, female leaders display distinctive communicative styles in debate, suggesting that the effects that such leaders have on other women may in part be driven by the ways that they interact with MPs during political exchanges.

While there is growing empirical evidence for the link between descriptive and substantive representation (Wängnerud, 2009), the mechanisms that connect increasing numbers of women in parliament to qualitative changes in political outcomes have not been fully articulated (Beckwith and Cowell-Meyers, 2007). One possibility is that it is not merely the number of women who gain elected office that matters for substantive representation, but also the heights to which those women rise once they have been elected. The findings here indicate a possible mechanism through which policy change may occur: female leaders promote increased participation and influence of other women in policymaking. Tracing out a full causal relationship between female leadership and policy outcomes that enhance the substantive representation of women is a difficult empirical task. However, the results here suggest that the appointment of women to high-office can have non-negligible effects on the behaviour of other legislators, and therefore provide empirical support for recent recommendations to extend the study of women’s legislative representation “from critical mass to critical actors” (Childs and Krook, 2009, 125).

The structure of political debate in Westminster is different from other legislatures, and so further study is required to establish whether these gender-based leadership effects hold elsewhere. In addition, future work should also consider the potential for legislative leadership effects for other disadvantaged groups. Historically, political elites have disproportionately shared characteristics of the dominant groups in society, and several groups remain significantly underrepresented in the policy process. It would be profitable in the US case, for example, to examine whether the elevation of African-American members to senior positions in the Congressional hierarchy is associated with a concomitant increase in the participation and influence of black legislators in policymaking.

Finally, a growing formal literature examines the consequences of leaders' communication strategies in collective decision making but the empirical literature on communication and leadership has lagged behind (Ahlquist and Levi, 2011). In part, this is due to the difficulty of operationalising reliable measures of spoken communication and establishing credible identification strategies that isolate the effects of leaders in observational settings. This paper makes progress on both fronts. First, the identification strategy I employ suggests that by exploiting variation over time in the identity of political leaders, it is possible to estimate causal effects of leadership on parliamentary outcomes. Second, the measures of influence and responsiveness introduced here could be profitably applied to other questions of rhetoric and parliamentary leadership. I leave such endeavours for future work.

REFERENCES

- Ahlquist, John S and Margaret Levi. 2011. "Leadership: What it means, what it does, and what we want to know about it." *Annual Review of Political Science* 14:1–24.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Asgari, Shaki, Nilanjana Dasgupta and Nicole Gilbert Cote. 2010. "When Does Contact with Successful Ingroup Members Change Self-Stereotypes?" *Social Psychology* 41(3):203–211.
- Ban, Pamela, Justin Grimmer, Jaclyn Kaslovsky and Emily West. 2018. "A Woman's Voice in the House: Gender Composition and its Consequences in Committee Hearings." *Working Paper* .
- Barnes, Tiffany D. and Diana Z. O'Brien. 2018. "Defending the Realm: The Appointment of Female Defense Ministers Worldwide." *American Journal of Political Science* .
- Beaman, Lori A, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande and Petia Topalova. 2008. Powerful women: does exposure reduce bias? Technical report National Bureau of Economic Research.
- Beaman, Lori, Esther Duflo, Rohini Pande and Petia Topalova. 2012. "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India." *science* 335(6068):582–586.
- Beckwith, Karen and Kimberly Cowell-Meyers. 2007. "Sheer numbers: Critical representation thresholds and women's political representation." *Perspectives on politics* 5(03):553–565.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller and Akitaka Matsuo. 2018. "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software* 3(30):774.
URL: <https://quanteda.io>
- Bertrand, Marianne, Sandra E Black, Sissel Jensen and Adriana Lleras-Muney. 2014. Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway. Technical report National Bureau of Economic Research.
- Bettinger, Eric P and Bridget Terry Long. 2005. "Do faculty serve as role models? The impact of instructor gender on female students." *American Economic Review* pp. 152–157.
- Bird, Karen. 2005. "Gendering parliamentary questions." *The British Journal of Politics & International Relations* 7(3):353–370.

- Blei, David and John Lafferty. 2006. "Correlated topic models." *Advances in neural information processing systems* 18:147.
- Boothroyd, Baroness Betty. 2013.
URL: <https://www.youtube.com/watch?v=uH6-lfwUeS4>
- Brajer, Victor and Andrew Gill. 2010. "Yakity-Yak: Who Talks Back? An Email Experiment*." *Social Science Quarterly* 91(4):1007–1024.
- Bullock, John G, Donald P Green and Shang E Ha. 2010. "Yes, but what's the mechanism? (don't expect an easy answer)." *Journal of Personality and Social Psychology* 98(4):550.
- Cameron, A Colin and Douglas L Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources* 50(2):317–372.
- Campbell, David E and Christina Wolbrecht. 2006. "See Jane run: Women politicians as role models for adolescents." *The Journal of Politics* 68(2):233–247.
- Catalano, Ana. 2009. "Women acting for women? An analysis of gender and debate participation in the British House of Commons 2005–2007." *Politics & Gender* 5(01):45–68.
- Chattopadhyay, Raghavendra and Esther Duflo. 2004. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72(5):1409–1443.
- Childs, Sarah. 2000. "The New Labour women MPs in the 1997 British Parliament: issues of recruitment and representation." *Women's History Review* 9(1):55–73.
- Childs, Sarah. 2004. "A feminised style of politics? Women MPs in the House of Commons." *The British Journal of Politics & International Relations* 6(1):3–19.
- Childs, Sarah and Mona Lena Krook. 2009. "Analysing women's substantive representation: From critical mass to critical actors." *Government and Opposition* 44(2):125–145.
- Csardi, Gabor and Tamas Nepusz. 2006. "The igraph software package for complex network research." *InterJournal, Complex Systems* p. 1695.
- Dasgupta, Nilanjana and Shaki Asgari. 2004. "Seeing is believing: Exposure to counter-stereotypic women leaders and its effect on the malleability of automatic gender stereotyping." *Journal of Experimental Social Psychology* 40(5):642–658.
- Dietrich, Bryce J., Ryan D. Enos and Maya Sen. 2017. "Gender Dynamics in Elite Political Contexts: Evidence from Supreme Court Oral Arguments." *Working Paper* .
- Eagly, Alice H and Blair T Johnson. 1990. "Gender and leadership style: A meta-analysis." *Psychological bulletin* 108(2):233.

- Erkan, Günes and Dragomir R Radev. 2004. “LexRank: Graph-based lexical centrality as salience in text summarization.” *Journal of Artificial Intelligence Research* pp. 457–479.
- Escobar-Lemmon, Maria and Michelle M Taylor-Robinson. 2009. “Getting to the top: Career paths of women in Latin American cabinets.” *Political Research Quarterly* .
- Fader, Anthony, Dragomir R Radev, Michael H Crespín, Burt L Monroe, Kevin M Quinn and Michael Colaresi. 2007. MavenRank: Identifying Influential Members of the US Senate Using Lexical Centrality. In *EMNLP-CoNLL*. pp. 658–666.
- Gilardi, Fabrizio. 2015. “The Temporary Importance of Role Models for Women’s Political Representation.” *American Journal of Political Science* .
- Grimmer, Justin and Brandon M Stewart. 2013. “Text as data: The promise and pitfalls of automatic content analysis methods for political texts.” *Political Analysis* 21(3):267–297.
- Hannah, Annette and Tamar Murachver. 1999. “Gender and conversational style as predictors of conversational behavior.” *Journal of Language and Social Psychology* 18(2):153–174.
- Hannah, Annette and Tamar Murachver. 2007. “Gender preferential responses to speech.” *Journal of Language and Social Psychology* 26(3):274–290.
- Hawkesworth, Mary. 2003. “Congressional enactments of race–gender: Toward a theory of raced–gendered institutions.” *American Political Science Review* 97(04):529–550.
- Heath, Roseanna Michelle, Leslie A Schwindt-Bayer and Michelle M Taylor-Robinson. 2005. “Women on the sidelines: Women’s representation on committees in Latin American legislatures.” *American Journal of Political Science* 49(2):420–436.
- Humphreys, Macartan, William A Masters and Martin E Sandbu. 2006. “The role of leaders in democratic deliberations: results from a field experiment in São Tomé and Príncipe.” *World Politics* 58(04):583–622.
- Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. “Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies.” *American Political Science Review* 105(04):765–789.
- Karakowsky, Leonard, Kenneth McBey and Diane L Miller. 2004. “Gender, perceived competence, and power displays examining verbal interruptions in a group context.” *Small Group Research* 35(4):407–439.
- Karpowitz, Christopher F. and Tali Mendelberg. 2014. *The Silent Sex*. Princeton University Press.
- Karpowitz, Christopher F, Tali Mendelberg and Lee Shaker. 2012. “Gender inequality in deliberative participation.” *American Political Science Review* 106(03):533–547.

- Kathlene, Lyn. 1994. "Power and Influence in State Legislative Policymaking: The Interaction of Gender and Position in Committee Hearing Debates." *American Political Science Review* 88(03):560–576.
- Kleinberg, Jon M. 1999. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46(5):604–632.
- Krook, Mona Lena and Diana Z O'Brien. 2012. "All the president's men? The appointment of female cabinet ministers worldwide." *The Journal of Politics* 74(03):840–855.
- Lovenduski, Joni. 2005. *Feminizing politics*. Polity.
- Mansbridge, Jane. 1999. "Should blacks represent blacks and women represent women? A contingent "yes"." *The Journal of politics* 61(03):628–657.
- Mendelberg, Tali, Christopher F Karpowitz and J Baxter Oliphant. 2014. "Gender inequality in deliberation: Unpacking the black box of interaction." *Perspectives on Politics* 12(01):18–44.
- Mendelberg, Tali, Christopher F Karpowitz and Nicholas Goedert. 2014. "Does Descriptive Representation Facilitate Women's Distinctive Voice? How Gender Composition and Decision Rules Affect Deliberation." *American Journal of Political Science* 58(2):291–306.
- Mihalcea, Rada. 2004. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*. Association for Computational Linguistics p. 20.
- Nixon, Lucia A and Michael D Robinson. 1999. "The educational attainment of young women: Role model effects of female high school faculty." *Demography* 36(2):185–194.
- O'Brien, Diana Z. 2015. "Rising to the Top: Gender, Political Performance, and Party Leadership in Parliamentary Democracies." *American Journal of Political Science* .
- Page, Lawrence, Sergey Brin, Rajeev Motwani and Terry Winograd. 1999. "The PageRank citation ranking: bringing order to the Web."
- Pearson, Kathryn and Logan Dancey. 2011. "Speaking for the underrepresented in the House of Representatives: Voicing women's interests in a partisan era." *Politics & Gender* 7(04):493–519.
- Sanders, Lynn M. 1997. "Against deliberation." *Political theory* 25(3):347–376.
- Sapiro, Virginia. 1981. "Research frontier essay: When are interests interesting? The problem of political representation of women." *American Political Science Review* 75(03):701–716.

- Studlar, Donley T and Gary F Moncrief. 1999. "Women's work? The distribution and prestige of portfolios in the Canadian provinces." *Governance* 12(4):379–395.
- Thomson, Rob, Tamar Murachver and James Green. 2001. "Where is the gender in gendered language?" *Psychological Science* 12(2):171–175.
- Wang, Mingzhu and Elisabeth Kelan. 2013. "The gender quota and female leadership: Effects of the Norwegian gender quota on board chairs and CEOs." *Journal of business ethics* 117(3):449–466.
- Wängnerud, Lena. 2009. "Women in parliaments: Descriptive and substantive representation." *Annual Review of Political Science* 12:51–69.
- Wolbrecht, Christina and David E Campbell. 2007. "Leading by example: Female members of parliament as political role models." *American Journal of Political Science* 51(4):921–939.

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.

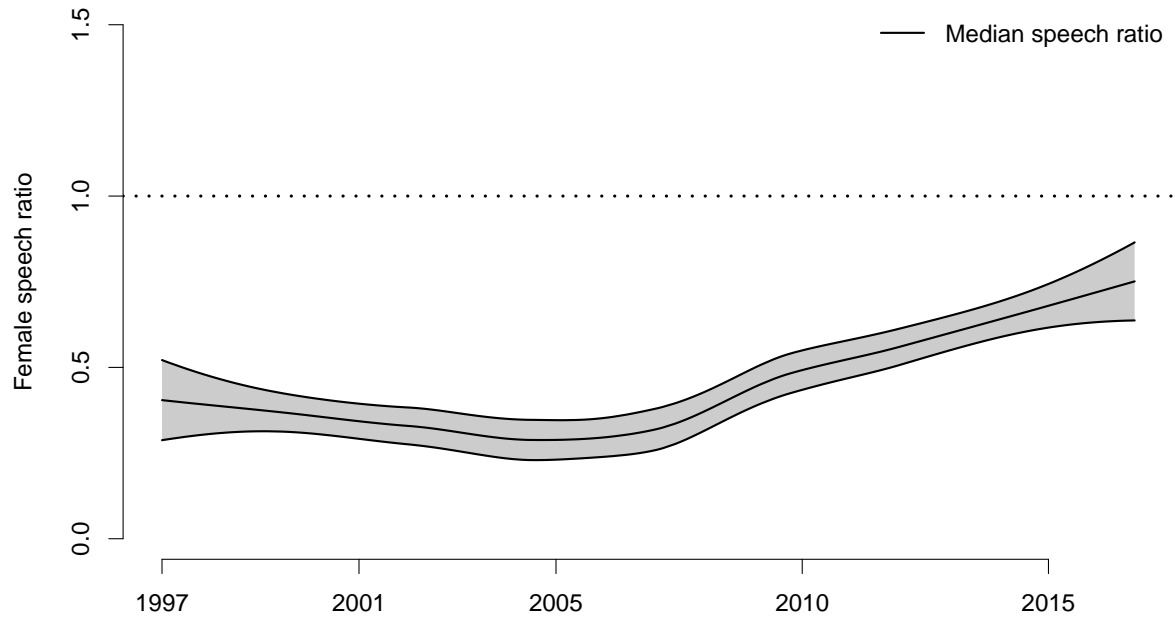
Figure S1 demonstrates how the median female speech ratio has changed over the 20 years of the study period. The solid black line is a loess curve of the median female speech ratio in each calendar month. The plot demonstrates that, even conditional on the fact that they are underrepresented numerically 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.

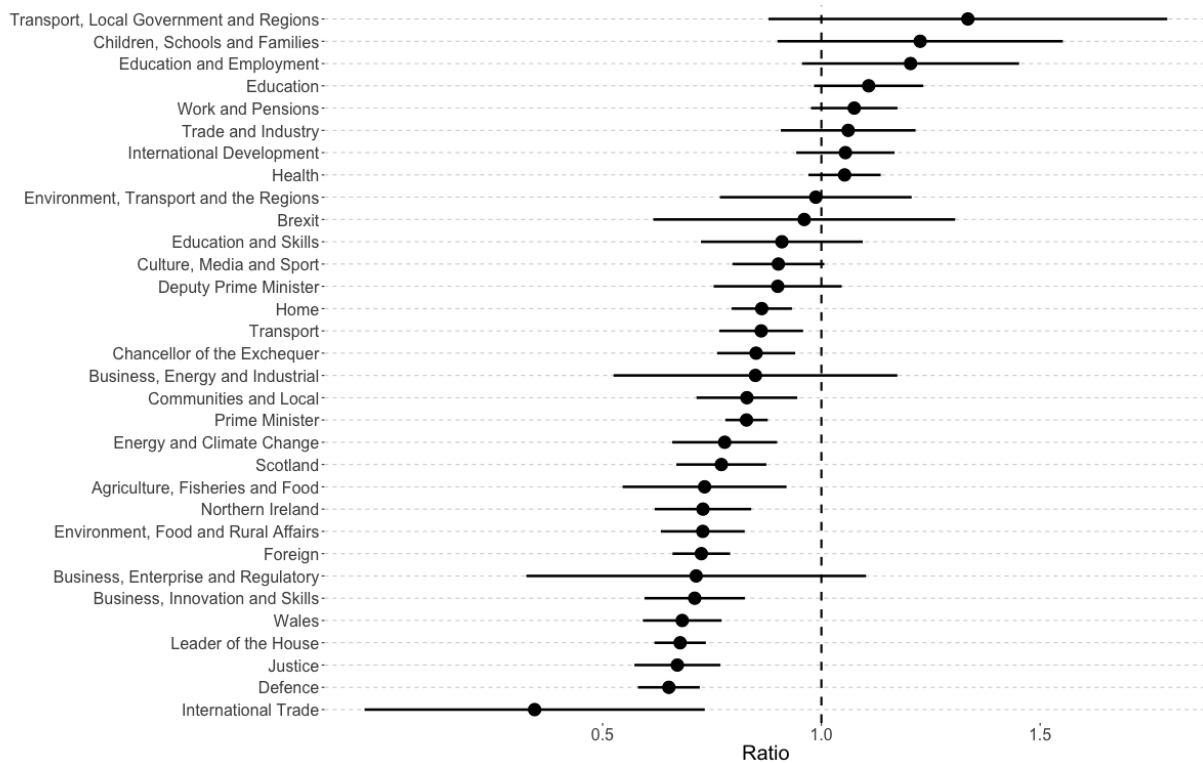
Figure S1: Female speech ratio, over time



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

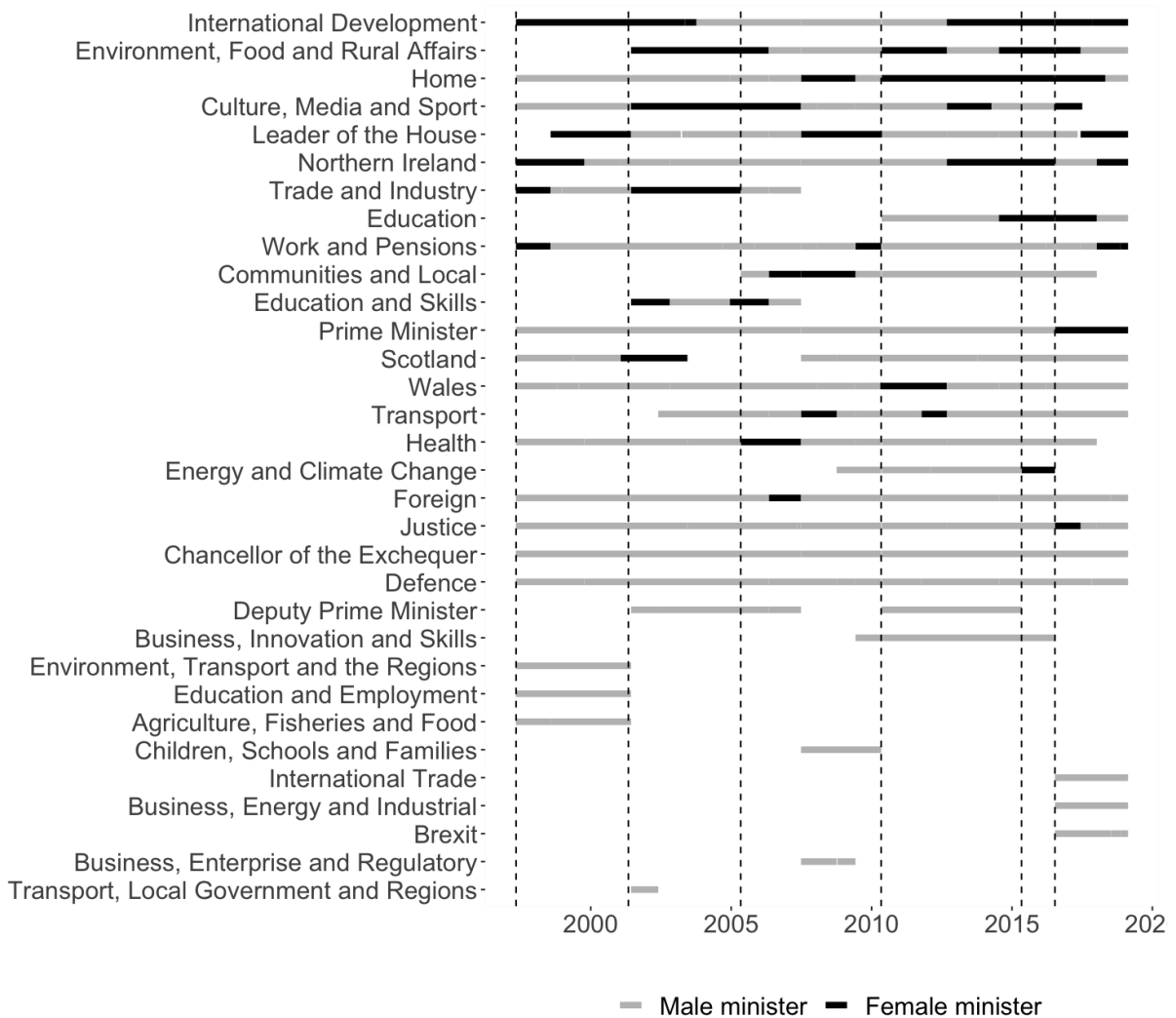


NOTE: The figure shows the average female speech ratio as defined in equation S1 for each ministry, pooled across all debates in the data. It is clear from the figure that some ministries are subject to greater levels of female participation than others.

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.

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

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

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

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

	<i>RatioWordsWomen</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.256*** (0.074)	0.253*** (0.068)	0.195*** (0.069)	0.188*** (0.053)	0.174*** (0.054)	0.174*** (0.047)	0.177*** (0.036)
Constant	0.806*** (0.045)	0.541** (0.262)	0.561*** (0.182)	0.227 (0.280)	0.264 (10.339)	5.517 (560.509)	0.811 (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	[14,50]	[15,48]	[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)

	<i>PropSpeeches Women</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.050*** (0.016)	0.043*** (0.011)	0.043*** (0.016)	0.034*** (0.008)	0.032*** (0.009)	0.032*** (0.008)	0.032*** (0.007)
Constant	0.179*** (0.009)	0.092* (0.052)	0.122* (0.071)	0.052 (0.052)	0.048 (2.559)	-1.740 (173.136)	0.555 (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	[11,45]	[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)

	<i>RatioSpeechesWomen</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.201*** (0.055)	0.198*** (0.054)	0.162*** (0.044)	0.156*** (0.037)	0.146*** (0.043)	0.144*** (0.038)	0.149*** (0.032)
Constant	0.819*** (0.036)	0.525* (0.283)	0.679*** (0.127)	0.299 (0.263)	0.275 (8.380)	-1.067 (599.295)	3.461 (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	[11,38]	[11,37]	[9,30]	[10,28]	[8,28]	[8,27]	[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

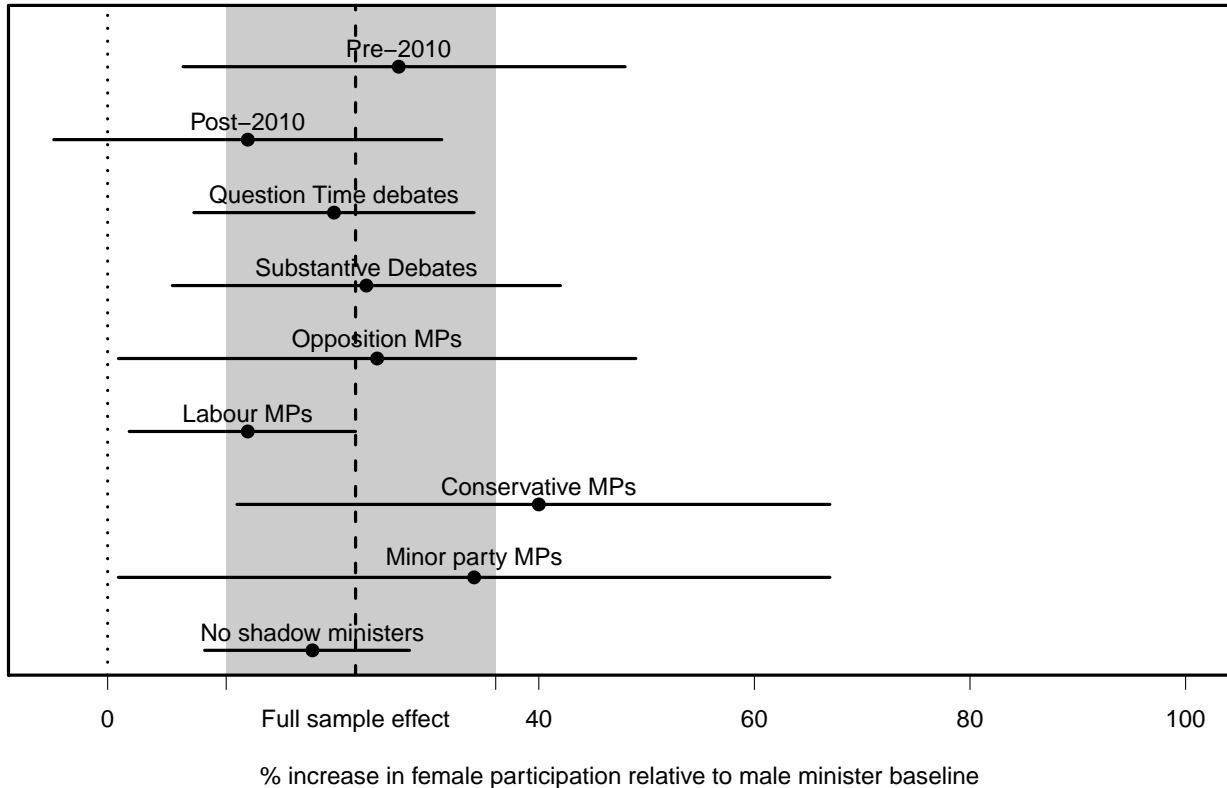
S4. SUBSET ANALYSES

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 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)

<i>Prop Words Women</i> (Labour MPs)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.059** (0.025)	0.050*** (0.017)	0.047 (0.030)	0.033** (0.015)	0.030*** (0.010)	0.024* (0.013)	0.021 (0.013)
Constant	0.263*** (0.012)	0.163*** (0.059)	0.198* (0.104)	0.137* (0.074)	0.094 (3.695)	10.657 (447.936)	1.012 (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	[4,41]	[6,32]	[-5,40]	[1,24]	[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)

	<i>Prop Words Women</i> (Conservative MPs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.052*** (0.015)	0.051*** (0.016)	0.046*** (0.012)	0.044*** (0.016)	0.045** (0.018)	0.048*** (0.018)	0.028*** (0.010)
Constant	0.112*** (0.009)	0.091 (0.074)	0.063* (0.037)	0.029 (0.061)	0.112 (4.837)	-3.305 (190.938)	-0.076 (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	[21,73]	[17,74]	[20,62]	[12,67]	[8,72]	[11,75]	[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 ?? 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

	<i>Prop Words Women</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.061*** (0.020)	0.055*** (0.014)	0.050** (0.020)	0.041*** (0.012)	0.038*** (0.012)	0.039*** (0.011)	0.044*** (0.008)
0-1 month after	0.018 (0.020)	0.007 (0.023)	0.022 (0.018)	0.006 (0.021)	-0.006 (0.020)	0.003 (0.018)	0.016 (0.021)
1-2 months after	0.015 (0.050)	0.036 (0.059)	0.020 (0.047)	0.028 (0.054)	0.024 (0.054)	0.024 (0.054)	0.040 (0.039)
2-3 months after	0.099** (0.047)	0.094* (0.049)	0.084** (0.037)	0.069 (0.050)	0.067 (0.056)	0.065 (0.056)	0.085** (0.035)
3-4 months after	0.077** (0.036)	0.049 (0.034)	0.065* (0.036)	0.033 (0.035)	0.024 (0.035)	0.028 (0.031)	0.042 (0.027)
4-5 months after	0.059 (0.037)	0.020 (0.031)	0.037 (0.032)	0.003 (0.028)	-0.006 (0.029)	-0.002 (0.029)	0.016 (0.027)
5-6 months after	0.175*** (0.011)	0.094* (0.052)	0.100 (0.082)	0.039 (0.051)	-2.214 (338.719)	1.173 (173.850)	0.056 (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	[12,56]	[15,47]	[6,51]	[10,37]	[8,35]	[9,34]	[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:

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

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 (0.073)	0.093 (0.088)
Constant	0.257 (0.037)	0.000 (0.316)
Ministry FEs	No	Yes
Observations	204	204
R ²	0.015	0.147
<i>Note:</i>	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

<i>PropWordsWomen</i> (Opposition ministers excluded)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.042** (0.017)	0.036*** (0.012)	0.041*** (0.015)	0.033*** (0.009)	0.029*** (0.009)	0.028*** (0.008)	0.024*** (0.007)
Constant	0.174*** (0.009)	0.099** (0.044)	0.113 (0.078)	0.050 (0.051)	0.010 (3.065)	0.693 (151.781)	0.253 (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	[5,44]	[7,34]	[6,41]	[9,29]	[6,27]	[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

	<i>Prop Words Women</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female shadow minister	0.046*** (0.016)	0.037* (0.019)	0.048*** (0.017)	0.042** (0.019)	0.040* (0.022)	0.032 (0.022)	0.039*** (0.015)
Constant	0.175*** (0.014)	0.390*** (0.134)	0.168*** (0.014)	0.417*** (0.155)	0.338* (0.179)	-1.695 (3.687)	-0.339 (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	[8,44]	[0,43]	[8,46]	[3,45]	[-2,47]	[-7,43]	[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)

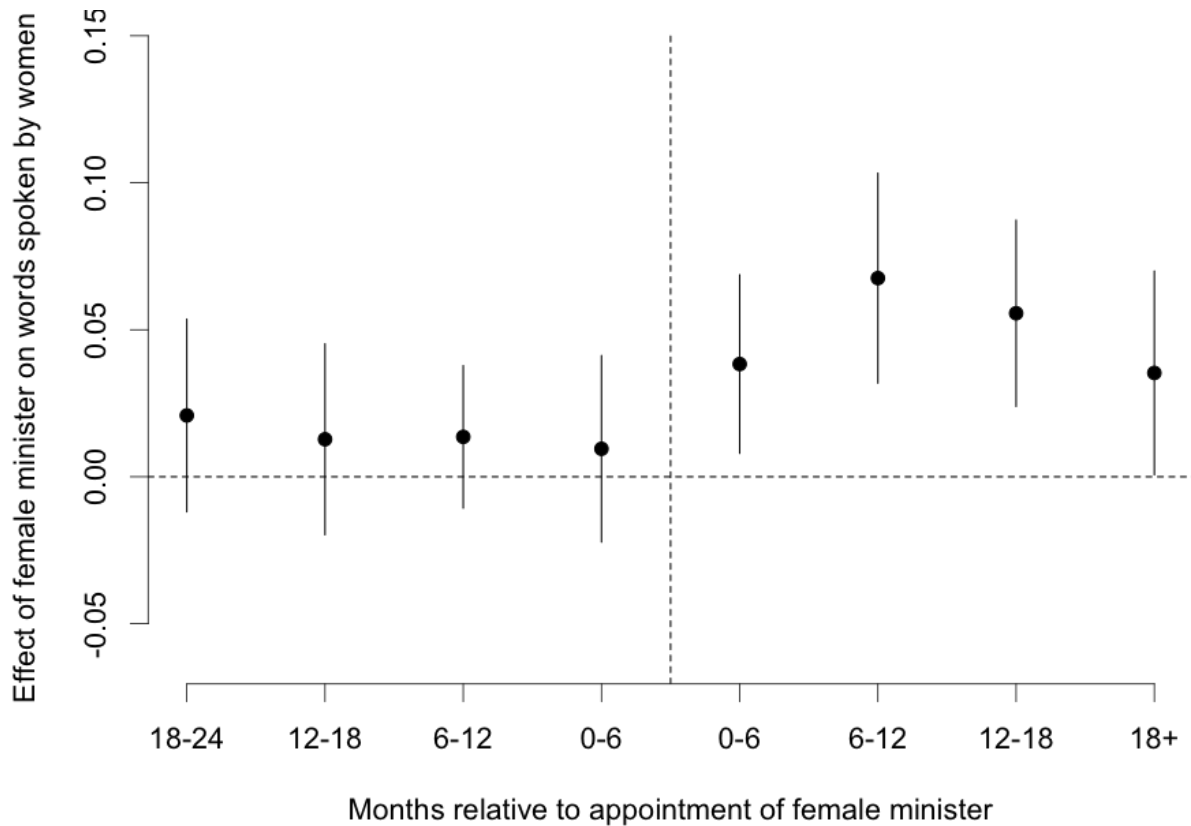
	<i>PropWordsWomen</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.061*** (0.018)	0.055*** (0.014)	0.049** (0.019)	0.041*** (0.012)	0.037*** (0.012)	0.037*** (0.011)	0.039*** (0.008)
Female Speaker	-0.043*** (0.009)	0.0001 (0.007)	-0.037*** (0.008)	-0.001 (0.006)	-0.004 (0.006)	-0.005 (0.006)	-0.005 (0.008)
Constant	0.184*** (0.011)	0.094* (0.050)	0.125* (0.073)	0.040 (0.051)	0.050 (3.222)	1.264 (155.882)	0.061 (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]	[16,46]	[6,49]	[10,37]	[8,34]	[9,33]	[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:

$$v_{ws} = tf_{ws} * \log\left(\frac{N}{n_w}\right) \quad (\text{S4})$$

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:

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

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 $\text{sim}(v_i, v_j)$ 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 is distributed to the speeches that it references:

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

Where $p(i)$ is the influence of speech i , $\text{adj}(i)$ is the set of speeches that have edges with i , and $\text{deg}(j)$ 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 ($\text{adj}(i)$), when the influence of the referencing speeches ($p(j)$) increases, and when the referencing speeches reference relatively few other speeches ($\text{deg}(j)$). 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:

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

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

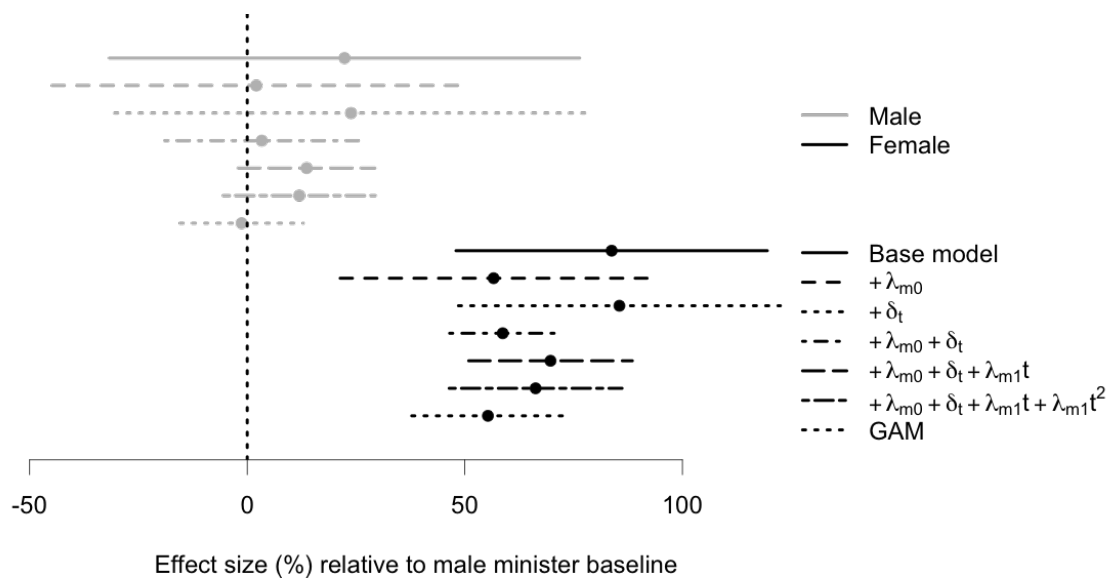
This section replicates the analysis in table 2 and figure 3 estimating P in equation S7 using Kleinberg’s (1999) HITS algorithm. As table S11 and figure S6 show, the results are substantively and statistically very similar regardless of the estimation approach.

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

	<i>Influence</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.191*** (0.030)	0.211*** (0.011)	0.374*** (0.041)	0.290*** (0.024)	0.324*** (0.034)	-3.139* (1.815)	3.662 (6.335)
Female minister	0.024 (0.029)	0.002 (0.025)	0.025 (0.029)	0.004 (0.012)	0.014* (0.008)	0.013 (0.009)	-0.001 (0.008)
Female MP	-0.012 (0.008)	-0.010 (0.008)	0.002 (0.008)	-0.001 (0.008)	-0.001 (0.008)	-0.0001 (0.008)	-0.007* (0.004)
Interaction	0.070*** (0.018)	0.061*** (0.012)	0.070*** (0.018)	0.062*** (0.013)	0.063*** (0.013)	0.061*** (0.013)	0.063*** (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

	<i>Influence (male MPs)</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.003 (0.005)	0.001 (0.003)	0.003 (0.005)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	-0.005*** (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

	<i>Influence (female MPs)</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female minister	0.014*** (0.005)	0.008*** (0.003)	0.016*** (0.005)	0.010*** (0.003)	0.008* (0.004)	0.007** (0.004)	0.005** (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 (1.034)	-1.276 (1.725)	-0.778 (0.760)	-1.295*** (0.433)	-0.797 (0.507)	-0.686 (0.532)	-1.366** (0.533)
Constant	27.066*** (0.855)	27.105*** (0.129)	33.333*** (6.097)	29.309*** (5.308)	29.614*** (5.107)	297.810** (119.837)	328.162 (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*** (1.206)	3.580*** (0.871)	5.598*** (1.761)	3.629** (1.616)	3.366*** (1.057)	4.709*** (1.202)	4.026*** (1.158)
Constant	26.207*** (2.504)	22.428*** (1.650)	38.575*** (12.888)	30.386** (13.645)	40.320*** (12.909)	-70.378 (153.288)	-6.208 (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

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

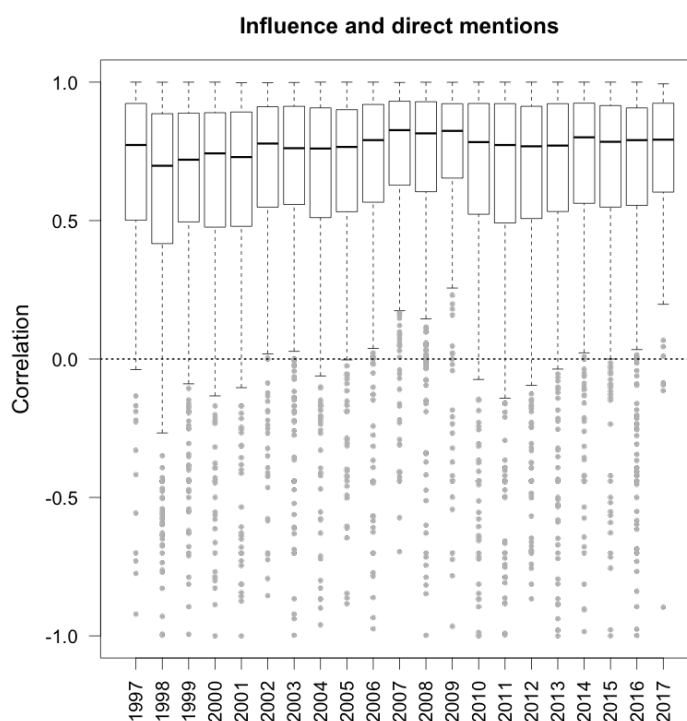
NOTE: OLS regressions where the outcome variable is *influence* as defined in equation S7, and the independent variables are “Minister” – an indicator that is equal to one when a speech is given by a government minister – and “Speaker” – an indicator that is equal to one when a speech is given by the 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.

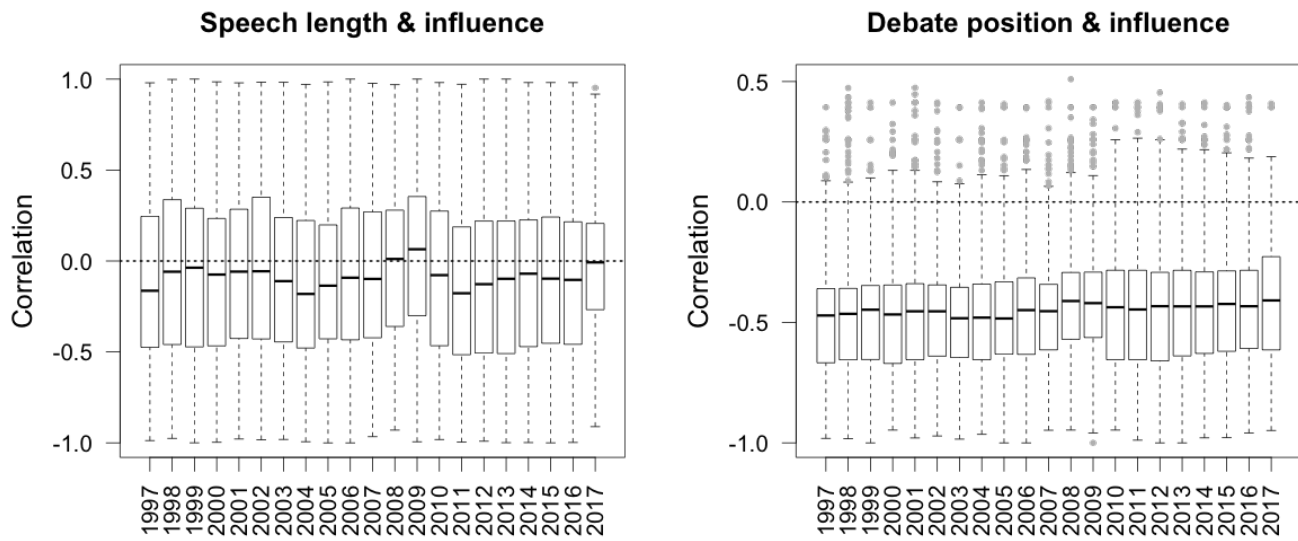
Deane” and so on. Constituency names are unique to each MP, and by searching for constituencies in the speech-texts it is possible to construct a count for the number of times any particular MP is directly mentioned by any other MP during the course of a debate. It seems clear that an MP who is directly 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

	<i>res</i>
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

	<i>res</i>
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)

	<i>Influence</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.047*** (0.004)	0.047*** (0.011)	0.055*** (0.018)	0.032 (0.022)	0.050 (0.460)	-0.717 (1.104)	-0.567 (1.077)
Female minister	0.003 (0.006)	0.001 (0.003)	0.003 (0.006)	0.001 (0.003)	0.001 (0.004)	0.0004 (0.002)	-0.002 (0.002)
Female MP	-0.001 (0.001)	-0.0004 (0.002)	0.001 (0.001)	0.001 (0.002)	0.001 (0.002)	0.001 (0.001)	0.001 (0.001)
Interaction	0.011*** (0.003)	0.009*** (0.003)	0.012*** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.009*** (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)

	<i>Responsiveness</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.629 (1.689)	-0.368 (1.633)	0.387 (1.608)	0.370 (1.558)	0.422 (1.635)	0.475 (1.601)	0.526 (0.361)
Female Minister	-0.772 (1.181)	-1.769 (1.548)	-0.873 (0.762)	-1.777*** (0.584)	-1.322** (0.611)	-0.912 (0.659)	-1.557*** (0.586)
Interaction	6.517*** (1.960)	6.445*** (1.767)	6.572*** (1.930)	6.452*** (1.826)	6.492*** (1.906)	6.458*** (1.752)	6.245*** (0.808)
Constant	25.171*** (0.778)	25.055*** (5.439)	32.456*** (6.551)	28.292*** (6.168)	29.904 (119.128)	214.074 (13,182.630)	255.959 (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 $T \times D$ matrix of topic proportions that describe the fraction of each ministerial statement $d \in \{1, 2, \dots, D\}$ that is from each topic $t \in \{1, 2, \dots, T\}$. 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.

$$Y_d = b_{k1}\theta_{k1d} + b_{k2}\theta_{k2d} + \dots + b_{kT}\theta_{kTd} + \epsilon_d \quad (\text{S8})$$

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:

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

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:

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

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

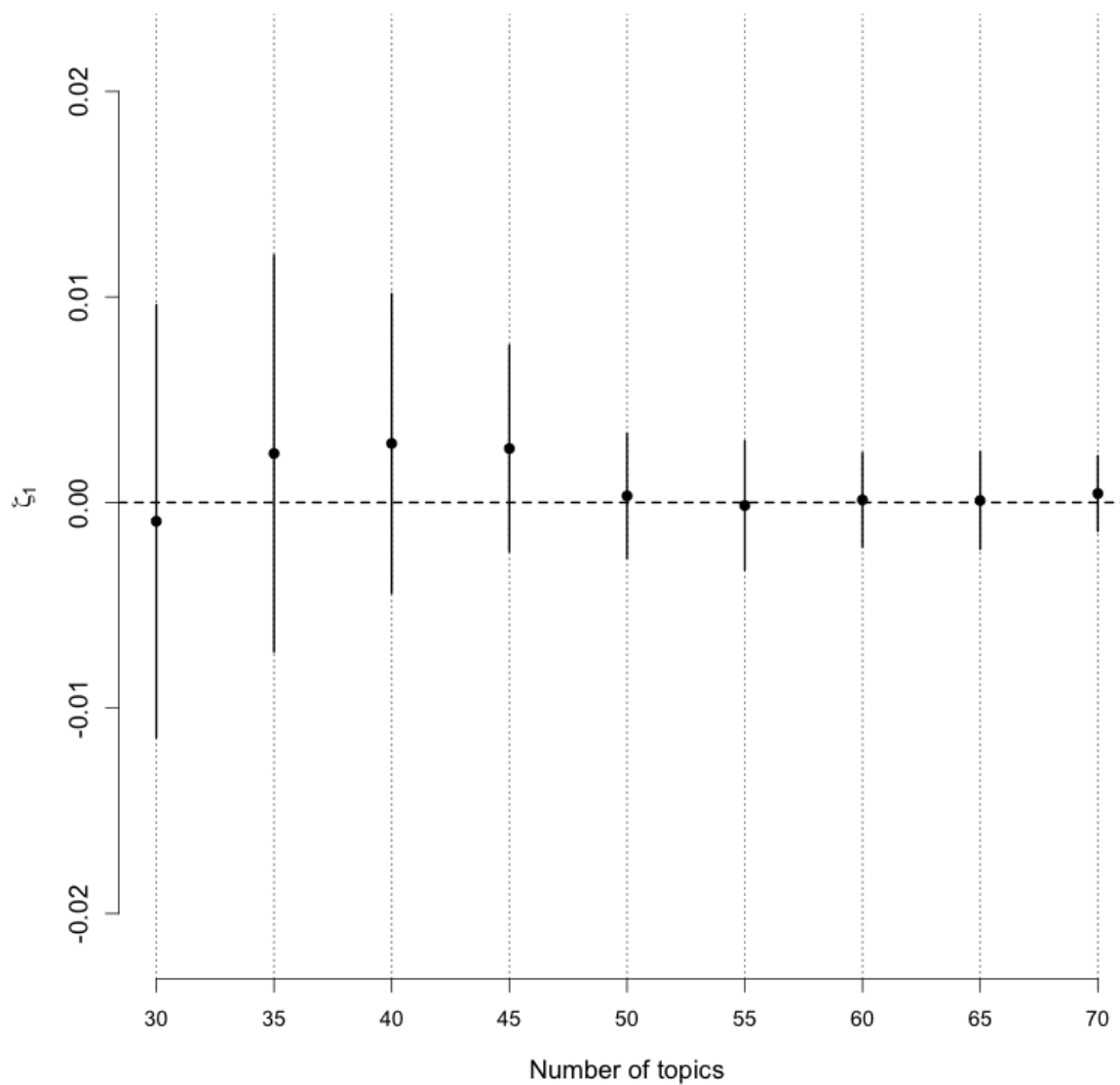
Topic	beta
labour_deficit_budget_left_coalit_shadow_let	1.116
doctor_hospit_trust_nhs_junior_patient_staff	0.880
school_teacher_pupil_educ_student_teach_children	0.723
pension_disabl_credit_child_incom_women_wage	0.678
eight_sport_olymp_game_lotteri_footbal_art	0.640
amend_bill_claus_legisl_provis_draft_order	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.