AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

How Politicians Learn from Citizens' Feedback: The Case of Gender on Twitter

Nikolas Schöll Universitat Pompeu Fabra
Aina Gallego Universitat de Barcelona
Gaël Le Mens Universitat Pompeu Fabra

Abstract: *This article studies how politicians react to feedback from citizens on social media. We use a reinforcement-learning framework to model how politicians respond to citizens' positive feedback by increasing attention to better received issues and allow feedback to vary depending on politicians' gender. To test the model, we collect 1.5 million tweets published by Spanish MPs over 3 years, identify gender-issue tweets using a deep-learning algorithm (BERT) and measure feedback using retweets and likes. We find that citizens provide more positive feedback to female politicians for writing about gender, and that this contributes to their specialization in gender issues. The analysis of mechanisms suggests that female politicians receive more positive feedback because they are treated differently by citizens. To conclude, we discuss implications for representation, misperceptions, and polarization.*

Verification Materials: The materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the American Journal of Political Science Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/DWYLME.

Does feedback from citizens on social media affect the issues that politicians choose to discuss? Recent research on issue responsiveness finds that when an issue becomes salient among citizens on social media, politicians quickly follow and become more likely to discuss it over the next days (Barberà et al. 2019). This finding raises the question of how politicians can learn and respond so rapidly to changes in public mood. In traditional dynamic representation models (Soroka and Wlezien 2010; Stimson, Mackuen, and Erikson 1995; Wlezien 1995), public policy adjusts to shifts in aggregate public opinion over much longer periods of time, typically years, and politicians learn about changes in public opinion through tools that require careful analysis, such as opinion polls (Druckman and Jacobs 2006), expert consensus (Stimson, Mackuen, and Erikson 1995), or by recording and analyzing information (Henderson et al. 2021). These approaches to detecting changes in public opinion do not seem applicable to the social media setting because they are impractical in settings in which new information is highly decentralized and spreads in minutes (Cagé, Hervé, and Viaud 2020). While online information is abundant, unbiased and up-to-date summaries about which issues are relevant for

Nikolas Schöll is a researcher and holds a PhD in Economics from the Universitat Pompeu Fabra (nikoschoell@gmail.com). Aina Gallego is Associate Professor of Political Science at the Universitat de Barcelona and affiliated to the Institut Barcelona d'Estudis Internacionals (aina.gallego@ub.edu). Gaël Le Mens is Full Professor at the Department of Economics and Business, Universitat Pompeu Fabra and affiliated to the Barcelona School of Economics and the UPF-Barcelona School of Management (gael.le-mens@upf.edu).

We thank Ruben Enikolopov, Maria Petrova, David Rossell, Tania Verge, Fabrizio Gilardi, Pablo Barbera, Elias Dinas, Elisa Mougin, and Cyrill Otteni, Mikhail Spektor, Arnout Van De Rijt, 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 2019 Quantitative Text Analysis Dublin Workshop. the 2019 American Political Science Association Annual Meeting, the 2019 European Political Science Association Conference, the 2021 Heidelberg Sampling Symposium, 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-PSI201675353, 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 and BBVA Foundation Grant G999088Q to Gaël Le Mens.

American Journal of Political Science, Vol. 0, No. 0, January 2023, Pp. 1–18

DOI: 10.1111/ajps.12772

citizens are not available. The strategies that politicians use to be responsive to citizens on social media thus remain unspecified to date.

This article focuses on how politicians use one source of information about the preferences of citizens that is continuously available on social media: feedback from citizens. We study how feedback affects subsequent decisions about which political issues to discuss. Contact with politicians has long been considered a relevant way in which citizens can influence politicians' issue agendas (Fenno 1977; Miller and Stokes 1963) but has been difficult to measure. Studying the impact of feedback from citizens is timely as the interactive features of social media have reduced the cost of two-way communication between politicians and citizens and made it more abundant (Jungherr, Rivero, and Gayo-Avello 2020).

In this article, we specify the process through which politicians respond to citizen feedback in terms of a "reinforcement learning" model grounded in research about how people learn from feedback (Holland 1992; Sutton and Barto 2018). We propose that after talking about an issue, politicians observe the amount of positive feedback from citizens, update their perceptions about the popularity of the issue, and respond by increasing attention to popular issues and decreasing attention to unpopular issues.

This simple strategy allows politicians to be continuously responsive, although only to the self-selected citizens who interact with them. A relevant characteristic of social media is that users, including politicians, are exposed to information environments that tend to match and possibly reinforce their preexisting views (Sunstein 2018; Zhuravskaya, Petrova, and Enikolopov 2020). To study the implications of exposure to such fragmented audiences, our reinforcement-learning model allows politicians of different social categories to be exposed to systematically different feedback from citizens. Specifically, we focus on the social category of gender—whether a politician is a female or a male—and the extent to which female and male politicians attend to gender issues. The model shows that, if female politicians receive more positive feedback for talking about gender as compared to male politicians, reinforcement learning creates a difference in attention to gender issues between female and male politicians. The model is general and can apply to other issues and to other social categories such as race or partisanship. It can also apply to offline settings.

We test the theory with rich social media data that record politician–citizen interactions over time and allow longitudinal analysis. We collected 1.5 million tweets published by elected representatives in national and regional assemblies, active during the 2016–19 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 language model which is sensitive to word dependencies, vastly outperforms standard bag-of-word models, and works well in multilingual contexts. We estimate the effect of citizen feedback on attention to gender issues by female and male 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 are responsive to citizen feedback on social media: after receiving more retweets for tweeting on gender issues, they increase attention to this issue. This is also the case with "likes." Moreover, we find that female and male politicians are exposed to systematically different feedback environments: female politicians receive relatively more retweets and likes for tweeting about gender issues. This leads them to talk more about gender issues. Our analyses of mechanisms also reveal that female politicians obtain more feedback because they are treated differently by citizens, and not because their messages are more engaging or because of differences in the composition of audiences.

Our study advances research on how politicians respond to changes in public opinion. It is most related to Barberà et al. (2019), who document issue responsiveness on social media but do not study the underlying mechanisms. More generally, theoretical models of dynamic representation remain unclear about how politicians learn about public opinion (e.g., Stimson, Mackuen, and Erikson 1995). We propose and test one specific learning process that allows politicians to be continuously responsive to the citizens with whom they interact. Methodologically, we develop an empirical approach that allows the analysis of actual interactions between politicians and citizens on social media, instead of relying on inferences from population-wide averages. Substantively, we document, for the first time, that the direct interactions between politicians and citizens influence the issues that politicians choose to discuss on social media and show that differential treatment from citizens leads politicians with different characteristics to diverge in issue attention.

We also contribute to the large literature on the political representation of women by connecting the gender-specific experiences of women in office to the rise of attention to gender issues. Theoretical work about descriptive representation argues that female representatives are more likely to talk about issues relevant to women because they have different experiences both in life and in office (Mansbridge 1999; Phillips 1995). Empirical research supports the claim that descriptive representation increases substantive representation (Clayton 2021; Lawless 2015; Lowande, Ritchie, and Lauterbach 2019; Wängnerud 2009), but it has more difficulties at examining specific mechanisms that link both. In particular, existing empirical studies have not been able to isolate the effects of specific gendered experiences in office on politicians' attention to gender issues. We demonstrate that exposure to systematically different feedback environments contributes to differences in attention to gender issues between female and male politicians beyond what can be explained by differences in intrinsic motivation or preexisting preferences. Furthermore, our analyses of mechanisms shed light on why this happens.

Learning from Feedback on Social Media

To be responsive to citizens, politicians first need to learn about citizens' preferences both in terms of issue salience and issue position.[1] How do politicians learn about the preferences of the public? Dynamic representation theory (Stimson, Mackuen, and Erikson 1995) provides one answer to this question. While recognizing that politicians cannot directly know the preferences of the public, this theory proposes that all politicians have access to a "consensus view" about the direction of change in preferences which is produced by a community of opinion leaders, including politicians, journalists, and academics. In a similar spirit, thermostatic models of public opinion (Wlezien 1995, 2004) assume that politicians are aware of directional changes in aggregate public opinion.

The assumption that all politicians have access to a shared view about the preferences of the public may be well-suited to homogeneous information environments, as was the case when a few broadcast television channels were dominant and thus everyone was exposed to similar information (Prior 2007). However, the low barriers to entry and the reliance on user-generated content have made online information environments much more fragmented than traditional media environments (Zhuravskaya, Petrova, and Enikolopov 2020). Politicians, like other users, are not exposed to content that is centrally produced by gatekeepers and similar for all users, but to content that depends on whom they choose to follow, which users choose to interact with them, and on how algorithms prioritize information. An additional challenge is that new topics appear and disseminate online at a high speed (Cagé, Hervé, and Viaud 2020). This reduces the usefulness of tools such as traditional opinion polls to track changes in public opinion. Since social media platforms do not provide systematic information about the average preferences of citizens on political issues, politicians must find other approaches to learn about them.

Research on how representation operates in practice finds that when politicians (or their aides) try to learn about citizens' preferences, they rarely use tools like surveys, which are often not available. Instead, they keep track of their communication with interest groups and regular citizens and make inferences based on this information (Fenno 1977; Henderson et al. 2021; Miller and Stokes 1963).

Information obtained through direct interactions with citizens, and in particular the feedback they provide, is particularly relevant in social media contexts because it is abundant, immediately available, and easy to use. Before the advent of social media, citizens communicated their opinions to politicians through actions such as writing letters or talking in public meetings which require civic skills and are relatively costly (Verba, Schlozman, and Brady 1995). The built-in interactive features of social media, such as the ability to provide feedback to other users through easily clickable buttons, have made two-way communication between citizens and politicians easier and more convenient (Jungherr, Rivero, and Gayo-Avello 2020). Moreover, feedback is obtained in real time. As Zhuravskaya, Petrova, and Enikolopov note, "Social media allows politicians to receive immediate feedback on policy actions, to discuss policy proposals, and to measure political discontent" (2020, 417). Finally, feedback in social media is more easily usable than traditional communication with constituents because it comes in a highly standardized quantitative form (such as the number of retweets, likes, or hearts), which makes it easy to compare how different statements fare. Thus, we expect that politicians use the feedback they obtain on social media to make inferences about citizens' preferences.

How do politicians use feedback? To address this question, 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

[1] We focus on issue salience in this article because the decision to talk about an issue or not is binary, and this facilitates empirical analysis. However, the logic applies to issue positions as well.

because they see themselves as delegates of the public.[2] However, politicians are uncertain about the popularity of the issues they might discuss.

We propose that politicians learn about the popularity of issues by observing how their messages are received by the public and that they increase attention to issues that obtain more positive feedback than expected and reduce attention to those that obtain less positive feedback than expected. In short, issues that obtain relatively more positive feedback are "reinforced." Prior research has shown that people frequently behave this way when they make repeated choices between options with uncertain payoffs and aim to obtain positive payoffs (Denrell 2005; Thorndike 1927) and that this behavior is often reasonable (Holland 1992; Le Mens and Denrell 2011; Sutton and Barto 2018). In the context of politicians writing on Twitter, the options consist in different political issues which they can choose to discuss in their next tweet. Feedback is the reaction of the citizens to the politicians' tweets, which can be more positive or negative than expected. 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. We analyze a formal model of this learning-from feedback process in the section entitled "Reinforcement Learning by an Individual Politician" and provide empirical estimates of the model parameters in the section entitled "Responsiveness to Issue-Specific Feedback."

A key drawback of relying on feedback as a source of information is that citizens who provide it are self-selected, and politicians cannot know in which way the preferences of their followers differ from the preferences of the population at large (or of other relevant groups, such as copartisans or voters in their districts). While politicians and their staff are aware that their online followers are not representative of the public (Henderson et al. 2021), they have no way to fully correct the ensuing biases.[3]

If politicians of different social categories, such as gender or race, are exposed to more positive feedback from the public when they talk about issues related to their social categories, they will form different perceptions of what the public wants and will ultimately be more likely to talk about issues related to their social category. Our study focuses on gender, which is a more politically relevant characteristic than race in the Spanish context. We expect that female politicians receive relatively more positive feedback from citizens when they talk about gender issues rather than on other issues—a difference in feedback that, from now on, we call the "gender-issue feedback advantage." There are several reasons why the gender-issue feedback advantage would be larger for female politicians than for male politicians.

First, female politicians may communicate more engagingly about gender issues because they are more knowledgeable and interested in these issues (Dolan 2011; Lawless 2015; Lowande, Ritchie, and Lauterbach 2019), and this more engaging style may in turn generate more positive reactions from citizens. There exists abundant evidence that female representatives have different positions on gender issues than male representatives (e.g., Lovenduski and Norris 2003), although whether female politicians communicate more engagingly about gender has not been rigorously assessed. We call this mechanism the "engagingness channel."

Second, female citizens may interact more with female politicians. This argument has been advanced most clearly by Mansbridge (1999, 641) who argues that politicians of traditionally marginalized groups provide better representation to ingroup members because they have "enhanced communication" with them. Empirical research finds that citizens are more likely to contact politicians of their race (Broockman 2014; Gay 2007), although there is less direct evidence about gender (for null results, see, for instance, Bush and Prather 2021; Haynes 1997). If female citizens "self-select" into interacting more with female politicians and female users are more likely to give feedback to tweets on gender issues, this could potentially explain the gender-issue feedback advantage. We call this mechanism related to the composition of audiences the "self-selection channel."

Third, citizens may believe that female politicians are more competent to talk about gender issues (Dolan 2010; Huddy and Terkildsen 1993) and, for this reason, may provide them with more positive feedback for tweeting on the topic even if there is no difference in the content of the gender-issue tweets written by female and male politicians. Recent research finds that partisanship or incumbency dominate gender stereotypes when citizens decide for which candidate to vote (Dolan 2014; Lawless 2015). But this does not rule out that voters reward female politicians for behaving according

[2]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 recognize that politicians sometimes 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 et al. 2011; Pereira 2021).

[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).

to stereotypes in social media contexts, where voters are not restricted in the amount of feedback they can provide and thus do not need to prioritize one consideration over others. Research in social psychology and sociology supports the claim that people tend to evaluate the behavior of others more positively if it is congruent with expectations related to their social categories (Eagly, Wood, and Diekman 2000; Hannan et al. 2019). If Twitter users expect female politicians to talk more about gender issues, they may react more positively when they do, because this is *congruent* with their expectations regarding the issues female politicians should attend to. These arguments imply that citizens are more likely to retweet tweets on gender issue when they are published by female politicians, rather than male politicians, even if there is no difference in tweet content. We call this mechanism the "congruity channel."

We empirically test for the differences in gender-issue feedback advantage feedback in the section entitled "Gender-Issue Feedback Advantage for Female and Male Politicians" and test for the three potential mechanisms in the section entitled "Mechanisms for the Difference in Gender-Issue Feedback Advantage between Female and Male Politicians."

Model

Reinforcement Learning by an Individual Politician

Consider a politician i who publishes a series of messages on policy issues. Without loss 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. In reinforcement-learning models, agents have latent "valuations" of each option, which they update based on feedback. The valuation of different policy issues can be interpreted as politicians' perception of the popularity of that issue. Politician i's valuation of the "gender issues" option at the time they decide on the issue of message m is $V_{i,m}{}^{\mathrm{GI}}$ and the valuation of the "other issues" option is $V_{i,m}{}^{\mathrm{other}}$. The politician is more likely to choose "gender issues" if the difference in valuations favors this issue, that is, 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":

$$A_{i,m}^{\mathrm{GI}} = \mathrm{Logit}\left(\pi_i^{\mathrm{GI}} + r\Delta V_{i,m}\right), \tag{1}$$

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

We denote by $V_{i,1}^{\mathrm{GI}}$ and $V_{i,1}^{\mathrm{other}}$ the initial valuations of the two issues. After every message m, the politician observes the feedback $FB_{i,m}^{k}$ 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 (see Appendix A, p. 2, in the online supporting information for a discussion of this assumption). Formally, if message m is on issue k, then

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

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

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:

$$\mathrm{FB}_{i,m}^{\mathrm{GI}} \sim N\left(\mu_i^{GI}, \sigma\right);\ \mathrm{FB}_{i,m}^{\mathrm{other}} \sim N\left(\mu_i^{\mathrm{other}}, \sigma\right). \tag{3}$$

It is possible to derive a formula for the long-run share of attention to gender issues, A_{∞}^{GI} (see proof in Appendix A, p. 1, in the online supporting information)

$$A_{\infty}^{GI} = \mathrm{Logit}\left(\pi_i^{GI} + r\Delta\mu_i\right), \tag{4}$$

where $\Delta\mu_i = \mu_i^{\mathrm{GI}} - \mu_i^{\mathrm{other}}$ is the difference between the means of the feedback distributions for the two issues ("gender" and "other"). This corresponds to what we call 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. It is noteworthy that the long-run attention to gender issue does not depend on the initial valuations. This means that our main result holds whether the politician initially believes average feedback for the two issues to be the same or different (see Appendix A, p. 2, in the online supporting information for further discussion of this).

Differences between Female and Male Politicians

Now consider two hypothetical politicians, F and M, who behave according to the reinforcement-learning model but are exposed to different feedback environments such that the gender-issue feedback advantage differs between the two politicians ($\Delta\mu_F \neq \Delta\mu_M$). Using Equation (4), we can derive a necessary and sufficient condition for a difference in long-run issue attention, such that attention to the gender issue is larger for F than for M:

$$A^{GI}_{F,\infty} > A^{GI}_{M,\infty} \Leftrightarrow \pi^{GI}_F + r_F \Delta\mu_F > \pi^{GI}_M + r_M \Delta\mu_M. \quad (5)$$

This difference in issue attention can emerge as the result of a difference in the feedback received by F and M.

A feedback-driven difference in valuations and issue attention can emerge even if F and M have identical baseline propensities for publishing tweets on gender issues ($\pi^{\mathrm{GI}}_F = \pi^{\mathrm{GI}}_M$) and are equally responsive to changes in issue valuations ($r_F = r_M$). 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 ($\Delta\mu_F > \Delta\mu_M$). We discuss model dynamics for different values of the initial valuations in Appendix A, p. 3, in the online supporting information.

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 to write about gender issues when the following condition holds:

$$r_F \Delta\mu_F > r_M \Delta\mu_M. \quad (6)$$

We test whether the condition in Equation (6) holds in the section entitled "Responsiveness to Issue-Specific Feedback."

Case, Data, and Measurement

To analyze whether and how citizen feedback affects politicians' issue attention, we collected the tweets published 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 Twitter usernames of 1,530 national or regional parliamentarians. 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 1,265 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). Additional summary statistics are reported in Appendix C, p. 9, in the online supporting information.

Measuring Attention to Gender Issues

The main empirical challenge consisted of identifying tweets related to gender issues. We used human-coded data to train and validate a text classifier based on a state-of-the-art deep-learning language model, BERT (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 or not. We chose this model, because it has been shown to perform much better than "bag-of-words" classifiers which are most often used in the social sciences (Grimmer and Stewart 2013). We recruited research assistants to code about 20,000 tweets as being on gender issues or not, and 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.

Our model achieved an excellent classification performance on our validation data: 90% of the tweets the model classified as gender-issue tweets are actually on gender issues, and 79% of gender issue tweets are classified as such. For comparison with the more traditional "bag-of-words" approach, we trained a naïve Bayes classifier. It produced three times more mistakes than our BERT classifier. We discuss the advantages of BERT, the coding details, how we fine-tuned the model and model accuracy in Appendix B, p. 5, in the online supporting information.

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: $A^{\mathrm{GI}}_{ip} = n^{\mathrm{GI}}_{ip}/N_{\mathrm{ip}}$.

There exists a large difference in attention to gender issues by female and male politicians. Over the entire sample period, female politicians devoted, on average, 11.2% of their tweets to gender issues whereas male

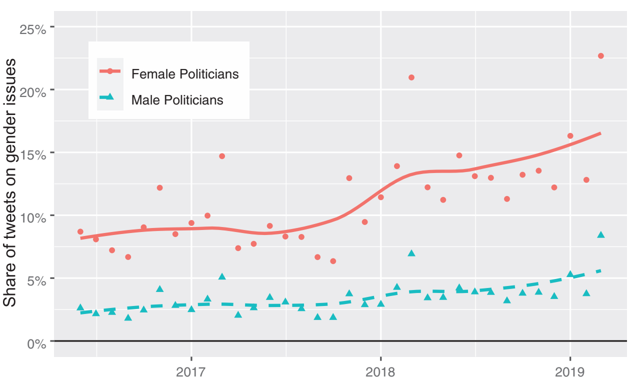
TABLE 1 Summary Statistics for Male and Female Politicians

	Female Politicians	Male Politicians
Number of politicians	554	711
Number of tweets (mean)	1,087.9	1,380.1
Number of tweets (median)	568.0	697.0
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
Average number of replies (mean)	3.4	7.5
Average number of replies (median)	0.5	0.7
Standard deviation retweets (mean)	52.1	95.9
Standard deviation retweets (median)	10.2	12.7
Standard deviation likes (mean)	38.6	80.2
Standard deviation likes (median)	8.4	10.4
Standard deviation replies (mean)	10.9	22.0
Standard deviation replies (median)	1.5	2.0

Notes: 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.

politicians only devoted 3.4% of their tweets to gender issues. Figure 1 depicts the average attention to gender issues by female and male politicians over the period studied.

FIGURE 1 Attention to Gender Issues by Male and Female Politicians



Notes: Line is a smoothed time trend. Points represent monthly averages.

Comparing the mean number of raw retweets that each politician received for tweets on gender issues and other issues reveals the existence of gender-issue retweet advantage for female politicians (see Table 2). Tweets on gender issues written by female politicians receive on average 18% more retweets in absolute terms than tweets on other issues. By contrast, male politicians receive about the same number of retweets for tweeting about gender issues and other issues. A similar asymmetry between female and male politicians holds for likes.

Measuring Issue-Specific Feedback

We construct our main measure of citizen feedback based on the number of *retweets*. Prior research has shown that a higher number of retweets implies approval (Metaxas et al. 2015). Consistent with the view that most retweets are instances of positive feedback, we observe in our data

TABLE 2 Summary Statistics: Retweets, Likes, and Replies

	Female Politicians			Male Politicians		
	Gender Issues	Other	Difference between Gender Issues/Other	Gender Issues	Other	Difference between Gender issues/Other
Number of tweets (mean)	123.5	966.4		49.6	1332.5	
Number of tweets (median)	46.0	506.0		23.0	683.0	
Average number of retweets (mean)	25.5	21.6	+18%	45.5	45.7	0%
Average number of retweets (median)	7.4	6.2	+19	7.4	7.4	0
Average number of likes (mean)	45.4	37.2	+22	83.9	79.9	+5
Average number of likes (median)	9.4	8.3	+13	10.1	10.3	−2
Average number of replies (mean)	3.9	3.3	+18	7.4	7.5	−1
Average number of replies (median)	0.4	0.5	−12	0.6	0.7	−25

Notes: 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. There were 554 observations for female politicians and 711 for male politicians.

that most of the retweets between politicians happen within parties (see Appendix D, p. 10, in the online supporting information). Rather than using the raw number of retweets as the measure of feedback to politician i about the tweet message m they published, we construct a feedback measure grounded in behavioral research on how past experience affects future decisions. We proceed in several steps.

First, we take the natural logarithm of the number of retweets. This transformation is motivated by research that shows that payoffs have declining marginal effects (Tversky and Kahneman 1992). Taking the logarithm also reduces the weight of instances of extremely large numbers of retweets which have the potential to drive the model estimation results.[4] Differences in logs express scale-invariant ratios of feedback, implying that the added utility of receiving 10% more *retweets* would be the same for a politician who usually receives 10 or 10,000 retweets.

Second, we take out a politician-specific time trend.[5] This step is motivated by research on learning from feedback that has shown that agents tend to evaluate outcomes with respect to a time-dependent "aspiration level" or reference point (Cyert and March 1963; March and Shapira 1992). In our context, the average number of retweets increases over time for most politicians, probably because the politicians' followership is growing. Thus, comparing the number of retweets received by tweets published many months or several years apart is not meaningful.

Finally, we proceed to within-politician z-score standardization. The relevant comparison for a given politician to learn about issue popularity is to compare the number of retweets they received for tweeting on a specific issue with the average level of retweets they receive themselves, rather than the number of retweets that other politicians received.[6] By construction, the distribution of feedback for each politician now has mean zero ($E[\mathrm{FB}_{i,m}] = 0$) and standard deviation one ($\sigma_{FBi,m} = 1$).

Our feedback measure can be interpreted as follows: a one-unit increase in feedback means that the tweet received one standard deviation more in "feedback utility units" relative to other tweets published by the same politician around the same point in time. We focus on retweets over likes because information about retweeters is more easily available on Twitter than information

[4]The number of retweets is strongly skewed and approximately follows a power-law distribution for each politician: the median tweet received four retweets, the mean is 58 retweets, and the maximum is almost 43,000 retweets.

[5]We regress log retweets$_{i,m}$ on the time t the tweet was posted using OLS and then take the residual:

$$\widehat{u_{i,m}} = \log \text{retweets}_{i,m} - \widehat{trend}\,(\log \text{retweets}_i) * t \tag{7}$$

[6]As a robustness check, we replicate our main analyses by omitting within-politician normalization in Appendix E, p. 53, in the online supporting information. Our main results remain.

about those who gave likes, and we use this information in some analyses. Results are similar with likes (Appendix E, p. 12, in the online supporting information). In ancillary analyses we also analyzed replies (Appendix E, p. 12).

Results

Gender-Issue Feedback Advantage for Female and Male Politicians

We estimate by ordinary least squares (OLS) a set of linear models with feedback as the dependent variable. In our baseline specification, the feedback received by politician i for tweet message m, $FB_{i,m}$, is regressed on politician gender and the issue of the tweet:

$$FB_{i,m} = \beta_{GI} GI_{i,m} + \beta_M M_i + \beta_{GI*M} GI_{i,m} * M_i + \epsilon_{i,m}, \quad (8)$$

where $GI_{i,m}$ is a dummy variable equal to 1 if tweet m published by politician i is on gender issue, M_i is a dummy equal to 1 if politician i is male and $\epsilon_{i,m}$ 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 it is negative, the gender-issue feedback advantage is stronger for female politicians. In most specifications, we include politician fixed effects to absorb the effect of politician characteristics which remain constant over time such as their gender, specialization of policy area, or political party. We also add day and hour of the day fixed effects to absorb the effect of temporal variations affecting all politicians such as general shifts in issue salience.

Estimation results are reported in Table 3. In all specifications, the gender-issue feedback advantage is stronger for female politicians ($\beta_{GI*M} < 0$, $p < .01$ in Model 1–3, $p < .05$ in Model 4). Model 1 is a basic specification without controls or fixed effects. We find that the gender-issue feedback advantage is larger for female politicians (+0.24 standard deviation) than for male politicians (+0.15 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-wing and right-wing politicians (see Appendix E, p. 12, in the online supporting information for details on coding).

Appendix F, p. 14, in the online supporting information reports the robustness checks.

Responsiveness to Issue-Specific Feedback

Do politicians increase attention to gender issues after obtaining relatively more positive feedback? To address this question, we estimate the parameters of the reinforcement-learning model described in the section entitled "Model Using Two-Way Fixed-Effect Logistic Panel Models."

To render the data amenable to analysis using panel models, we discretize it into fixed-length time periods p. We use the calendar month as the time period 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).

To estimate the latent-issue valuations, 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, that is, $V_{ip}^k = V_{i,m}^k\ [m =$ first message in period $p]$. We take the valuation at 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. Hence, assuming feedback still affects issue attention after several months does not seem realistic. Therefore, we restrict our analysis 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 published by the politician in the respective month (N_{ip}) as regression weights. Each tweet thus receives the same weight in our estimations.

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^G. 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 the equation with month fixed effects, τ_p, politician fixed-effects, π_i, and time-varying control variables. Global shifts in issue attention over time are captured by the month fixed effect. For example, around March 8, 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 capture other time-invariant confounds such as their gender, party, region, policy focus, etc., as well as time-invariant

TABLE 3 Female Politician Receive More Positive Feedback for Tweeting on Gender Issues

	(1)	(2)	(3)	(4)
Gender issues	0.24**	0.29**	0.28**	0.27**
	(0.00)	(0.02)	(0.02)	(0.05)
Gender issues ∗ Male politician	−0.09**	−0.12**	−0.12**	−0.12*
	(0.01)	(0.02)	(0.02)	(0.05)
Part of thread			0.29†	0.29†
			(0.15)	(0.15)
Tweets on day by politician			−0.01**	−0.01**
			(0.00)	(0.00)
Gender issues ∗ Left				0.02
				(0.06)
Gender issues ∗ Male politician ∗ Left				0.00
				(0.06)
Male politician	0.02**			
	(0.00)			
(Intercept)	−0.03**			
	(0.00)			
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

Notes: Linear regressions of tweet feedback on politicians' gender and issue of the tweet (variations of Equation 8). Standard errors are clustered by politician in specifications with fixed effects: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

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 must be estimated jointly with the issue-attention equation. The full model thus consists of two equations, jointly estimated as a generalized linear model using GLS:

$$\begin{cases} V_{i,m}^{k} = (1-\gamma)\ V_{i,m-1}^{k} + \gamma FB_{i,m-1}^{k} \\ A_{ip}^{GI} = \text{ Logit}\left(\pi_i^{GI} + r{*}\Delta V_{ip} + \tau_p + \epsilon_{ip}\right). \end{cases} \tag{9}$$

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.

TABLE 4 Reinforcement Learning Model: Results

	(1)	(2)	(3)	(4)	(5)	(6)
Difference in Valuation	0.09** (0.02)					
Difference in Valuation $_{\text{Female politician}}$		0.11** (0.03)				
Difference in Valuation $_{\text{Male politician}}$		0.07** (0.02)				
Valuation$^{\text{gender issues}}$ Female politician			0.14** (0.04)	0.12** (0.04)	0.13** (0.04)	0.14** (0.04)
Valuation$^{\text{gender issues}}$ Male politician			0.13** (0.03)	0.13** (0.03)	0.13** (0.03)	0.13** (0.03)
Valuation$^{\text{other}}$ Female politician			−0.08** (0.03)	−0.03 (0.03)	−0.06* (0.03)	−0.08** (0.03)
Valuation$^{\text{other}}$ Male politician			−0.03 (0.03)	−0.03 (0.03)	−0.03 (0.03)	−0.03 (0.03)
Individual trend				5.52** (0.44)		
Lagged DV					1.01** (0.10)	
Social influence						−0.05 (0.86)
$\hat{\gamma}$ (to calculate valuation)	0.07	0.07	0.07	0.08	0.07	0.07
Fixed effects						
Politician	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Squared correlation	0.57	0.57	0.57	0.60	0.58	0.57
Observations	18,482	18,482	18,482	18,482	18,482	18,482

Notes: Logistic regression of the share of tweets written on gender issues on the valuation of gender issues and other issues (see model in Equation 9). ΔValuation is defined as the difference in valuation of gender issues and other issues. $\hat{\gamma}$ is the estimated updating parameter used to calculate issue valuations. All regressions use cell-size regression weights, i.e., number of tweets published by politician i in month p (N_{ip}). Standard errors are clustered by politician: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

Estimation results are reported in Table 4. Model 1 corresponds to Equation (9). 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 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 in attention to gender issues of 7.7%.[7] 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_{ipF} the valuation difference if politician i is female and ΔV_{ipM} 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 8.3% (1.00 percentage points) whereas male politicians' issue attention increases by 6.8% (0.25 percentage points). The difference between these two estimates is not statistically significant ($p > .1$).

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 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. Again, we do not find significant differences between female and male politicians ($p > .1$).

We report robustness checks in Appendix F, p. 14, in the online supporting information. Our main results persist when controlling for politician-specific trajectories in issue attention, serial correlation, or peer effects (Model 4, 5, and 6 in Table 4). They are also robust to alternative specifications that employ different feedback measures (based on likes, replies, or a retweet-based feedback measure with no within-politician normalization), a different time period to compute issue attention (weeks instead of month), a different weighting scheme of observations, or a more substantive reference category instead of "other." We do not find statistically significant differences in responsiveness between female and male politicians in any of the specification (always $p > .05$).

Finally, we conduct a placebo test by randomly swapping politicians' issue valuations with the feedback-based valuation of another politician of the same gender (male or female), for the same issue, and in the same month. Using another politician's valuation leads to null results across all specifications. Thus, the robustness checks confirm that politicians are responsive to feedback. Clearly, politicians adjust their attention to gender issues in response to the feedback they receive for their tweets on this issue.

In the section entitled "Model," we specified conditions for feedback to contribute to a difference in attention to gender issues between female and male politicians (Equation 6). The model most apt to this comparison is Model 2 in 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 female and male politicians (see Model 3 in Table 3), we obtain:

$$r_F \Delta \mu_F = 0.11 * 0.28 > r_M \Delta \mu_M = 0.07 * 0.16. \quad (10)$$

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

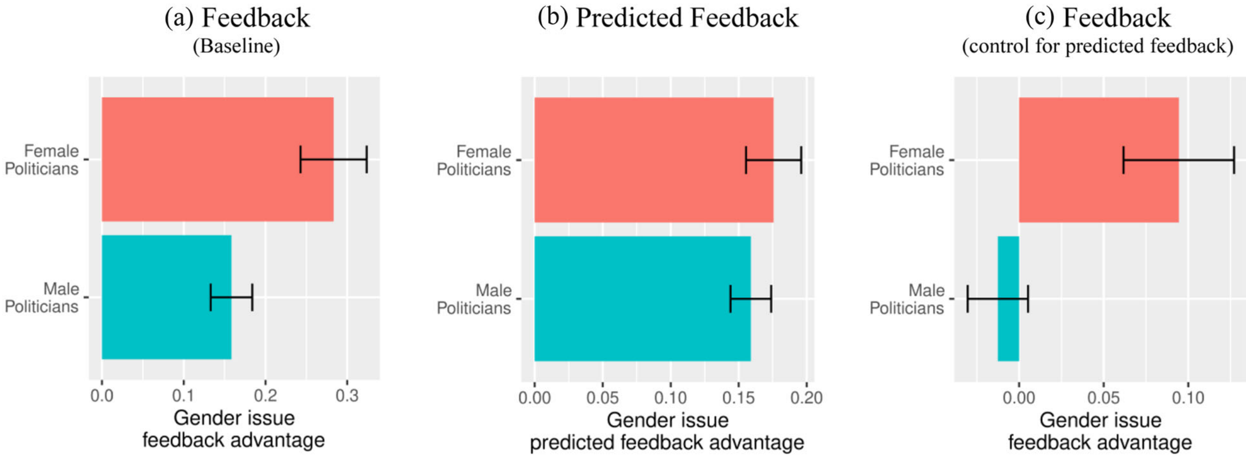
Mechanisms for the Difference in Gender-Issue Feedback Advantage between Female and Male Politicians

In this section, we report empirical tests of the three potential mechanisms for the difference in gender-issue feedback advantage discussed in the section entitled "Differences between Female and Male Politicians."

The Engagingness Channel. The "engagingness channel" posits that female politicians write relatively more engaging tweets on gender issues compared to male politicians. We measured how engaging is a tweet by predicting the retweet-based feedback solely on the text of the tweet. For this, we trained a BERT model to predict the feedback received by a tweet just based on its content. Importantly, the model does not take any information about the identity of the politician who published a tweet as input. As such, the *predicted feedback* is a measure of how engaging is a tweet, independent of the social category of the politician who published it (including their gender—see Appendix B, p. 8, in the online supporting information).

[7] 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): $\widehat{AME} = \frac{1}{N}\sum_{i=1}^{I}\sum_{p=1}^{P} N_{ip}(\mathrm{Logit}(\widehat{\pi_i^{GI}} + \hat{r}*1 + \hat{\tau}_p) - \mathrm{Logit}(\widehat{\pi_i^{GI}} + \hat{r}*0 + \hat{\tau}_p))$.

FIGURE 2 Gender-Issue Feedback Advantage between Female and Male Politicians and Tweet Engagingness



Notes: Panel (a) shows that the gender-issue feedback advantage is larger for female politicians than for male politicians ($\Delta^{F-M} = 0.12^{**}$). Panel (b) shows that tweets by female and male politicians are similarly engaging ($\Delta^{F-M} = 0.02$). We measured how engaging is a tweet as the predicted feedback based on the text of the tweet. Panel (c) shows that the difference in gender-issue feedback advantage between female and male politicians remains almost the same when controlling for predicted feedback ($\Delta^{F-M} = 0.11^{**}$). Black bars represent 95% confidence intervals.

Consider the difference in *predicted feedback* for gender-issue tweets and tweets on other issues. We call it the "gender-issue *predicted feedback* advantage." If the "engagingness channel" operates, we expect this difference to be larger for female politicians than for male politicians.

Furthermore, we expect the gap in gender-issue feedback advantage between female and male politicians would disappear once we control for how engagingly tweets are written.

Figure 2 describes the key results based on model estimations reported in Table SI7.7 in the online supporting information; Models 2 and 3. Panel (a) reports the gender-issue feedback advantage for female and male politicians according to the baseline model (Model 3 in Table 3). Panel (b) shows that whether the politician is female or male hardly affects how engaging is a tweet: the average predicted feedback is almost the same for tweets of female and male politicians. Finally, panel (c) shows that the difference in gender-issue feedback advantage is almost the same when controlling for *predicted feedback* as with the baseline model. These two findings imply that *predicted feedback* does not explain the difference in gender-issue advantage.

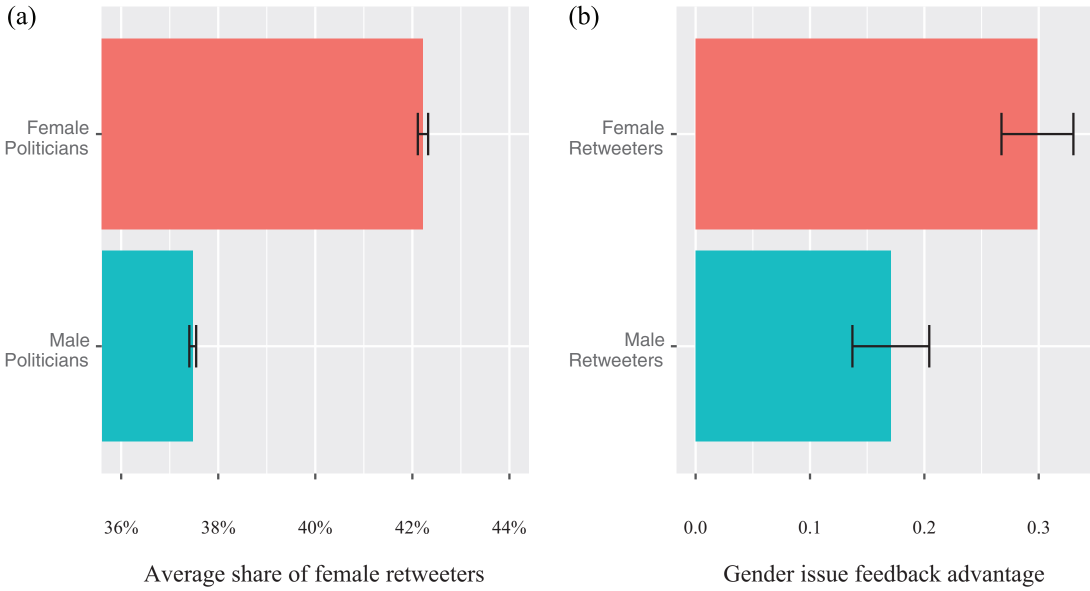
In ancillary analyses, we use stylistic features as another measure of how engaging is a tweet. We code for sentiment (from negative to positive), the number of words (tokens), hashtags, mentions, emojis, and if a tweet contains a link or a graphic element (picture or video) as alternative mediators. Model 4 in Table SI7.7 in the online supporting information shows that the coefficient of the interaction term, β_{GI*M}, remains similar to that obtained in the baseline model when controlling for stylistic features. Hence, stylistic features do not contribute much to the difference in gender issue advantage (see Appendix G, p. 20, in the online supporting information).

In conclusion, we do not find support for the "engagingness channel."

The Self-Selection Channel. The "self-selection channel" posits that Twitter users are more likely to provide feedback to politicians of the same gender as them and that female Twitter users are more likely to provide feedback on gender-issue tweets as compared to tweets on other issues.

We first provide evidence for the hypothesis that Twitter users are more likely to provide feedback to politicians of the same gender. For this, we classified retweeters as female or male by applying a name-recognition algorithm to their Twitter username (see Appendix G, p. 21, in the online supporting information). We find differences in the gender composition of Twitter users who provide feedback to female or male politicians. The average share of female retweeters is 4.7 percentage points larger for female politicians (see Figure 3a). This difference holds for gender issue tweets (+11.5 percentage points) and for tweets on other issues (+4.9

FIGURE 3 Gender-Issue Feedback Advantage between Female and Male Politicians and Gender Composition of the Set of Retweeters



Notes: Panel (a) shows that tweets from female politicians attract a larger share of female retweeters. Panel (b) shows that female Twitter users more likely to retweet gender issue tweets. Black bars represent 95% confidence intervals.

percentage points). Hence, we find evidence that female citizens self-select into interaction more with female politicians.

To show that female Twitter users are relatively more likely to retweet tweets on gender issues, we construct the same reference-dependent standardized measure of feedback as described in the section entitled "Measuring Issue-Specific Feedback," separately for female and male retweeters. This allows us to analyze how female and male retweeters react to gender-issue tweets versus tweets on other issues. We find that the gender-issue feedback advantage is almost twice as strong among female Twitter users compared to male Twitter users (see Figure 3b, and Models 6 and 7 in Table SI7.7 in the online supporting information).

In conclusion, we find evidence in support of the self-selection channel.

The Congruity Channel. The "congruity channel" posits that, when Twitter users decide whether to retweet a tweet, their decision is affected by the interaction of the gender of the politician who published the tweet and the issue of the tweet, such that users are more likely to retweet a gender-issue tweet if it was published by a female politician rather than a male politician, even after controlling for tweet content.

This mechanism differs from the "self-selection channel" in that the latter focuses on the composition of the audience of a politician whereas the "congruity channel" focuses on the behavior of the audience members. Accordingly, to test the hypothesis that the "congruity channel" operates, we change the perspective from the politicians to the Twitter user as the unit of analysis. We assembled a sample of individual Twitter users and their retweeting behavior. For each user u, we take the set of tweets published by all politicians whom the user followed and test if a given user is more likely to retweet a tweet about gender if it was published by a female politician. We include a series of fixed effects to control for the general propensity of the user to retweet gender-issue tweets and the user's general propensity to retweet a given politician—independently of the policy issue of the tweet.

To be able to include these fixed effects, we focus on users who follow multiple politicians. For computational reasons, we focus on a subsample of the most active retweeters.[8] More specifically, we estimate the following logistic regression:

$$\text{retweet}_{i,u,m} = \text{Logit}(\beta_{GI*M} * GI_m * \text{M}_i + GI_m \times user_u FE + politician_i \times user_u FE + \epsilon_{i,u,m}). \quad (11)$$

[8]We selected the 1,000 male and 1,000 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 Gender of Politician

	(1)	(2)	(3)	(4)
Gender issues ∗ Male politician	-0.11^{**}	-0.13^{**}		
	(0.02)	(0.02)		
Gender issues ∗ Male politician ∗ Female user			-0.14^{**}	-0.15^{**}
			(0.03)	(0.03)
Gender issues ∗ Male politician ∗ Male user			-0.09^{**}	-0.10^{**}
			(0.03)	(0.03)
Predicted feedback		1.03^{**}		1.03^{**}
		(0.01)		(0.01)
Fixed effects				
Gender issues ∗ Retweeter	Yes	Yes	Yes	Yes
Retweeter ∗ Politician	Yes	Yes	Yes	Yes
Fit statistics				
Squared correlation	0.14	0.15	0.14	0.15
Observations	4,276,978	4,276,978	4,276,978	4,276,978

Notes: Logistic regression of the retweeting probability on the interaction of a dummy regarding the issue of the tweet (gender issues or other) and the social group of the politician (male or female). Standard errors in parenthesis are clustered at the levels of the fixed effects $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

The dependent variable retweet$_{i,u,m}$ is a dummy equal to 1 if tweet message m published by politician i was retweeted by user u. The main coefficient of interest is the interaction between the politician being male and the tweet being on gender issues, β_{GI*M}. Under the hypothesis that the congruity channel operates, we expect a negative coefficient.

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 issue dummy, $GI_m \times user_u$, and we control for all time-invariant aspects of the politicians–user interaction (general propensity to retweet a given politician) by including a set of politician-by-user fixed, $politician_i \times user_u$.

Estimation results are reported in (Table 5). In Models 1 and 2, the coefficient of the interaction term, β_{GI*M}, is negative and strongly significant. The marginal effect implies that a given user is 9% less likely to retweet a tweet on gender issues if it was published by a male politician. Models 3 and 4 reveal that the effect is similar for female and male users (difference not statistically significant, $p > .05$).

In summary, we find clear evidence for the "congruity channel."

Discussion

In this article, we advance the understanding of how politicians interact with citizens on social media by studying how the feedback that politicians obtain from citizens affects their issue attention through the lens of a reinforcement-learning model. We show that politicians respond to feedback by adjusting issue attention and that politicians from different social groups are exposed to different feedback. 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 than male politicians. Our analyses suggest that this difference in feedback exists because citizens treat politicians differently depending on their gender ("self-selection" and "congruity" channels), and not because female politicians approach the issue in a more engaging way. The difference in feedback environments to which female and male politicians are exposed leads them to focus on different issues.

Reinforcement learning allows politicians to be responsive, but only to the self-selected set of citizens who choose to interact with them. Being responsive to other entities, such as the median voter, may be more desirable from a normative perspective, but reinforcement learning is not conducive to responsiveness to such entities because politicians lack information about the preferences of citizens they do not see and cannot perfectly adjust for biases in the feedback they receive. 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). Yet, information about the average views of the public is 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. Uncovering when politicians rely on reinforcement learning versus public opinion polls or other tools to form perceptions of public opinion 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 (Zhuravskaya, Petrova, and Enikolopov 2020). Our results imply that politicians shift attention to issues relevant to citizens with whom they personally interact. 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. Our approach could be combined with advances in text-scaling methods to code the "extremity" of tweets and study citizen-driven political polarization.

In the analysis of mechanisms, we demonstrate that citizens treat men and women politicians differently while we do not find any direct evidence that female politicians write more engaging tweets. Even if we made every effort to measure quantifiable differences in the content and style of the tweets written by male and female politicians, our analyses may have missed more subtle differences. Specifically, a possibility worth further investigation is that male and female politicians frame gender issues differently and that such framing differences affect citizens' reaction to tweets.

The study of politicians' behavior on Twitter is important in its own right since this behavior has real consequences (Jungherr 2016). Still, an important next step would study the extent to which feedback on Twitter affects politicians' offline behavior. Furthermore, we suspect that the mechanism we study in this article generalizes to any setting in which politicians interact with an audience. The situation we describe, in which politicians can choose to talk about many different issues but are uncertain about which issues are better received, does not only exist in social media settings, but is pervasive in political life. There are many instances in which politicians can get impressions about which statements fare well with an audience, such as through the volume of applause in campaign meetings or TV shows. Whether politicians respond to feedback in a reinforcement-learning fashion in offline settings 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 stronger gender-issue feedback advantage for female politicians strengthens the case for descriptive representation. Our findings 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 group-based specialization and the relegation of representatives from underrepresented social categories to niche issues. We show that experiences in social media push Spanish female representatives to specialize in gender issues. Such pressure is likely to exist in other countries and to apply to other types of representatives, such as ethnic and racial minority representatives or LGBT representatives. Future work should aim to uncover if the differences in the feedback environments of politicians from different social categories affect their political careers.

References

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.

Broockman, David E. 2014. "Distorted Communication, Unequal Representation: Constituents Communicate Less to Representatives Not of Their Race." *American Journal of Political Science* 58(2):307–21

Bush, Sarah Sunn, and Lauren Prather. 2021. "Islam, Gender Segregation, and Political Engagement: Evidence from an Experiment in Tunisia." *Political Science Research and Methods* 9(4):728–44.

Butler, Daniel M., and David W. Nickerson. 2011. "Can Learning Constituency Opinion Affect how Legislators Vote? Results from a Field Experiment." *Quarterly Journal of Political Science* 6(1):55–83.

Cagé, Julia, Nicolas Hervé, and Marie-Luce Viaud. 2020. "The Production of Information in an Online World." *The Review of Economic Studies* 87(5): 2126–64.

Campillo, In´es. 2019. "'If We Stop, the World Stops': The 2018 Feminist Strike in Spain." *Social Movement Studies* 18(2): 252–8.

Clayton, Amanda. 2021. "How Do Electoral Gender Quotas Affect Policy?" *Annual Review of Political Science* 24: 235–52.

Conover, M. D., J. Ratkiewicz, and M. Francisco. 2011. "Political Polarization on Twitter." *Proceedings of the Fifth International Conference on Weblogs and Social Media* 133(26): 89–96.

Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ. pp. 169–187.

Denrell, Jerker. 2005. "Why Most People Disapprove of Me: Experience Sampling in Impression Formation." *Psychological Review* 112(4): 951–78.

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* 32(1): 69–88.

Dolan, Kathleen. 2011. "Do Women and Men Know Different Things? Measuring Gender Differences in Political Knowledge." *The Journal of Politics* 73(1): 97–107.

Dolan, Kathleen A. 2014. *When Does Gender Matter?: Women Candidates and Gender Stereotypes in American Elections*. USA: Oxford University Press.

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):453–76.

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, Richard F. 1977. "US House Members in their Constituencies: An Exploration." *American Political Science Review* 71(3):883–917.

Fiedler, Klaus. 2012. "Meta-Cognitive Myopia and the Dilemmas of Inductive-Statistical Inference." *Psychology of Learning and Motivation* 57: 1–55.

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.

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): 267–97.

Hannan, Michael T., Gaël Le Mens, Greta Hsu, Balazs Kovács, Giacomo Negro, László Pólos, Elizabeth Pontikes, and Amanda J. Sharkey. 2019. *Concepts and Categories: Foundations for Sociological and Cultural Analysis*. New York: Columbia University Press.

Haynes, Elizabeth. 1997. "*Women and Legislative Communication*." Harvard University. Cambridge, Massachusetts: Typescript.

Henderson, Geoffrey, Alexander Hertel-Fernandez, Matto Mildenberger, and Leah C. Stokes. 2021. "Conducting the Heavenly Chorus: Constituent Contact and Provoked Petitioning in Congress." *Perspectives on Politics* : 1–18. First view. https://doi.org/10.1017/S1537592721000980C

Holland, John Henry. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, Massachusetts: MIT press.

Huddy, Leonie, and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and Female Candidates." *American Journal of Political Science* 37(1):119–47.

Jungherr, Andreas, Gonzalo Rivero, and Daniel Gayo-Avello. 2020. *Retooling Politics: How Digital Media are Shaping Democracy*. Cambridge, Massachusetts: Cambridge University Press.

Jungherr, Andreas. 2016. "Twitter Use in Election Campaigns: A Systematic Literature Review." *Journal of Information Technology & Politics* 13(1):72–91.

Lawless, Jennifer L. 2015. "Female Candidates and Legislators." *Annual Review of Political Science* 18:349–66.

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):379–92.

Lovenduski, Joni, and Pippa Norris. 2003. "Westminster Women: The Politics of Presence." *Political Studies* 51(1):84–102.

Lowande, Kenneth, Melinda Ritchie, and Erinn Lauterbach. 2019. "Descriptive and Substantive Representation in Congress: Evidence from 80,000 Congressional Inquiries." *American Journal of Political Science* 63(3):644–59.

Mansbridge, Jane. 1999. "Should Blacks Represent Blacks and Women Represent Women? A Contingent "yes"." *The Journal of Politics* 61(3):628–57.

Mansbridge, Jane. 2003. "Rethinking Representation." *American Political Science Review* 97(4):515–28.

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 Proceedings of the International AAAI Conference on Web and Social media. pp. 658–61.

Miller, Warren E., and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1):45–56.

Pereira, Miguel M. 2021. "Understanding and Reducing Biases in Elite Beliefs About the Electorate." *American Political Science Review* 115(4):1308–24.

Phillips, A. 1995. *The Politics of Presence*. Cambridge, Massachusetts: Oxford University Press.

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press.

Rama, José, Lisa Zanotti, Stuart J. Turnbull-Dugarte, and Andrés Santana. 2021. *VOX: The Rise of the Spanish Populist Radical Right*. Routledge.

Soroka, Stuart N., and Christopher Wlezien. 2010. *Degrees of Democracy: Politics, Public Opinion, and Policy*. Cambridge University Press.

Stimson, James A., Michael B. Mackuen, and Robert S. Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89(3):543–65.

Sunstein, Cass R. 2018. *# Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Sutton, R. S., and A. G. Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press.

Thorndike, E. L. 1927. "The Law of Effect." *The American Journal of Psychology* 39(1): 212–22.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5(4): 297–323.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Harvard University Press.

Wängnerud, Lena. 2009. "Women in Parliaments: Descriptive and Substantive Representation." *Annual Review of Political Science* 12:51–69.

Wlezien, Christopher. 1995. "The Public as Thermostat: Dynamics of Preferences for Spending." *American Journal of Political Science*: 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, et al. 2020. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12: 415–38.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Model - Proof and Additional Analyses
Appendix B: Deep Learning Tweet Issue Classifier
Appendix C: Descriptive Statistics
Appendix D: Evidence for Retweets as Positive Feedback
Appendix E: Gender Issue feedback Advantage – Robustness
Appendix F: Responsiveness to Feedback – Robustness
Appendix G: Mechanisms – Additional results