Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economic Working Paper
Series Working Paper No. 1769

Politician-citizen interactions and dynamic representation: Evidence from Twitter

Nikolas Schöll
Aina Gallego
Gaël Le Mens

January 2021

Politician-Citizen Interactions and Dynamic Representation: Evidence from Twitter[*]

Nikolas Schöll[†], Aina Gallego[‡] and Gaël Le Mens[§]

January 29 2021

Abstract

We study how politicians learn about public opinion through their regular interactions with citizens and how they respond to perceived changes. We model this process within a reinforcement learning framework: politicians talk about different policy issues, listen to feedback, and increase attention to better received issues. Because politicians are exposed to different feedback depending on their social identities, being responsive leads to divergence in issue attention over time. We apply these ideas to study the rise of gender issues. We collected 1.5 million tweets written by Spanish MPs, classified them using a deep learning algorithm, and measured feedback using retweets and likes. We find that politicians are responsive to feedback and that female politicians receive relatively more positive feedback for writing on gender issues. An analysis of mechanisms sheds light on why this happens. In the conclusion, we discuss how reinforcement learning can create unequal responsiveness, misperceptions, and polarization.

Keywords: Political responsiveness, representation, social media, gender.

Word count: 9461

[*] **Acknowledgements:** We thank Ruben Enikolopov, Maria Petrova, David Rossell, Tania Verge, Fabrizio Gilardi, Pablo Barberá, Elias Dinas, Elisa Mougin, and Cyrill Otteni, Mikhail Spektor, seminar participants at Universitat Pompeu Fabra, Wissenschaftszentrum Berlin, European University Institute, King's College London, Online Political Economy Seminar Series, Instituto de Empresa, and conference attendants at the European Political Science Association for helpful comments and suggestions. Javier Beltran, Nicolas Bicchi, Yeimy Ospina, and Guillem Pros Rius provided excellent research assistance. The research leading to these results has received financial support from the project "Local politicians: selection and performance (LEADERS)" (CSO2016-79569-P) to Aina Gallego and from the grants AEI/FEDER UE-PSI2016-75353, PID2019-105249GB-I00/AEI/10.13039/501100011033, #RYC-2014-15035 to Gaël Le Mens, funded by the Spanish Ministry for the Economy, Industry and Competitiveness. It also benefited from funding from ERC Consolidator Grant #772268 to Gaël Le Mens from the European Commission.

[†] Universitat Pompeu Fabra, IPEG and BGSE. E-mail: nikolas.schoell@upf.edu

[‡] Institut Barcelona d'Estudis Internacionals and IPEG. E-mail: agallego@ibei.org

[§] Universitat Pompeu Fabra, BGSE and UPF-BSM. E-mail: gael.le-mens@upf.edu

1 Introduction

In order to be responsive to public opinion, politicians need to track the preferences of the public, sense changes in them, and adjust their behavior accordingly. A large literature provides evidence that politicians are rather responsive: Public policy changes following shifts in public opinion (Stimson, Mackuen and Erikson, 1995; Hobolt and Klemmensen, 2008; Caughey and Warshaw, 2018; Adams, 2012) and politicians talk more about issues that are important to citizens (Jones and Baumgartner, 2004; Jones, Larsen-Price and Wilkerson, 2009; Klüver and Spoon, 2016; Wagner and Meyer, 2014). However, it is also well documented that politicians are more responsive to rich citizens (Gilens and Page, 2014; Giger, Rosset and Bernauer, 2012) and copartisans (Barberá et al., 2019), and that they misperceive public opinion with conservative politicians holding particularly biased views (Broockman and Skovron, 2018). One key mechanism that can explain both unequal influence and asymmetric misperceptions are direct interactions between politicians and self-selected citizens (Bartels, 2018; Broockman and Skovron, 2018). Empirical research about how citizens influence politicians' behavior, however, has been held back by difficulties in recording interactions between politicians and citizens and identifying how such interactions affect politicians' subsequent behavior.

This paper presents a theory of how politicians learn from and respond to their interactions with citizens and tests it with rich empirical data. We propose that politicians use a simple strategy: After talking about an issue, they observe whether the feedback from their audience is positive or negative, update their perceptions about the popularity of the issue in question, and respond by increasing attention to popular issues and decreasing attention to unpopular issues.[1] We specify this process in terms of a 'reinforcement learning' model (Holland, 1992; Sutton and Barto, 2018). Importantly, the model allows for the possibility that politicians

[1] In this paper, we focus on issue salience because, as discussed below, it facilitates empirical analysis. However, the logic applies to issue positions as well.

are exposed to different audiences and that feedback depends on their visible characteristics (Broockman, 2014). To study if differential exposure produces divergence in perceptions of issue popularity and, ultimately, in issue attention, we focus on the case of gender, a highly visible social identity that shapes the feedback politicians receive, and attention to gender issues.

Observing how individual politicians respond to citizen feedback is challenging, because doing so requires longitudinal data about what each politician talks about and the feedback he or she receives for addressing different issues. We overcome this difficulty by relying on social media data. We collected 1.5 million tweets published by elected representatives in national and regional assemblies, active during the 2016 to 2019 election cycle in Spain. We measured the reception of each tweet in terms of 'retweets' and 'likes' and use these data to estimate politicians' responsiveness to feedback. To code gender issues, we rely on 'BERT' (Devlin et al., 2018), a deep learning algorithm for natural language understanding, which takes into account word dependencies, vastly outperforms simpler bag-of-word models, and works well in multi-lingual contexts. We estimate the effect of citizen feedback on attention to gender issues by male and female politicians using two-way fixed effect panel models, which allows us to control for all factors that are constant for a given politician or for a given point in time.

We find that politicians respond to feedback from citizens: after receiving more positive feedback for tweets on gender issues, they increase attention to them. We also find that female and male politicians are exposed to different feedback environments: female politicians receive relatively more feedback for tweeting about gender issues which motivates them to talk more about gender issues. Our analyses of mechanisms reveal that female politicians obtain more feedback because they are treated differently by citizens, and not because their messages are intrinsically more appealing.

Our paper contributes to the literature on dynamic representation and issue responsiveness.

Most of the empirical research implicitly assumes that all politicians are exposed to similar information environments and measures public opinion at the macro level. For instance, research about dynamic responsiveness typically measures public opinion through survey questions aggregated over all voters (e.g. Stimson, Mackuen and Erikson, 1995) or subsets of voters (e.g. Gilens and Page, 2014). Even Barberá et al. (2019), who use rich social media data, aggregate these data across politicians to test the relationship between the issue attention of politicians and different groups of citizens. Methodologically, we extend previous research by developing an empirical approach that allows the analysis of actual interactions on social media, rather than making inferences from population-wide averages. Substantively, we document, for the first time, that interactions between politician and citizens influence the issues that politicians choose to discuss and show that unequal treatment from citizens motivates politicians with different visible social identities to diverge in issue attention. The learning mechanism we describe is psychologically realistic and can help explain several observed phenomena.

Our findings also contribute to the literature on the political representation of women. Existing theories emphasize that female representatives are more likely to talk about issues that are relevant to women because they have different experiences both in life and in office (Mansbridge, 1999; Phillips, 1995; Lovenduski and Norris, 2003), but previous empirical research has not connected specific experiences in office to attention to gender issues. We demonstrate that exposure to systematically different feedback environments contributes to differences in attention to gender issues beyond what can be explained by differences in intrinsic preferences.

2 How do politicians learn about and respond to public opinion?

In order to be responsive to their constituents, politicians first need to find out what citizens want. In standard dynamic representation theory (Stimson, Mackuen and Erikson, 1995) this step in the public opinion-policy nexus is relatively unproblematic. While this theory recognizes that politicians often do not know the preferences of the public on specific aspects of policy design, it proposes that all politicians have access to a "consensus view" about the direction of change in global types of preferences which is created by a community of politicians, journalists, academics, etc. Similarly, thermostatic models of public opinion (Wlezien, 1995, 2004) assume that politicians are aware of directional changes in public opinion.

Recent work, however, calls into question the view that politicians are accurately and homogeneously informed about public opinion and has sparked interest in the information and cognitive constraints in which they operate. Broockman and Skovron (2018) demonstrate that US politicians often misperceive public preferences, and that conservative politicians have more biased views. Arceneaux et al. (2016) argue that partisan news networks shape politicians' perceptions of constituents' preferences and document the way in which the arrival of partisan networks made elected representatives more likely to align with their parties in roll-call votes. Butler and Dynes (2016) demonstrate that politicians are biased in how they process information and systematically discount opinions from citizens with opposing views.

Interactions with citizens have long been seen as an essential source of information about public opinion, but also as a source of distortion in politicians' perceptions and, consequently, of unequal representation (e.g. Fenno, 1977; Miller and Stokes, 1963). This is because citizens who interact with politicians are not necessarily representative of the views of the public.

Broockman and Skovron (2018) provide suggestive evidence that politicians perceive public opinion to be more conservative than it really is because they are more frequently contacted by conservative activists. Bartels (2018) shows that the gap in political influence between high- and low-income citizens is reduced by one third when controlling for contact-weighted public opinion. These studies point to a key mechanism why politicians have biased and heterogeneous perceptions of public opinion. However, they do not directly examine if interactions with citizens influence politicians' behavior and the implications of being exposed to different feedback environments.

2.1 Responsiveness through reinforcement learning

This paper advances a theory about how politicians use interactions with citizens to sense changes in public mood and how they respond to perceived changes which also allows for the possibility that politicians are exposed to different feedback. We assume that when making decisions about which issues to discuss and which positions to take, politicians aim to choose popular topics and positions. This could be because they believe that consistently doing so will increase support for themselves or their parties or because they see themselves as delegates of the public.[2] However, they are uncertain about the views of the public.

In what follows, we analyse the process according to which politicians choose which policy issues to address, observe how messages are received by the public, and respond to feedback by increasing attention to issues that obtained more positive feedback than expected and reduce attention to those that obtained less positive feedback than expected. In short, we expect that issues that obtained relatively more positive feedback in the past are 'reinforced.'

[2] We note that in some conceptions of representation, such as gyroscopic or trustee representation (Mansbridge, 2003), politicians do not need to be responsive to represent the public. We also recognize that politicians sometimes have incentives or motivations to deviate from public opinion, but we assume that in general they are motivated to be responsive to citizens, as suggested by recent research which demonstrates that politicians change their votes when they receive information about the preferences of voters (see Butler, Nickerson et al., 2011).

Prior research on human and animal behavior and on adaptive systems has shown that such 'reinforcement learning' behavior is often an effective strategy in settings where an agent learns about the outcomes of available choice alternatives while simultaneously wanting to avoid negative outcomes (e.g. an animal wanting to avoid going to an area with little food, or a politician wanting to avoid saying things that generate indifference) (Holland, 1992; Sutton and Barto, 2018; Thorndike, 1927). Moreover, it has been shown that reinforcement learning is a realistic description of how people choose among activities, interaction partners, or investment styles (Denrell, 2005; Denrell and Le Mens, 2007; Malmendier and Nagel, 2011).

Yet, reinforcement learning can give rise to systematic judgment biases (for a review, see Denrell and Le Mens, 2011). What politicians learn about public opinion depends on the nature of the feedback they receive. If politicians operate in different feedback environments, they will develop different beliefs about public opinion. For example, politicians of different social identities might receive systematically different feedback for messages on the same issue because citizens self-select into interacting with politicians of shared social identity (Broockman, 2014; Gay, 2007). Politicians may also be exposed to different feedback environments if citizens expect politicians to talk about issues congruent with their visible social identity and reward those who do so with more positive feedback. Even if politicians suspect that the feedback they receive is unrepresentative of public opinion at large, they are unlikely to be able to fully correct for biases in their feedback environment.[3] Self-selection into giving feedback and differential treatment of politicians by audience members can lead politicians of distinct social identities to form different perceptions about public opinion. This might, in turn, lead to a divergence in issue attention.

[3] There exists evidence that when producing population estimates, people go beyond the information they obtain from their immediate social environments, yet they do not fully correct for the biases already present in their information sample (Galesic, Olsson and Rieskamp (2018), see also Fiedler (2012)).

2.2 Model

We introduce a simple reinforcement learning model to clarify our perspective on how politicians' learn from and respond to citizen feedback. The main behavioral assumption of the model is that politicians tend to repeat actions that produced positive feedback and to avoid actions that produced negative feedback (Thorndike, 1927). Depending on the context, positive and negative feedback can take the form of approving or disapproving reactions (e.g. praise or criticism from parents or media pundits) or of a larger or smaller amount of reactions (e.g. more or less applause, money, or invitations to social events). In reinforcement learning models, agents have latent 'valuations' of each option, which they update based on feedback. Positive feedback increases the valuation of the option, while negative feedback decreases it. Agents tend to choose options with higher valuation.

In our context, the options are different policy issues which the politician can choose to address. Feedback is the reaction of the public, which can be more positive or negative than expected. The valuation of different policy issues can be interpreted as politicians' perception of the popularity of that issue. Politicians are responsive to feedback if they tend to choose issues that obtained positive feedback in the past and hence are perceived as more popular.

Consider a politician i who publishes a series of messages on policy issues. Without lack of generality, we assume that there are only 'gender issues' and 'other issues' and denote them by GI and $other$. We refer to the first message by $m = 1$, the second message by $m = 2$, etc. Politician i's perception of the popularity of gender issues (valuation) at the time they decide on the issue of message m is V_{im}^{GI} and the perceived popularity of other issues is V_{im}^{other}. The politician is more likely to choose the gender issue for message m when the difference in valuations favors this issue, i.e. they perceive it as more popular. We specify the probability that the politician chooses issue k as a logistic function of the difference in valuations of the two issues. We call this quantity the ‘attention to the gender issue’:

$$A_{im}^{GI} = Logit(\pi_i^{GI} + r\Delta V_{im}), \tag{1}$$

where $\Delta V_{im} = V_{im}^{GI} - V_{im}^{other}$ is the valuation difference, r denotes the responsiveness of issue attention to perceived popularity, and π_i^{GI} characterizes the baseline tendency to write about gender issues. This latter construct can be thought of as the intrinsic motivation to address the issue.

After every message m, the politician observes the feedback FB_{im}^k and updates their valuation of the issue of the message. Following research on how people update valuation based on experience (see Denrell (2005) for a review), we assume that the new valuation of an issue is a weighted average of the previous valuation of that issue and the last feedback instance on that issue. More formally, if message m is on issue k, then

$$V_{i,m+1}^k = (1-\gamma)V_{im}^k + \gamma FB_{im}^k. \tag{2}$$

If message m is not on issue k, the valuation of issue k does not change: $V_{i,m+1}^k = V_{i,m-1}^k$.

We assume that feedback is normally distributed, with common standard deviation σ, and with means μ_i^{GI} and μ_i^{other} that differ between issues:

$$FB^{GI} \sim N(\mu_i^{GI}, \sigma); \qquad FB^{other} \sim N(\mu_i^{other}, \sigma)$$

It is possible to derive a formula for the long-run share of attention devoted to gender issues, A_∞^{GI}.[4]

$$A_\infty^{GI} = Logit(\pi_i^{GI} + r\Delta\mu_i), \tag{3}$$

[4] The proof is in Appendix SI6.

where $\Delta\mu_i = \mu_i^{GI} - \mu_i^{other}$ is the difference between the means of the feedback distributions for the two issues ('gender' and 'other'). We call this construct the 'gender issue feedback advantage.' Unsurprisingly, the long-run attention to gender issues increases with the 'gender issue feedback advantage.' This feedback effect is stronger when the issue responsiveness parameter, r, is larger.

2.3 Social identity and divergence

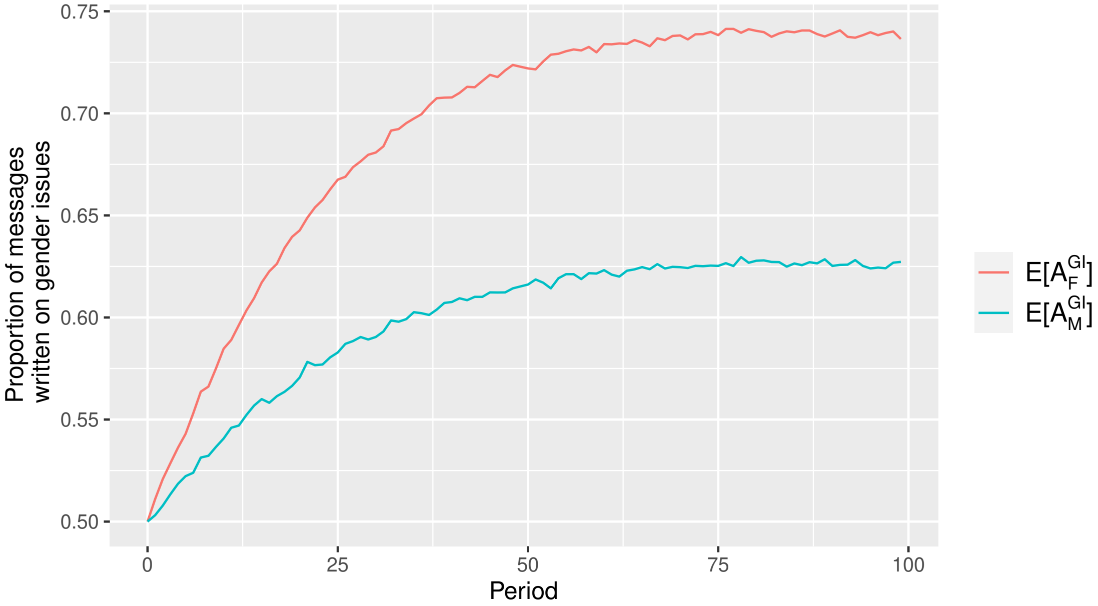
Now consider two hypothetical politicians, F and M who behave according to the reinforcement learning model. The gender issue feedback advantage potentially differs between the two politicians ($\Delta\mu_F \neq \Delta\mu_M$) if citizens provide different feedback depending on visible social identities. Using equation 3, we can derive a necessary and sufficient condition for the emergence of a feedback-driven difference in long-run issue attention such that attention to the gender issue is larger for F than for M:

$$A_{F,\infty}^{GI} > A_{M,\infty}^{GI} \iff \pi_F^{GI} + r_F\Delta\mu_F > \pi_M^{GI} + r_M\Delta\mu_M. \tag{4}$$

Feedback-driven divergence in valuations and issue attention can emerge even if male and female politicians have identical intrinsic preferences for writing about gender issues. Suppose that the baseline propensities to tweet about gender issues are the same for the two politicians ($\pi_F^{GI} = \pi_M^{GI}$) and that they are equally responsive to changes in perceived popularity of issues ($r_F = r_M$). In this case, politician F will devote a larger attention to the gender issue whenever the gender issue feedback advantage is stronger for F than for M ($\Delta\mu_F > \Delta\mu_M$).

In the general case, feedback contributes to the difference in issue attention between politicians F and M beyond what could be explained just by a difference in baseline propensities

Figure 1: Simulating the dynamics of issue attention for F and M politicians



Note: Figure based on 100,000 simulations with $\Delta\mu_F = 1$, $\mu_M = .5$, $r = 1$, $\gamma = .1$ and $\sigma = 1$.

to write about gender issues when the following condition holds:

$$r_F \Delta\mu_F > r_M \Delta\mu_M. \tag{5}$$

As an illustration, we simulate the model with a larger gender issue feedback advantage for the F politician: $\Delta\mu_F = 1 > \mu_M = .5$. We also assumed $r = 1$, $\gamma = .1$, $\sigma = 1$. To emphasize the fact that the asymmetry in feedback for messages on gender issues can be a sufficient cause for the emergence of an asymmetry in issue attention between politicians F and M, we assume that the initial valuations of the two issues are the same for the two politicians (set to 0). Initially, both politicians devote the same attention to gender issues, but as the number of periods grows, an asymmetry in issue attention emerges.

In Section 4, we report empirical estimates of all parameters of this learning model based on our Twitter data.

3 Case, data and measurement

To examine if politicians use reinforcement learning to choose which issues to discuss based on the feedback they obtain from citizens, we collected the tweets written by all politicians who served in the national parliament of Spain or any of its regional parliaments between the start and the end of the national legislature (from July 2016 to March, 2019).

Spain is a relevant case to study the rise of gender issues. Gender evolved from being a relatively niche issue into a major topic during the time covered by our study, culminating in a general strike in March 2018, which was probably the largest women's strike in history (Campillo, 2019). Spain is a fairly typical consolidated democracy. It has a proportional representation system and closed party lists. It is also a decentralized state, with regional governments holding significant powers. Therefore, both national and regional representatives are relevant for the political process. Social media use is high. We collected the user names of 1530 accounts, and estimate than more than 80% of the politicians who were in office for some time during this period had a Twitter account. They posted more than 1.5 million original tweets in this period.

The set of 'original' tweets consists of tweets politicians posted on their own wall and replies to other users' tweets. We included all tweets with at least two words published by politicians who were active Twitter users (writing on average at least one original tweet per month). We only consider the first tweet of a thread of tweets. The resulting data contains the tweets of 1265 politicians (554 females and 711 males).

In comparison to male politicians, female politicians were less active and their tweets received fewer retweets and likes (Table 1). The difference between the means and medians of the distributions of the average number of retweets and likes received by each politician reveals that some politicians receive many more retweets and likes than others. Additional summary statistics are reported in Appendix SI1.

Table 1: Female politicians post fewer tweets than male politicians and receive fewer retweets and likes.

	Female politicians	Male politicians
Number of politicians	554	711
Number of tweets (mean)	1087.9	1380.1
Number of tweets (median)	568	697
Average number of retweets (mean)	22.4	45.8
Average number of retweets (median)	6.3	7.4
Average number of likes (mean)	38.6	80.2
Average number of likes (median)	8.4	10.4
Standard deviation retweets (mean)	52.1	95.9
Standard deviation retweets (median)	10.2	12.7

Note: Data is aggregated in two steps. First, we calculate average values per politician. Second, we calculate the mean or median value of those averages for female and male politicians.

3.1 Identifying tweets on gender issues

The main empirical challenge consisted of identifying tweets related to gender issues. We recruited research assistants to code about twenty thousand tweets as being on gender issues or not (see SI4.1 for details). We used these human-coded data to train and validate a text classifier based on a state-of-the-art machine learning algorithm for language representation, the BERT model (Devlin et al., 2018). This consists of an artificial neural network with many layers (a 'deep neural network') that takes the text of a tweet as an input and labels it as being about gender issues or not.

BERT-based text classifiers offer three advantages over other classifiers. First, they perform better than 'bag-of-words' classifiers which are most often used in the social sciences (Grimmer and Stewart, 2013). By contrast to the latter, BERT is sensitive not only to word frequencies or word sequences but also to context effects. The mathematical representation of a word depends on the other words that come before and after in the text. BERT performs so well because of this sensitivity to bi-directional dependency in word meaning. Second, BERT is pre-trained on a vast amount of data (the text of all Wikipedia articles) to learn a rich language representation but can then be 'fine-tuned' for specific tasks such as classification. Most text classifiers based on machine-learning techniques are trained from scratch on a particular dataset. If the data is of limited size, performance suffers. Classifiers that are pre-trained on large amounts of text but cannot be fine-tuned are limited by the fact that word representations are not adapted to the particular task at hands (in our case identifying tweets on gender issues). Our BERT-based approaches overcomes the limitations of these two earlier approaches. Third, there exists a multi-lingual version of the BERT model that can be used with text written in more than 100 languages. This implies that it is not necessary to translate the texts before inputting them into the model. This was vital for us, as Spanish politicians regularly tweet in Spanish (Castilian), Basque, Catalan and Galician.

We fine-tuned our BERT classifier to optimize its classification performance on our data. We used 10-fold cross validation to identify the optimal training parameters.[5] Our model achieved an excellent classification performance on our test data: 90% of the tweets the model classified as being on gender issues are actually on gender issues and 79 % of gender issue tweets are classified as being on gender issues. For comparison, we trained a naïve Bayes classifier as a benchmark for more traditional 'bag-of-words' approaches. It produced three times more mistakes than our BERT classifier.[6]

3.2 Measuring issue-specific feedback

We measure citizen feedback for each tweet through reactions from Twitter users. We construct our main feedback measure based on the number of *retweets* and use *likes* for robustness checks. The advantage of retweets over likes is that we can obtain information about the feedback givers (e.g. if they are male or female). We assume that politicians interpret retweets as positive feedback. Metaxas et al. (2015) show that most retweets imply

[5] See Appendix SI4.2 for details.
[6] See Table SI4.9 in Appendix SI4.2.

an approval. Conover, Ratkiewicz and Francisco (2011) show that most retweets between politicians occur within partisan groups and thus are used as a form of support. Consistent with this assumption, Figure SI2.1 reveals that, in our data, most retweets between politicians happen within parties.

Prior research on learning from feedback has shown that outcomes have decreasing marginal effects and that agents tend to evaluate outcomes with respect to a time-dependent 'aspiration level' or reference point (Cyert and March, 1963; Greve, 1998; March and Shapira, 1992). In order to implement the assumption of decreasing marginal utility of additional feedback, we take the natural logarithm of the number of retweets obtained by tweet message m published by politician i. This also reduces the importance of extremely large instances of retweets and makes the distribution of feedback more comparable across politicians. Differences in logs express scale-invariant ratios of feedback, i.e., the added utility of receiving 10% more *retweets* would be the same for a politician who usually receives 10 or 10000 retweets. In order to incorporate time-dependent aspirations, we take out a politician-specific time trend.[7] We then proceed to within-politician z-score standardization to obtain our feedback measure. By construction, the distribution of feedback for each politician has mean zero ($E[FB_{im}] = 0$) and standard deviation one ($\sigma_{FB,i} = 1$).

The feedback measure can thus be interpreted in the following way: a one-unit increase in feedback means that the tweet received one standard deviation more in "feedback utility units" relative to other tweets written by the same politician around a similar point in time. The resulting measure is psychological meaningful and comparable across politicians.

[7] We regress $log\,\text{retweets}_{im}$ on the time t the tweet was posted using OLS and then take the residual:

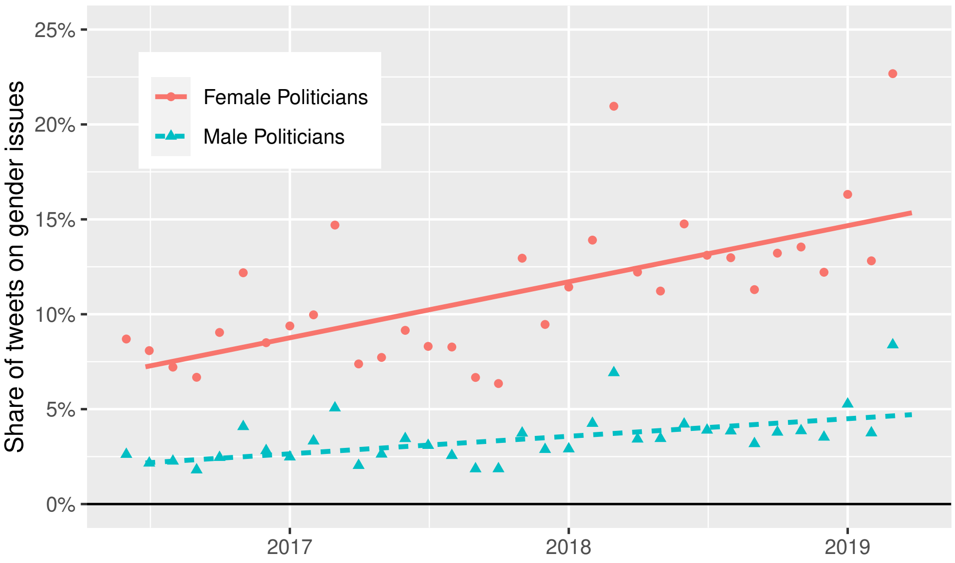
$$\widehat{u_{im}} = log\,\text{retweets}_{im} - \widehat{trend}(log\,\text{retweets}_i) * t \tag{6}$$

4 Results

4.1 Attention to gender issues by female and male politicians

We define politician i's attention to gender issues in period p as the proportion of gender issue tweets posted by this politician over that period: $A_{ip}^{GI} = \frac{n_{ip}^{GI}}{N_{ip}}$. Over the entire sample period, female politicians devoted, on average, 11.2% of their tweets to gender issues whereas male politicians only devoted 3.4% of their tweets to gender issues. Figure 2 shows that attention to gender issues increased by 2.5% for male politicians and by 8.1% for female politicians.

Figure 2: Attention to gender issues over time



Note: The lines represent linear trends based individual tweets (n=1,583,000). Points represent monthly averages.

4.2 Issue-specific feedback received by female and male politicians

Female politicians receive on average 18% more retweets for tweeting on gender issues compared to tweeting on other issues. By contrast, male politicians receive about the same number of retweets for tweeting about gender issues and other issues. Similar findings hold when focusing on the median number of retweets and the mean and median numbers of likes (see Table 2).

Table 2: Female politicians get relatively more positive feedback for posting on gender issues.

	Female Politicians (N= 554)			Male Politicians (N=711)		
	GI	other	$\Delta^{GI/other}$	*GI*	other	$\Delta^{GI/other}$
Number of tweets (mean)	123.5	966.4		49.6	1332.5	
Number of tweets (median)	46	506		23	683	
Averge number of retweets (mean)	25.5	21.6	18%	45.5	45.7	-0%
Averge number of retweets (median)	7.4	6.2	19%	7.4	7.4	0%
Averge number of likes (mean)	45.4	37.2	22%	83.9	79.9	5%
Averge number of likes (median)	9.4	8.3	13%	10.1	10.3	-2%

Note: To aggregate the data, we first calculate average values per politician and then the mean or median value of those averages for female and male politicians.

In order to confirm this finding, we analyze the gender issue feedback advantage of female and male politicians using the reference-dependent measure of feedback introduced in Section 3.2 and a set of linear regressions estimated by OLS. In our baseline specification, our measure of feedback, FB_{im}, is regressed on the social identity of the politician i and the issue of tweet message m:

$$FB_{im} = \beta_{GI} GI_{im} + \beta_M M_i + \beta_{GI*M} GI_{im} * M_i + \epsilon_{im}, \tag{7}$$

where GI_{im} is a dummy variable equal to 1 if tweet m written by politician i is on gender issue and M_i is a dummy equal to 1 if politician i is male. ϵ_{im} is an error term.

We are most interested in the coefficient of the interaction term, β_{GI*M}, which measures how the gender issue feedback advantage differs between female and male politicians. If said coefficient is negative, the gender issue feedback advantage is stronger for female politicians. In additional specifications, we include politician fixed effects in order to absorb the effect of politician characteristics which remain constant over time such as their social identity, specialization of policy area, political party, etc. We also add day and hour of the day fixed effects in order to absorb the effect of systematic temporal variations affecting all politicians and reflect the general level of activity on Twitter.

Estimation results are reported in Table 3. In all specifications, the gender issue feedback advantage is stronger for female politicians ($\beta_{GI*M} < 0$). Model 1 is a basic specification without controls or fixed-effects. We find that the gender issue feedback advantage is larger for female politicians (+0.23 standard deviation) than for male politicians (+0.14 standard deviations). The pattern remains similar when politician and day fixed effects are included (Model 2) as well as when additional time-varying control variables are included, such as the hour of the day the tweet was published, the number of tweets published by the politician on that day, and the length of the thread of the tweet (Model 3). Model 4 shows that the effect is similar for left- and right-wing politicians (see Appendix SI3.1 for details). Our main finding is robust to using likes rather than retweets to construct the feedback measure (see Table SI3.3 in Appendix SI3.1).

4.3 Responsiveness to issue-specific feedback

In this section, we estimate the parameters of the reinforcement learning model described in Section 2.2 using two-way fixed-effect logistic panel models via GLS.

To render the data amenable to panel models, we discretize it into fixed-lengths time periods p. We use months as period length since this provides a compromise between two goals: having a precise estimate of the attention given to gender issues (longer time intervals) and having more observations (shorter time intervals).[8] To estimate issue valuations (measures of politicians' perception of issue popularity), we update valuations with every tweet m and then 'freeze' the valuations at the beginning of each period to make them conform to our panel data structure, i.e. $V_{ip}^{k} = V_{im}^{k}[m = \text{first message in period}\, p]$. We take the valuation at

[8] Ancillary analyses with weeks as periods yield similar results (See Appendix SI3.2).

Table 3: Linear regressions of tweet feedback on politicians' social identity and policy issue of the tweet

Dependent variable:	Tweet-level standardized feedback, FB_{im}			
Model:	(1)	(2)	(3)	(4)
GI	0.2359***	0.2860***	0.2834***	0.2682***
	(0.0041)	(0.0203)	(0.0200)	(0.0513)
GI × Male politician	-0.0904***	-0.1221***	-0.1250***	-0.1228**
	(0.0068)	(0.0235)	(0.0232)	(0.0550)
Part of thread			0.2920*	0.2921*
			(0.1514)	(0.1513)
Tweets on day by politician			-0.0072***	-0.0072***
			(0.0015)	(0.0015)
GI × Left				0.0199
				(0.0551)
GI × Male politician × Left				0.0009
				(0.0606)
Male politician	0.0213***			
	(0.0017)			
(Intercept)	-0.0264***			
	(0.0014)			
Fixed-effects				
Politician		Yes	Yes	Yes
Day		Yes	Yes	Yes
Hour of day			Yes	Yes
Fit statistics				
Squared Correlation	0.003	0.008	0.018	0.018
Observations	1,583,917	1,583,917	1,583,917	1,583,917

Note: Estimations of variations of equation 7. Some main effects of the interactions (e.g. the main effect of politicians from a left party) are dropped due to the fixed effects. Standard error are reported in parentheses are clustered according to the fixed effects *p<0.1; **p<0.05; ***p<0.01

the beginning of the month (rather than the average valuation for example) to avoid feedback endogeneity issues. Some politicians have breaks in their Twitter activity and assuming that feedback still affects issue attention after several months does not seem realistic. Therefore, we restrict our attention to politician-month cells where the valuation of each issue was updated at least once during the previous month. Furthermore, we use the number of tweets written by the politician in the respective month (N_{ip}) as regression weights. Each tweet thus receives the same weight in our estimations.[9]

[9] Estimation results where all politician-month cells are weighted equally, independently of the number of

In accordance with the reinforcement learning model, we estimate a logistic regression of issue attention, A_{ip}^{GI}, on the difference in valuations of gender issues and other issues, ΔV_{ip}, and politician fixed effects, π_i. To account for factors that affect issue attention in our empirical setting but that, for parsimony, were left out of the formal model, we augment it with month fixed effects, τ_p, and time-varying control variables. Global shifts in issue attention over time are captured by the month fixed effect. For example, around March 8^{th}, the International Women's Day, politicians tweet more on gender issues. Beyond accounting for differences in baseline attention to gender issues, the politician fixed-effects, π_i, also capture other time-invariant confounders such as their social identity, party, region, policy focus, etc., as well as time-invariant characteristics of their followers (e.g. level of interest in gender issues).

Issue valuations are not directly observable in our data. They are latent variables constructed based on the feedback received by tweets on the issues. Therefore, the valuation updating equations have to be estimated jointly with the issue attention equation. The full model thus consists of two equations, jointly estimated as a generalized linear model:

$$\begin{cases} V_{im}^{k} = (1-\gamma)V_{i,m-1}^{k} + \gamma FB_{i,m-1}^{k} \\ A_{ip}^{GI} = Logit(\pi_i^{GI} + r * \Delta V_{ip} + \tau_p + \epsilon_{ip}). \end{cases} \tag{8}$$

Because standard software packages do not include readily available commands for the estimation of such models, we performed a grid search for the updating parameter γ. For each possible value of $\gamma \in (0, 1]$ (step size = 0.01), we construct the issue valuations and the valuation difference ΔV_{it}, estimate the parameters of the responsiveness model and select the updating parameter γ with best model fit (lowest mean squared error). The exact value of γ depends on the model specification but estimates are close to 0.07 in all cases, meaning that the issue valuation is revised by approximately 7% with each tweet on the issue.

underlying tweets are reported in Appendix SI3.2.

Estimation results are reported in Table 4. Model 1 corresponds to equation 8. The combination of a positive coefficient for the valuation difference ΔV_{it} and the positive valuation updating weight γ reveals that an increase in feedback to gender issue tweets (or a decrease in feedback for tweets on other issues) is associated with an increase in attention to gender issues. A one unit increase in the difference in valuation between gender issues and other issues is associated with an average marginal increase attention to gender issues of 7.8% (0.55 percentage points).[10] We interpret this as a substantial effect given that our fixed effect specification likely leads to conservative estimates since it focuses on within-politician, within-month variation.

In Model 2, we examine the difference in how female and male politicians learn from feedback by introducing separate valuation difference coefficients for female and male politicians. We denote by ΔV_{ip_F} the valuation difference if politician i is female and ΔV_{ip_M} if i is male. Estimates reveal that politicians of both genders are responsive to valuation differences. The weighted average marginal effect implies that an additional standard deviation in valuation difference ($+1\Delta V$) increases female politicians' attention to gender issues by 8.5% whereas male politicians' issue attention increases by 6.5%. The difference between these two estimates is not statistically significant.

Two mechanisms could explain why the valuation difference might affect issue attention. An increase in feedback for addressing gender issues could motivate politicians to talk more about them or an increase in the feedback for addressing other issues, diminishing ΔV, could crowd out attention to gender issues. We separate these two mechanisms in Model 3. We find evidence for both mechanisms, but effect sizes differ: the positive effect size for the

[10] To account for differences in the number of tweets across months, we weight for the number of tweets written in a month (N_{ip}) when calculating the average marginal effect (AME):

$$\widehat{AME} = \frac{1}{N}\sum_{i=1}^{I}\sum_{p=1}^{P} N_{ip}\big(\text{Logit}(\widehat{\pi_i^{GI}} + \widehat{r} * 1 + \widehat{\tau}_p) - \text{Logit}(\widehat{\pi_i^{GI}} + \widehat{r} * 0 + \widehat{\tau}_p)\big)$$

valuation of gender issues is larger than the negative effect size for the valuation of other issues. This suggests that crowding out is of secondary importance.

Table 4: Reinforcement learning model: politicians are responsive to feedback

Dependent Variable:	Monthly share of tweets written on GI, $A_{ip}^{GI} = \frac{n_{ip}^{GI}}{N_{ip}}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
ΔV	0.0936***					
	(0.0201)					
ΔV_F		0.1108***				
		(0.0287)				
ΔV_M		0.0677***				
		(0.0243)				
V_F^{GI}			0.1386***	0.1215***	0.1293***	0.1386***
			(0.0436)	(0.0376)	(0.0399)	(0.0437)
V_M^{GI}			0.1247***	0.1297***	0.1249***	0.1247***
			(0.0335)	(0.0304)	(0.0321)	(0.0334)
V_F^{other}			-0.0864***	-0.0363	-0.0698**	-0.0863***
			(0.0296)	(0.0279)	(0.0276)	(0.0296)
V_M^{other}			-0.0235	-0.0210	-0.0235	-0.0236
			(0.0337)	(0.0339)	(0.0333)	(0.0338)
Indiv. trend				5.518***		
				(0.4425)		
Lagged DV					1.011***	
					(0.1013)	
Social Influence						-0.0449
						(0.8642)
$\widehat{\gamma}$ (to calc. valuation)	0.07	0.07	0.07	0.07	0.07	0.07
Fixed-effects						
Politician	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Squared Correlation	0.573	0.573	0.573	0.601	0.578	0.573
Observations	18,482	18,482	18,482	18,482	18,482	18,482

Note: Estimation of the model in equation 8. All regressions use cell-size regression weights, ie. number of tweets written by politician i in period p (N_{ip}). Standard errors are clustered at the level of politicians and months. *p<0.1; **p<0.05; ***p<0.01

Our main results persist when controlling for politician specific trajectories in issue attention, serial correlation, or peer effects (Model 4, 5, & 6 of Table 4, see Section SI3.2 for details).

We report additional robustness checks in Section SI3.2 of the Appendix. Our main results persist with alternate specifications such as a different feedback measure (likes instead of retweets), a different time period to compute issue attention (weeks instead of month), or a different weighting scheme of observations. Finally, we conduct a placebo test by randomly swapping politicians' issue valuations with the feedback-based valuation of another politician of the same social identity (male or female), for the same issue, and in the same month. Using another politician's valuation leads to a series of null results across all specifications (see Table SI3.7 in the Appendix). Thus, the robustness checks confirm that politicians are responsive to feedback. They adapt their attention to gender issues in response to the feedback they receive for their tweets on this issue.

In Section 2.2, we specified conditions for feedback to contribute to a difference in attention to gender issues between female and male politicians (eq. 4). The model most apt to this comparison is Model 2 of Table 4 because it relies on differences in issue valuations and includes separate responsiveness coefficients for female and male politicians. Combining these estimates with the estimates of the gender issue feedback advantage for male and female politicians (see Model 3 in Table 3), we obtain:

$$r_F \Delta\mu_F = .11 * .29 > r_M \Delta\mu_M = .07 * .16 \tag{9}$$

The empirical evidence thus supports the claim that the different feedback that female and male politicians obtain in their interactions with citizens contribute to a divergence in attention to gender issues.

5 Why do female politicians receive more positive feedback for tweeting on gender issues?

We examine three potential mechanisms that can explain why the gender issue feedback advantage is stronger for female politicians than for male politicians.

First, politicians could communicate more engagingly about issues relevant to their own social group and, as a consequence, receive more feedback. Research on descriptive representation finds that politicians are more knowledgeable and more intrinsically motivated when talking about issues that are relevant to their social group (Broockman, 2013). We examine if female politicians talk more engagingly (perhaps they send more enthusiastic or detailed messages) when addressing gender issue. We call this the 'quality channel.'

Second, citizens may be more likely to interact with politicians who share their social identity, particularly when they talk about issues that are relevant to that identity. Differences in the composition of the citizens who choose to interact with them would then shape politicians' perceptions about what the public wants. For instance, citizens are more likely to communicate with representatives of the same race (Broockman, 2014). We assess if female citizens are more interested in gender issues and more likely to interact with female politicians. This is the 'self-selection channel.'

Third, the public may treat politicians differently depending on their social identities. Citizens might reward politicians for behaving in ways that are congruent with their identities (Eagly, Wood and Diekman, 2000). Citizens might perceive female politicians to be more competent to talk about gender issues because they can rely on their personal experience (Dolan, 2010). Accordingly, we examine if the same citizens give more feedback to similar gender issue tweets when they are written by female politicians. We call this the 'congruity channel.'

We address these three channels in turns.

5.1 Tweet quality channel

To test for differences in the quality of tweets about gender issues written by female and male politicians, we differentiate between style and content. We measure various stylistic elements of the tweets, such as sentiment (from negative to positive), the number of words (tokens), hashtags, mentions, emojis, and if a tweet contains a link or a graphic element (picture or video).

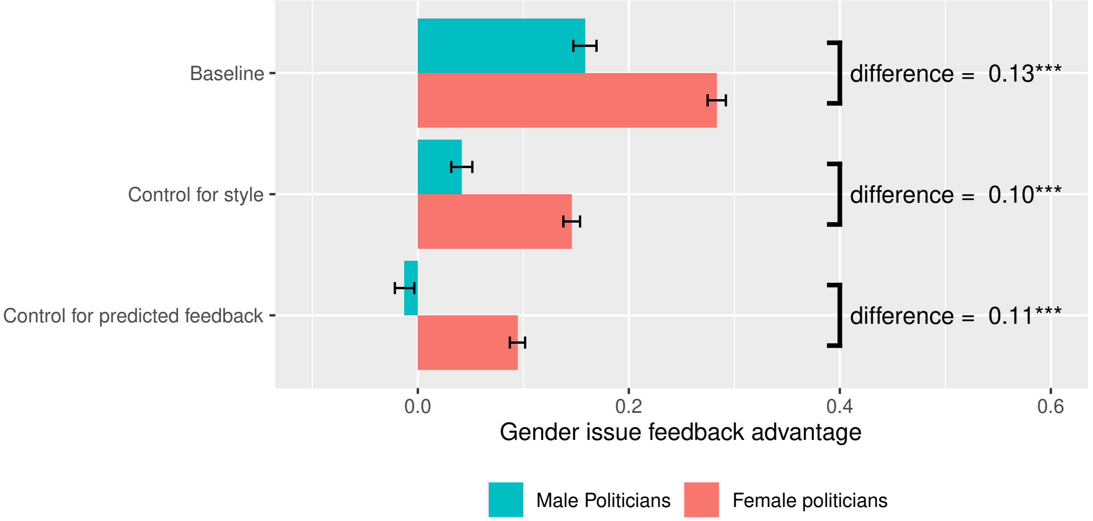
First, we note that both female and male politicians use different styles when writing about gender issues, as compared to tweets on other issues, and style affects the amount of feedback a tweet receives. However, female politicians do not use more engaging features in their gender tweets compared to male politicians (see Figure SI3.2 in the Appendix).

Next, we test more rigorously if differences in style can explain the larger gender issue feedback advantage for female politicians. We estimate a regression model based on the baseline specification from the feedback analysis (see Equation 7), augmented with a vector of stylistic elements as controls. We find that style explains about half of the gender issues feedback advantage, but it hardly affects the gap in gender issue feedback advantage between female and male politicians (see Figure 3).

Next, we analyze tweet content. The quality of tweet content is hard to measure directly because this requires a deep understanding of the meaning of each tweet. Instead, we measure content quality indirectly by retraining our deep learning algorithm to predict feedback based only on the text of the tweets.[11] During training, the deep neural network learns to identify elements of text that contribute to more feedback. If the difference in the gender issue

[11] See Appendix SI4.4 for the details of the feedback prediction algorithm.

Figure 3: Do female politicians write more engaging tweets on gender issues?



Note: The graph plots the standardized amount of feedback politicians receive for tweeting on gender issues relative to tweeting on other issues. Controlling for the style or content of the tweet does not explain the difference in feedback. Black bar represent 95% confidence interval. See Table SI3.8 in SI for the full results.

feedback advantage between female and male politicians disappears when controlling for predicted feedback, we will conclude that differences in the content of tweets written by female and male politicians play a role.

The third row of Figure 3 shows that controlling for predicted feedback completely eliminates the gender issue feedback advantage when considering tweets of female and male politicians jointly. This shows that the feedback prediction algorithm is very accurate. It picks up the popularity of gender issues. Nevertheless, the difference in the gender issue feedback advantage between female politicians and male politicians does not decrease when controlling for predicted feedback. This suggests that it is not the content of the tweets which explains the difference in the gender issue feedback advantage.

In summary, it seems that politicians of both social identities write similarly appealing tweets on gender issues. Hence, we do not find evidence for the quality channel.

5.2 Self-selection channel

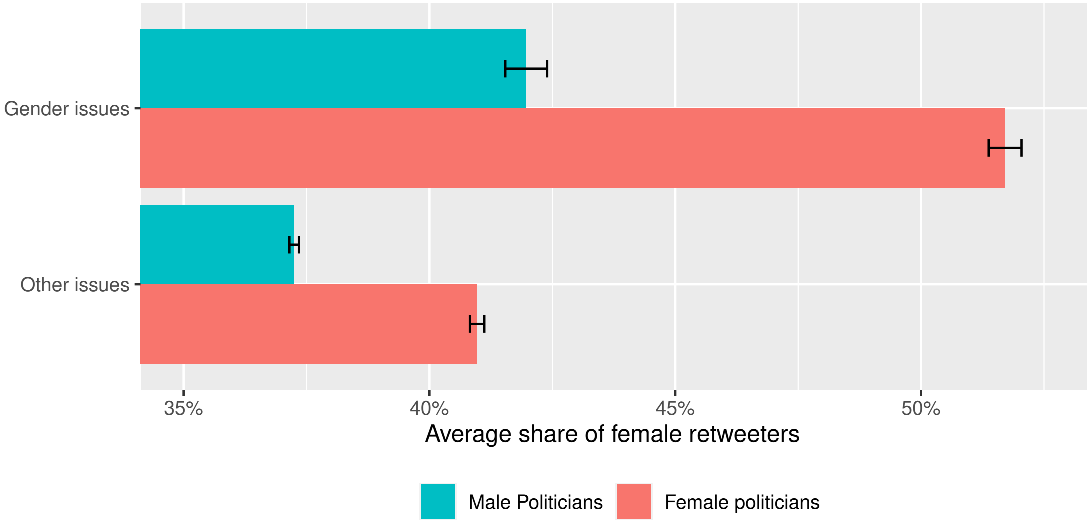
We test the hypothesis that the gap in feedback could be explained by self-selection of citizens who choose to interact with a given politician. Put differently, are female politicians more exposed to the feedback of female citizens who like gender issues better? We differentiate between female and male followers of each politician and female and male retweeters of each tweet by applying a name recognition algorithm to their Twitter username (see Appendix SI5 for details).

We do not find that female politicians have a higher proportion of female followers (see Figure SI3.3 in the Appendix). Female and male politicians have an average of one third of female followers. By contrast, there exist differences in the composition of Twitter users who provide feedback to politicians (see Figure 4). First, the share of female retweeters is larger for tweets published by female politicians than for tweets published by male politicians (43% versus 37%).

This is consistent with existing evidence showing that citizens are more likely to interact with politicians of shared social identity (see Broockman, 2014; Gay, 2007). Second, the share of female retweeters is larger for tweets on gender issues than for tweets on other issues (48% versus 39%). Moreover, there exists an interaction between the social identity of the politician and issue of the tweet. For both female and male politicians, the share of female retweeters is larger when they tweet about gender issues, but the effect is substantially stronger for female politicians. These findings provide evidence for the self-selection channel.

Next, we try to disentangle if this effect is driven by female citizens reacting more or male citizens reacting less when a tweet is on gender issues. To do so, we distinguish between retweets from female or male users and create the same reference-dependent standardized

Figure 4: Share of female retweeters



Note: Black bars represent 95% confidence intervals.

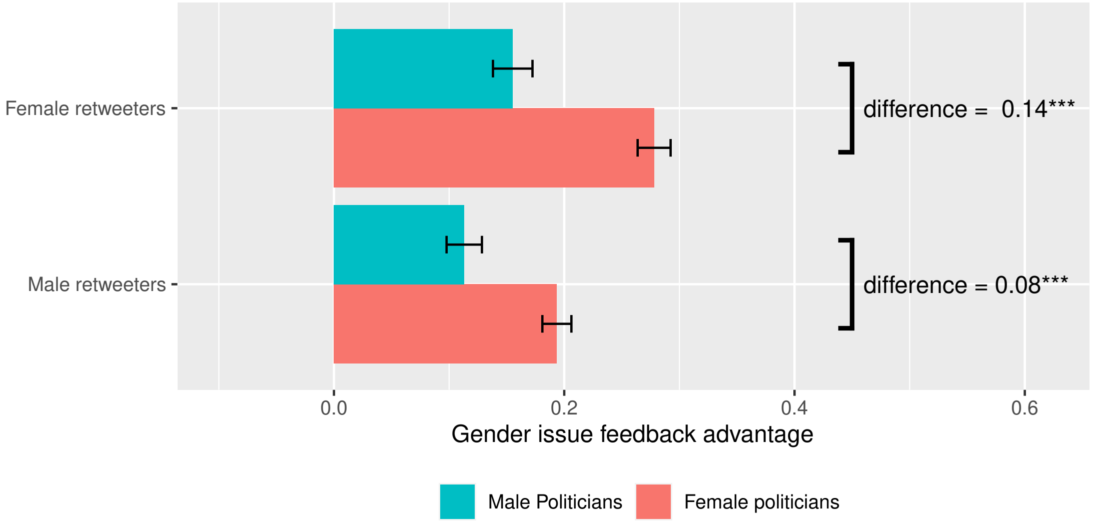
measure of feedback as described in Section 3.2 to make results comparable (see Appendix SI5 for details).

First, we note that there is a positive gender issue feedback advantage when considering only male retweeters, demonstrating that male citizens do not react less when a tweet is on gender issues (see Figure 5). Yet, the gender issue feedback advantage is substantially stronger among female retweeters. This shows that the larger share of female retweeters for tweets on gender issues is driven by female users responding more to gender issues tweets, not male users responding less.

Second, we analyze the gender issue feedback advantage for each combination of citizen's and politician's social identity separately (i.e. female users retweeting female politicians, female users retweeting male politicians, etc.). In Section 4.2 we found that the gender issue feedback advantage is stronger for female politicians than for male politicians. This difference holds when considering female retweeters (+0.14 standard deviations) and male retweeters (+0.08 standard deviations) separately.

Still, the fact that among both female and male retweeters the gender issue feedback advantage is stronger for female politicians suggests that the self-selection channel is not the only operating mechanism. Retweeters seem sensitive to the congruity between policy issue of the tweet and politician's social identity. In the next section, we scrutinize this further.

Figure 5: Is the gap driven by female retweeters?



Note: The graph plots gender issue feedback advantage. Only tweets after 2018 with retweeter information considered (n=473,000). Black bar represent 95% confidence interval. See Table SI3.8 for the full results. *p<0.1; **p<0.05; ***p<0.01

5.3 Congruity channel

To test the hypothesis that citizens react more strongly to tweets congruent with politicians' social identities, we analyze the retweeting behavior of Twitter users who follow multiple politicians. This allows us to hold constant the user's personal characteristics (e.g. their level of interest in gender issues) through a user fixed effect. For each user u, we take the set of tweets written by all politicians whom the user follows and test if the same user is more likely to retweet a gender issues tweet if it was written by a female politician, holding constant the general propensity of the user to retweet gender issues tweets and the user's propensity to retweet a given politician - independently of the policy issue of the tweet.

For computational reasons, we focus on a subsample of the most active retweeters.[12] More specifically, we estimate the following logistic regression:

$$\text{retweet}_{i,u,m} = Logit(\beta * GI_m * M_i + GI_m \times user_u FE + politician_i \times user_u FE + \epsilon_{i,u,m}) \quad (10)$$

The dependent variable $\text{retweet}_{i,u,m}$ is a dummy equal to 1 if tweet m written by politician i was retweeted by user u. The main coefficient of interest is the interaction of a tweet being on gender issues and a politician being male, $GI_m * M_i$. We control for the average propensity of each user to retweet tweets on gender issues by including a set of user fixed effects interacted with the tweet being on gender issues, $GI_m \times user_u$, and a set of politician-by-user fixed, $politician_i \times user_u$, to control for all time-invariant aspects of the politicians, the users, and their relationship.

Model 1 in Table 5 shows that the retweeting probability of gender issue tweets by a given user is generally lower if the tweet was written by male politicians. The average marginal effect is -0.55 percentage points with a baseline retweeting probability of gender tweets of 5.2% (among the sampled users). The average marginal effect implies that a given user is 10.5% less likely to retweet a tweet on gender issues if it was written by a male politician. Controlling for the predicted feedback does not affect the results (Model 2). In Model 3, we estimate separate coefficients for female and male users to see if congruity is equally important to female and male users. Female users are more sensitive to congruity than male users. Female users are 12.7% less likely to retweet a tweet on gender issues if it was written by a male politician. Male users are also less likely to retweet male politicians' tweets on gender issues, albeit only by 8.4%. The difference is not statistically significant.

[12] We selected the 1000 male and 1000 female most retweeting users, and drew a 10% random sample of the tweets of the politicians they follow. This yielded 4.4 million potential retweets.

Table 5: Retweeting Probabilities by Identity of Politician

Dependent Variable:	Dummy = 1 if tweet is retweeted by user			
Model:	(1)	(2)	(3)	(4)
GI * Male politician	-0.1352***	-0.1396***		
	(0.0225)	(0.0224)		
Predicted feedback		1.031***		1.031***
		(0.0127)		(0.0127)
GI * Male politician * Female user			-0.1649***	-0.1752***
			(0.0322)	(0.0323)
GI * Male politician * Male user			-0.1080***	-0.1066***
			(0.0314)	(0.0312)
Fixed-effects				
Retweeter × GI	Yes	Yes	Yes	Yes
Retweeter × Politician	Yes	Yes	Yes	Yes
Fit statistics				
Squared Correlation	0.138	0.152	0.138	0.152
Observations	4,396,339	4,396,339	4,396,339	4,396,339

Note: Logistic regression of retweets on tweet issue, politicians' and users' social identity. Standard errors in parenthesis are clustered as the fixed effects. *p<0.1; **p<0.05; ***p<0.01

6 Discussion

In this article, we advance the understanding of political responsiveness by studying how on-going interactions between politicians and citizens affect politicians' behavior through the lens of a reinforcement learning model. We show that politicians respond to issue-specific feedback by adjusting issue attention. Using gender as an important case study, we demonstrate that female politicians receive systematically more positive feedback from the public when they address issues related to gender issues than male politicians. Our analysis of mechanisms suggests that this difference in feedback exists because citizens treat politicians differently depending on their social identity, and not because female politicians approach the issue in a different way. Our findings suggest that being exposed to different feedback leads politicians from different identities to diverge in the issues they discuss.

Standard dynamic representation models (Stimson, Mackuen and Erikson, 1995; Wlezien,

2004) claim that politicians are responsive to average public opinion to increase their re-election chances. These theories are inconsistent with the finding that politicians are more responsive to rich citizens (Gilens and Page, 2014; Bartels, 2018) and copartisans (Barberá et al., 2019) than to the general public, and that they have heterogeneous and incorrect views about what the public wants (Broockman and Skovron, 2018). Our model is grounded on a psychologically realistic and well-tested view about how humans learn, and it is consistent with these three key findings.

Our theory implies that politicians are responsive to the public opinion they "see" (Fenno, 1977, 883). Reinforcement learning is a strategy prone to the emergence of biases in judgments and beliefs (Denrell, 2005; Fiedler, Juslin et al., 2006; Le Mens and Denrell, 2011; Le Mens, Kareev and Avrahami, 2016); it allows politicians to be responsive, but only to the self-selected set of citizens who choose to interact with them. Politicians lack information about the preferences of citizens they do not see, and cannot perfectly adjust for biases in the feedback they receive, likely because they, as other humans, lack the cognitive abilities to do so (Fiedler, 2012). Being responsive to other entities such as the median voter, all voters in an electoral constituency, all partisans, or national public opinion may be more desirable from a normative perspective, but reinforcement learning is not conducive to responsiveness to such entities.[13] Of course, politicians do not learn about public opinion between elections only through interactions with the public via Twitter or in other settings. They also rely on other strategies such as opinion polls (Druckman and Jacobs, 2006), which can provide accurate information about the average views of the public. Yet, public opinion polls are not available continuously and for all issues, while the learning strategy we describe in this article is readily available to politicians who want to test the popularity of different issues. We cannot test when politicians rely on bias-prone strategies such as reinforcement learning to form perceptions of public opinion and when they rely on less biased strategies such as

[13] Elections could in principle be interpreted through the lens of reinforcement learning, but they are held too far apart and send a signal that is too blurred to be useful to learn about public opinion on specific issues.

public opinion polls, but this is an interesting avenue for future research.

Another relevant extension of this research would consist in applying our reinforcement learning approach to study whether the rise of Twitter and social media has increased polarization among politicians. Our results imply that politicians shift attention to issues relevant to citizens they personally interact with. Hence, if politicians are frequently exposed to views from one extreme of the political spectrum on social media while seeing less moderate or opposing views, reinforcement learning could contribute to polarization of politician's discourse and behavior. Consistent with this conjecture, Barberá and Rivero (2015) showed that politically active Twitter users are more partisan and more polarized than the average voter. Applying our framework to the study of polarization requires overcoming empirical challenges related to the coding of the 'extremity' of tweets. One reason to focus on issue salience rather than on issue positions or extremity is that talking about an issue or not is a binary decision and thus more straightforward to measure than issue extremity. Yet, our approach can be combined with advances in text scaling methods to study citizen-driven political polarization.

Another important question concerns the extent to which the dynamics we observed in the Twitter context affect politicians' offline behavior and whether these dynamics apply to other settings. We are persuaded that the study of politicians' behavior on Twitter is important on it's own right as politicians' tweeting behavior has real consequences: The large majority of politicians uses Twitter frequently to announce their priorities, many citizens and opinion leaders follow politicians on Twitter, and journalists closely monitor Twitter to inform the public about ongoing debates (Jungherr, 2016). In this paper we do not address the question of whether social media feedback affects offline behavior and public policy, but the feedback that politicians receive on social media could be linked to the oral questions they ask in parliament and to roll call votes. This is a promising avenue for future research. Moreover, we suspect that the mechanism we study in this article generalizes to other settings such as other social networks, campaign meetings (applause is a clear source of feedback), or any setting in which a politicians interact with an audience, and this could be tested empirically.

Finally, more work is needed to clarify the implications of our findings for the political representation of historically underrepresented groups. On the one hand, the fact that politicians experience a feedback advantage when addressing policy issues related to their own social identity strengthens the case for descriptive representation. Our results imply that there would be less attention to gender issues if there were fewer female politicians. On the other hand, the mechanism we describe could perpetuate identity-based specialization and the relegation of representatives from under-represented roles to niche issues. For example, female politicians might specialize on gender issues and be less inclined to address other issues which might be important to advance their careers. Future work should aim to uncover if the kind of identity-based reinforcement we found affects the careers of women and other marginalized groups in politics.

References

Adams, James. 2012. "Causes and electoral consequences of party policy shifts in multiparty elections: Theoretical results and empirical evidence." *Annual Review of Political Science* 15:401–419.

Arceneaux, Kevin, Martin Johnson, René Lindstädt and Ryan J Vander Wielen. 2016. "The influence of news media on political elites: Investigating strategic responsiveness in Congress." *American Journal of Political Science* 60(1):5–29.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost and Joshua A. Tucker. 2019. "Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data." *American Political Science Review* 113(4):883–901.

Barberá, Pablo and Gonzalo Rivero. 2015. "Understanding the Political Representativeness of Twitter Users." *Social Science Computer Review* 33(6):712–729.

Bartels, Larry M. 2018. *Unequal democracy: The political economy of the new gilded age.* Princeton University Press.

Broockman, David E. 2013. "Black Politicians Are More Intrinsically Motivated to Advance Blacks' Interests: A Field Experiment Manipulating Political Incentives." *American Journal of Political Science* 57(3):521–536.

Broockman, David E. 2014. "Distorted Communication, Unequal Representation: Constituents Communicate Less to Representatives Not of Their Race." *American Journal of Political Science* 58(2):307–321.

Broockman, David E. and Christopher Skovron. 2018. "Bias in perceptions of public opinion among political elites." *American Political Science Review* .

Butler, Daniel M and Adam M Dynes. 2016. "How politicians discount the opinions of constituents with whom they disagree." *American Journal of Political Science* 60(4):975–989.

Butler, Daniel M, David W Nickerson et al. 2011. "Can learning constituency opinion affect how legislators vote? Results from a field experiment." *Quarterly Journal of Political Science* 6(1):55–83.

Campillo, Inés. 2019. "'If we stop, the world stops': the 2018 feminist strike in Spain." *Social Movement Studies* 18(2):252–258.

Caughey, Devin and Christopher Warshaw. 2018. "Policy preferences and policy change: Dynamic responsiveness in the American States, 1936-2014.".

Conover, M D, J Ratkiewicz and M Francisco. 2011. "Political polarization on twitter." *Icwsm* 133(26):89–96.

Cyert, Richard M and James G March. 1963. *A behavioral theory of the firm.* Vol. 2.

Denrell, Jerker. 2005. "Why most people disapprove of Me: Experience sampling in impression formation." *Psychological Review* 112(4):951–978.

Denrell, Jerker and Gaël Le Mens. 2007. "Interdependent sampling and social influence." *Psychological review* 114(2):398.

Denrell, Jerker and Gaël Le Mens. 2011. Social Judgments from Adaptive Samples. In *Social*

Judgment and Decision Making, ed. Joachim I Krueger. Psychology Press pp. 151–169.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Technical report.

Dolan, Kathleen. 2010. "The impact of gender stereotyped evaluations on support for women candidates." *Political Behavior* .

Druckman, James N and Lawrence R Jacobs. 2006. "Lumpers and splitters: The public opinion information that politicians collect and use." *International Journal of Public Opinion Quarterly* 70(4):453–476.

Eagly, Alice H, Wendy Wood and Amanda B Diekman. 2000. "Social role theory of sex differences and similarities: A current appraisal." *The developmental social psychology of gender* 12:174.

Fenno, R F. 1977. "US House members in their constituencies: An exploration." *American Political Science Review* .

Fiedler, Klaus. 2012. Meta-cognitive myopia and the dilemmas of inductive-statistical inference. In *Psychology of Learning and Motivation*. Vol. 57 Elsevier pp. 1–55.

Fiedler, Klaus, Peter Juslin et al. 2006. *Information sampling and adaptive cognition.* Cambridge University Press.

Galesic, Mirta, Henrik Olsson and Jörg Rieskamp. 2018. "A sampling model of social judgment." *Psychological Review* 125(3):363.

Gay, Claudine. 2007. "Spirals of Trust? The Effect of Descriptive Representation on the Relationship between Citizens and Their Government." *American Journal of Political Science* 46(4):717.

Giger, Nathalie, Jan Rosset and Julian Bernauer. 2012. "The poor political representation of the poor in a comparative perspective." *Representation* 48(1):47–61.

Gilens, Martin and Benjamin I Page. 2014. "Testing theories of American politics: Elites, interest groups, and average citizens." *Perspectives on Politics* 12(3):564–581.

Greve, Henrich R. 1998. "Performance, aspirations, and risky organizational change." *Administrative Science Quarterly* pp. 58–86.

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.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *Springer Series in Statistics*. Vol. 27.

Hobolt, Sara Binzer and Robert Klemmensen. 2008. "Government responsiveness and political competition in comparative perspective." *Comparative Political Studies* .

Holland, John Henry. 1992. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* MIT press, Cambridge, Massachusetts.

Jones, Bryan D. and Frank R. Baumgartner. 2004. "Representation and agenda setting." *Policy Studies Journal* 32(1):1–24.

Jones, Bryan D, Heather Larsen-Price and John Wilkerson. 2009. "Representation and American governing institutions." *Journal of Politics* 71(1):277–290.

Jungherr, Andreas. 2016. "Twitter use in election campaigns: A systematic literature review." *Journal of information technology & politics* 13(1):72–91.

Klüver, Heike and Jae Jae Spoon. 2016. "Who Responds? Voters, Parties and Issue Attention." *British Journal of Political Science* 46(3):633–654.

Landis, J. Richard and Gary G Koch. 1977. "The measurement of observer agreement for categorical data." *Biomedics* pp. 159–174.

Le Mens, Gaël and Jerker Denrell. 2011. "Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases." *Psychological review* 118(2):379.

Le Mens, Gaël, Jerker Denrell, Balázs Kovács and Hülya Karaman. 2019. "Information Sampling, Judgment, and the Environment: Application to the Effect of Popularity on Evaluations." *Topics in Cognitive Science* 11(2):358–373.

Le Mens, Gaël, Yaakov Kareev and Judith Avrahami. 2016. "The Evaluative Advantage of Novel Alternatives: An Information-Sampling Account." *Psychological Science* 27(2):161–168.

Lovenduski, Joni and Pippa Norris. 2003. "Westminster women: The politics of presence." *Political Studies* 51(1):84–102.

Malmendier, Ulrike and Stefan Nagel. 2011. "Depression babies: do macroeconomic experiences affect risk taking?" *The Quarterly Journal of Economics* 126(1):373–416.

Mansbridge, Jane. 1999. "Should Blacks Represent Blacks and Women Represent Women? A Contingent "Yes"." *The Journal of Politics* 61(3):628–657.

Mansbridge, Jane. 2003. “Rethinking representation.” *American political science review* pp. 515–528.

March, James G and Zur Shapira. 1992. “Variable risk preferences and the focus of attention.” *Psychological review* 99(1):172.

Metaxas, Panagiotis Takis, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O’Keefe and Samantha Finn. 2015. What Do Retweets Indicate? Results from User Survey and Meta-Review of Research. In *ICWSM.* pp. 658–661.

Miller, Warren E. and Donald E. Stokes. 1963. “Constituency Influence in Congress.” *American Political Science Review* .

Phillips, A. 1995. *The politics of presence.* Oxford University Press.

Stimson, James A., Michael B. Mackuen and Robert S. Erikson. 1995. “Dynamic Representation.” *American Political Science Review* 89(3):543–565.

Sutton, RS and AG Barto. 2018. *Reinforcement learning: An introduction.* MIT Press.

Thorndike, E L. 1927. “The law of effect.” *The American journal of psychology* .

Wagner, Markus and Thomas M Meyer. 2014. “Which issues do parties emphasise? Salience strategies and party organisation in multiparty systems.” *West European Politics* 37(5):1019–1045.

Wlezien, Christopher. 1995. “The public as thermostat: Dynamics of preferences for spending.” *American journal of political science* pp. 981–1000.

Wlezien, Christopher. 2004. “Patterns of representation: Dynamics of public preferences and policy.” *The Journal of Politics* 66(1):1–24.

Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R’emi Louf, Morgan Funtowicz and Jamie Brew. 2019. “HuggingFace’s Transformers: State-of-the-art Natural Language Processing.” *ArXiv* abs/1910.03771.

Supplementary Information

Contents

SI1 Descriptive Statistics

Table SI1.1: Summary Statistics: Tweets

Statistic	N	Min	Median	Mean	Max	St. Dev.
Tweet is on gender issues	1,583,917	0	0	0.06	1	0.24
Writer is female politician	1,583,917	0	0	0.38	1	0.49
Writer is left-wing politician	1,583,917	0	1	0.56	1	0.50
Number of retweets	1,583,917	0	3	56.37	42,244	385.18
Number of likes	1,583,917	0	5	96.50	70,085	709.03
Feedback measure (FB_{im})	1,583,917	−6.81	−0.11	0.00	14.99	1.00
Share of female retweeters	472,959	0.00	0.38	0.39	1.00	0.27
Thread length	1,583,917	1	1	1.00	5	0.02
Tokens	1,583,917	3	21	22.34	95	11.84
Hashtags	1,583,917	0	0	0.45	30	0.92
Mentions	1,583,917	0	1	1.09	50	1.84
Emojis	1,583,917	0	0	0.28	140	1.07
Contains picture	1,583,917	0	0	0.31	1	0.46
Contains link	1,583,917	0	0	0.40	1	0.49
Sentiment score	1,582,931	0.00	0.19	0.27	1.00	0.26

Note: Tokens are words or other symbols (mentions, emojis, etc.). Mentions are references to other Twitter users. Share of female retweeters only calculated for tweets starting from 2018 with at least one identified retweeter. Sentiment score could not be computed for approximately 1000 tweets.

Table SI1.2: Summary Statistics: Politicians

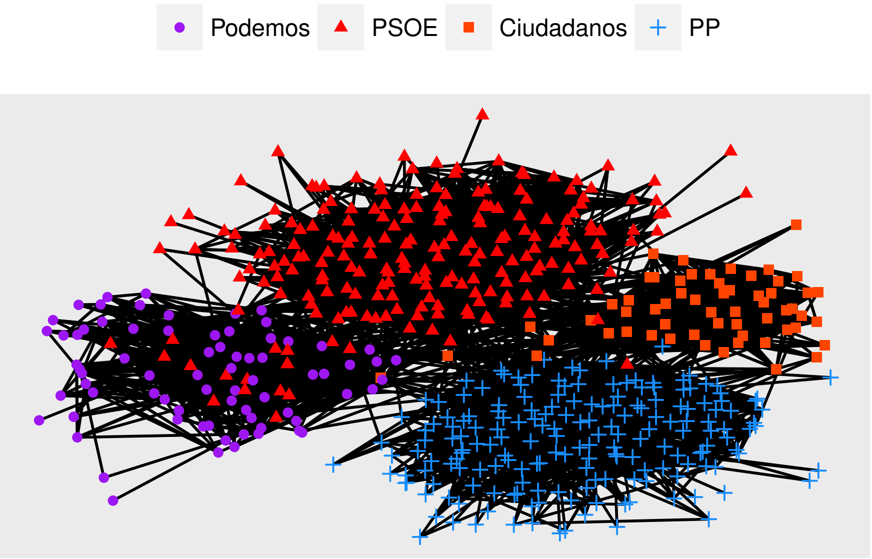
Statistic	N	Min	Median	Mean	Max	St. Dev.
Share of tweets written on gender issues	1,265	0.00	0.04	0.07	0.60	0.08
Female	1,265	0	0	0.44	1	0.50
Left-wing	1,265	0	1	0.57	1	0.50
Followers	1,223	49	3,127	23,094.08	2,390,647	118,166.20
Following	1,223	7	1,121	1,749.19	98,465	3,710.73
Tweets written since joining Twitter	1,223	97	7,977	12,875.22	134,620	15,036.53
Tweets written in sample period	1,265	33	652	1,252.11	29,172	2,059.92
Average number of retweets	1,265	0.13	6.87	35.54	2,651.50	136.06
Standard deviation of retweets	1,265	0.37	11.45	76.72	3,660.48	234.62
Average number of likes	1,265	0.32	9.41	62.00	5,040.03	274.30
Standard deviation of likes	1,265	0.82	15.45	131.20	6,687.89	456.71
Average number of tokens	1,265	7.36	22.63	23.04	45.14	5.47
Average share of female retweeters	1,257	0.00	0.38	0.39	1.00	0.13

SI2 Evidence for Retweets as Positive Feedback

Figure SI2.1 plots the network of retweets between Members of Parliament of Spain's four major parties (n=527). Each politician represents one vertex. An edge exists if one politician retweeted another politician in our sampling period or vice versa. This constitutes an undirected graph. The figure shows that most retweets happen within parties. We interpret this as evidence that retweets are used as positive feedback.

Members of Parliaments from one of the smaller parties were excluded to facilitate visualization.

Figure SI2.1: Retweeting Network between Politicians



SI3 Robustness Checks in Detail

SI3.1 Differences in Feedback for Tweeting on Gender

This section provides further details on Model 4 of Table 3 and replicates all four specification using likes instead of retweets to construct our measure of feedback (see Table SI3.3).

SI3.1.1 Details on Model 4 of Table 3

Model 4 tests if the gender issue feedback advantage is driven by one side of the political spectrum. We could conjecture that left-leaning politicians might receive more positive feedback for addressing gender issues or that a stronger feedback advantage for female politicians might be more pronounced among right-leaning politicians. However, when we interact our variables of interest (GI_{im}, M_i) with a dummy equaling 1 if politician i belongs to a left-leaning party L_i, we do not find that our effects depend on the politician's ideological leaning.

We coded parties into left- and right-leaning with the following coding scheme:

Left-leaning parties: ASG, AVANCEM, Bildu, BNG, CHA, Coalición Caballas, COMPROMIS, CpM, CUP, EM, ERC, Eusko Alkartasuna, GENTxFORMENTERA+PARTIT SOCIALISTA DE LES ILLES BALEARS, Geroa Bai, ICV, INDEPENDENT, MDyC, MÉS PER MALLORCA-PSM-ENTESA-INICIATIVAVERDS, MÉS PER MENORCA, NCa, Podemos, PRC, PSOE, UPL

Right-leaning parties: CCa-PNC, Ciudadanos, EAJ-PNV, EL PI-PROPOSTA PER LES ILLES, Foro Asturias, JxCat, PAR, PDECAT, PP, PPL, UPN, VOX

SI3.1.2 Likes instead of retweets

Table SI3.3 replicates the Table 3 using likes instead of retweets to construct our feedback measure. In line with the main results female politicians have a larger gender issue feedback advantage than male politicians in all specifications.

Table SI3.3: Additional Feedback for GI tweets based on likes

Dependent Variable:	Tweet-level standardized likes, FB_{im}			
Model:	(1)	(2)	(3)	(4)
GI	0.2191***	0.2652***	0.2644***	0.2655***
	(0.0042)	(0.0196)	(0.0195)	(0.0499)
GI × Male politician	-0.0672***	-0.0970***	-0.1011***	-0.1137**
	(0.0068)	(0.0229)	(0.0227)	(0.0531)
Part of thread			0.2535	0.2535
			(0.1620)	(0.1620)
Tweets on day by politician			-0.0083***	-0.0084***
			(0.0019)	(0.0019)
GI × Left				-0.0014
				(0.0537)
GI × Male politician × Left				0.0200
				(0.0587)
Male politician	0.0192***			
	(0.0017)			
(Intercept)	-0.0245***			
	(0.0014)			
Fixed-effects				
Politician		Yes	Yes	Yes
Day		Yes	Yes	Yes
Hour of day			Yes	Yes
Fit statistics				
Squared Correlation	0.002	0.011	0.021	0.021
Observations	1,583,917	1,583,917	1,583,917	1,583,917

Note: The dependent variable is FB_{im} which is based on the number of likes. Effects can be interpreted in standard deviations of feedback. Estimations of variations of equation 7. Model 1 is without fixed effects. Model 2 included fixed-effects. Model 3 adds controls. Model 4 includes triple interaction of gender issues, female and left-wing. Standard error are reported in parentheses are clustered according to the fixed effects *p<0.1; **p<0.05; ***p<0.01

SI3.2 Responsiveness to Feedback

This section extends the discussion about the robustness checks of the responsiveness models offered in the text. The full regression tables are found at the end of the section.

SI3.2.1 Additional specifications in main table

Ancillary analyses reveal that coefficient estimates are stable to the inclusion of additional factors. We first account for individual trajectories in politicians' attention to gender issues over time. This is relevant since female politicians increase their attention to gender issues more than male politicians during our study period. As explained in Section 3.2, the feedback measure (on which issue valuations are based) already includes a politician-specific trend. This makes valuations more comparable over periods even when politicians are on different time trends. We do the same for issue attention by including a linear time trend for each politician (Model 4). The coefficient for the trend is highly significant and increases the fit of the model. Yet, the estimated coefficients for the issue valuations hardly change. This is noteworthy as the trend is arguably endogenous to feedback: politicians who consistently receive more positive feedback for tweeting on gender issues will be on a steeper trend.

Next, we want to dispel concerns that serial correlation might bias our estimates. Accordingly, we include the lagged dependent variable (the share of tweets written on gender issues in the last month) as a control (Model 5). The additional variable has a large positive coefficient and the model fit increases, but the estimated coefficients for the issue valuations remain similar to Model 3.

Finally, we address the possibility that issue attention is influenced by peer effects. Even though the month fixed effect already captures common patterns in issue attention that affect all politicians equally, it could be that politician are more strongly affected by the behavior of politicians who they share a social identity with. In Model 6, we include the average attention to gender issues by politicians of the same social identity (male or female) as a control. The estimated coefficient is imprecisely estimated. This suggests that this sort of peer effects does not play an important role.

SI3.2.2 Alternative Feedback Measure

We chose retweets over likes to construct our feedback measure because it allowed us to learn about the feedback givers' identities. Still, our theory of reinforcement learning should also apply to likes as a form of feedback. Likes have the advantage that they unambiguously stand for positive feedback. Hence, if our responsiveness results replicate using likes instead of retweets, it provides further evidence that politicians are responsive to positive feedback.

As can be seen in Table SI3.4, we find that politicians' valuations are updated with a similar speed if feedback is based on likes. Again, the issue valuation is revised by approximately 7% with each tweet on the issue. Responsiveness coefficients all have the same sign as in the main results and are significant. Point estimates are somewhat attenuated, but close.

We conclude from this analysis that politicians are also responsive to likes as a form of feedback.

SI3.2.3 Alternative Time Period

To show that our main results do not depend on the particular choice of time period for computing issue attention (months), we replicate the specification using weeks instead of months. Table SI3.5 shows that our results hold. Again, both male and female politicians are generally responsive to feedback. A higher valuation of gender issues increases the attention to the issue whereas a higher valuation of other issues can lead to a crowding out. Estimated valuation updated parameters γ are slightly larger here (9% revision with every new tweet on the issue in most specifications).

SI3.2.4 Alternative Weights

To avoid concerns that our main results could be driven by the specific weighting scheme we used in the model estimations, we replicate our analyses by weighting each politician month cell equally, independently of the actual number of tweets written in the politician-month cell. Our main result holds, yet, there are some differences.

For female politicians, the estimated responsiveness parameters remain stable. Table SI3.6 shows that a higher valuation of gender issues is associated with a higher share of tweets written on the issue among female politicians. However, for male politicians we cannot recover the effect with this specification. We believe that this makes sense, considering that many male politicians write few tweets on gender issues and we need to focus on the set of politician-month cells containing more tweets to find significant effects. Giving equal weight to cells with too little underlying tweets induces too much noise.

SI3.2.5 Tables

Table SI3.4: Responsiveness to Likes

Dependent Variable:	Monthly share of tweets written on GI, $A_{ip}^{GI} = \frac{n_{ip}^{GI}}{N_{ip}}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
ΔV	0.0771***					
	(0.0205)					
ΔV_F		0.0869***				
		(0.0291)				
ΔV_M		0.0624**				
		(0.0259)				
V_F^{GI}			0.1250***	0.1065***	0.1199***	0.1249***
			(0.0440)	(0.0376)	(0.0412)	(0.0440)
V_M^{GI}			0.1134***	0.1094***	0.1146***	0.1133***
			(0.0374)	(0.0315)	(0.0361)	(0.0374)
V_F^{other}			-0.0571*	-0.0182	-0.0510	-0.0571*
			(0.0333)	(0.0305)	(0.0314)	(0.0333)
V_M^{other}			-0.0292	-0.0316	-0.0308	-0.0294
			(0.0342)	(0.0330)	(0.0339)	(0.0342)
Indiv. trend				5.535***		
				(0.4390)		
Lagged DV					1.022***	
					(0.1029)	
Social Influence						-0.0890
						(0.8666)
$\widehat{\gamma}$ (to calc. valuation)	0.07	0.07	0.06	0.07	0.06	0.06
Fixed-effects						
Politician	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Squared Correlation	0.573	0.573	0.573	0.601	0.578	0.573
Observations	18,482	18,482	18,482	18,482	18,482	18,482

Note: Estimation of the model in equation 8. All regressions use cell-size regression weights, ie. number of tweets written by politician i in period p (N_{ip}). Standard errors are clustered at the level of politicians and months. *p<0.1; **p<0.05; ***p<0.01

Table SI3.5: Responsiveness based on weeks

Dependent Variable:	Weekly share of tweets written on GI, $A_{ip}^{GI} = \frac{n_{ip}^{GI}}{N_{ip}}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
ΔV	0.1026***					
	(0.0174)					
ΔV_F		0.1225***				
		(0.0240)				
ΔV_M		0.0724***				
		(0.0221)				
V_F^{GI}			0.1507***	0.1210***	0.1476***	0.1508***
			(0.0329)	(0.0235)	(0.0322)	(0.0329)
V_M^{GI}			0.1043***	0.0982***	0.1089***	0.1042***
			(0.0317)	(0.0265)	(0.0317)	(0.0315)
V_F^{other}			-0.0980***	-0.0582***	-0.0833***	-0.0980***
			(0.0252)	(0.0219)	(0.0247)	(0.0252)
V_M^{other}			-0.0488*	-0.0432*	-0.0511*	-0.0487*
			(0.0282)	(0.0250)	(0.0286)	(0.0281)
Indiv. trend				5.561***		
				(0.4433)		
Lagged DV					0.5830***	
					(0.0494)	
Social Influence						-0.0434
						(0.5231)
$\widehat{\gamma}$ (to calc. valuation)	0.09	0.09	0.09	0.13	0.08	0.09
Fixed-effects						
Politician	Yes	Yes	Yes	Yes	Yes	Yes
Week	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Squared Correlation	0.362	0.362	0.362	0.375	0.38	0.362
Observations	74,588	74,588	74,588	74,588	72,103	74,588

Note: Estimation of the model in equation 8. All regressions use cell-size regression weights, ie. number of tweets written by politician i in period p (N_{ip}). Standard errors are clustered at the level of politicians and weeks. *p<0.1; **p<0.05; ***p<0.01

Table SI3.6: Responsiveness weighted by cells

Dependent Variable:	Monthly share of tweets written on GI, $A_{ip}^{GI} = \frac{n_{ip}^{GI}}{N_{ip}}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
ΔV	0.0898***					
	(0.0273)					
ΔV_F		0.1041***				
		(0.0364)				
ΔV_M		0.0573*				
		(0.0312)				
V_F^{GI}			0.1749***	0.1318***	0.1618***	0.1749***
			(0.0468)	(0.0356)	(0.0433)	(0.0468)
V_M^{GI}			0.0739*	0.0670*	0.0729*	0.0738*
			(0.0436)	(0.0389)	(0.0425)	(0.0436)
V_F^{other}			-0.0356	0.0218	-0.0267	-0.0356
			(0.0449)	(0.0344)	(0.0417)	(0.0449)
V_M^{other}			-0.0416	-0.0346	-0.0410	-0.0419
			(0.0356)	(0.0340)	(0.0350)	(0.0357)
Indiv. trend				5.655***		
				(0.4119)		
Lagged DV					0.6485***	
					(0.1291)	
Social Influence						-0.1639
						(0.7411)
$\widehat{\gamma}$ (to calc. valuation)	0.07	0.07	0.08	0.09	0.08	0.08
Fixed-effects						
Politician	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Squared Correlation	0.591	0.591	0.591	0.617	0.596	0.591
Observations	18,482	18,482	18,482	18,482	18,482	18,482

Note: Estimation of the model in equation 8. All cells are weighted equally, independently of the number of underlying tweets. Standard errors are clustered at the level of politicians and months. *p<0.1; **p<0.05; ***p<0.01

Table SI3.7: Responsiveness Placebo Test

Dependent Variable:	Monthly share of tweets written on GI, $A_{ip}^{GI} = \frac{n_{ip}^{GI}}{N_{ip}}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
ΔV	0.0041					
	(0.0119)					
ΔV_F		-0.0064				
		(0.0182)				
ΔV_M		0.0196				
		(0.0189)				
V_F^{GI}			-0.0057	-0.0157	0.0111	0.0052
			(0.0170)	(0.0170)	(0.0175)	(0.0191)
V_M^{GI}			-0.0056	0.0231	0.0137	-0.0272
			(0.0208)	(0.0157)	(0.0196)	(0.0218)
V_F^{other}			0.0431	-0.0365	0.0606^{**}	-0.0082
			(0.0269)	(0.0278)	(0.0252)	(0.0275)
V_M^{other}			0.0279	-0.0216	0.0117	0.0411
			(0.0249)	(0.0348)	(0.0273)	(0.0321)
Indiv. trend				5.545^{***}		
				(0.4368)		
Lagged DV					1.029^{***}	
					(0.1056)	
Social Influence						-0.1017
						(0.8722)
$\widehat{\gamma}$ (to calc. valuation)	0.07	0.07	0.07	0.07	0.07	0.07
Fixed-effects						
Politician	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Squared Correlation	0.572	0.572	0.572	0.601	0.577	0.572
Observations	18,481	18,481	18,481	18,481	18,481	18,481

Note: Estimation of the model in equation 8. Valuations are randomly swapped with valuations of other politicians but of the same social identity, the same issue and from the same month. Updating parameter $\widehat{\gamma}$ used from main specifications (see Table 4). All regressions use cell-size regression weights, ie. number of tweets written by politician i in period p (N_{ip}). Standard errors are clustered at the level of politicians and months. *p<0.1; **p<0.05; ***p<0.01

SI3.3 Mechanism for Differences in Feedback

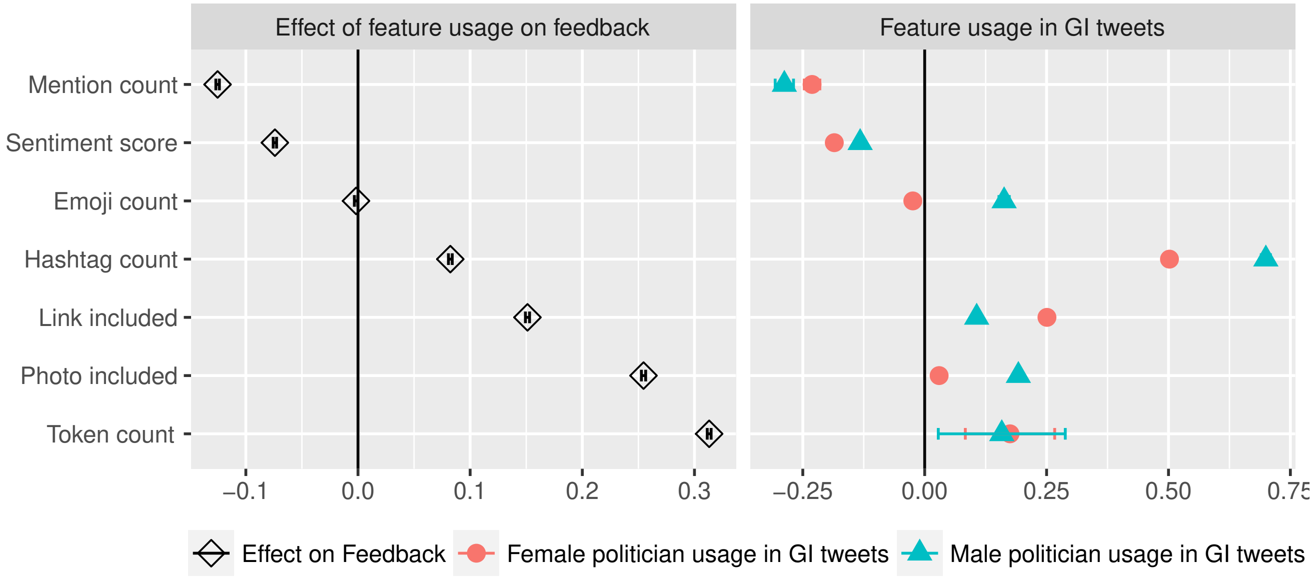
This section provides further detail on our analysis why the gender issue feedback advantage is larger for female politicians. First, Figure SI3.2 shows that, compared to male politicians, female politicians do not systematically use more features that attract positive feedback in their gender issue tweets. Second, Table SI3.8 provides the the regression details underlying Figure 3 and 4 of the main text.

More specifically, Figure SI3.2 shows in the left panel the effect of different features on standardized feedback and in the right panel the average usage of those features in gender issue tweets (compared to other tweets) separately for female and male politicians. For example, including an additional mention of another Twitter user decreases feedback by 0.12 standard deviations (black diamond in left panel). When female politicians tweet on gender, they use 24% fewer mentions to other users as compared as tweets on other issues (red circle in right panel). Male politicians use 28% fewer mentions (blue triangle in right panel). However, female politicians do not use systematically a style that attract more feedback when tweeting on gender issues since it is not the case that female politicians use popular features more often in their tweets on gender issues, relative to male politicians.

Table SI3.8 analyzes the gender issue feedback advantage. Dependent variable is feedback measure as explained in Section 3.2. Effects can be interpreted in standard deviations of feedback. Model 1 repeats the finding from our main results (see Model 3 of Table 3). Model 2 controls for the style of the tweet. Model 3 controls for predicted feedback. It reveals that the gender issue feedback advantage remains stable around 0.1 standard deviations. This is the basis of Figure 3 in the main text. Model 4 only considers tweets with feedback giver identity information (after 2018). Model 5 considers feedback from female users and 6 only feedback from male citizens.[14] The results reveal that the gender issue feedback advantage for female politicians is stronger is only the feedback from female Twitter users is considered. This is the basis for Figure 4 in the main text.

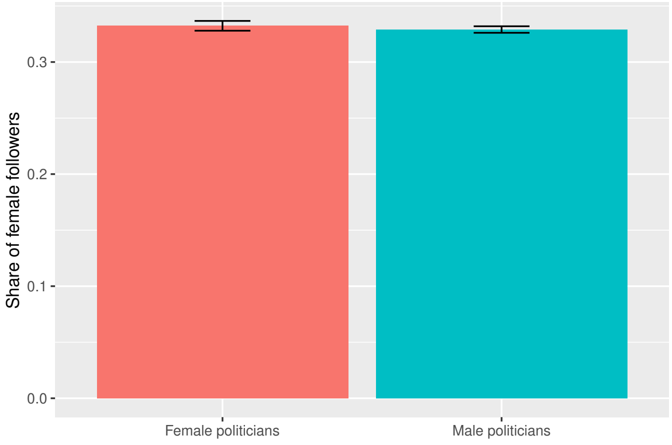
[14] See Appendix SI5 for an explanation how feedback variable was calculated for female and male Twitter users separately.

Figure SI3.2: Effect of Feature usage on Feedback, Difference between Male and Female Politicians on Feature Usage in Gender Issues Tweets



Note: Left panel plots effects of feature usage on feedback. Right panel shows feature usage in gender issue tweets relative to other tweets, separately for female and male politicians. Bars represent 95% confidence interval (confidence intervals sometimes invisible because they are close to zero).

Figure SI3.3: Share of Female Followers



Note: The graph shows that there is no difference in the average share of female followers between female and male politicians. The gender identity of the followers was inferred from their username displayed on Twitter. Black bars represent the 95% confidence interval.

Table SI3.8: Mechanism: Differences in Feedback

	Dependent variable:					
	Retweets				Retweets from female users	Retweets from male users
	(1)	(2)	(3)	(4)	(5)	(6)
GI	0.283***	0.146***	0.094***	0.169***	0.353***	0.194***
	(0.021)	(0.017)	(0.017)	(0.020)	(0.024)	(0.027)
GI * Male Politician	−0.125***	−0.104***	−0.107***	−0.094***	−0.144***	−0.080***
	(0.023)	(0.019)	(0.019)	(0.023)	(0.026)	(0.030)
Tweets on day by Politician	−0.007***	−0.005***	−0.003***	−0.010***	−0.010***	−0.011***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Part of thread	0.292*	0.189*	−0.035	0.169*	0.357***	0.291**
	(0.150)	(0.114)	(0.090)	(0.095)	(0.136)	(0.138)
Predicted Feedback		0.026***				
		(0.001)				
Sentiment scor		0.088***				
		(0.009)				
Token count		−0.068***				
		(0.004)				
Hashtag count		−0.002				
		(0.003)				
Mention count		0.549***				
		(0.019)				
Emoji count		0.307***				
		(0.025)				
Politician FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour of day FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,583,917	1,582,931	1,583,917	472,959	643,475	643,511

Note: Model (4-6) only consider tweets starting from 2018 with retweeter information. Standard errors are clustered by politician and by day. *p<0.1; **p<0.05; ***p<0.01

SI4 Deep Learning Tweet Issue Classifier

This section describes how we used the deep-learning classifier BERT to identify tweets on gender issues. First, we describe how we created a set of hand-labeled tweets. Second, it describes how BERT can be employed and fine-tuned. Third, we describe the accuracy of our classification. Finally, we describe how we adjusted the algorithm to predict feedback based on the text of tweets.

SI4.1 Human Coding Stage

Supervised machine learning algorithms require a set of tweets which are correctly labeled as being on gender issues or not. We manually classified tweets as follows. First, we developed coding guidelines by creating a list of issues related to gender based on previous literature (Phillips, 1995). Second, we selected a random sample of 19,377 tweets from that topic to be the training set and another 1975 tweets as a test set. To maximize the information contained in the training set, we over-sampled tweets on gender issues using an unsupervised topic model (LDA). We sampled tweets from one of the topics constructed by the model which contained many of words related to gender issues. The test set was sampled without over-sampling, to be representative of the whole sample. Third, we trained research assistants to code tweets independently, and resolved inter-coder disagreement or ambiguous cases by discussing with them the tweets on which such disagreement occurred. Based on a pilot study, we decided that each tweet was to be coded by two research assistants and in case of disagreement, we would search for a consensus solution. They reached an inter-coder reliability of 0.89 measured as Fleiss' Kappa which is considered a very high agreement (Landis and Koch, 1977). Disagreement occurred in only 5.2% of cases.

SI4.2 Fine-tuning BERT-based artificial neural network models

To fine-tune the algorithm we use a 10-fold cross validation (for an introduction see Hastie, Tibshirani and Friedman, 2009). This was implemented with Python relying on the Pytorch machine learning library by adapting publicly available code provided as part of the 'Transformers' library of language models (Wolf et al., 2019), available at https://github.com/huggingface/transformers.

We created our main script by editing the provided 'run_glue.py'. We used all the default training parameters except for the following parameters which we found would lead to higher performance on the kind of data we are using: `per_gpu_train_batch_size=64, learning_rate= 2e-5, warmup_steps=.1, max_grad_norm=1.0, num_train_epochs=1.0`.

The model was trained using a distributed training procedure on a GPU equipped workstation configured to perform `fp16` computations (NVidia RTX 3090).

SI4.3 Accuracy of Classification

Our model achieved an excellent classification performance. More precisely, it obtained a precision of .90 and a recall of .79. This means that 90% of tweets the model classified as being on gender issues are actually on gender issues and that 79% of gender issue tweets are classified as being on gender issues. For comparison, we also trained a model that adopts the 'bag-of-words' approach, the naïve Bayes classifier.[15] Our fine-tuned BERT classifier produces about one third of the mistakes produced by the naïve Bayes classifier (39 vs. 140). The confusion matrices for the predictions of our fine-tuned model and of the naive Bayes classifier on the test data are reported in Table SI4.9.

To develop an intuition for the quality of the model predictions, we computed the coefficient of inter-rater reliability (Fleiss' kappa) by assuming the fine-tuned BERT model is a rater, and human categorization by the research assistants is another rater. The obtained coefficient is .83, which is a level generally considered as 'almost perfect agreement.' The same coefficient for the naïve Bayes classifier is .55, which is generally considered as 'moderate agreement.'

SI4.4 BERT-based regression model for tweet feedback prediction

To predict feedback, we relied on an artificial neural network based on BERT Multilingual-cased, just as for identifying tweets on gender issue. The main difference is that the output layer in this case is not a classification layer, but a linear regression layer which takes as an input the 768

[15] Implemented using the Multinomial Naive Bayes model of the scikit-learn machine learning Python package. For details see: https://scikit-learn.org/stable/modules/naive_bayes.html.

Table SI4.9: Confusion matrices for the BERT classifier and the naïve Bayes classifier on the validation dataset (N=1974).

		Model Prediction		
		No	Yes	Total
Human Coding	No	1832	11	1843
	Yes	28	103	131
	Total	1860	114	

(a) BERT Multilingual Cased Classifier

		Model Prediction		
		No	Yes	Total
Human Coding	No	1735	108	1843
	Yes	32	99	131
	Total	1767	207	

(b) Naïve Bayes Classifier

dimension vector output by BERT and outputs predicted feedback as a linear combination of the vector elements.

We split our dataset of tweets into two sets of approximately the same size, by creating a random split of politicians such that all the tweets written by a given politician would fall in one of the two sets (call them set A and set B). We adopted this politician-level split of the data to prevent the algorithm from learning about the communication style of individual politicians and the popularity of the topic of gender issues is with their followers - which it could theoretically do even though no explicit pointers to politicians form part of the input data.

We constructed the measure of feedback for a given tweet by starting the with the number of retweets, taking out the politician level time-trend (eq. 6), taking out day fixed effects, and then normalizing within politicians. The difference between this measure and the measure used in the main analyses (FB_{im}) is the additional inclusion of day fixed effects. Inclusion of period fixed effects was not necessary in the construction of the main measure because these could be included *post-hoc* in the regression analyses. Yet, such *post-hoc* inclusion of fixed effects is not possible for the present purpose because we aim to use the trained model for out-of-sample predictions and thus need to remove the effects of day to day variations at the training stage.

We used all the tweets in set A to fine-tune the model (for one epoch) and applied the resulting model to predict the success for the tweets in set B. We used the same process on the other half of the data: we use the tweets in set B to fine-tune the model (for one epoch) and applied the trained model to predict feedback for the tweets in set A. This procedure allowed us to produce out-of-sample predictions of the amount of feedback expected by a tweet, just based on its semantic content. We would like to emphasize that no features of the tweet author were included as inputs, just the text of the tweet. The correlation between out-of-sample prediction and true feedback was about .50 (.52 for the model fine-tuned on one half of the data and .49 for the model fine-tuned on the other half of the data).

The technical details pertaining to model training are the same as for fine-tuning the BERT-based classifier.

SI5 Differentiate Between Male and Female Followers and Retweeters

We predict if a given Twitter user is male or female based on the Twitter username. For this, we use Genderize.io, a commercial online service that predicts if a name is male or female.

We do this for all followers of politician and hence even if not all user names can be identified as typical female or male, the large number of followers allows us to obtain a clear picture of the share of female or male followers of each politician.

Regarding the identity of retweeters, Twitter only allows to retrospectively download information about up to 100 retweeters per tweet. Furthermore, some of their Twitter user names were not indicative if the retweeter was male or female. Still, for the average tweet in our sample, we obtained a classification for 61% of the retweeters. We estimate the absolute number of female or male retweeters by multiplying the absolute number of retweeters with the estimated share of female and male retweeters of each tweet.

Finally, we apply the same steps of feedback normalization (see Section 3.2) to the retweets of female and male retweets. This is the basis of Figure 5 and Table SI3.8.

SI6 Proof of Equation 3

Lemma 1. *The reinforcement learning model described in Section 2.2 defines a stochastic process for* $\left(V_{im}^{GI}, V_{im}^{other}\right)_{m\geq 1}$ *that has a unique stationary distribution characterized by the following density:*

$$h\left(V^{GI}, V^{other}\right) = e^{-\frac{r^2\sigma^2\gamma}{2(2-\gamma)}} \frac{e^{-rV^{GI}-\pi^{GI}} + e^{-rV^{other}}}{e^{-r\mu^{GI}-\pi^{GI}} + e^{-r\mu^{other}}} g_{GI}(V^{GI}) g_{other}(V^{other}), \tag{11}$$

where, for $k \in \{GI, other\}$*,* $g_k(\cdot)$ *is a normal density with mean* μ^k *and variance* $\sigma^2\gamma/(2-\gamma)$*.*

Proof. This follows from Lemma 2 in Le Mens et al. (2019). □

Proof. The asymptotic probability of choosing the gender topic is obtained by integration of the choice probability (equation 1) with respect to the joint asymptotic density described in Lemma 1.

$$A_\infty^{GI} = \int_{V^{GI}, V^{other}} \frac{1}{1 + e^{-(\pi^{GI} + r(V^{GI} - V^{other}))}} dV^{GI} dV^{other} \tag{12}$$

$$= \frac{e^{-\frac{r^2\sigma^2\gamma}{2(2-\gamma)}}}{e^{-r\mu^{GI}-\pi^{GI}} + e^{-r\mu^{other}}} \int_{V^{GI}, V^{other}} e^{-rV^{other}} g_{GI}(V^{GI}) g_{other}(V^{other}) dV^{GI} dV^{other}, \tag{13}$$

$$= \frac{e^{-\frac{r^2\sigma^2\gamma}{2(2-\gamma)}}}{e^{-r\mu^{GI}-\pi^{GI}} + e^{-r\mu^{other}}} \int_{V^{other}} e^{-rV^{other}} g_{other}(V^{other}) dV^{other}. \tag{14}$$

Noting that $\int_{V^{other}} e^{-rV^{other}} g_{other}(V^{other}) dV^{other}$ is the moment generating function of the distribution $g_{other}(\cdot)$, evaluated at $-r$, we have:

$$\int_{V^{other}} e^{-rV^{other}} g_{other}(V^{other}) dV^{other} = e^{-r\mu^{other} + \frac{r^2\sigma^2\gamma}{2(2-\gamma)}}. \tag{15}$$

We finally obtain:

$$A_{\infty}^{GI} = \frac{e^{-r\mu^{other}}}{e^{-r\mu^{GI} - \pi^{GI}} + e^{-r\mu^{other}}} \tag{16}$$

$$= \frac{1}{1 + e^{-(\pi^{GI} + r(\mu^{GI} - \mu^{other}))}}. \tag{17}$$

□