Supporting Information for:
Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech

Bryce J. Dietrich* Matthew Hayes† Diana Z. O'Brien‡

Contents

*Corresponding Author. Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242 (bryce-dietrich@uiowa.edu, http://www.brycejdietrich.com).

†Department of Political Science, Rice University, 105 Herzstein Hall, 6100 Main St., Houston, TX 77005 (matthew.hayes@rice.edu, http://www.matthewjhayes.com).

‡Department of Political Science, Rice University, 105 Herzstein Hall, 6100 Main St., Houston, TX 77005 (dzobrien@rice.edu, http://dianaobrien.com).

S1 Description of Floor Speech Data

In this section we describe the *HouseLive* video archives, as well as our approach to extracting the text and vocal pitch of legislators' floor speeches from these archives. We also outline our approach for analyzing legislators' vocal pitch.

S1.1 House Video Archive

We chose to use *HouseLive*[1] as our data source because of their easily accessible audio and video archives. To conserve space and reduce processing time, we use their `mp3` files instead of `mp4` files. In total, we downloaded 863 `mp3` files spanning from January 6, 2009 to August 4, 2014.

To identify both the speakers and text of floor speeches, we used the closed-captioning information provided by *HouseLive*. Unlike the *Congressional Record*, closed-captioning information has the advantage of reporting verbatim what is said on the House floor.

One drawback to this approach is that since closed-captions are produced in real-time, typographical errors may be a concern. In email correspondence, the company that performs the closed-captioning service for the House of Representatives asserts that their transcribers are generally 95 percent accurate, meaning that 95 percent of the time the words that are transcribed are the words actually spoken on the House floor. This assessment is based on yearly evaluations, in which the company randomly selects a certain number of transcripts from each of their transcribers and determines the degree to which those transcripts capture the floor debate for that day. For this study, we transcribed 100 randomly selected speeches. When these speeches were compared to the closed-captioning information, regardless of the similarity measure one used, the closed-captions were essentially the same as the transcripts. Based on these results and our communication with the closed-captioning company, we are confident that the closed-captioning found on *HouseLive* is an accurate reflection of what is said in the U.S. House of Representatives.

The individual `mp3` files we collected averaged 7 hours and 27 minutes in length. Using the open-source software `ffmpeg`, we split these longer audio files into individual speeches using the time stamps found in the closed-captioning information and converted these into `wav` files. This resulted in 152,117 `wav` files. Due to the large number of extremely short speeches, we restricted our analysis to those speeches that had at least 50 words. This resulted in text and audio for 74,158 speeches.

[1] `http://houselive.gov/`

S1.2 Extracting Vocal Pitch

From this corpus of `wav` files, we used the software *Praat*[2] to extract vocal pitch.[3] This software implements the algorithm outlined by Boersma (1993). Similar to other algorithms that focus on time-domain periodicity, *Praat* estimates the fundamental frequency by dividing the autocorrelation of a windowed signal by the autocorrelation of the window itself. One must assume the signal is stationary within each window, which is why the algorithm divides the audio file into small segments (around 60ms), then takes the average.[4]

As explained in the main text (and below), women typically speak at a higher average vocal pitch than men. In order to make inferences across male and female MCs, one thus has to standardize each speaker's vocal pitch by their baseline. We accomplish this by subtracting the speaker's mean vocal pitch across all speeches from their pitch in a given speech and then dividing by the standard deviation of the speaker's pitch.[5] For example, Linda Sánchez's (D-CA) mean vocal pitch is 216.71Hz with a standard deviation of 26.35Hz. If she gave a speech with a vocal pitch of 250Hz, our standardized measure would be $\frac{250-216.71}{26.35} = 1.26$, suggesting for that speech her vocal pitch was a little over one standard deviation higher than her baseline.

After computing standardized vocal pitch measures for all available speeches, we added several additional controls. Party identification and ideology were obtained from *Voteview*. Race and seniority were obtained from *GovTrack*, and we identified committee chairs using the data provided by Stewart and Woon (2016).[6] We also included two dummy variables. One of these variables returns a 1 when the speech was given in an election year. The other variable returns a 1 when the speech was less than one-minute in length. Along these lines, we also included the length of the speech in minutes.

S2 Vocal Pitch and Emotional Intensity

One of the central claims of this paper is that vocal pitch can be used as an indicator of emotional intensity. Since this is a novel measure in political science, it is important to validate that pitch can, indeed, be used to measure emotional activation. To this end, in

[2] `http://www.fon.hum.uva.nl/praat`

[3] We extract vocal pitch using only voiced speech. Generally speaking, an utterance could be composed of (1) voiced speech, (2) unvoiced speech, and (3) silence. Although there is some debate over whether to use unvoiced speech when estimating the fundamental frequency (for review, see Hess 2007), for the most part scholars tend to use only voiced speech.

[4] Specifically, to use this software, one has to set five parameters: the pitch floor, pitch ceiling, window length, window shape, and voicing threshold. For the pitch floor and ceiling, we used *Praat* suggested settings, meaning for men, we set the pitch floor to 75Hz and the ceiling to 300Hz. For women, we used a pitch range of 100 to 500Hz. For both the window shape and voicing threshold we used the default settings.

[5] Dietrich, Enos and Sen (2019) also explain why standardization can help account for any unsystematic measurement errors associated with the algorithm used to extract the fundamental frequency.

[6] `http://web.mit.edu/17.251/www/data_page.html`

this section we provide a more thorough theoretical definition of emotional intensity and its link to pitch, and offer several validation exercises aimed at verifying that our measure of vocal pitch corresponds to changes in emotional intensity.

S2.1 The Circumplex Model of Affect and Vocal Pitch as a Measure of Emotional Activation

Our paper draws extensively from the work of James Russell (e.g., Russell 1980, 2003). The circumplex model of affect from that work posits that all affective states arise from two neurophysiological systems, one related to a pleasure-displeasure continuum (called "valence") and the other related to alertness (called "arousal" or "activation"). According to Russell (2003), at any given moment, one's emotional disposition is a single blend of these two dimensions. The horizontal dimension ranges from one extreme (e.g., agony) through a neutral point to its opposite extreme (e.g., ecstasy). For our purposes, we are interested in the vertical dimension, which ranges from a deactivated emotional state, such as being sleepy, to an activated emotional state, ultimately culminating in "frenetic excitement" (Russell 2003, 148). In the context of legislative speech, we call this "emotional intensity."

We offer vocal pitch as a reasonable measure of this arousal/intensity dimension. Specifically, in their review of emotional measurements Mauss and Robinson (2009) state:

> The assessment of vocal characteristics appears to be especially useful in understanding levels of emotional arousal, with higher levels of pitch and amplitude associated with higher levels of arousal (Table 1). By contrast, attempts to link emotional valence or discrete emotions to vocal characteristics have been met with mixed success at best, although more sophisticated methods may be capable of doing so in the future. Thus, we conclude that vocal characteristics are primarily reflective of the dimension of emotional arousal (225-226).

Generally speaking, the relationship between vocal pitch and emotional intensity is due to an automatic physiological reaction in which our muscles – including our vocal cords – naturally tighten when we are emotionally activated. According to Posner, Russell and Peterson (2005), when sensory stimuli are present, emotional arousal is likely relayed to the reticular formation (RF) through the amygdalorecticular pathways (Koch and Ebert 1993; Rosen et al. 1991). This broadly increases activity in the cerebral cortex (Heilman, Watson and Valenstein 2011; Jones 2003), which triggers changes in muscle tone and in the sweat glands (Jones 2003), both of which are associated with subjective ratings of emotional arousal (Lang et al. 1993). This increased blood flow to the muscles also causes vocal cords to contract naturally, raising the fundamental frequency (F_0) of one's voice.

Vocal pitch is not the only way to measure emotional intensity, nor is it the only audio variable that scholars should study. Rather, we simply suggest that vocal pitch is a reasonable measure of emotional intensity and should be seriously considered by those interested in both speech-as-data and audio-as-data approaches. We introduce vocal pitch as an important audio feature, especially for those interested in understanding elite emotional expression.

Although we believe vocal pitch can be an important measure for political scientists, we recognize that it will be unfamiliar to many in the field. For that reason, we offer here a number of validation exercises to justify the use of vocal pitch as a measure of emotional intensity. In the section that follows, we provide four additional validation exercises which collectively provide strong evidence that vocal pitch is measuring the activation dimension of Russell (2003)'s two-dimensional model.

S2.2 Additional Validation #1: Data from Goudbeek and Scherer (2010)

Our first validation exercise involves presenting some of the results from Goudbeek and Scherer (2010). In this study, the authors use the Geneva Multimodal Emotion Portrayals (GEMEP) corpus to understand how vocal characteristics are related to arousal and valance. The GEMEP corpus "contains a large set of systematically controlled portrayals of emotional expressions and is ideally suited for research on emotional response patterning" (Goudbeek and Scherer 2010, 1323). These include 12 emotional states grouped into "high" and "low" arousal and categorized as either "positive" or "negative" in valence, meaning actors were asked to portray twelve emotional states and Goudbeek and Scherer (2010) then categorized those emotions as being either an activated or deactivated emotional state.

It is important to note that in keeping with the literature, the data from Goudbeek and Scherer (2010) rely on actor portrayals of emotions. To create these data, trained actors are asked to portray different emotional states which the authors define as being either more or less intense. These actors are generally *not* specifically prompted to increase their vocal pitch. As explained in Bänziger and Scherer (2007), actors are provided with short definitions of the emotional states and scenarios they are to perform. For those concerned with the use of actor portrayals, Scherer (2013) examined the use of actor portrayals versus exogenous inducement of emotions. "[T]he data demonstrated that under both procedures (acted vs. induced emotions), the expression in the voice was almost the same on measures such as... F_0," as well as speech rate, energy, spectral, and temporal parameters." This "rejects the claim that acted or portrayed emotion expressions are artificial, exaggerated, and falsely prototypical when compared with induced emotional expressions" (Scherer 2014, 226).

As a validation exercise, we reprint the second panel of Figure 1 from Goudbeek and Scherer (2010), which provides strong evidence that mean vocal pitch is related to emotional intensity. As we show in Figure S1, their data demonstrate the relationship between vocal pitch and emotional intensity. In Panel A, we highlight high-intensity positive emotions, which include "Amusement," "Pride," and "Joy," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the lowest mean pitch (for "Pride"). In Panel B, we highlight low-intensity positive emotions, which

Table S1: The Twelve Emotions Included in Goudbeek and Scherer (2010) and Their Abbreviations

	Valence	
Arousal	Positive	Negative
High	Elation (joy)	Hot anger/rage (ang)
	Amusement (amu)	Panic fear (fea)
	Pride (pri)	Despair (des)
Low	Pleasure (ple)	Cold anger/irritation (irr)
	Relief (rel)	Anxiety/worry (anx)
	Interest (int)	Sadness/depression (sad)

Note: Reproduction of Table 1 from Goudbeek and Scherer (2010); boldface highlighting associated with high potency/control emotions added to aid interpretation.

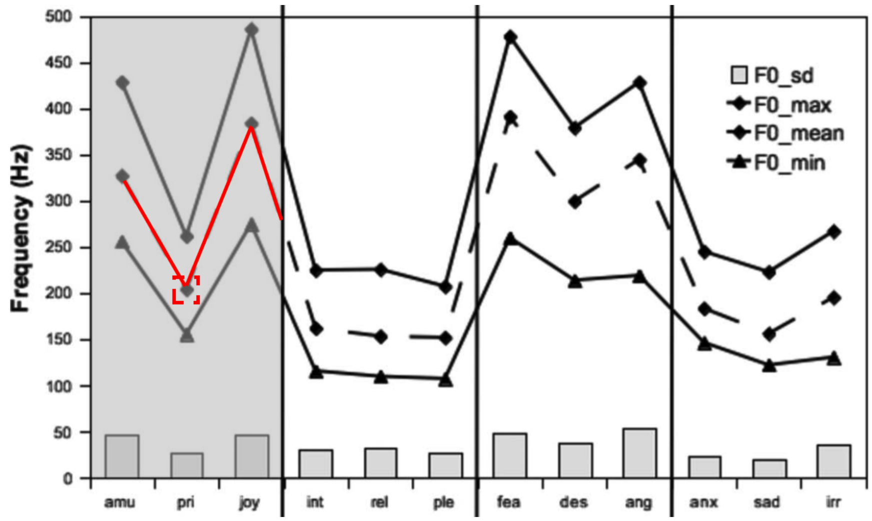
include "Interest," "Relief," and "Pleasure," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the highest mean pitch (for "Interest"). Comparing across these two panels, it is clear that the mean vocal pitch for these high-intensity positive emotions is consistently higher than the mean vocal pitch for low-intensity positive emotions. In fact, by comparing the lowest vocal pitch for high-intensity emotions with the highest vocal pitch for low-intensity emotions (identified with red boxes), we can see that the mean vocal pitch is *always* higher for these high-intensity emotions. This suggests that mean vocal pitch can reasonably discriminate between high- and low-intensity positive emotions.

In Figure S2, we demonstrate a similar finding for the mean vocal pitch for negative emotions. Here, again we re-print the second panel of Figure 1 in Goudbeek and Scherer (2010). In Panel A, we highlight the high-intensity negative emotions, which include "Fear," "Despair," and "Rage," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the lowest mean pitch (for "Despair"). In Panel B, we highlight the low-intensity negative emotions, which include "Anxiety," "Sadness," and "Irritation," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the highest mean pitch (for "Irritation"). As with positive emotions, we see that low-intensity negative emotions always have a lower vocal pitch than high-intensity negative emotions. Taken together, this provides strong evidence from past research that vocal pitch should offer some validity in assessing the intensity of emotions.

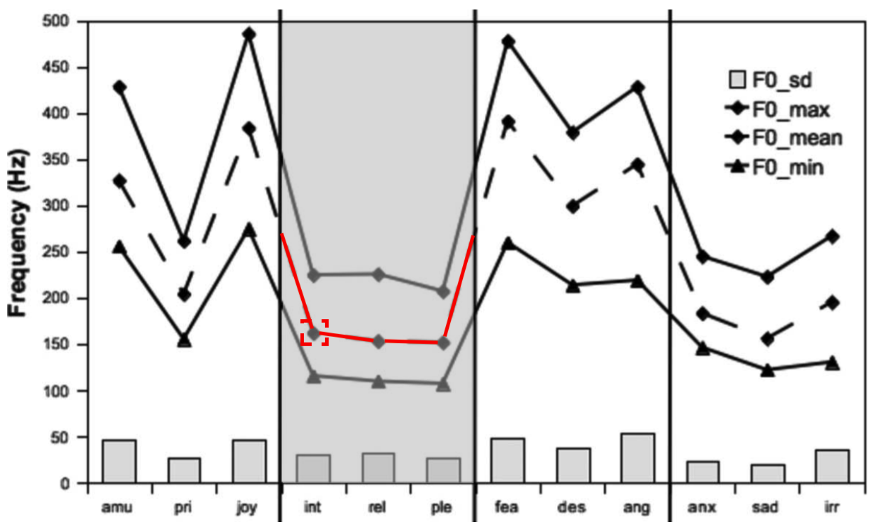
S2.3 Additional Validation #2: Data from Laukka (2004)

Our second validation exercise also uses actor portrayals. However, unlike GEMEP, Laukka (2004) considers only five base emotions: anger, disgust, fear, happiness, and sadness. He

Figure S1: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010) (Positive Emotions)



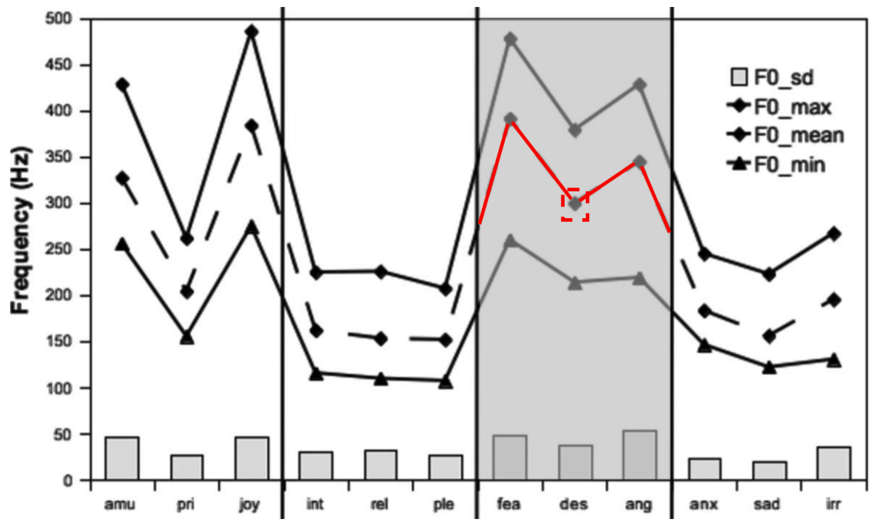
(a) High Intensity



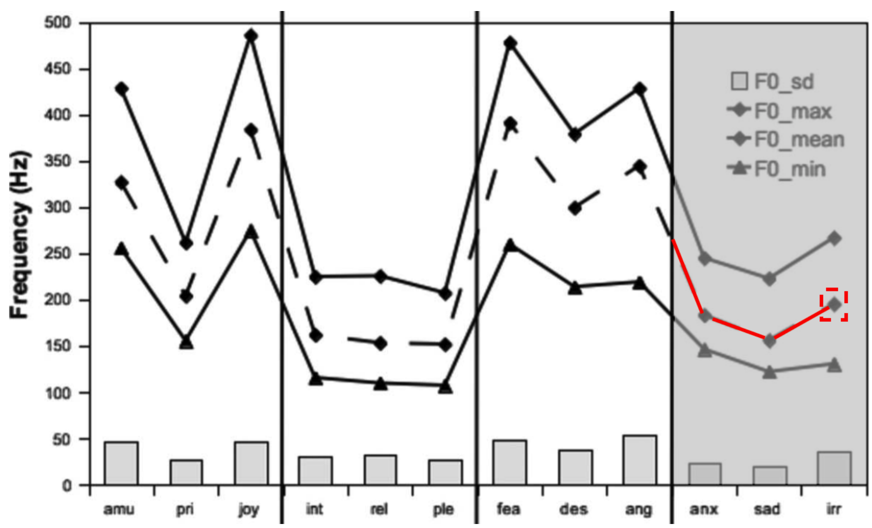
(b) Low Intensity

Note: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010). Additional grey shading and red highlights added to aid interpretation. In Panel A, red highlighting indicates mean vocal pitch for high intensity positive emotions, with a red box indicating the lowest mean vocal pitch. In Panel B, red highlighting indicates mean vocal pitch for low intensity positive emotions, with a red box indicating the highest mean vocal pitch.

Figure S2: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010) (Negative Emotions)



(a) High Intensity



(b) Low Intensity

Note: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010). Additional grey shading and red highlights added to aid interpretation. In Panel A, red highlighting indicates mean vocal pitch for high intensity negative emotions, with a red box indicating the lowest mean vocal pitch. In Panel B, red highlighting indicates mean vocal pitch for low intensity negative emotions, with a red box indicating the highest mean vocal pitch.

measures emotional intensity by simply asking the actors to display anger with more/less intensity. This is distinct from Goudbeek and Scherer (2010), who asked actors to display "Rage" (high-intensity negative emotion) and "Irritation" (low-intensity negative emotion), yet the results are largely the same. More specifically, Laukka (2004) asks actors to vary their level of intensity, whereas Goudbeek and Scherer (2010) simply coded specific emotions as being either more or less intense.

Figure S3 reprints the first panel of Laukka (2004)'s Figure 1. Here, we show the mean vocal pitch when actors are asked to display "high" and "low" intensity emotions. Comparing Panel A to Panel B, it is clear that vocal pitch is significantly higher when actors are displaying emotions with "high intensity." Indeed, not only is the mean vocal pitch highlighted in Panel A always higher than the mean vocal pitch highlighted in Panel B, but none of the confidence intervals overlap. This provides additional evidence that vocal pitch can reasonably differentiate between more and less intense emotional expressions.

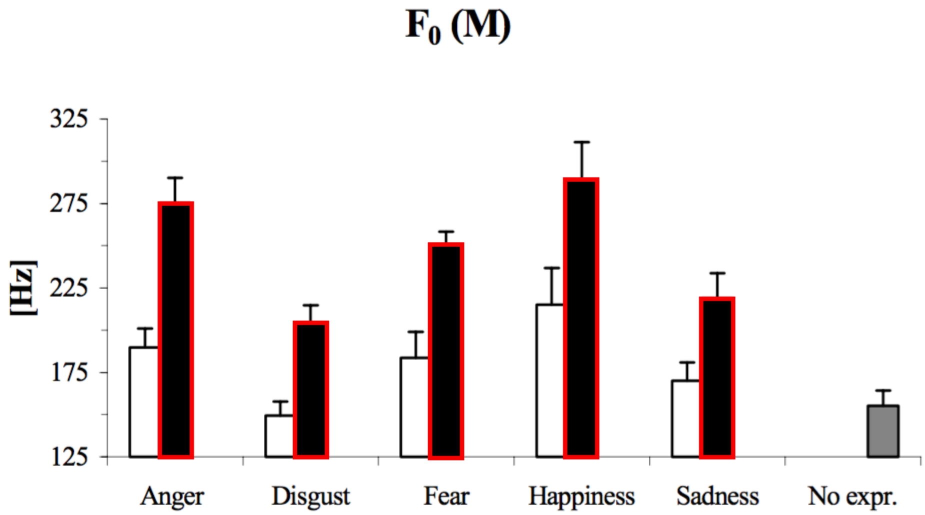
S2.4 Additional Validation #3: New Analysis of Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

In our third validation exercise, we analyze the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). This corpus contains 7,356 high-quality recordings of emotionally-neutral statements, spoken and sung with a range of emotions by 24 American male and female actors. The speech data consists of 8 emotional expressions: neutral, calm, happy, sad, angry, fearful, surprise, and disgust, each of which were delivered at two levels of emotional intensity: normal and strong. Similar to Laukka (2004), the actors were asked to vary their level of intensity prior to speaking, whereas Goudbeek and Scherer (2010) declared whether emotions were intense based on previous literature. Even though audio and video data is available, we restrict our analyses to the audio-only corpus, leaving us with 2,452 unique vocalizations. This valuable data set represents the most readily accessible source of a large number of unique emotional vocalizations. Perhaps more importantly, the emotional vocalizations have been validated by a team of 297 independent coders, making these data very well-suited for validating our measure of vocal pitch as an indicator of emotional intensity.

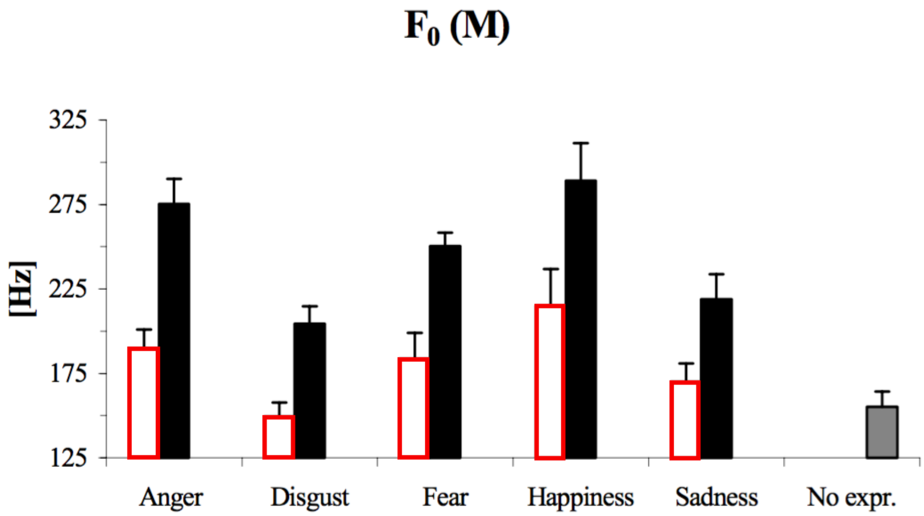
Our validation results can be found in Figure S4. Dark grey bars indicate the actor was portraying an emotion with strong (or high) emotional intensity, whereas light grey bars indicate the actor was asked to portray the same emotion with normal (or low) emotional intensity. We provide 95% confidence intervals for each estimate. Because women tend to speak at a higher vocal pitch than men, we also subdivided the data by male and female speakers.

Our findings are similar for both men and women. First, when all emotional categories are combined (bars labeled "All" and highlighted in red), high-intensity emotions are delivered at a significantly higher vocal pitch. This is the most important result for our study, since it provides additional corroborating evidence that mean vocal pitch can be used to

Figure S3: Reproduction of Figure 1, Panel 1 from Laukka (2004)



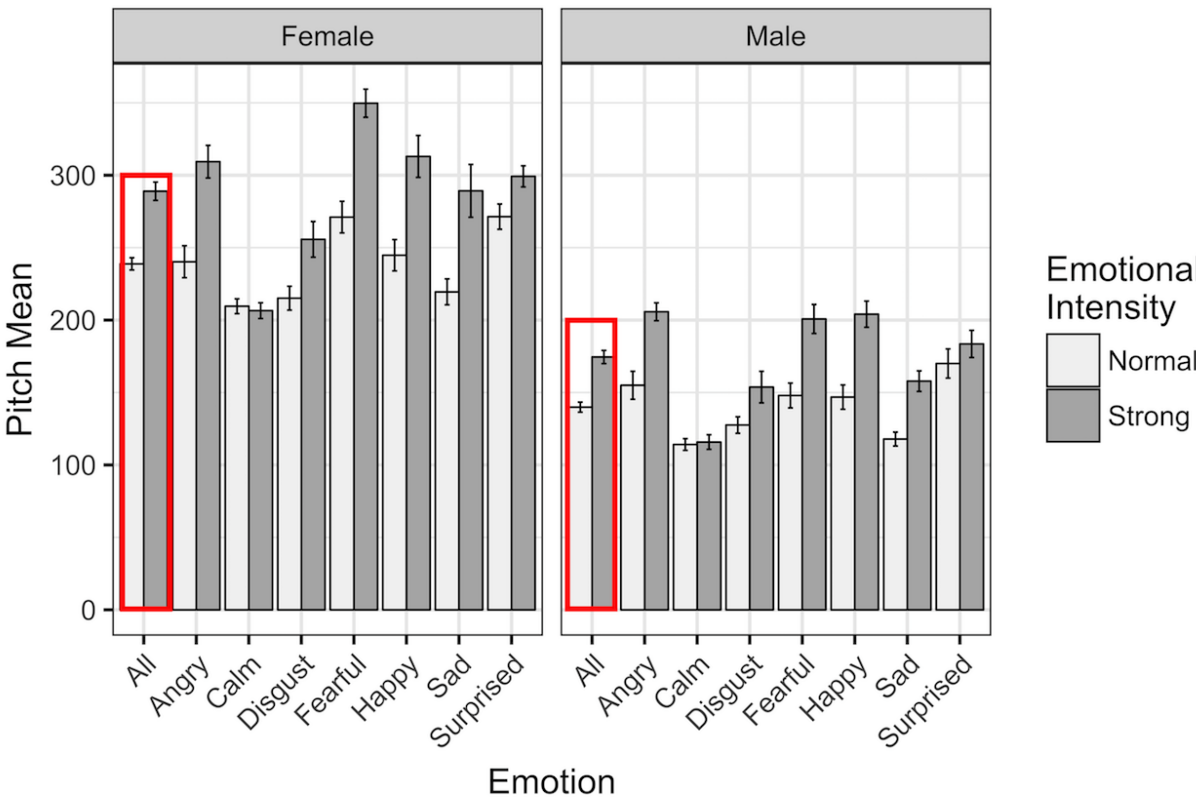
(a) High Intensity



(b) Low Intensity

Note: Reproduction of Figure 1, Panel 1 from Laukka (2004). Red boxes added to aid interpretation. In Panel A, red boxes indicate emotions expressed with high intensity. In Panel B, red boxes indicate emotions expressed with low intensity.

Figure S4: Emotions with “Strong” Intensity Delivered at a Higher Vocal Pitch Than Emotions with “Normal” Intensity (RAVDESS)



Note: Figure uses audio data from Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). Light grey boxes indicate emotions displayed with “normal” intensity. Dark grey boxes indicate emotions displayed with “strong” intensity. Base emotions displayed on the x-axis. Red boxes highlight the trend across all emotions.

discriminate between different levels of emotional intensity. Second, vocal pitch is higher for high-intensity emotions in every category, except for "Calm." We think this is telling. Not only is it difficult to think of what "high intensity" calmness looks like, but Russell (2003) actually describes emotional activation as "frenetic excitement" (148). Given that, we contend calmness is essentially synonymous with a less activated emotional state which is why it is not surprising that the speakers conveyed the high and low versions of this emotion using similar vocalizations. Outside of calmness, vocal pitch is significantly higher across all the emotional categories when those emotions are delivered with high intensity.

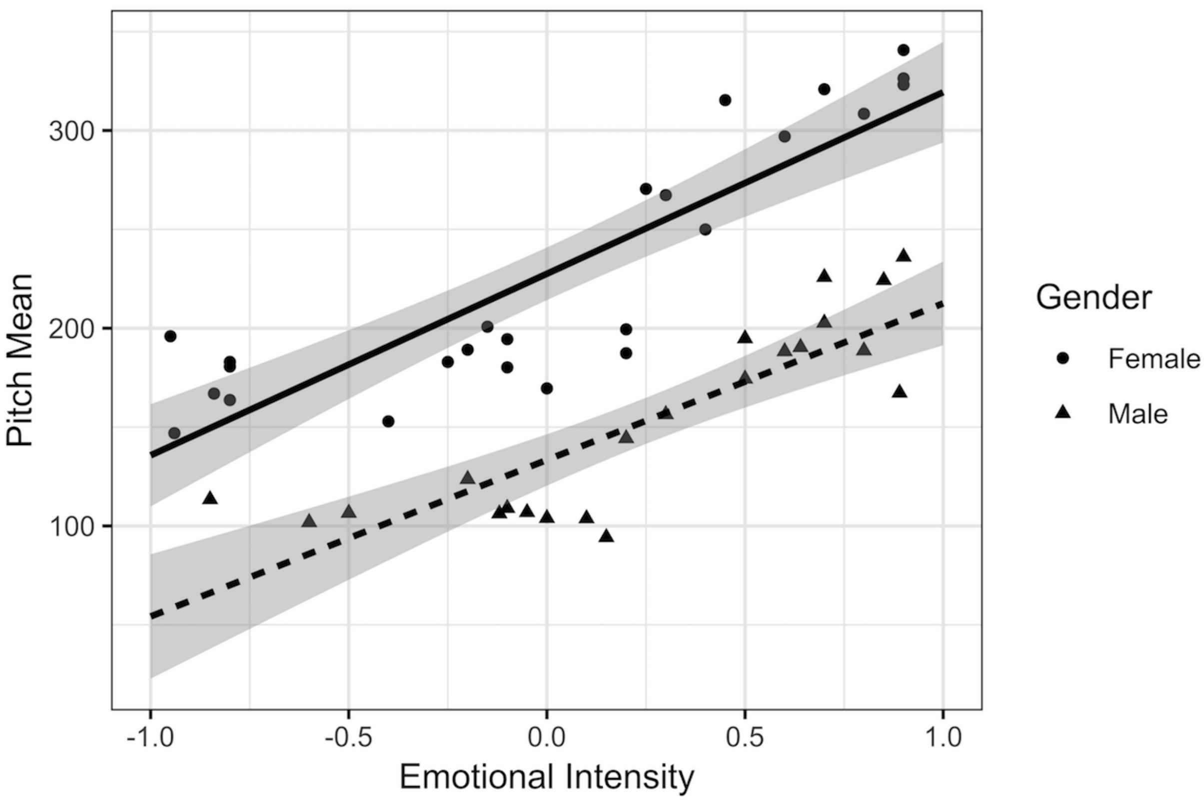
S2.5 Additional Validation #4: New Analysis of Giannakopoulos and Pikrakis (2014) Annotated Speaker Data

In our fourth validation exercise, we analyzed data provided by Giannakopoulos and Pikrakis (2014). This corpus consists of 47 audio clips collected from the Berlin Database of Emotional Speech.[7] Each clip is scored using a continuous scale ranging from low (-1) to high (1) emotional valence and activation which is distinctly different from both Goudbeek and Scherer (2010) and Laukka (2004). In terms of the former, Giannakopoulos and Pikrakis (2014) did not simply say discrete emotions were more/less intense, instead they had coders go through and listen to the audio files and code the degree to which the speaker was more/less activated. Similarly, unlike Laukka (2004) who asked actors to portray emotions with more/less intensity and then assume they did so, Giannakopoulos and Pikrakis (2014) actually code whether the speakers seem more/less activated when they are speaking. With that said, there is little information about the degree to which the coding of valence and activation have been validated. For this reason, these results should be viewed as a supplement to the validation exercises we report above.

Figure S5 displays the results of our analysis. The x-axis represents the emotional intensity of the recordings, as coded by Giannakopoulos and Pikrakis (2014). The y-axis represents the mean vocal pitch of these recordings as extracted by *Praat*. Circles represent emotional portrayals from male actors, and triangles represent portrayals from female actors. We also present the result of simple linear models in which mean vocal pitch is regressed on the level of emotional intensity separately by speaker sex, with grey bands representing the respective 95% confidence intervals. Not only does this plot show a strong linear trend, but emotional intensity is correlated with the mean vocal pitch at the 0.84 level for men ($t = 8.66, df = 23, p < 0.001$) and 0.87 level for women ($t = 7.04, df = 20, p < 0.001$). This data set supports our contention that vocal pitch can be used as an indicator of underlying emotional intensity. As with our previous validations, we consistently see higher mean vocal pitch when emotional intensity is higher. Moreover, mean vocal pitch is *not* significantly correlated with the valence dimension. This too was coded from negative (-1) to positive (1), but unlike emotional intensity there was no significant correlation found between emotional valence and mean vocal pitch ($t = 1.32, df = 45, p > 0.19$). This suggests that vocal pitch

[7] Unlike the RAVDESS data, these utterances are all German phrases.

Figure S5: Emotions with "Strong" Intensity Delivered at a Higher Vocal Pitch Than Emotions with "Normal" Intensity (Giannakopoulos and Pikrakis)



Note: Figure uses data provided by Giannakopoulos and Pikrakis (2014). Solid and dashed lines represent simple linear regression lines for male and female speakers, respectively. Perceived emotional activation (or intensity) is shown on the x-axis. This variable ranges from (-1) deactivated to (1) activated. The speaker's raw vocal pitch is shown on the y-axis.

is useful as an indicator of emotional intensity, rather than simply capturing valence (or positivity/negativity).

S2.6 Vocal Pitch: A Useful, but Not Exhaustive Measure

The analyses we present above demonstrate a consistent relationship between vocal pitch and emotional intensity (or activation). A reasonable concern, however, is that vocal pitch might also be influenced by the other dimension of emotions: valence. We are not arguing that there is absolutely no relationship between pitch and valence. Rather, we are arguing that pitch is more indicative of intensity than valence. This point has consistently been argued in the literature (e.g., Mauss and Robinson 2009). Thus, we feel safe in our assertion that vocal pitch can be used to measure emotional intensity.

Although we have shown vocal pitch to be a consistent indicator of emotional intensity, it is certainly not the only such indicator. We offer three reasons for the use of vocal pitch in our analysis over other measures. First, the present study is primarily interested in broadening audio-as-data approaches in political science. We use vocal pitch in order to achieve this end. Not only are there well-established theoretical arguments for what vocal pitch captures, but it is actually easy for researchers to use. Just as text analysis has a place for dictionary methods, audio analysis should have a place for specific features – like vocal pitch. Such features are not only theoretically interesting in and of themselves, but we must better understand those features before we begin to utilize more advance methods, like supervised and unsupervised learning algorithms.

Second, vocal pitch also has a well-established literature *within* political science. Scholars like Casey Klofstad (e.g., Klofstad 2016) have been working on understanding the role vocal pitch has to play in social science research for quite some time. These scholars have made considerable strides in understanding vocal pitch using small-n studies, but we are the first to apply such techniques to a large corpus of audio. Just as "all quantitative models of language are wrong – but some are useful," (Grimmer and Stewart 2013, 269) we view vocal pitch in a similar light, especially when it is used at scale. With large audio corpora, we can actually get a good sense of a speaker's baseline vocal pitch, which makes vocal pitch a much more useful measure because it can be scaled relative to what we expect.

Finally, "there is no globally best method for automated text analysis" (Grimmer and Stewart 2013, 270). We argue the same can be said for audio data. Indeed, much of the debate surrounding vocal pitch "can be resolved simply by acknowledging that there are different research questions and designs that imply different types of models will be useful" (Grimmer and Stewart 2013, 270). In this study, we are interested in understanding whether women exhibit higher vocal pitch when speaking about women and whether that seems to be generally good or bad for the advancement of women's interests. We think vocal pitch is a reasonable place to start this line of inquiry. Just as we hope this is not the last study on vocal pitch, this is not the only audio variable we plan to use in our own broader research agenda. In many ways, we view this study as the beginning, rather than the end,of audio-as-data approaches in the study of legislative speech, which is why the broader contribution is not vocal pitch, but the use of audio itself.

As an illustration of the important information contained in audio data that might be missed by text-as-data approaches, consider Figure S6, which shows two speeches from Rep. Rosa DeLauro. In the first speech (Panel A), Rep. DeLauro mentions "women" only once. In the second speech (Panel B), she references "women" eight times. Using existing measures, which would simply count references to women, we would conclude that the second speech is more women-focused than the first and move on. Using our method, in contrast, reveals that in the first speech Rep. DeLauro is speaking at 1.55 standard deviations above her baseline vocal pitch (i.e., with greater intensity), while in the second speech she speaks at her baseline. Investigating these speeches further, we find that in the first speech—which references women only once but was given with high intensity—Rep. DeLauro recalls her own experience with ovarian cancer and relates it to the experiences of other women who have been denied insurance because of a pre-existing condition. In the second speech, she is offering statistics related to women's health and the Affordable Care Act. Comparing these two speeches underscores the information lost when ignoring the non-verbal content of legislative speech.

S3 Descriptive Statistics

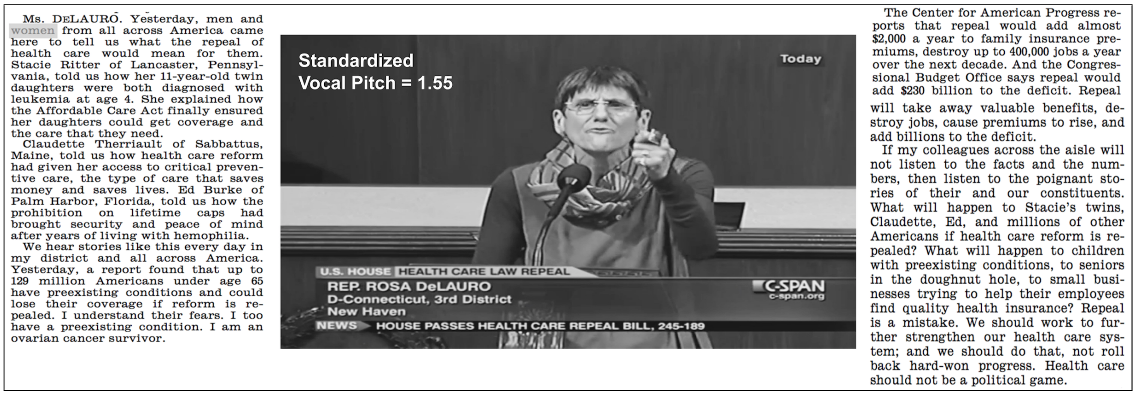
In this section, we report on several descriptive statistics on our measure of vocal pitch. Since women tend to speak at a higher vocal pitch than men, we subset our data by gender in the tables below.

S3.1 Male and Female MCs' Vocal Pitch when Speaking about Women

We first consider whether men and women in the U.S. House speak with higher vocal pitch when talking about women. To establish whether the speaker addressed women we create a binary variable indicating whether the speech used any of the Pearson and Dancey (2011*b*) dictionary terms related to women. These include "women," "woman's," "women's," "girl," "girl's," "girls," "girls'," "female," "female's," "females," "females'," "servicewoman," "servicewoman's," "servicewomen" and "servicewomen's." If a speech contains any of these terms it is coded as a 1, otherwise 0.

Beginning with Table S2, we report on the raw vocal pitch for men and women. We find male MCs did not speak with a significantly different vocal pitch when they were using one of the Pearson and Dancey (2011*b*) terms (t = 0.02, df = 58281, $p > .05$). This suggests that men do not tend to become more or less emotionally activated when talking about women. This is not the case for female MCs. Not only did they talk at a significantly higher vocal pitch when referencing women ($t = 4.14$, df = 12917, $p \leq .001$), but this result

Figure S6: The Importance of Vocal Pitch



Ms. DELAURO. Yesterday, men and women from all across America came here to tell us what the repeal of health care would mean for them. Stacie Ritter of Lancaster, Pennsylvania, told us how her 11-year-old twin daughters were both diagnosed with leukemia at age 4. She explained how the Affordable Care Act finally ensured her daughters could get coverage and the care that they need.

Claudette Therriault of Sabbattus, Maine, told us how health care reform had given her access to critical preventive care, the type of care that saves money and saves lives. Ed Burke of Palm Harbor, Florida, told us how the prohibition on lifetime caps had brought security and peace of mind after years of living with hemophilia.

We hear stories like this every day in my district and all across America. Yesterday, a report found that up to 129 million Americans under age 65 have preexisting conditions and could lose their coverage if reform is repealed. I understand their fears. I too have a preexisting condition. I am an ovarian cancer survivor.

The Center for American Progress reports that repeal would add almost $2,000 a year to family insurance premiums, destroy up to 400,000 jobs a year over the next decade. And the Congressional Budget Office says repeal would add $230 billion to the deficit. Repeal will take away valuable benefits, destroy jobs, cause premiums to rise, and add billions to the deficit.

If my colleagues across the aisle will not listen to the facts and the numbers, then listen to the poignant stories of their and our constituents. What will happen to Stacie's twins, Claudette, Ed, and millions of other Americans if health care reform is repealed? What will happen to children with preexisting conditions, to seniors in the doughnut hole, to small businesses trying to help their employees find quality health insurance? Repeal is a mistake. We should work to further strengthen our health care system; and we should do that, not roll back hard-won progress. Health care should not be a political game.

(a) Rep. DeLauro More Emotionally Intense

Ms. DELAURO. I yield myself 2 minutes.

I rise in opposition to this concurrent resolution. It has nothing to do with the budget and everything to do with ideology.

This is an attempt to turn back the clock on women's health and basic rights. The majority wants to impose their traditional view of a woman's role and take us back to a day when family planning was not available. With this resolution, the majority aims to exclude one specific health care provider, Planned Parenthood, from all Federal resources. This will needlessly put lives in danger.

Planned Parenthood carries out millions of lifesaving preventative and primary care services every year. They deliver immunizations, routine gynecological exams, nearly 1 million screenings for cervical cancer, 830,000 breast exams, and nearly 4 million tests and treatments for sexually transmitted infections like HIV every single year. If this resolution passes, all of these services would be lost.

Seventy-five percent of their more than 3 million patients live at or below 150 percent of the poverty level, make less than $33,000 for a family of four. One of every five women in America has gone to Planned Parenthood for access to health care. Sixty percent of these women consider Planned Parenthood their main source of care. And, in fact, even the number of men Planned Parenthood serves has doubled over the past decade. All of these women and men would lose access to these services if this should pass.

This resolution guts a primary source of care for millions of American families. We all know this has nothing to do with Federal funding of abortion. Federal funds are already banned from going towards abortion services under the Hyde amendment.

We should not be playing political games with women's lives. I urge my colleagues to oppose this dangerous resolution and to stand for women's health and, above all, to trust women to make the right decisions.

(b) Rep. DeLauro Less Emotionally Intense

Note: In Panel A, we show a frame and the text from a speech delivered by Rep. DeLauro (D-CT) on January 19, 2011 in which she only mentions women once, but her vocal pitch suggests she is speaking more intensely about women. In Panel B, we show a frame and the text from a speech delivered by Rep. DeLauro on April 14, 2011 in which she mentions women eight times, but her vocal pitch suggests she is speaking less intensely about women. The Pearson and Dancey (2011*b*) dictionary terms are highlighted in grey.

Table S2: Average Vocal Pitch and Standard Deviation for Male and Female MCs by Party

	"Women" Mentioned		"Women" Not Mentioned	
	Pitch Mean	Pitch SD	Pitch Mean	Pitch SD
Male				
Republican	151.11	24.28	150.95	24.51
Democrat	151.94	24.29	152.17	25.65
All	151.50	24.28	151.49	25.03
Female				
Republican	207.02	30.27	203.11	30.52
Democrat	205.68	25.64	203.35	28.25
All	206.01	26.87	203.27	28.99

Note: Measurements of vocal pitch are in Hertz (Hz). In the first two columns, we restricted our data to speeches which used at least one of the terms outlined by Pearson and Dancey (2011*b*). In the last two columns, we restricted our data to speeches which did not use any of these terms. Rows correspond to indicated groups. For example, the average vocal pitch for all speeches delivered by Republican men mentioning women was 151.11Hz. Averages for each column can be found in the "All" rows.

holds across both Republican and Democratic women. Indeed, Democratic and Republican women tended to talk at 205.68Hz and 207.02Hz respectively when they spoke about women. Moreover, Democratic women's vocal pitch was lower when they were not using any of the Pearson and Dancey (2011*b*) terms ($t = 3.12$, df = 8967, $p \leq .01$), a result mirrored among Republican women ($t = 2.81$, df = 3948, $p \leq .01$). Taken together, this suggests that women in Congress of both parties tend to speak with higher vocal pitch when they are speaking about women.

In Table S3, we report on the same descriptive statistics as above, but using our measure standardized by a speaker's baseline vocal pitch. Positive values signify MCs are speaking above their mean or baseline vocal pitch. The first column shows female MCs speak at a significantly higher vocal pitch when using one of the Pearson and Dancey (2011*b*) terms, both compared to their baseline as well as compared to their male counterparts ($t = 3.01$, df = 7484, $p \leq .01$). Indeed, not only do Congresswomen tend to speak *above* their baseline when talking about women, they actually speak *below* their mean vocal pitch when they are not referencing women. This difference is highly significant ($t = 4.76$, df = 12916, $p \leq .001$), suggesting female MCs' vocal pitch increases when they reference women. The same cannot

Table S3: Average Vocal Pitch and Standard Deviation for Male and Female MCs (Standardized) by Party

	"Women" Mentioned		"Women" Not Mentioned	
	Pitch Mean	Pitch SD	Pitch Mean	Pitch SD
Male				
Republican	0.02	0.96	-0.00	1.00
Democrat	0.02	0.96	-0.00	1.00
All	0.02	0.96	-0.00	1.00
Female				
Republican	0.06	1.02	-0.01	0.99
Democrat	0.10	0.91	-0.02	1.01
All	0.09	0.93	-0.02	1.01

Note: For each MC, we converted vocal pitch to standard deviations above and below his or her average vocal pitch. In the first two columns, we restricted our data to speeches which used at least one of the terms outlined by Pearson and Dancey (2011*b*). In the last two columns, we restricted our data to speeches which did not use any of these terms. Rows correspond to indicated groups. Averages for indicated groups can be found in the "All" rows.

be said for male MCs, whose vocal pitch remains essentially unchanged when referencing women ($t = 1.34$, df $= 58278$, $p > 0.05$).

S3.2 Most and Least Emotionally Activated Female MCs

To identify the most and least emotionally activated MCs when talking about women, we first calculated the average vocal pitch when the legislator used at least one of the terms outlined by Pearson and Dancey (2011*b*). We also calculated the lawmaker's average vocal pitch when not using any of these terms. We then subtracted the latter from the former, yielding a measure which is positive when the MC spoke at a *higher than average* vocal pitch when using terms related to women.

In Table 3 in the main text, we report the 25 women in the U.S. House who were most/least emotionally activated when talking about women. We also show the average score those women received from 24 prominent women's groups obtained from *Project Vote*

Smart.[8] Our findings indicate women who speak with greater emotional intensity about women tend to receive higher scores from women's groups. This suggests that the heightened emotional intensity with which female MCs reference women is also reflected in their voting patterns.

Table S4 lists the 24 interest groups from *Project Vote Smart* we included to compute women's interest group scores. The *Project Vote Smart* group identification numbers and categories are listed in columns 1 and 2, respectively. All groups were listed under category 68, or "women" in the *Project Vote Smart* data. Column 3 reports the group name. Not only are these groups all the women's groups indexed by *Project Vote Smart*, but they are generally representative of the main groups that advance women's interests. Given that we used all *Project Vote Smart* women's interest groups, we view this as a fairly comprehensive list for this validation exercise. Please refer to Section in the main text for more discussion of these results and their broader implications.

S3.3 Most and Least Emotionally Activated Democrats and Republicans

In the main text, we also report on a validation of vocal pitch as a measure of emotional intensity by analyzing whether Democrats and Republicans are more emotionally intense on their party's "owned" issues (see Section). To investigate whether party members tended to be more emotionally intense on issues owned by their party, we created a dummy variable capturing whether the speech was a "party speech." To compute this, we first calculated the average proportion of a speech dedicated to party issues based on the closed captioning of the speech. These issues were identified using our STM, and are listed in Table S15. If a speech contained a greater proportion of party issues than the mean, then we coded it as a party speech (1). Otherwise, it was not considered a party speech (0). For all MCs, we calculated their mean vocal pitch when they were and were not giving a party speech. The difference between these indicates the extent to which MCs were emotionally intense when giving party speeches, with positive values indicating MCs that were more emotionally intense in speeches on their own party's owned issues.

As an additional analysis, we calculated each legislator's distance from the median DW-NOMINATE score for each party. We argue that those legislators closest to the party median should be most engaged with party-owned issues.

Table S5 lists the 25 most- and least-activated MCs when talking about party issues. In accordance with expectations, we find that the 25 legislators most activated when talking about party issues had an average DW-NOMINATE distance of 0.11 from their party median, compared to a distance of 0.19 for the 25 least activated legislators. Although this difference is slight, it is still statistically significant at the 0.05-level ($t = 3.24$, $df = 48$, $p < 0.01$). This general result holds for both Democrats ($t = 2.43$, $df = 30$, $p < 0.03$) and Republicans ($t =$

[8] https://votesmart.org

Table S4: Women's Interest Groups from *Project Vote Smart*

ID	Category	Group Name
164	68	American Association of University Women
38	68	American Congress of Obstetricians and Gynecologists
143	68	Business and Professional Women USA
134	68	Concerned Women for America
2154	68	Concerned Women PAC
1189	68	Emily's List
1343	68	Federally Employed Women
1906	68	Feminist Majority Political Action Committee
2332	68	Jewish Women International
1833	68	League of Women Voters
2493	68	LPAC
1930	68	Maggie's List PAC
1475	68	National Organization for Women
1654	68	National Women's Political Caucus
2340	68	Right Now Women PAC
1946	68	Susan B. Anthony List
671	68	The Woman Activist
319	68	United States Women's Chamber of Commerce
2243	68	Voices of Conservative Women
1860	68	Women Employed
1197	68	Women's Action for New Directions (WAND) and WILL
1910	68	Women's Campaign Fund
2339	68	Women Under Forty Political Action Committee
2336	68	Young Women's Christian Association (YWCA)

Note: List of all interest groups identified by *Project Vote Smart* as advancing women's issues. The full list can be found here: `https://votesmart.org/interest-groups/NA/68#.WuYGS9PwbVo` (Accessed on 4/29/2018).

1.96, $df = 16$, $p < 0.07$), even though the latter difference is only statistically significant at the 0.07-level. These results provide additional evidence that MCs become more emotionally intense when talking about issues to which we suspect they have deeper policy commitments. This provides another piece of predictive validity for our use of vocal pitch as an indicator of emotional intensity.

S4 Praat Floor and Ceiling

When estimating vocal pitch the most important *Praat* settings are the pitch floor and ceiling. Unfortunately, there is very little guidance in the literature about which settings should be used. For example, the *Praat* default is to set the pitch floor at 75Hz and the pitch ceiling at 500Hz, but the *Praat* online manual[9] also says "For a male voice, you may want to set the floor to 75 Hz, and the ceiling to 300 Hz," and "for a female voice, set the range to 100-500 Hz instead." We initially set our pitch floor and ceiling using the settings suggested by Re et al. (2012). They used a range of 50Hz to 300Hz for male voices and 100Hz to 600Hz for female voices. Based on this standard, we initially used a range of 50-600Hz to cover both male and female voices in earlier drafts of this manuscript. Upon further review, we decided to use the *Praat* suggested settings for our main analyses. This means we used a range of 75-300Hz for male MCs and 100-500Hz for female MCs. In order to ensure that our results are robust to the choice of pitch window, below we report on several replications of our results using alternative specifications of the minimum and maximum pitch settings in *Praat*.

S4.1 Replicating Results Using Three Different Pitch Windows

To eliminate the possibility that our results are dependent on choice of *Praat* settings, we re-estimated all of the models we report in the main text using different *Praat* window parameters. In these tables, the "Praat Default" column reports the results with a pitch floor of 75Hz and a pitch ceiling of 500Hz. The "Literature Suggested" column uses the range we used in our preliminary analyses (50-600Hz) and the "Praat Suggested" column uses a range of 75-300Hz for men and 100-500Hz for women. We find the results are substantively similar across these different *Praat* settings, suggesting our results are robust to different pitch windows.

Beginning with Table S6, we can see the substantive interpretation is the same when using any of the three settings. The results from Table 1 in the main text hold across *Praat* floor and ceiling choices.

Table S7 replicates the results from Table 5 in the main text using different *Praat* settings. The substantive interpretation of our results holds across settings. For the main effect of

[9] http://www.fon.hum.uva.nl/praat/manual/Intro_4_2__Configuring_the_pitch_contour.html

Table S5: Members of Congress Who Speak with Emotional Intensity About Party Issues Tend to Vote More With Their Party

(a) *Most* Activated

	Party Issue	No Party Issue	Pitch Diff.	|DW Diff.|
Name				
Sánchez (D-CA)	280.45	247.53	32.92	0.05
Carney (D-PA)	166.85	136.27	30.58	0.29
Wexler (D-FL)	202.62	173.03	29.59	0.00
Tauscher (D-CA)	220.03	192.17	27.86	0.10
Herrera (R-WA)	236.86	209.76	27.10	0.00
Meek (D-FL)	168.63	142.01	26.62	0.08
Holden (D-PA)	145.89	119.50	26.39	0.17
Napolitano (D-CA)	207.24	182.36	24.87	0.08
Roybal-Allard (D-CA)	209.44	188.53	20.91	0.01
Tiberi (R-OH)	152.14	131.24	20.89	0.08
Scott (R-GA)	165.65	145.42	20.23	0.04
Davis (D-IL)	137.95	117.80	20.15	0.06
Owens (D-NY)	173.91	153.82	20.09	0.23
Bonner (R-AL)	179.73	160.96	18.77	0.26
Schauer (D-MI)	151.64	133.03	18.61	0.06
Thompson (D-MS)	154.29	136.10	18.19	0.04
Bono (R-CA)	220.32	202.32	18.00	0.03
Wasserman (D-FL)	199.16	182.19	16.97	0.03
Amash (R-MI)	194.68	177.87	16.81	0.25
Myrick (R-NC)	225.12	208.92	16.20	0.04
Dingell (D-MI)	193.83	177.81	16.03	0.01
Kilpatrick (D-MI)	212.90	196.92	15.97	0.06
Rodriguez (D-TX)	150.02	134.16	15.86	0.11
Shuler (D-NC)	153.84	138.05	15.79	0.35
Lee (R-NY)	157.19	141.47	15.72	0.19
Groups				
All	186.41	165.17	21.24	0.11
Democrats	184.04	161.84	22.20	0.10
Republicans	191.46	172.24	19.21	0.11

(b) *Least* Activated

	Party Issue	No Party Issue	Pitch Diff.	|DW Diff.|
Name				
Amodei (R-NV)	133.02	160.13	-27.11	0.10
Murtha (D-PA)	147.66	174.50	-26.85	0.18
Green (D-TX)	155.33	181.73	-26.40	0.09
Cramer (R-ND)	153.89	175.37	-21.48	0.27
LoBiondo (R-NJ)	154.23	174.76	-20.54	0.25
Space (D-OH)	147.24	166.05	-18.81	0.25
Pomeroy (D-ND)	155.89	172.92	-17.02	0.18
Clarke (D-MI)	176.30	193.27	-16.97	0.08
Melancon (D-LA)	125.94	142.88	-16.94	0.24
Giffords (D-AZ)	204.09	220.54	-16.45	0.31
McHugh (R-NY)	107.96	123.97	-16.02	0.38
Cooper (D-TN)	106.02	121.33	-15.31	0.13
DelBene (D-WA)	181.40	196.69	-15.29	0.14
Adler (D-NJ)	141.06	156.33	-15.27	0.27
Heck (R-NV)	147.91	163.13	-15.23	0.08
Noem (R-SD)	212.15	227.20	-15.05	0.23
Cook (R-CA)	164.12	178.99	-14.87	0.12
Baird (D-WA)	132.08	146.79	-14.71	0.13
Moore (D-KS)	124.83	139.40	-14.58	0.18
Kelly (R-PA)	168.05	182.37	-14.32	0.29
Markey (D-MA)	154.66	168.85	-14.19	0.12
Gibson (R-NY)	141.05	155.21	-14.15	0.21
Miller (R-CA)	131.68	145.16	-13.48	0.13
Halvorson (D-IL)	237.54	251.01	-13.47	0.12
Peters (D-CA)	129.57	142.87	-13.29	0.25
Groups				
All	153.35	170.46	-17.11	0.19
Democrats	154.64	171.68	-17.04	0.18
Republicans	151.41	168.63	-17.22	0.20

Note: Measurements of vocal pitch are in Hertz (Hz). To make this table comparable to Table 3 in the main text, we created a dummy variable which equals 1 when a MC's speech contained more than the average number of party references. We called these "Party Speeches." In the first column, we restricted our data to party speeches. In the second column, we restricted our data to speeches which contained less than the average number of party references. The "Pitch Difference" column (abbreviated "Pitch Diff.") is the difference between these two columns. The 25 *most* (see Panel A) and *least* activated (see Panel B) MCs had the highest and lowest "Pitch Difference," respectively. The absolute difference between the MC's DW-Nominate score and the median DW-Nominate score for the MC's party is included in the "DW Difference" (abbreviated "DW Diff.") column. Higher values imply the MC's ideology was further away from the party median. Column averages for Democrats and Republicans can be found in the "Groups" section.

Table S6: Female MCs More Likely to Talk About Women, with Greater Intensity (Different Pitch Windows)

	Dependent variable:					
	Standardized Vocal Pitch					
	Praat Default Settings		*Literature Suggested Settings*		*Praat Suggested Settings*	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	-0.001	0.139^{***}	-0.0004	0.128^{***}	-0.002	0.151^{***}
	(0.004)	(0.024)	(0.004)	(0.024)	(0.004)	(0.024)
Female	-0.020^{*}	-0.033^{***}	-0.020^{*}	-0.033^{***}	-0.017	-0.032^{***}
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
"Women" Mentioned	0.007	-0.064^{***}	0.005	-0.064^{***}	0.020	-0.054^{***}
	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)
Female × "Women" Mentioned	0.110^{***}	0.131^{***}	0.111^{***}	0.132^{***}	0.090^{***}	0.112^{***}
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Controls		✓		✓		✓
N_1	71,203	71,203	71,245	71,245	71,198	71,198
N_2	613	613	613	613	613	613
Log Likelihood	$-100,726.400$	$-99,746.740$	$-100,786.400$	$-99,903.080$	$-100,720.100$	$-99,645.100$
AIC	201,464.800	199,521.500	201,584.800	199,834.200	201,452.100	199,318.200

Note: Models are identical to Table 1, Models 3 and 4 except we use different Pratt settings. Control variables are excluded to save space Full models available upon request. The dependent variable is the speaker's vocal pitch in standard deviations above or below the speaker's baseline. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S7: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women (Different Pitch Windows)

	Dependent variable:					
	"Women" Mentioned					
	Praat Default Settings		*Literature Suggested Settings*		*Praat Suggested Settings*	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−2.693***	−2.234***	−2.695***	−2.237***	−2.692***	−2.235***
	(0.041)	(0.220)	(0.041)	(0.220)	(0.041)	(0.221)
Female Speeches	0.055***	0.060***	0.056***	0.060***	0.055***	0.060***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Female Pitch	−0.121***	−0.129***	−0.124***	−0.129***	−0.127***	−0.135***
	(0.031)	(0.032)	(0.031)	(0.032)	(0.031)	(0.032)
Female Speeches × Female Pitch	0.009*	0.012**	0.009	0.011**	0.011*	0.014**
	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Controls		✓		✓		✓
N_1	50,235	50,235	50,235	50,235	50,235	50,235
N_2	619	619	619	619	619	619
Log Likelihood	−14,735.990	−13,950.320	−14,735.510	−13,950.230	−14,735.630	−13,949.720
AIC	29,481.990	27,930.630	29,481.010	27,930.460	29,481.260	27,929.440

Note: Models are identical to Table 5, Models 1 and 2 except we use different Pratt settings. Control variables are excluded to save space. Full models available upon request. Dependent variable equals 1 if the speech included any of the Pearson and Dancey (2011*b*) terms, 0 otherwise. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

"Female Speeches" and "Female Pitch" the coefficients and levels of statistical significance are essentially unchanged, regardless of the pitch floor and ceiling, as is the interaction term between these two variables. Unlike the results we reported in our preliminary analyses using a pitch window of 50-600Hz, when either the *Praat* default or recommended settings are used the interaction term is statistically significant even when no controls are included. This suggests our results are quite robust to different pitch windows.

The robustness of our results is also shown in Table S8. Here, we replicated the results from Table 6 in the main text using three different *Praat* settings. Again, the results are largely the same. Again, this provides strong evidence that the results presented in the main text cannot be attributed to the *Praat* settings.

We find essentially the same results in Table S9 which replicates the results from Table 7 in the main text using different *Praat* settings. Again, we are primarily interested in the interaction between "Female Speeches" and "Female Pitch," which does not change substantially from one model to the next. It is always positive and statistically significant with a coefficient that only varies by 0.001 when different *Praat* settings are used. Altogether, these replications give us a high degree of confidence that our results are not sensitive to the selection of pitch windows within *Praat*.

S4.2 Replicating Results Using Multiple Permutations

The lack of empirical guidance in terms of the *Praat* floor and ceiling is one of the reasons why we emphasize mean vocal pitch in this study. Not only is vocal pitch a useful measure of emotional intensity, but Dietrich, Enos and Sen (2019) demonstrate that it is also a fairly robust measure. Hess (2007) emphasizes the challenges of properly estimating the vocal pitch track, which is why summary measures can be especially useful. Unlike estimating a single value of the vocal pitch track (e.g., minimum), the mean and median are much more resistant to errors in pitch estimation. For this reason, scholars using median or mean vocal pitch should be able to estimate models that are quite robust to the selection of *Praat* floor and ceiling parameters. As a demonstration of this, we now turn to a replication of our results using floor and ceiling parameters that are outside the bounds of those recommended by either the literature or *Praat* software itself.

To demonstrate the robustness of summary statistics such as mean vocal pitch to the bounds set in *Praat*, we re-estimated our main results using arbitrary pitch floor and ceiling settings. We selected pitch floor values of 50, 75, and 100Hz, and pitch ceiling values of 300, 350, 400, 450, 550, and 600Hz for this analysis. Recall that the default *Praat* settings are 50Hz and 600Hz for the floor and ceiling respectively. We are thus significantly shrinking the pitch window used by *Praat* to demonstrate that summary statistics such as mean vocal pitch generate reliable findings even with arbitrarily smaller pitch windows. The replicated results for Table 1, Model 3 (from the main text) are shown in Figure S7. In this figure, each panel plots the coefficient estimates derived using different pitch floor and ceiling settings. The panels represent pitch floors of 50Hz, 75Hz, and 100Hz from left to right. Variations

Table S8: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch (Different Pitch Windows)

	Dependent variable:		
	Male Vocal Pitch		
	Praat Default Settings	*Literature Suggested Settings*	*Praat Suggested Settings*
	(1)	(2)	(3)
Fixed Effects			
Constant	−0.019***	−0.022***	−0.028***
	(0.007)	(0.007)	(0.007)
"Women" Mentioned	0.010	0.026	0.013
	(0.022)	(0.022)	(0.022)
Female Speeches	0.003***	0.003***	0.005***
	(0.001)	(0.001)	(0.001)
Female Pitch	0.046***	0.077***	0.097***
	(0.009)	(0.009)	(0.009)
"Women" Mentioned × Female Speeches	−0.003	−0.003	−0.004
	(0.003)	(0.003)	(0.003)
"Women" Mentioned × Female Pitch	0.029	0.004	−0.009
	(0.031)	(0.031)	(0.031)
Female Speeches × Female Pitch	0.012***	0.012***	0.008***
	(0.002)	(0.002)	(0.002)
"Women" Mentioned × Female Speeches × Female Pitch	0.015***	0.016***	0.018***
	(0.005)	(0.005)	(0.006)
Random Effects			
MC	0.000	0.000	0.000
N_1	49,919	49,914	49,962
N_2	506	506	506
Log Likelihood	−70,715.350	−70,478.580	−70,726.420
AIC	141,450.700	140,977.200	141,472.800

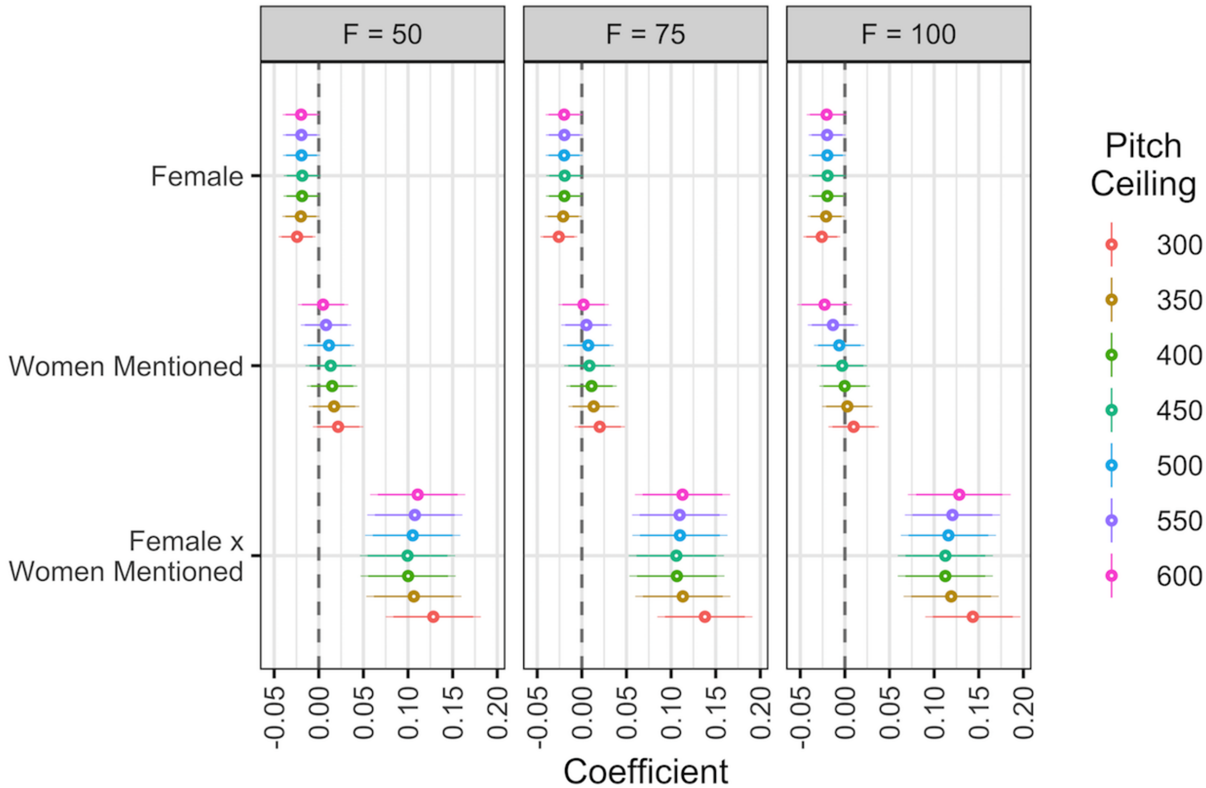
Note: The dependent variable is the speaker's vocal pitch in standard deviations above or below the speaker's baseline. In Model 1 the pitch floor and ceiling are set to 75Hz and 600Hz. In Model 2 the pitch floor and ceiling are set to 50Hz and 600Hz. In Model 3 the pitch floor and ceiling are set to 75Hz and 300Hz. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S9: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (Different Pitch Windows)

	Dependent variable:					
	Male Votes Cast					
	Praat Default Settings		*Literature Suggested Settings*		*Praat Suggested Settings*	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	0.020	0.100**	0.019	0.100**	0.018	0.095*
	(0.015)	(0.051)	(0.015)	(0.051)	(0.015)	(0.051)
Female Speeches	0.001	−0.00002	0.001	0.0001	0.001	0.0003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Pitch	−0.189***	−0.179***	−0.187***	−0.177***	−0.179***	−0.167***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Female Speeches × Female Pitch	0.015***	0.012***	0.015***	0.012***	0.014***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Controls		✓		✓		✓
N_1	21,920	21,920	21,920	21,920	21,920	21,920
N_2	485	485	485	485	485	485
Log Likelihood	−28,118.740	−28,102.000	−28,122.730	−28,105.190	−28,128.860	−28,112.830
AIC	56,249.470	56,234.010	56,257.460	56,240.370	56,269.720	56,255.660

Note: Models are identical to Table 7, Models 1 and 2 except we use different Pratt settings. Control variables excluded to save space. Full models available upon request. Outcome is the proportion of time male MCs voted with women, as described on pages S36–S41. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Figure S7: Female MCs More Likely to Talk About Women, with Greater Intensity (No Controls)



Note: Replicating the results from Table 1 (main text), Model 3 using a variety of *Praat* pitch floor and ceiling settings. More specifically, in the panels labeled "F = 50," "F = 75," and "F = 100," the pitch floor is set to 50Hz, 75Hz, and 100Hz, respectively. The colored points indicate the pitch ceiling which ranges from 300Hz (red) to 600Hz (magenta). On the x and y axises we show the coefficients and the variable labels, respectively. Thinner lines represent 95% confidence intervals. Thicker lines represent 90% confidence intervals. The vertical dashed line represents zero, meaning any interval that overlaps this line is statistically indistinguishable from zero.

in pitch ceiling are indicated by different colored circles within each panel. As you can see, the coefficient estimates vary only slightly across arbitrarily set pitch settings. Even with extreme *Praat* settings (i.e., a pitch floor of 100Hz and a ceiling of 300Hz), the substantive interpretation of results is essentially unchanged from our initial analysis, with the coefficient for our interaction term remaining positive, statistically significant, and substantively nearly identical to what we report in the main text.

Our coefficient estimates are similarly robust after the inclusion of additional control variables. Figure S8 replicates Table 1, Model 4 from the main text using a variety of pitch windows. As before, not only does the substantive interpretation of results remain unchanged with arbitrarily defined pitch windows, but the coefficient estimates themselves are virtually identical. This is likely due to the law of large numbers. Each point on the pitch contour is being drawn from the same distribution (the speaker's vocal range), meaning with a large enough sample the center of that distribution (the speaker's fundamental frequency) can be estimated using the sample mean (the speaker's mean vocal pitch). In *Praat*, the pitch floor and ceiling are important because they affect the sample size, but even when the pitch window is restrictive there will still be hundreds of pitch samples, which is likely enough to estimate the speaker's average fundamental frequency. Future work should be done to explore the bounds of sample size and pitch windows necessary to achieve robust results, but these results suggests that scholars should have a high degree of confidence in the use of mean vocal pitch with the *Praat* software, at least with sufficiently large data sets.

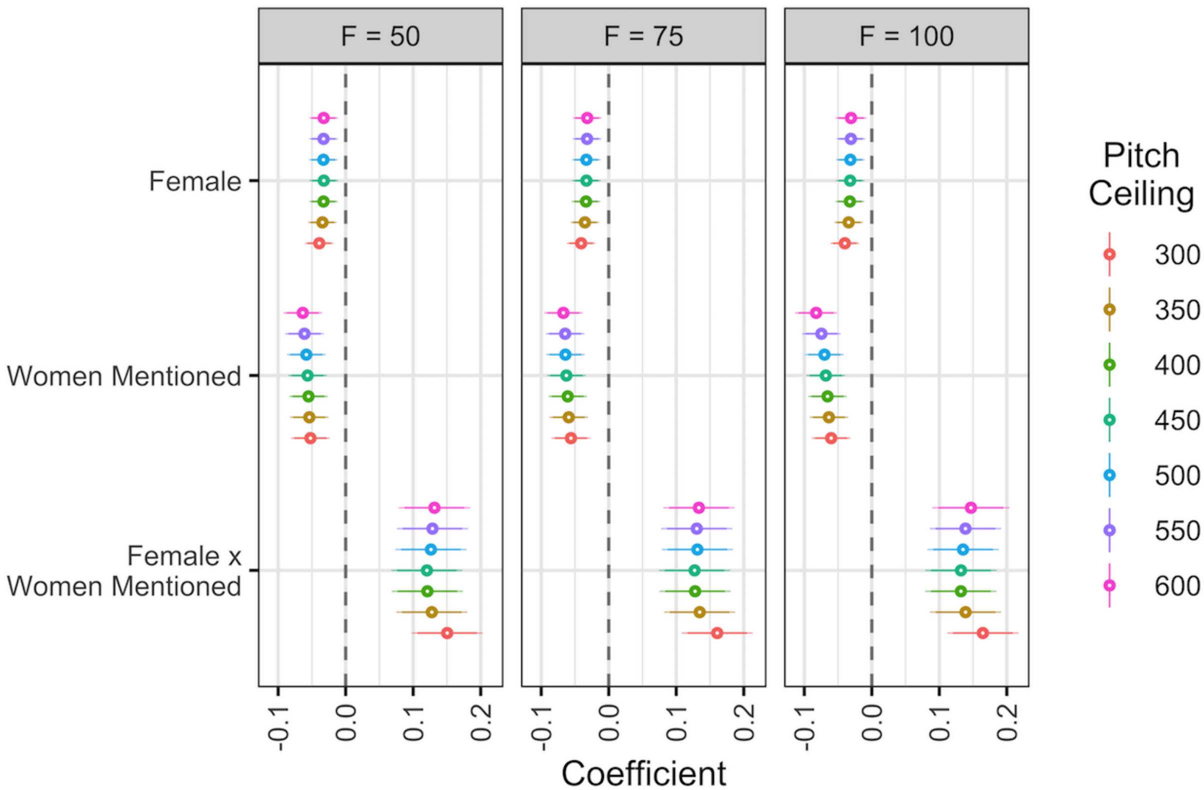
S5 Measuring Legislative Speech About Women

In addition to concerns about software choices and the estimation of mean pitch, one might be concerned about the sensitivity of our results to how we define speeches about women. We see at least three possible strategies for identifying legislative speech about women. The first, which we employ in the main text, is to identify whether a speech included any of the Pearson and Dancey (2011*b*) dictionary terms. If a speech used any of the terms in this dictionary, it was coded as a 1, otherwise it was coded as a 0. This is an established measure in the literature, so we are confident of its utility.

A second measure could consider the proportion of words in the speech drawn from dictionary terms. This measure would capture the degree to which the speech was referencing women, while also controlling for the fact that these references are more likely in longer speeches (thus eliminating the possibility that we are simply capturing speech length). We examine whether our results are robust to this alternative operationalization of the dependent variable (see discussion and Tables S11, S12, and S13 below).

Third, a more empirically-derived measure could be constructed from the floor speech texts themselves. Namely, one could use a Structural Topic Model (STM) (Roberts et al. 2013; Roberts, Stewart and Tingley 2014; Roberts et al. 2014) to capture speeches about women. This approach does not rely on a dictionary, which previous scholars have shown

Figure S8: Female MCs More Likely to Talk About Women, with Greater Intensity (with Controls)



Note: Replicating the results from Table 1 (main text), Model 4 using a variety of *Praat* pitch floor and ceiling settings. More specifically, in the panels labeled "F = 50," "F = 75," and "F = 100," the pitch floor is set to 50Hz, 75Hz, and 100Hz, respectively. The colored points indicate the pitch ceiling which ranges from 300Hz (red) to 600Hz (magenta). On the x and y axises we show the coefficients and the variable labels, respectively. Thinner lines represent 95% confidence intervals. Thicker lines represent 90% confidence intervals. The vertical dashed line represents zero, meaning any interval that overlaps this line is statistically indistinguishable from zero.

to be unreliable under certain conditions (Grimmer and Stewart 2013). Unlike standard Latent Dirichlet Allocation (LDA) models, STM allows researchers to use their substantive knowledge of the corpus to better "structure" topic identification (for additional details, see Roberts et al. 2014, 4). In the section below, we include two covariates in our structural topic model: the date of the speech and the speaker's ideology.

We assume that each legislative day is restricted to a handful of topics, meaning that representatives are likely to deliver similar speeches on the same day. Words appearing on the same day are thus more likely to be associated with one another. This variable was measured in days since the first date in the data set – January 1, 2009. We also assume that representatives who are on the same side of the ideological spectrum are more likely to speak about similar issues. We measure ideology using DW-Nominate scores which range from -1 ("liberal") to 1 ("conservative") (Poole and Rosenthal 2001). Using these covariates, we estimated a 30-topic STM, the results of which can be found in Table S15. We choose to focus on Topic 14, as it appears to be the most directly comparable to the set of terms chosen by Pearson and Dancey (2011*b*). More specifically, this topic includes word stems like "women," "children," "famili," "live," "life," and "children" which are all consistent with references to women.

S5.1 Replicating Results Using Different "Women" Operationalizations

Table S10 shows the comparability of each strategy for identifying speeches about women. The first column defines speeches about women by whether they included any of the Pearson and Dancey (2011*b*) terms. The second and third columns use the proportion of words in the speech that included these terms and whether the speech was classified under Topic 14 of our STM, respectively. As shown below, each of these measures yield substantively identical results. Not only are all the coefficients related to the speaker's gender (see `Female`) positive, but they are all highly significant. This suggests that irrespective of how we measure speech about women, female MCs are more likely to reference women on the House floor.

The results concerning women's vocal pitch are also consistent across the three measures. Table S11 reports these results. Here, the dependent variable is the standardized vocal pitch with positive values indicating MCs are speaking above their baseline. Our primary variable of interest is the interaction between a speaker's gender and whether the speaker mentions women. Again, in these models, the results are the same regardless of how the variable of interest is measured. This provides us with added confidence about the robustness of the findings we report in the main text concerning female legislators' increased amount and emotional intensity of speech referencing women.

To demonstrate the robustness of our findings about the responses of male legislators, we replicate our analyses using these alternative measures of speech about women. Our dependent variables in Table S12 capture whether a male MC referenced women in his speech using the same three definitions described above. In these models, our primary

Table S10: Number of Speeches About Women Across Different Dependent Variables

	Dependent variable:					
	"Women" Mentioned		"Women" Percent		"Women" Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−2.427***	−2.218***	0.050***	0.083***	2.446***	2.837***
	(0.035)	(0.184)	(0.003)	(0.017)	(0.072)	(0.366)
Female	0.866***	0.790***	0.126***	0.115***	2.310***	2.051***
	(0.078)	(0.081)	(0.008)	(0.008)	(0.172)	(0.173)
Controls		✓		✓		✓
N_1	74,151	74,151	74,151	74,151	74,150	74,150
N_2	619	619	619	619	619	619
Log Likelihood	−23,909.700	−22,786.800	−18,098.780	−18,101.540	−234,275.800	−234,221.500
AIC	47,825.410	45,595.610	36,205.570	36,227.090	468,559.700	468,466.900

Note: Models are identical to Table 1, Models 1 and 2 except we use different operationalizations of references to women (description on page S32). Control variables excluded to save space. Full models available upon request. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

Table S11: Vocal Pitch of Speeches about Women Across Different Independent Variables

	Independent variable:					
	"Women" Mentioned		"Women" Percent		"Women" Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−0.002	0.151***	−0.001	0.143***	0.015***	0.172***
	(0.004)	(0.024)	(0.004)	(0.024)	(0.005)	(0.024)
Female	−0.017	−0.032***	−0.011	−0.027***	−0.006	−0.022**
	(0.011)	(0.011)	(0.010)	(0.010)	(0.011)	(0.011)
Talking About "Women"	0.020	−0.054***	0.021	0.011	−0.007***	−0.008***
	(0.014)	(0.014)	(0.018)	(0.018)	(0.001)	(0.001)
Female ×	0.090***	0.112***	0.051**	0.058**	0.005***	0.004***
Talking About "Women"	(0.027)	(0.027)	(0.024)	(0.024)	(0.001)	(0.001)
Controls		✓		✓		✓
N_1	71,198	71,198	71,198	71,198	71,197	71,197
N_2	613	613	613	613	613	613
Log Likelihood	−100,720.100	−99,645.100	−100,721.800	−99,645.990	−100,700.700	−99,606.080
AIC	201,452.100	199,318.200	201,455.600	199,320.000	201,413.300	199,240.200

Note: Models are identical to Table 1, Models 3 and 4 except we use different operationalizations of references to women (description on page S32). Control variables excluded to save space. Full models available upon request. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

Table S12: The Effect of Quantity and Intensity of Women's Speech on Frequency of Men's Speeches on Women Across Different Dependent Variables

	Dependent variable:					
	"Women" Mentioned		"Women" Percent		"Women" Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−2.695***	−2.237***	0.021***	0.061***	1.997***	2.642***
	(0.041)	(0.220)	(0.002)	(0.013)	(0.069)	(0.365)
Female Speeches	0.056***	0.060***	0.006***	0.006***	0.109***	0.108***
	(0.003)	(0.003)	(0.0003)	(0.0003)	(0.006)	(0.006)
Female Pitch	−0.124***	−0.129***	−0.007***	−0.008***	−0.387***	−0.360***
	(0.031)	(0.032)	(0.002)	(0.002)	(0.046)	(0.046)
Female Speeches × Female Pitch	0.009	0.011**	0.002***	0.002***	0.022**	0.016
	(0.005)	(0.006)	(0.0004)	(0.0004)	(0.010)	(0.010)
Controls		✓		✓		✓
N_1	50,235	50,235	50,235	50,235	50,234	50,234
N_2	509	509	509	509	509	509
Log Likelihood	−14,735.510	−13,950.230	2,065.148	2,044.574	−154,044.400	−153,962.700
AIC	29,481.010	27,930.460	−4,118.297	−4,057.148	308,100.800	307,957.400

Note: Models are identical to Table 5, Models 1 and 2 except we use different operationalizations of references to women (description on page S32). Control variables excluded to save space. Full models available upon request. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

independent variable is the interaction between `Female Speeches` and `Female Pitch`. This interaction effect captures the total number of speeches (`Female Speeches`) and the average vocal pitch for female MCs (`Female Pitch`) on a given legislative day. Moving from left to right, our results hold regardless of the measure of speech about women. This demonstrates the robustness of our finding that men talk about women more often when female legislators give more, and more intense, speeches about women.

Our finding that men's vocal pitch increases when talking about women in response to female MCs' quantity and intensity of speech also holds across measures, as shown in Table S13. Here, we predict the standardized vocal pitch for male MCs using the interaction between `Female Speeches`, `Female Pitch`, and whether the speaker references "women" (`Talking About Women`). As explained in the main text, these models determine whether the vocal pitch of a male MC changes with differences in female speaking behavior. As before, the results are substantively identical regardless of the measure of speeches about women we employ.

Finally, in Table S14 we replicate our results while also including controls for party identification, ideology, seniority, committee position, race, whether the speech was given in an election year, women's issue bills, and the number of CQ bills on a given legislative

Table S13: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch Across Different Independent Variables

	Independent variable:		
	"Women" Mentioned	"Women" Percent	"Women" Topic
	(1)	(2)	(3)
Fixed Effects			
Constant	-0.022^{***}	-0.017^{**}	-0.008
	(0.007)	(0.007)	(0.007)
Talking About "Women"	0.026	-0.052^{*}	-0.007^{***}
	(0.022)	(0.029)	(0.001)
Female Speeches	0.003^{***}	0.002^{**}	0.004^{***}
	(0.001)	(0.001)	(0.001)
Female Pitch	0.077^{***}	0.069^{***}	0.065^{***}
	(0.009)	(0.009)	(0.010)
Talking About "Women" × Female Speeches	-0.003	0.009^{***}	0.0001
	(0.003)	(0.003)	(0.0001)
Talking About "Women" × Female Speeches	0.004	0.148^{***}	0.003^{*}
	(0.031)	(0.046)	(0.002)
Female Speeches × Female Pitch	0.012^{***}	0.015^{***}	0.015^{***}
	(0.002)	(0.002)	(0.002)
Talking About "Women" × Female Speeches × Female Pitch	0.016^{***}	-0.011^{**}	-0.0003
	(0.005)	(0.006)	(0.0003)
Random Effects			
MC	0.000	0.000	0.000
N_1	49,914	49,914	49,913
N_2	506	506	506
Log Likelihood	−70,478.580	−70,477.020	−70,468.530
AIC	140,977.200	140,974.000	140,957.100

Note: Outcome is the vocal pitch of male speakers scaled to standard deviations above their baseline. Column labels (e.g., "Women Mentioned") indicate how `Talking About ‘‘Women’’` was measured. Please refer to page S32 for descriptions of the different measures. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

day. Excepting a few minor differences, the interpretation of our results remain unchanged. Given that our findings are robust across three different dependent variables and hold with the inclusion of a number of control variables, we are confident that our measurement choice on women's speeches had little substantive effect on the results we report in the main text.

S5.2 Party Topics

In the main text we examine whether Democratic and Republican MCs speak with heightened pitch on issues traditionally owned by their respective parties. To conduct the analyses found in Table 4 in the main text – and Table S5 in the Supplemental Information – we had to identify issues that were owned by Democrats and Republicans. These are presented in Table S15. "Democratic issues" fell into three general categories: (1) social welfare, (2) land management/transportation, and (3) civil rights. These includes topic numbers 1, 2, 6, 13, 19, 22, 25, 28, and 29. For Republicans, we identified (1) defense, (2) immigration, and (3) tax/budget policy as "Republican issues." These include topic numbers 5, 8, 9, 10, 12, 15, 23, 26, and 30. Although there is no universally-accepted definition of which issues are owned by the Democratic and Republican parties, we feel that these selections are in keeping with general perceptions of the parties' issue ownership. Moreover, as we report in the main text, Democrats and Republicans dedicate a larger proportion of their speeches to topics owned by their parties ($t = 23.32, df = 74148, p < 0.001$ and $t = 30.46, df = 74148, p < 0.001$, respectively).

S6 Measuring Potential Backlash Effects

Below we discuss the standardized and unstandardized vote measures used to estimate the relationship between women's speech and men's voting behavior.

S6.1 Standardized Vote Measure

To determine whether male MCs were generally responding positively or negatively to women's speaking behavior, we created a measure which captures whether a given male MC votes with female speakers more than we would expect on the average legislative day. This is the variable we report in Table 7 of the main text.

For the purpose of illustration, we walk through the construction of this variable for a single, hypothetical legislative day (e.g., January 20, 2010). We refer to this as the "day of interest." To do so, we take the following steps:

1. Find women who gave speeches using any of the Pearson and Dancey (2011*b*) terms on the day of interest. Let's assume there are two women who gave such speeches: F1 and F2.

Table S14: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch Across Different Independent Variables (with Additional Controls)

	Independent variable:		
	"Women" Mentioned	"Women" Percent	"Women" Topic
	(1)	(2)	(3)
Fixed Effects			
Constant	0.108***	0.109***	0.127***
	(0.030)	(0.030)	(0.030)
Talking About "Women"	−0.042*	−0.056**	−0.008***
	(0.022)	(0.028)	(0.001)
Female Speeches	0.003***	0.002*	0.003***
	(0.001)	(0.001)	(0.001)
Female Pitch	0.079***	0.072***	0.066***
	(0.009)	(0.009)	(0.010)
Talking About "Women" × Female Speeches	−0.003	0.009***	0.0001
	(0.003)	(0.003)	(0.0001)
Talking About "Women" × Female Pitch	−0.0003	0.134***	0.003*
	(0.030)	(0.045)	(0.002)
Female Speeches × Female Pitch	0.010***	0.013***	0.014***
	(0.002)	(0.002)	(0.002)
Talking About "Women" × Female Speeches × Female Pitch	0.016***	−0.011*	−0.0003
	(0.005)	(0.006)	(0.0002)
Controls	✓	✓	✓
N_1	49,914	49,914	49,913
N_2	506	506	506
Log Likelihood	−69,839.170	−69,843.180	−69,823.470
AIC	139,718.300	139,726.400	139,686.900

Note: Replicates results from Table 6 from the main text and Table S13 from the SI with the same battery of controls we have included in other models. Controls not shown to save space. Full models available upon request. Please refer to page S32 for descriptions of the different measures. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01.

Table S15: Democratic and Republican Issues Identified by the Structural Topic Model (STM) Outlined in Section S5

Topic	Word 1	Word 2	Word 3	Word 4	Word 5	Label	Proportion
1	court	case	justic	judg	law	law 1	0.01
2	right	peopl	constitut	american	freedom	rights	0.03
3	colleagu	support	today	like	new	collegiality 1	0.05
4	work	make	need	peopl	can	values	0.07
5	war	militari	afghanistan	forc	defens	middle east 1	0.02
6	school	educ	student	colleg	communiti	education	0.02
7	republican	american	democrat	will	pass	party	0.04
8	busi	small	regul	cost	will	business	0.02
9	budget	spend	cut	year	debt	spending cut	0.04
10	secur	nation	inform	protect	agenc	security	0.03
11	energi	oil	gas	will	price	energy	0.02
12	state	unit	texa	border	come	immigration 2	0.02
13	care	health	insur	will	cost	health care	0.04
14	women	children	famili	live	life	children	0.03
15	nuclear	israel	iran	world	peac	middle east 2	0.02
16	job	economi	creat	american	econom	jobs	0.04
17	peopl	get	thing	talk	got	discursive 1	0.05
18	honor	year	great	serv	first	collegiality 2	0.04
19	transport	build	new	system	air	transportation	0.02
20	financi	credit	loan	bank	street	financial	0.02
21	will	side	pass	floor	debat	procedural	0.05
22	water	land	area	communiti	nation	land	0.03
23	law	immigr	enforc	victim	crime	immigration	0.02
24	say	think	know	want	one	discursive 2	0.08
25	fund	program	million	provid	billion	welfare 1	0.04
26	tax	govern	pay	feder	american	tax	0.03
27	administr	quot	report	obama	public	administration	0.03
28	act	requir	author	law	provis	law 2	0.05
29	famili	food	benefit	million	cut	welfare 2	0.02
30	servic	veteran	nation	serv	support	veterans	0.03

Note: Blue and Red topics were flagged as Democratic and Republican issues, respectively. Top-5 words and labels from the ($k = 30$) STM outlined in Section S5 also included. The labels are not returned by the software. They were added after reviewing the top-5 words and other related output.

2. Find men who also gave speeches on the day of interest. Let's assume there are two men who gave such speeches: M1 and M2.

3. Using all the votes from the appropriate Congress, determine how often each male and female pair voted in the same direction. Convert this to a mean and standard deviation for the average legislative day.

To continue this example, let's assume the mean and standard deviation (see Step 3) for every male/female dyad was the following:

$\mu_{\texttt{M1,F1}} = .90$, $\sigma_{\texttt{M1,F1}} = .10$

$\mu_{\texttt{M1,F2}} = .80$, $\sigma_{\texttt{M1,F2}} = .10$

$\mu_{\texttt{M2,F1}} = .90$, $\sigma_{\texttt{M2,F1}} = .20$

$\mu_{\texttt{M2,F2}} = .80$, $\sigma_{\texttt{M2,F2}} = .10$

Thus, on the average legislative day `M1` and `F1` are likely to vote in the same direction 90 percent of the time with a standard deviation of 10 percent. This was calculated by subsetting the vote data using each legislative day in the appropriate Congress. In this example, January 20, 2010 is in the 111th Congress, so we would cycle through each day in the 111th Congress and calculate the percentage of time `M1` and `F1` voted together. Once done, we then take the mean and standard deviation across all the days, giving us an expectation of how often `M1` and `F1` should vote together on the average legislative day. We call this the "dyadic mean and standard deviation."

For the purposes of the example, let us assume there were only three days in the 111^{th} Congress. On these days, `M1` and `F1` voted in the following way:

$p1_{\texttt{M1,F1}} = .90$ (Day 1)

$p2_{\texttt{M1,F1}} = .80$ (Day 2)

$p3_{\texttt{M1,F1}} = 1$ (Day 3)

Thus, on Day 1 `M1` voted in the same direction as `F1` 90 percent of the time. On Day 2, `M1` voted in the same direction as `F1` 80 percent of the time. On Day 3, `M1` voted in the same direction as `F1` 100 percent of the time. If we took the mean and standard deviation of these values, we would get .90 and .10, respectively.

With these baseline measures in hand, we will now add the following step:

4. Find all the votes that occurred on the day of interest and determine the degree to which the male and female speakers cast votes in the same direction. Again, we restrict the female speakers to only those who delivered speeches using any of the Pearson and Dancey (2011*b*) terms.

For example, assume there were only two votes, `V1` and `V2`. To make things easier, we will report the votes associated with each dyad. These can be found here:

`M1` = `Yes`, `F1` = `Yes` (Vote 1)

`M1` = `Yes`, `F2` = `No` (Vote 1)

`M2` = `Yes`, `F1` = `Yes` (Vote 1)

`M2` = `Yes`, `F2` = `Yes` (Vote 1)

`M1` = `Yes`, `F1` = `Yes` (Vote 2)

`M1` = `Yes`, `F2` = `Yes` (Vote 2)

`M2` = `Yes`, `F1` = `Yes` (Vote 2)

`M2` = `Yes`, `F1` = `Yes` (Vote 2)

We can see that `M1` votes with `F1` 100 percent of the time, whereas he votes with `F2` 50 percent of the time. Conversely, `M2` votes with both `F1` and `F2` 100 percent of the time.

These percents are then standardized using the baseline measures calculated in Step 3. This yields the next step:

5. Standardize the percentage of instances male and female speakers vote in the same direction using the corresponding dyadic means and standard deviations. Again, restricting the female speakers to those who used at least one of the Pearson and Dancey (2011*b*) terms.

For example, when this is done for all the male (`M1` and `M2`) and female (`F1` and `F2`) speakers, we get the following:

`M1`, `F1` $= \frac{1-.90}{.10} = 1$

`M1`, `F2` $= \frac{.50-.80}{.10} = -3$

`M2`, `F1` $= \frac{1-.90}{.20} = 0.50$

`M2`, `F2` $= \frac{1-.80}{.10} = 2$

Finally, to convert these standardized percentages to an overall score for each male speaker, we simply take the average. This yields the final step:

6. Take the average of the standardize percentages created in Step 5. This average represents a male speaker's willingness to vote with female speakers who reference women on the day of interest, accounting for his baseline willingness to vote with those same female speakers.

When this is done for the male (`M1` and `M2`) speakers, we get the following:

$\texttt{M1} = \frac{1-3}{2} = -1$

$\texttt{M2} = \frac{2+.50}{2} = 1.25$

In Table 7, we used the average of the standardized scores as our dependent variable with two important caveats. First, we only used "yea" or "nay" votes. Excluding "present" votes and abstentions is standard practice in the literature, and we follow this norm for our study.

Second, as is standard practice in the literature, we only considered passage votes on House bills and resolutions. For these, we downloaded the roll call "Description" files from *Voteview*. The dependent variable in Table 7 only considers legislation that began with either "HR" or "HRES."

S6.2 Unstandardized Vote Measure

As a robustness check, we re-estimated the models in Table 7 using raw percentages. These results can be found in Table S16. Here, we restricted the analysis to male MCs who spoke on the same day as the female MCs who are used to generate the `Female Speeches` and `Female Pitch` variables. As shown in Table S16, the interaction between `Female Speeches` and `Female Pitch` is statistically significant and the coefficients are in the same direction as those found in Table 7. The same is true for the main effects.

We also estimated separate models for Democratic and Republican men with both our standardized and raw percentage measures. For both Democratic and Republican men, the interaction between `Female Speeches` and `Female Pitch` is positive and statistically significant in all models, irrespective of the measure used. The same can be said for the main effects.

Finally, it is important to note that we use these analyses mostly to rule out the possibility that men's increased attention to, and intensity in talking about, women is evidence of a backlash effect. Thus, the results outlined in Table 7 in the main text and Table S16 in the Supplemental Information should not be considered in isolation. Instead, they should be seen as preliminary evidence that the reaction of male MCs is likely a net positive for the advancement of women's interests in the House of Representatives. We hope this finding sparks future work on this topic.

S7 Comparing Democratic and Republican Members of Congress

Below we re-estimate the models from the main text while splitting the sample by partisanship.

Table S16: The Effect of Women's Speech Amount and Intensity on Men's Voting Patterns (Raw Percent)

	All		Democrats		Republicans	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	0.634***	0.786***	0.912***	0.873***	0.427***	0.787***
	(0.015)	(0.033)	(0.008)	(0.036)	(0.013)	(0.168)
Female Speeches	−0.004***	−0.003***	0.002***	0.002***	−0.010***	−0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Pitch	−0.040***	−0.035***	0.022***	0.028***	−0.093***	−0.087***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)
Female Speeches ×	0.005***	0.004***	−0.001	−0.003***	0.010***	0.009***
Female Pitch	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Controls		✓		✓		✓
N_1	11,943	11,943	5,314	5,314	6,629	6,629
N_2	482	482	215	215	267	267
Log Likelihood	−4,229.062	−3,823.784	529.471	551.688	−3,379.338	−3,286.319
AIC	8,470.125	7,677.567	−1,046.942	−1,073.377	6,770.677	6,602.639

Note: Models are identical to Table 7 except the outcome is the percentage of time male MCs voted with women. Controls are not shown to save space. Full models available upon request. Column labels indicate the subset of data used (all men, Democratic men, or Republican men respectively). Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

Table S17: Female MCs' References to Women, by Party

			Dependent variable:			
			"Women" Mentioned			
	All		Democrats		Republicans	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−2.427***	−2.364***	−2.322***	−2.171***	−2.519***	−1.693***
	(0.035)	(0.116)	(0.050)	(0.181)	(0.050)	(0.641)
Female	0.866***	0.795***	0.784***	0.794***	0.905***	0.875***
	(0.078)	(0.081)	(0.093)	(0.097)	(0.147)	(0.144)
Controls		✓		✓		✓
hline						
N_1	74,151	74,151	36,190	36,190	37,961	37,961
N_2	619	619	314	314	305	305
Log Likelihood	−23,909.700	−22,787.330	−12,713.380	−12,172.230	−11,192.220	−10,589.030
AIC	47,825.410	45,594.650	25,432.760	24,364.460	22,390.440	21,198.070

Note: Models are identical to Table 1, Models 1 and 2. However, we subset our data by party identification. Controls not shown to save space. Full models available upon request. Column labels indicate the subset of data used (all MCs, Democrats, or Republicans respectively). Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

S7.1 Party Effects Among Female MCs

First, we re-estimated the results outlined in Table 1 in the main text separately for Democratic and Republican women. In Table S17, the dependent variable is whether a female MC used any one of the Pearson and Dancey (2011*b*) terms. In the first two columns, we replicated the results reported in Table 1, Models 1 and 2 from the main text. In the next two columns, we re-estimated these same models using only Democratic women. In the final two columns, we did the same for Republican women. Our results do not change as one moves from left to right, suggesting the results reported in Table 1 of the paper cannot be attributed to a single party (e.g., Democrats). Indeed, it seems both Democratic and Republican women are more likely to reference women as compared to their male counterparts. This is also consistent with the results outlined in Section S3.

Similar results are found for Table 1, Models 3 and 4. These models from the main text are re-estimated in Table S18. Here, the dependent variable is standardized vocal pitch. In the first two columns, we replicated the results reported in Table 1. In the next two columns, we re-estimated these same models using only Democratic women. In the final two columns, we did the same for Republican women. Unlike the previous table, the results vary by party. Indeed, while the interaction term is in the same direction, the effect seems to be more pronounced for Democratic women. To ensure that our main results are not being driven solely by the behavior of Democratic women, the results we report in Table 1 control

Table S18: Female MCs' Standardized Vocal Pitch When Referencing Women, by Party

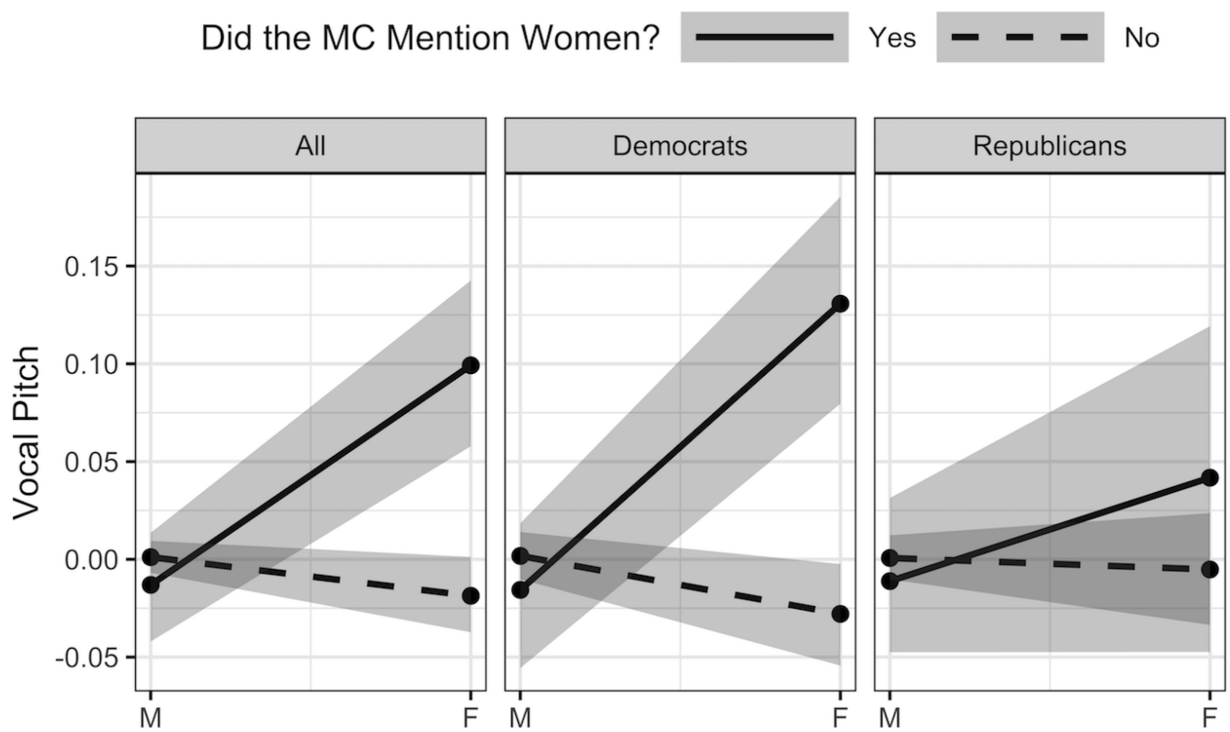
	Dependent variable: Standardized Vocal Pitch					
	All		Democrats		Republicans	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−0.002	0.126***	−0.002	0.104***	−0.001	0.164
	(0.004)	(0.015)	(0.007)	(0.026)	(0.006)	(0.126)
Female	−0.017	−0.032***	−0.021	−0.033**	−0.008	−0.026
	(0.011)	(0.011)	(0.013)	(0.013)	(0.018)	(0.018)
"Women" Mentioned	0.020	−0.054***	0.024	−0.049**	0.017	−0.061***
	(0.014)	(0.014)	(0.021)	(0.021)	(0.020)	(0.020)
Female × "Women" Mentioned	0.090***	0.112***	0.101***	0.124***	0.051	0.065
	(0.027)	(0.027)	(0.034)	(0.034)	(0.049)	(0.049)
Controls		✓		✓		✓
N_1	71,198	71,198	34,837	34,837	36,361	36,361
N_2	612	612	310	310	302	302
Log Likelihood	−100,720.100	−99,643.370	−49,277.870	−48,791.520	−51,453.090	−50,878.630
AIC	201,452.100	199,312.700	98,567.740	97,609.050	102,918.200	101,783.300

Note: Models are identical to Table 1, Models 3 and 4. However, we subset our data by party identification. Controls not shown to save space. Full models available upon request. Column labels indicate the subset of data used (all men, Democratic men, or Republican men respectively). Levels of significance are reported as follows: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Standard errors are reported in parentheses.

for the party identification of the speaker, meaning the general relationship we discuss in the paper cannot solely be attributed to the differential party effect outlined in this table. To emphasize this point, we plotted predicted values for Table S18, Models 1, 3, and 5 in Figure S9.

When we compare across the three panels ("All", "Democrats", and "Republicans"), we see a consistent positive relationship between gender and vocal pitch. This relationship is strongest for Democratic women. Indeed, although female Republicans tend to speak at a higher vocal pitch when referencing women, the 95-percent confidence intervals overlap considerably. Of course, this finding could be influenced by the comparatively smaller number of women in the Republican caucus, thus making it harder to detect an effect. To ensure that our main results are not being driven solely by the behavior of Democratic women, the results we report in Table 1 control for the party identification of the speaker, meaning the general relationship we discuss in the paper cannot solely be attributed to the differential party effect outlined in this plot.

Figure S9: Female MCs' Standardized Vocal Pitch When Referencing Women, by Party



Note: Predicted vocal pitch derived from Table 1, Model 2 in the main text. A dashed line indicates the speech mentioned "women." A solid line indicates all other speeches. In the x-axis we set the speaker's gender to either male or female. The y-axis has the predicted vocal pitch in standard deviations above or below the speakers' baseline. All other variables held constant.

Table S19: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women, by Party

	Dependent variable: "Women" Mentioned					
	All		Democrats		Republicans	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−2.695***	−2.458***	−2.650***	−2.139***	−2.735***	−1.800**
	(0.041)	(0.146)	(0.056)	(0.205)	(0.059)	(0.702)
Female Speeches	0.056***	0.060***	0.068***	0.072***	0.045***	0.051***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.005)
Female Pitch	−0.124***	−0.129***	−0.109**	−0.117**	−0.131***	−0.131***
	(0.031)	(0.032)	(0.046)	(0.048)	(0.042)	(0.043)
Female Speeches × Female Pitch	0.009	0.011**	0.010	0.015*	0.005	0.006
	(0.005)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
Controls		✓		✓		✓
N_1	50,235	50,235	22,207	22,207	28,028	28,028
N_2	509	509	234	234	275	275
Log Likelihood	−14,735.510	−13,951.110	−6,782.614	−6,452.685	−7,936.107	−7,469.518
AIC	29,481.010	27,930.230	13,575.230	12,933.370	15,882.210	14,967.040

Note: Models are identical to Table 5, but we subset our data by party identification. Controls not shown to save space. Full models available upon request. Column labels indicate the subset of data used (all men, Democratic men, or Republican men respectively). Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

S7.2 Party Effects Among Male MCs

We also re-estimated Table 5 in the main text separating Democratic and Republican men (see Table S19). Tables S20 and S21 reports similar models for Tables 6 and 7. Although the results associated with Republican men seem to be more pronounced, when predicted values are plotted in Figure S10, we found very little difference across the parties. Indeed, the plots themselves are nearly identical.

The solid lines suggest Republican and Democratic men respond similarly to changes in female speaking behavior. Indeed, as both `Female Speeches` and `Female Pitch` increase, both Republican and Democratic men seem to become more emotionally activated when referencing women. These findings give us confidence that the results we present in the main text cannot be attributed to one party. Instead, male MCs from both parties increase their vocal pitch when a large number of female MCs deliver speeches on women with heightened vocal pitch.

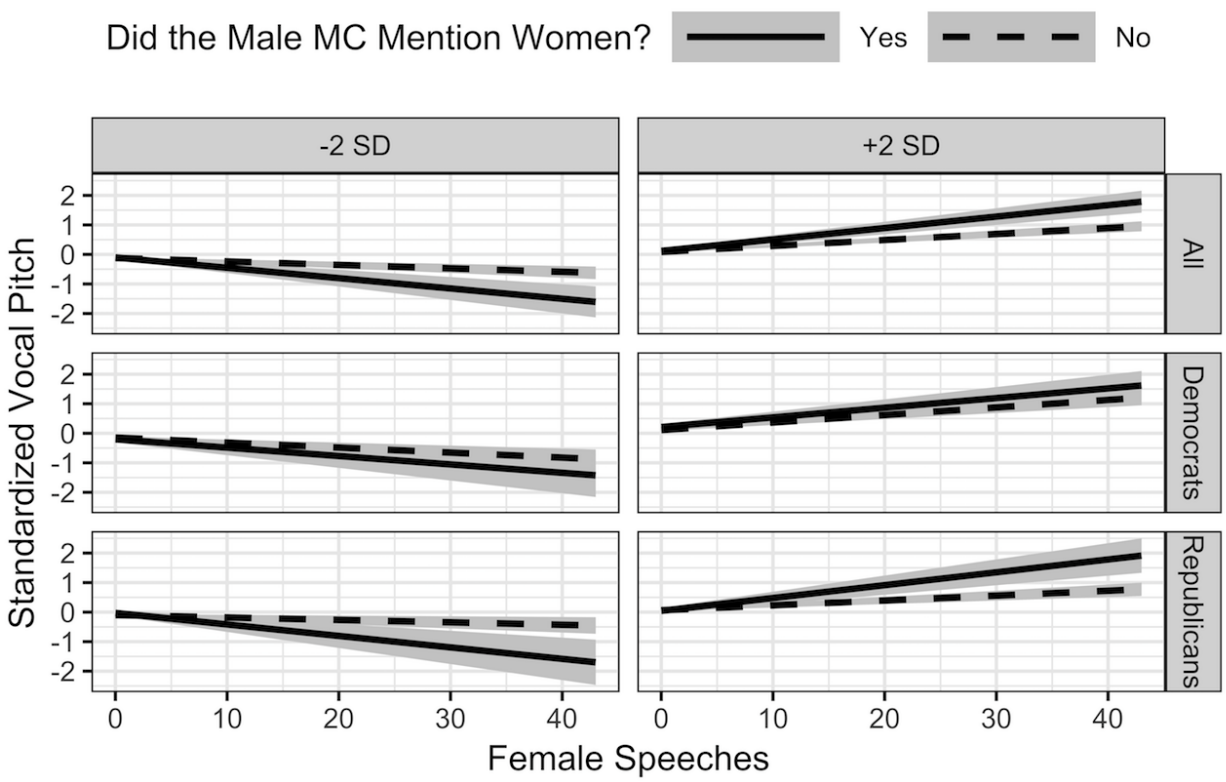
We also re-estimated the models in Table 7 from the main text using only Democratic

Table S20: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch by Party

	Dependent variable: Standardized Vocal Pitch		
	All	Democrats	Republicans
	(1)	(2)	(3)
Fixed Effects			
Constant	−0.022***	−0.024**	−0.021**
	(0.007)	(0.010)	(0.009)
"Women" Mentioned	0.026	0.019	0.027
	(0.022)	(0.033)	(0.030)
Female Speeches	0.003***	0.003*	0.004**
	(0.001)	(0.002)	(0.002)
Female Pitch	0.077***	0.095***	0.061***
	(0.009)	(0.014)	(0.012)
"Women" Mentioned × Female Speeches	−0.003	−0.003	−0.003
	(0.003)	(0.004)	(0.004)
"Women" Mentioned × Female Pitch	0.004	0.058	−0.035
	(0.031)	(0.046)	(0.041)
Female Speeches × Female Pitch	0.012***	0.016***	0.009***
	(0.002)	(0.003)	(0.003)
"Women" Mentioned × Female Speeches × Female Pitch	0.016***	0.007	0.021***
	(0.005)	(0.007)	(0.008)
Random Effects			
MC	0.000	0.000	0.000
N_1	49,914	22,080	27,834
N_2	506	233	273
Log Likelihood	−70,478.580	−31,110.330	−39,385.840
AIC	140,977.200	62,240.650	78,791.680

Note: Outcome is the vocal pitch of male speakers scaled to standard deviations above his baseline. `''Women'' Mentioned` indicates whether the speech used any of the Pearson and Dancey (2011*b*) terms. Column labels indicate the subset of data used (all men, Democratic men, or Republican men respectively). Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Figure S10: Estimated Effect of Quantity and Intensity of Women's Speeches on Men's Emotional Intensity by Party



Note: Predicted male vocal pitch derived from Table 6, Model 1 in the main text. Solid lines indicate a speech using a Pearson and Dancey (2011*b*) term, dashed lines indicate all other speeches. The y-axis displays the standardized vocal pitch of male speeches. On the x-axis `Female Speeches` is allowed to vary from its minimum (0) to maximum (43). The left panel shows `Female Pitch` set to two standard deviations below the mean. The right panel shows `Female Pitch` set to two standard deviations above the mean. Gray ribbons represent 90% confidence intervals. 95% confidence intervals look very similar, but the wide range in predicted values makes it difficult to see differences.

Table S21: The Effect of Women's Speech Amount and Intensity on Men's Voting Patterns (Standardized Percent)

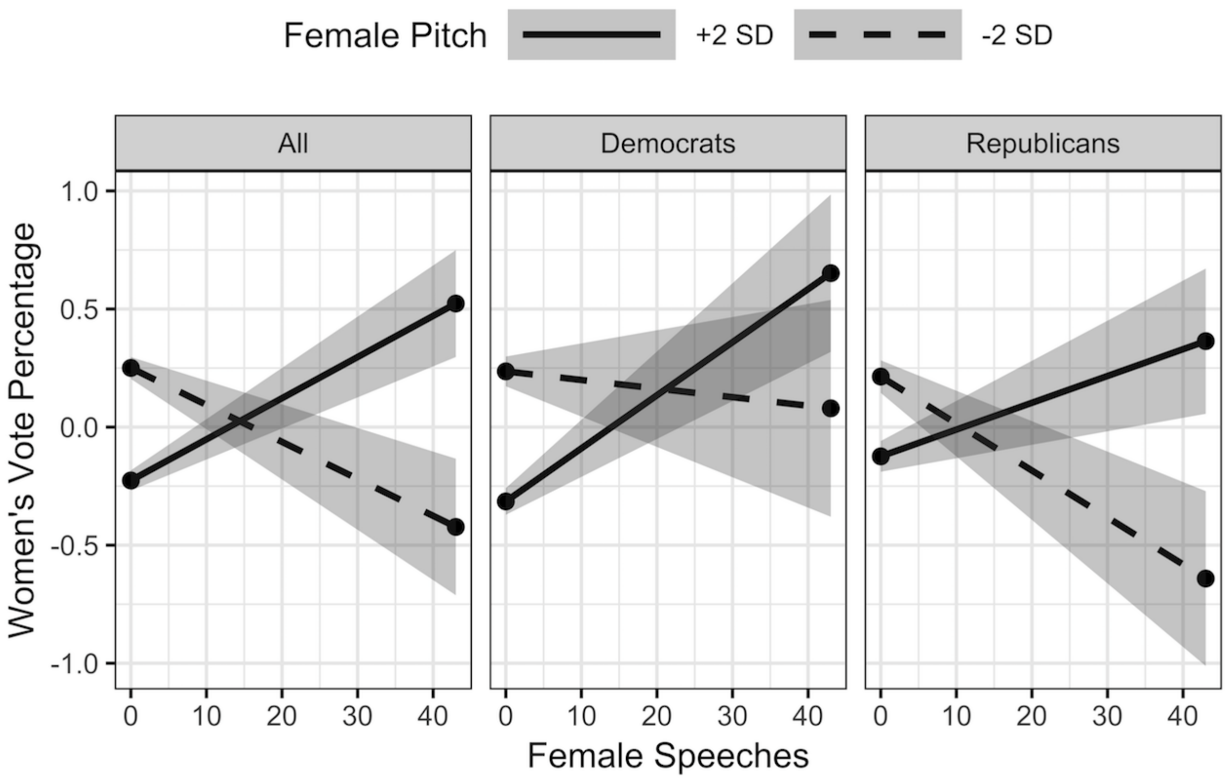
	Dependent variable: Male Votes Cast					
	All		Democrats		Republicans	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	0.019	0.100**	−0.026	0.127*	0.053**	0.503*
	(0.015)	(0.051)	(0.019)	(0.070)	(0.022)	(0.260)
Female Speeches	0.001	0.0001	0.011***	0.009***	−0.005***	−0.005***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Female Pitch	−0.187***	−0.177***	−0.221***	−0.205***	−0.120***	−0.126***
	(0.013)	(0.013)	(0.017)	(0.017)	(0.019)	(0.020)
Female Speeches × Female Pitch	0.015***	0.012***	0.012***	0.010***	0.013***	0.012***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Controls		✓		✓		✓
N_1	21,920	21,920	11,619	11,619	10,301	10,301
N_2	485	485	221	221	264	264
Log Likelihood	−28,122.730	−28,105.190	−15,270.430	−15,220.490	−12,800.010	−12,800.760
AIC	56,257.460	56,240.370	30,552.870	30,470.980	25,612.030	25,631.520

Note: Outcome is the percentage of time male MCs voted with women, as described on pages S36–S41. Control variables not shown to save space. Full models available upon request. Column labels indicate the subset of data used (all men, Democratic men, or Republican men respectively). Levels of significance are: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

or Republican men (see Table S21). Moving from left to right, it is readily apparent that our main results cannot be attributed to a single political party. Not only is the interaction between `Female Speeches` and `Female Pitch` positive and statistically significant, but the main effects of both variables are also the same in each model. To emphasize this point, we plotted predicted values for Table S21, Models 1, 3, and 5 in Figure S11.

In the first panel (see "All") we find a plot identical to Figure 3 in the main text. Here, we use the coefficients from Table S21, Model 1 to show male MCs tend to vote more with women as both `Female Speeches` and `Female Pitch` increase. In the second (see "Democrats") and third (see "Republicans") panels, we replicate the main result reported in Table 7 of the main text using only Democrats and Republicans. Not only is there generally a positive effect associated with increases in `Female Speeches` and `Female Pitch`, but the change is nearly identical for both groups. This suggests that men generally respond favorably to female speeches about women when they are delivered in large numbers and with a heightened vocal pitch.

Figure S11: Estimated Effect of Quantity and Intensity of Women's Speeches on Men's Voting Behavior by Party



Note: Predicted male voting behavior from Models 1 ("All"), 3 ("Democrats"), and 5 ("Republicans") in Tables S21. Solid and dashed lines indicate `Female Pitch` was set to two standard deviations above and below the mean respectively. On the x-axis `Female Speeches` is allowed to vary from its minimum (0) to maximum (43). The y-axis has the percentage of time the male MC voted with women, as described on pages S36–S41.

S8 Alternative Model Specifications

Below we present a series of alternative model specifications.

S8.1 Random Intercept: Speaker and Day

In the models we report in the main text, we view each speech as being a single observation that we can use to measure the emotional intensity a MC uses when speaking about women. Since we should expect multiple observations from the same MC to be related, we nest each speech within the MC. This not only accounts for potential speaker-level clustering, but is also consistent with our broader conceptualization of legislative speech.

One concern is that speeches are clustered not only at the legislator-level, but also at the day-level. In order to address the possibility that there is residual day-level clustering that we are not accounting for with the inclusion of controls for number of bills on women's interests and CQ bills, below we discuss several strategies for addressing day-level clustering.

The first possibility is to include a random intercept for each day. However, this is problematic due to the nature of legislative speech in the U.S. House. Many legislators give only a single speech on any given legislative day. Thus, we would be nesting both legislators and speeches within the same upper-level unit. This makes it nearly impossible for these models to converge, making this option methodologically intractable.

A second approach is to include a random intercept for each speaker-day pairing. This type of model would assume that MCs bring a certain level of emotional intensity towards discussing women on each legislative day. Similar to a repeated-measure design, this modeling strategy views each speech as being nested within each legislator, but unlike the modeling strategy used in the main text, this random intercept structure assumes that the level of emotional intensity in discussing women varies by day within each legislator.

We replicated the results reported in the main text using this modeling strategy of nesting speeches within each speaker-day. These results are reported in Tables S22-S25. Of these, Tables S22 and S23 are the most relevant. The former replicates Table 1, Models 1 and 2 including a random intercept for each speaker-day instead of a random intercept for each speaker. In each of these models women are found to speak significantly more about women, regardless of the measure one uses to operationalize references to women. The latter replicates Table 1, Models 3 and 4. Here too we find the results remain essentially unchanged. Most importantly, the interaction between `Female` and `Talking about ‘‘Women’’` is positive and statistically significant, suggesting that women tend to speak with more emotional intensity when talking about women.

Our findings concerning male MCs' responses are also consistent when using this alternative modeling strategy. In Table S24 we show that male MCs are more likely to talk about women when a large number of female speeches are delivered about women with emotional intensity, precisely what we reported in Table 5 in the main text. Somewhat more mixed

Table S22: Female MCs More Likely to Talk About Women (Speaker and Day)

	Dependent variable:					
	"Women" Mentioned		"Women" Percent		"Women" Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	-8.525^{***}	-5.192^{***}	0.052^{***}	0.108^{***}	2.598^{***}	3.806^{***}
	(0.081)	(0.001)	(0.002)	(0.010)	(0.033)	(0.185)
Female	0.943^{***}	0.523^{***}	0.133^{***}	0.122^{***}	2.175^{***}	1.922^{***}
	(0.095)	(0.081)	(0.004)	(0.004)	(0.073)	(0.075)
Controls		✓		✓		✓
N_1	74,151	74,151	74,151	74,151	74,150	74,150
N_2	36,240	36,240	36,240	36,240	36,240	36,240
Log Likelihood	−21,943.300	−21,149.460	−15,882.600	−15,824.220	−232,303.700	−232,153.300
AIC	43,892.590	42,320.920	31,773.200	31,672.430	464,615.300	464,330.700

Note: Re-estimated models found in Table S10 including a random intercept for the speaker-day, rather than a random intercept for just the speaker. Controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S23: Female MCs More Likely to Talk About Women with Greater Emotional Intensity (Speaker and Day)

	Dependent variable:					
	"Women" Mentioned		"Women" Percent		"Women" Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	0.056^{***}	0.227^{***}	0.059^{***}	0.219^{***}	0.075^{***}	0.245^{***}
	(0.005)	(0.030)	(0.005)	(0.030)	(0.006)	(0.030)
Female	−0.019	-0.033^{**}	−0.011	-0.027^{**}	−0.008	−0.021
	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)	(0.013)
Talking About "Women"	0.033^{**}	-0.046^{***}	0.014	0.006	-0.006^{***}	-0.007^{***}
	(0.014)	(0.014)	(0.018)	(0.017)	(0.001)	(0.001)
Female ×	0.088^{***}	0.110^{***}	0.055^{**}	0.059^{**}	0.005^{***}	0.004^{***}
Talking About "Women"	(0.027)	(0.026)	(0.024)	(0.024)	(0.001)	(0.001)
Controls		✓		✓		✓
N_1	71,198	71,198	71,198	71,198	71,197	71,197
N_2	36,240	36,240	36,240	36,240	36,240	36,240
Log Likelihood	−21,943.300	−21,149.460	−15,882.600	−15,824.220	−232,303.700	−232,153.300
AIC	−97,074.120	−95,799.470	−97,082.070	−95,801.030	−97,065.360	−95,771.050

Note: Re-estimated models found in Table S11 including a random intercept for the speaker-day, rather than a random intercept for just the speaker. Controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S24: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women (Speaker and Day)

	Dependent variable:					
	"Women" Mentioned		"Women" Percent		"Women" Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed Effects						
Constant	−8.559***	−4.564***	0.021***	0.072***	2.078***	3.012***
	(0.214)	(0.0002)	(0.002)	(0.009)	(0.048)	(0.217)
Female Speeches	0.068**	0.082***	0.007***	0.007***	0.126***	0.123***
	(0.029)	(0.0002)	(0.0003)	(0.0003)	(0.008)	(0.008)
Female Pitch	−0.130	−0.179***	−0.010***	−0.010***	−0.565***	−0.532***
	(0.292)	(0.0002)	(0.003)	(0.003)	(0.063)	(0.063)
Female Speeches × Female Pitch	0.010	0.028***	0.002***	0.002***	0.029**	0.026**
	(0.050)	(0.0002)	(0.001)	(0.001)	(0.013)	(0.013)
Controls		✓		✓		✓
N_1	50,235	50,235	50,235	50,235	50,234	50,234
N_2	23,413	23,413	23,413	23,413	23,413	23,413
Log Likelihood	−13,639.910	−13,243.050	2,649.738	2,662.739	−152,781.600	−152,668.600
AIC	27,289.830	26,516.100	−5,287.477	−5,293.478	305,575.200	305,369.200

Note: Re-estimated models found in Table S12 including a random intercept for the speaker-day, rather than a random intercept for just the speaker. Controls excluded to save space. Full models available upon request. All models converged, except for Models 1 and 2. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

Table S25: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch by Party (Speaker and Day)

	Independent variable:		
	"Women" Mentioned	"Women" Percent	"Women" Topic
	(1)	(2)	(3)
Constant	0.031***	0.038***	0.048***
	(0.009)	(0.008)	(0.009)
"Women" Mentioned	0.041*	−0.038	−0.006***
	(0.021)	(0.028)	(0.001)
Female Speeches	0.003**	0.002	0.003**
	(0.001)	(0.001)	(0.001)
Female Pitch	0.087***	0.081***	0.076***
	(0.012)	(0.011)	(0.012)
"Women" Mentioned × Female Speeches	−0.003	0.006**	0.0001
	(0.003)	(0.003)	(0.0001)
"Women" Mentioned × Female Pitch	0.018	0.124***	0.002
	(0.030)	(0.044)	(0.002)
Female Speeches × Female Pitch	0.012***	0.014***	0.015***
	(0.002)	(0.002)	(0.002)
"Women" Mentioned × Female Speeches × Female Pitch	0.012**	−0.007	−0.0003
	(0.005)	(0.006)	(0.0002)
N_1	49,914	49,914	49,913
N_2	23,323	23,323	23,323
Log Likelihood	−67,951.430	−67,951.580	−67,947.370
AIC	135,922.900	135,923.200	135,914.700

Note: Re-estimated models found in Table S13 including a random intercept for the speaker-day, rather than a random intercept for just the speaker. Outcome is the vocal pitch of male speakers scaled to standard deviations above their baseline. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

results are found in Table S25, which uses men's vocal pitch as the dependent variable.

Model 1 replicates our findings from the main text, but the interaction terms in Models 2 and 3 are less consistent. We think this can be attributed in part to the complexity of the interaction term and the restrictiveness of the speaker-day random intercept. More specifically, many legislators give only one speech on a given day. In these cases, the speaker-day random intercept is perfectly correlated with the day-level female speech and vocal pitch variables. Indeed, we cannot re-estimate the vote models we report in the main text (see Table 7) because in many cases the speaker-day random intercept was essentially identical to the dependent variable, which was a day-level measure of vote congruity. For these reasons, we believe that our original modeling strategy is more methodologically sound, even though we fully acknowledge that it does not perfectly address day-level clustering.

S8.2 Speaker Order: Dyadic Model

To address concerns about the temporal ordering of speeches when evaluating men's response to female MCs' speeches, we also estimated several dyadic models. By taking into account the temporal ordering of speeches, these models help us investigate whether there is an immediate backlash to an individual female MC's emotionally intense speech. Of course, since we are still relying on observational data, we are unable to test the causal mechanism linking female MCs' speeches to male MCs' subsequent behavior. At the same time, these models do allow us to have greater confidence in our findings concerning the absence of backlash effects.

In the first dyadic model specification, we simply see whether women speaking more about women at a higher vocal pitch can increase the likelihood that the next speaker (1) talks more about women, and (2) votes more with the preceding female speaker. That is, we examine the effect of the percent of the current speech which uses any of the Pearson and Dancey (2011*b*) terms (`"Women" Percent`$_t$) on the percent of the subsequent speech that uses any of the Pearson and Dancey (2011*b*) terms (`"Women" Percent`$_{t+1}$). In the second dyadic model specification, the dependent variables are the same, but we change the independent variable from whether the previous female speaker raises her vocal pitch above her baseline to whether the female speaker raised her pitch higher than the previous speaker. That is, we compare `Vocal Pitch`$_t$ to `Vocal Pitch`$_{t-1}$ to determine if that female MC's speech referencing women was more emotionally intense than the speech preceding it. In this way, we account for whether it is women's speeches that are elevating the emotional intensity of the chamber, or if both men's and women's speeches are responding to the broader emotional environment on the floor.

In Table S26, we present the results from our dyadic models. In Panel A, we report the standardized vocal pitch of the female MC (`Vocal Pitch`$_t$), whereas in Panel B we report the standardized vocal pitch of the female MC relative to the previous speech (`Vocal Pitch`$_t$ `-Vocal Pitch`$_{t-1}$). To account for likely clustering within dyads, we included a random intercept for each dyad. The results in Table S26 are consistent with Table 5 in the main

Table S26: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women (Dyads #1)

(a) Vocal Pitch$_t$

	Dependent variable:	
	"Women" Percent$_{t+1}$	
	(1)	(2)
Constant	0.0002**	0.001***
	(0.0001)	(0.0005)
"Women" Percent$_t$	1.240***	1.240***
	(0.022)	(0.022)
Vocal Pitch$_t$	0.00003	0.0001
	(0.0001)	(0.0001)
|Ideology$_{t+1}$ - Ideology$_t$|		−0.0003
		(0.0002)
|Duration$_{t+1}$ - Duration$_t$|		0.0001***
		(0.00003)
Same Race		−0.00002
		(0.0003)
Same Chair		−0.001***
		(0.0003)
Election Year		0.0001
		(0.0002)
"Women" Percent$_t$ × Vocal Pitch$_t$	0.070***	0.071***
	(0.024)	(0.024)
N	6,807	6,807
Log Lik	22,344.870	22,316.340
AIC	−44,677.740	−44,610.680

(b) Vocal Pitch$_t$ -Vocal Pitch$_{t-1}$

	Dependent variable:	
	"Women" Percent$_{t+1}$	
	(3)	(4)
Constant	0.0002**	0.001***
	(0.0001)	(0.0005)
"Women" Percent$_t$	1.246***	1.247***
	(0.021)	(0.021)
Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$	0.00004	0.0001
	(0.0001)	(0.0001)
|Ideology$_{t+1}$ - Ideology$_t$|		−0.0003
		(0.0002)
|Duration$_{t+1}$ - Duration$_t$|		0.0001***
		(0.00003)
Same Race		−0.00003
		(0.0003)
Same Chair		−0.001***
		(0.0003)
Election Year		0.0001
		(0.0002)
"Women" Percent$_t$ × (Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$)	0.053***	0.053***
	(0.019)	(0.019)
N	6,802	6,802
Log Lik	22,325.240	22,296.620
AIC	−44,638.480	−44,571.250

Note: Results are from dyadic models in which a female MC (t) speaks before a male MC ($t+1$). In all models the dependent variable is the total number of times the male MC ($t+1$) used one of the Pearson and Dancey (2011*b*) terms divided by the total number of words in the speech (`''Women'' Percent`$_{t+1}$). In Panel A, we interact the total number of times the female MC used one of the Pearson and Dancey (2011*b*) terms divided by the total number of words in the speech (`''Women'' Percent`$_t$) with the female MC's standardized vocal pitch (`Vocal Pitch`$_t$). In Panel B, we interact `''Women'' Percent`$_t$ with the female MC's standardized vocal pitch minus the standardize vocal pitch from the previous speaker `Vocal Pitch`$_{t-1}$. Ultimately, this variable (`Vocal Pitch`$_t$-`Vocal Pitch`$_{t-1}$) captures whether the female MC was more emotionally intense than the previous speaker. These models report the results from a multilevel linear regression. All models also include a randomly varying intercept for each dyad. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

text. In Panel A, when a female MC uses more Pearson and Dancey (2011*b*) terms (`''Women'' Percent`$_t$) in a speech with greater emotional intensity (`Vocal Pitch`$_t$), a speech by a male MC given immediately after her speech will tend to include a greater percentage of words about women. This result holds when we introduce our control variables (Model 2). Panel B shows evidence that male MCs' speeches are responding to changes in the vocal pitch of women's speeches, rather than the two rising in response to a previous speech. When a female MC uses more Pearson and Dancey (2011*b*) terms with a higher standardized vocal pitch than the speaker preceding her (`Vocal Pitch`$_t$ `-Vocal Pitch`$_{t-1}$), a subsequent speech by a male MC is likely to devote a higher percentage of the speech to talking about women. This evidence reinforces our findings in the main text. When female MCs talk more about women, and with greater intensity, male MCs become more likely to talk about women.

Table S27 addresses the much more difficult question of vote choice. As explained in the main text, it is difficult to imagine that a single female speech will influence male voting behavior. Instead, we argue that it is when a large number of female MCs take to the floor and give emotionally intense speeches about women that we should see men's voting patterns change. Moreover, unlike references to women, it is even more difficult to measure the effect of a single female speech on male voting behavior immediately after the female speech has concluded, since we cannot easily identify when votes occurred relative to speeches. With those caveats in mind, we now present a supplementary analysis using dyads to explore whether female MCs' speeches about women affect men's voting patterns. Instead of focusing on the degree to which a given male MC votes with all female speakers we are only going to consider the degree to which a male MC votes with the female speaker who directly preceded him. In our first set of analyses, we focus on the raw number of votes cast in the same direction. In our second set of analyses, we examine the percentage of votes cast in the same direction.

As before, we report the results of two models. The first, in Panel A, reports the effect of the standardized vocal pitch of the female MC (`Vocal Pitch`$_t$). The second, in Panel B, reports the standardized vocal pitch of the female MC relative to the previous speech (`Vocal Pitch`$_t$ `-Vocal Pitch`$_{t-1}$) to help isolate the effect of a single woman's speech. In both Panels, we see evidence that a female MC's emotionally intense speech about women has a positive effect on the likelihood that the subsequent male speaker votes in the same direction as her. The large and statistically significant interaction term, `''Women'' Percent`$_t$ $\times$ `Vocal Pitch`$_t$, shows this effect. As the percent of Pearson and Dancey (2011*b*) terms and standardized vocal pitch in a female MC's speech increase, an immediately following male speaker is significantly more likely to vote with that female MC on that day. This result holds with the introduction of control variables (Model 2), as well as when we consider whether the female MC's standardized vocal pitch was higher than the speaker preceding her (Models 3 and 4). Although these results are insufficient for demonstrating a causal relationship, they are consistent with the argument that the content and emotional intensity of women's speeches are linked to men's behaviors in the U.S. House.

Table S28 replicates these results using the percentage of votes cast in the same direction

Table S27: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (Dyads #2)

(a) Vocal Pitch_t

	Dependent variable:	
	Same Votes	
	(1)	(2)
Constant	2.248***	3.500***
	(0.030)	(0.101)
"Women" Percent_t	−10.754**	−8.411*
	(5.387)	(4.874)
Vocal Pitch_t	−0.111***	−0.061**
	(0.028)	(0.026)
$\lvert\text{Ideology}_{t+1}$ - $\text{Ideology}_t\rvert$		−1.690***
		(0.050)
$\lvert\text{Duration}_{t+1}$ - $\text{Duration}_t\rvert$		−0.002
		(0.006)
Same Race		0.001
		(0.062)
Same Chair		−0.135*
		(0.074)
Election Year		0.270***
		(0.049)
"Women" Percent_t × Vocal Pitch_t	13.833**	12.884**
	(5.909)	(5.331)
N	4,854	4,854
Log Lik	−9,887.093	−9,386.306
AIC	19,786.190	18,794.610

(b) Vocal Pitch_t -Vocal Pitch_{t-1}

	Dependent variable:	
	Same Votes	
	(3)	(4)
Constant	2.232***	3.500***
	(0.030)	(0.101)
"Women" Percent_t	−9.499*	−7.237
	(5.359)	(4.842)
Vocal Pitch_t - Vocal Pitch_{t-1}	−0.106***	−0.091***
	(0.021)	(0.019)
$\lvert\text{Ideology}_{t+1}$ - $\text{Ideology}_t\rvert$		−1.688***
		(0.050)
$\lvert\text{Duration}_{t+1}$ - $\text{Duration}_t\rvert$		−0.002
		(0.006)
Same Race		0.002
		(0.062)
Same Chair		−0.134*
		(0.074)
Election Year		0.285***
		(0.049)
"Women" Percent_t × (Vocal Pitch_t - Vocal Pitch_{t-1})	13.956***	13.914***
	(4.942)	(4.484)
N	4,851	4,854
Log Lik	−9,875.795	−9,370.803
AIC	19,763.590	18,763.610

Note: Results are from dyadic models in which a female MC (t) speaks before a male MC ($t+1$). In all models the dependent variable is the total number of times the male MC ($t+1$) cast the same vote as the female MC (t) on the day of the dyadic interaction. In Panel A, we interact the total number of times the female MC used one of the Pearson and Dancey (2011*b*) terms divided by the total number of words in the speech (`''Women'' Percent`$_t$) with the female MC's standardized vocal pitch (`Vocal Pitch`$_t$). In Panel B, we interact `''Women'' Percent`$_t$ with the female MC's standardized vocal pitch minus the standardize vocal pitch from the previous speaker `Vocal Pitch`$_{t-1}$. Ultimately, this variable (`Vocal Pitch`$_t$-`Vocal Pitch`$_{t-1}$) captures whether the female MC was more emotionally intense than the previous speaker. These models report the results from a multilevel linear regression. All models also include a randomly varying intercept for each dyad. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

Table S28: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (Dyads #3)

(a) Vocal Pitch$_t$

	Dependent variable:	
	Same Percent	
	(1)	(2)
Constant	0.605***	0.988***
	(0.006)	(0.017)
"Women" Percent$_t$	−2.264**	−1.464*
	(0.990)	(0.797)
Vocal Pitch$_t$	−0.025***	−0.012***
	(0.005)	(0.004)
|Ideology$_{t+1}$ - Ideology$_t$|		−0.476***
		(0.008)
|Duration$_{t+1}$ - Duration$_t$|		0.001
		(0.001)
Same Race		−0.026**
		(0.010)
Same Chair		−0.019
		(0.012)
Election Year		−0.002
		(0.008)
"Women" Percent$_t$ × Vocal Pitch$_t$	1.285	0.939
	(1.094)	(0.873)
N	4,854	4,854
Log Lik	−1,769.599	−617.486
AIC	3,551.198	1,256.972

(b) Vocal Pitch$_t$ -Vocal Pitch$_{t-1}$

	Dependent variable:	
	Same Percent	
	(3)	(4)
Constant	0.602***	0.986***
	(0.006)	(0.017)
"Women" Percent$_t$	−2.201**	−1.451*
	(0.988)	(0.794)
Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$	−0.012***	−0.008**
	(0.004)	(0.003)
|Ideology$_{t+1}$ - Ideology$_t$|		−0.477***
		(0.008)
|Duration$_{t+1}$ - Duration$_t$|		0.001
		(0.001)
Same Race		−0.025**
		(0.010)
Same Chair		−0.019
		(0.012)
Election Year		−0.001
		(0.008)
"Women" Percent$_t$ × (Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$)	1.779**	1.505**
	(0.903)	(0.735)
N	4,851	4,854
Log Lik	−1,775.689	−617.998
AIC	3,563.377	1,257.997

Note: Results are from dyadic models in which a female MC (t) speaks before a male MC ($t+1$). In all models the dependent variable is the percent of time the male MC ($t+1$) cast the same vote as the female MC (t) on the same day as the dyadic interaction. In Panel A, we interact the total number of times the female MC used one of the Pearson and Dancey (2011*b*) terms divided by the total number of words in the speech (``Women'' Percent$_t$) with the female MC's standardized vocal pitch (Vocal Pitch$_t$). In Panel B, we interact ``Women'' Percent$_t$ with the female MC's standardized vocal pitch minus the standardize vocal pitch from the previous speaker Vocal Pitch$_{t-1}$. Ultimately, this variable (Vocal Pitch$_t$-Vocal Pitch$_{t-1}$) captures whether the female MC was more emotionally intense than the previous speaker. These models report the results from a multilevel linear regression. All models also include a randomly varying intercept for each dyad. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

instead of the total number of identical votes. We find less consistent evidence of an effect on voting behavior here. When using a female MC's standardized vocal pitch (Models 1 and 2), we find a positive but statistically insignificant interaction effect between vocal pitch and percent of terms related to women. When considering whether a female MC's standardized vocal pitch is higher than the speaker preceding her, our results are consistent with those reported above: a subsequent male MC casts a greater percentage of his votes with the female MC preceding him (t) when she speaks with intensity (Vocal Pitch$_t$) about women ("Women" Percent$_t$).

Taken together, these dyadic models are generally supportive of the results we present in the main text. When we account for the temporal ordering of speeches, in six of our eight models we reach the same conclusions as our main findings. Given that our theoretical expectations focus on large numbers of female MCs speaking with intensity about women – and modeling men's responses in a dyadic framework is thus a conservative test of the impact of female speech – we find these results encouraging.

Because we are using observational data, we cannot make causal claims about the effect of female MCs' emotionally intense speeches about women on their male colleagues. We have attempted to rule out, however, a backlash to these speeches from male MCs. The dyadic models presented above leverage the temporal ordering of speeches and show that there is a consistent, and positive, relationship between the emotional intensity of female MCs' speeches about women and subsequent male MCs' behavior. Below, we present a series of placebo tests to help rule out alternative explanations for this correlation. Our first two placebo tests attempt to account for the possibility that men who take the floor after an emotionally intense female speech would likely vote in line with the preceding female speaker irrespective of the intensity of her speech. We thus estimate new dyadic models that replicate those described above, but predict men's voting behavior using their own vocal pitch and references to women, rather than that of the preceding female MC. Our final placebo test predicts male MCs' behavior based on the emotional intensity of a *subsequent* female MCs' speech about women.[10]

Table S29 reports the first of our placebo tests. Models 1 and 2 in Panel A show the results we obtained in Table S27, Model 1 and Table S28, Model 1, respectively. Recall that these are the predicted effect of a female MC's emotional intensity (`Vocal Pitch`$_t$) and references to women (`"Women" Percent`$_t$) on a subsequent male speaker's voting behavior. Models 3 and 4 in Panel B show our re-estimated placebo test. Here, we predict a male MC's number and percentage of votes cast in the same direction as the preceding female speaker based on that MC's emotional intensity (`Vocal Pitch`$_{t+1}$) and references to women (`"Women" Percent`$_{t+1}$). Based on this model, we find that a male MC's vocal pitch and references to women does not predict either the number of votes cast (Model 3) or percentage of votes cast (Model 4) with the female MC speaking immediately before him. This suggests that the dyadic results we presented in Tables S27 and S28 are not simply due to men who are already intense in their discussion of women taking to the floor following an emotionally

[10] We thank an anonymous reviewer for suggesting this test.

Table S29: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (Placebo #1)

(a) Vocal Pitch_t

	Dependent variable:	
	Same Votes	Same Percent
	(1)	(2)
Constant	2.248***	0.605***
	(0.030)	(0.006)
"Women" Percent_t	−10.754**	−2.264**
	(5.387)	(0.990)
Vocal Pitch_t	−0.111***	−0.025***
	(0.028)	(0.005)
"Women" Percent_t × Vocal Pitch_t	13.833**	1.285
	(5.909)	(1.094)
N	4,854	4,854
Log Lik	−9,887.093	−1,769.599
AIC	19,786.190	3,551.198

(b) Vocal Pitch_{t+1}

	Dependent variable:	
	Same Votes	Same Percent
	(3)	(4)
Constant	2.222***	0.601***
	(0.030)	(0.006)
"Women" Percent_{t+1}	2.684	−0.166
	(2.342)	(0.419)
Vocal Pitch_{t+1}	0.040	−0.002
	(0.028)	(0.005)
"Women" Percent_{t+1} Vocal Pitch_{t+1}	1.307	0.160
	(2.572)	(0.467)
N	4,853	4,853
Log Lik	−9,895.375	−1,785.202
AIC	19,802.750	3,582.405

Note: In Panel A, `Same Votes` is the number of times the subsequent male MC ($t + 1$) voted in the same direction as the female MC (t). In `Same Percent`, we divide `Same Votes` by the total number of votes on that legislative day. Results are reprinted from Model 1 in Tables S27 and S28. In Panel B, the dependent variables are the same, but we re-estimate those models using the total number of times the male MC used one of the Pearson and Dancey (2011*b*) terms divided by the total number of words in the speech (`''Women'' Percent`$_{t+1}$) with the male MC's standardized vocal pitch (`Vocal Pitch`$_{t+1}$). Ultimately, this tests whether the voting patterns outlined in Tables S27 and S28 can be attributed to the male MC generally talking about women with emotional intensity. These models report the results from a multilevel linear regression. All models also include a randomly varying intercept for each dyad. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

Table S30: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (Placebo #2)

(a) Vocal Pitch$_t$ -Vocal Pitch$_{t-1}$

	Dependent variable:	
	Same Votes	Same Percent
	(1)	(2)
Constant	2.232***	0.602***
	(0.030)	(0.006)
"Women" Percent$_t$	−9.499*	−2.201**
	(5.359)	(0.988)
Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$	−0.106***	−0.012***
	(0.021)	(0.004)
"Women" Percent$_t$ × (Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$)	13.956***	1.779**
	(4.942)	(0.903)
N	4,851	4,851
Log Lik	−9,875.795	−1,775.689
AIC	19,763.590	3,563.377

(b) Vocal Pitch$_{t+1}$ -Vocal Pitch$_{t-1}$

	Dependent variable:	
	Same Votes	Same Percent
	(3)	(4)
Constant	2.227***	0.601***
	(0.030)	(0.006)
"Women" Percent$_{t+1}$	2.616	−0.191
	(2.311)	(0.412)
Vocal Pitch$_{t+1}$ - Vocal Pitch$_{t-1}$)	−0.026	0.00003
	(0.021)	(0.004)
"Women" Percent$_{t+1}$ × (Vocal Pitch$_{t+1}$ - Vocal Pitch$_{t-1}$)	3.436	0.576
	(2.308)	(0.411)
N	4,850	4,850
Log Lik	−9,889.605	−1,785.051
AIC	19,791.210	3,582.103

Note: In Panel A, `Same Votes` is the number of times the subsequent male MC ($t+1$) voted in the same direction as the female MC (t). In `Same Percent`, we divide `Same Votes` by the total number of votes on that legislative day. Results are reprinted from Model 3 in Tables S27 and S28. In Panel B the dependent variables are the same, but we re-estimate those models using the total number of times the male MC used one of the Pearson and Dancey (2011*b*) terms divided by the total number of words in the speech (``Women'' Percent$_{t+1}$) with the male MC's standardized vocal pitch (Vocal Pitch$_{t+1}$) minus the standardized vocal pitch from the speaker who preceeded the female MC (Vocal Pitch$_{t-1}$). Ultimately, this tests whether the voting patterns outlined in Tables S27 and S28 can be attributed to the male MC generally talking about women with emotional intensity. These models report the results from a multilevel linear regression. All models also include a randomly varying intercept for each dyad. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

intense speech given by a female colleague.

We show a second placebo test in Table S30. In this test, we reproduce the results from Model 3 of Tables S27 and S28 in columns 1 and 2. These models are the predicted effect of a female MC's emotional intensity (Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$) and references to women relative to a previous speaker (``Women'' Percent$_t$ - ``Women'' Percent$_{t-1}$) on a subsequent male speaker's voting behavior. Columns 3 and 4 report our placebo test. These models predict a male speaker's number and percentage of votes cast in the same direction as the preceding female speaker based on his emotional intensity (Vocal Pitch$_{t+1}$) and references to women (``Women'' Percent$_{t+1}$), relative to those of the speaker before last. As before, we find no significant effects of male MC's own vocal pitch or references to women on their likelihood of voting with a preceding female speaker.

We present a third set of placebo tests in Table S31. The dependent variable in Models 1,

Table S31: Effect of Subsequent Women's Speech on Prior Men's Behavior (Placebo #3)

(a) Vocal Pitch$_t$

	Dependent variable:			
	Same Votes		Same Percent	
	(1)	(2)	(3)	(4)
Constant	2.567***	3.927***	0.623***	0.953***
	(0.037)	(0.124)	(0.007)	(0.019)
"Women" Percent$_t$	−0.620	−1.732	0.019	−0.196
	(7.287)	(6.466)	(1.267)	(0.967)
Vocal Pitch$_t$	−0.088**	−0.064*	−0.024***	−0.016***
	(0.037)	(0.033)	(0.006)	(0.005)
"Women" Percent$_t$ × Vocal Pitch$_t$	8.644	8.517	0.964	0.741
	(8.688)	(7.709)	(1.512)	(1.153)
Controls		✓		✓
N	3,441	3,441	3,441	3,441
Log Lik	−7,325.411	−6,919.179	−1,312.693	−396.814
AIC	14,662.820	13,860.360	2,637.386	815.628

(b) Vocal Pitch$_t$ -Vocal Pitch$_{t-1}$

	Dependent variable:			
	Same Votes		Same Percent	
	(5)	(6)	(7)	(8)
Constant	2.550***	3.911***	0.619***	0.951***
	(0.037)	(0.124)	(0.007)	(0.019)
"Women" Percent$_t$	1.775	0.932	0.165	−0.030
	(7.068)	(6.266)	(1.230)	(0.938)
(Vocal Pitch$_t$ - Vocal Pitch$_{t-1}$)	−1.685	−0.224	−0.161	0.068
	(5.949)	(5.272)	(1.035)	(0.789)
"Women" Percent$_t$ × (Vocal Pitch$_t$ -Vocal Pitch$_{t-1}$)	−1.685	−0.224	−0.161	0.068
	(5.949)	(5.272)	(1.035)	(0.789)
Controls		✓		✓
N	3,437	3,437	3,437	3,437
Log Lik	−7,317.752	−6,908.878	−1,315.319	−399.691
AIC	14,647.500	13,839.750	2,642.637	821.381

Note: `Same Votes` is the number of times the previous male MC ($t-1$) voted in the same direction as the female MC (t). In `Same Percent`, we divide `Same Votes` by the total number of votes on that legislative day. In Panel A, the percentage of the female MC's speech dedicated to women references (`"Women" Percent`$_t$) is interacted with her standardized vocal pitch (`Vocal Pitch`$_t$). In Panel B, the dependent variables are the same, but we subtract the previous male MC's standardized vocal pitch (`Vocal Pitch`$_{t-1}$). All models are multilevel linear regressions with randomly varying intercepts for each dyad. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

2, 5, and 6 (`Same Votes`) is the raw number of times the male speaker at time $t-1$ cast the same vote as a female speaker at time t on that legislative day. In Models 3, 4, 7, and 8 the dependent variable is the raw votes cast in the same direction divided by the total number of votes on that day (`Same Percent`). In both models, we predict male voting behavior at time $t-1$ using the percent of references to women and the vocal pitch of a female MC's speech given at time t. In this way, we are predicting a male MC's behavior using the subsequent speech of a female MC. We present in Table S31 both models with and without the set of control variables used in Tables S26–S28.

Because there is no means for a speaker to affect a temporally prior speaker, we should expect these results to be null. If our models predict a statistical relationship between subsequent female MCs' speeches and prior male MCs' behavior, then it would suggest that our findings may be influenced by a spurious correlation. Said differently, combined with Placebo Test #1 and Placebo Test #2, this placebo test should detect whether male MCs would have voted in line with a female MC regardless of her speech's content. With our observational data we are unable to provide a direct test of any causal link between female MCs' speeches and men's behavior, but this placebo test should help us understand if there is substantial cause for concern about the interpretation of our main results.

Panel A presents our results predicting the previous male MC's voting behavior using a female MC's speech he has yet to hear. Beginning with Model 1, we see the interaction between the current female MC's references to women (``"Women" Percent``$_t$) and her emotional intensity (`Vocal Pitch`$_t$) is not statistically distinguishable from zero. This finding is robust to the inclusion of controls (Models 2 and 4) as well as to an alternative construction of our dependent variable (Models 3 and 4). Together, this suggests that prior male speakers are not being influenced by the female MCs that follow them. This stands in contrast to Model 1 in Table S27. In that model, we found a strong relationship between a female MC's emotionally intense speech about women and a subsequent male MC's voting behavior. In particular, the male MC was significantly more likely to vote in the same direction as the female speaker. Together, these findings are consistent with male MCs responding (and responding favorably) to previous female MCs who delivered emotionally intense speeches about women. They are not consistent with those male MCs simply being more likely to vote in line with female speakers regardless of their speech's content or emotional intensity.

We extend this placebo test in Panel B of Table S31. For this extension, rather than using the standardized vocal pitch for a subsequent female MC, we instead consider whether that female MC was speaking with more emotional intensity than the previous male MC. This helps to account for whether that female MC gave a particularly emotionally intense speech relative to the context in which it was delivered. It also parallels Model 3 in Table S28. Our results for this placebo test echo those presented in Panel A. There is no statistically significant relationship between a female MC's emotionally intense speech about women and prior male speakers' behavior. This again suggests that our main dyadic results are not simply picking up a spurious correlation between female MCs' emotionally intense speeches and male MCs' voting behavior, but are instead consistent with a potential response to female MCs' speeches.

Since we are relying exclusively on observational data, we cannot provide a direct test of a causal mechanism linking female MCs' emotionally intense speeches about women to male MCs' behavior. This is especially relevant for our analysis of voting, since the literature has long suggested that floor speeches have a limited (if any) effect on voting behavior. Our results here are not meant to contradict this finding. Instead, our aim is simply to determine whether there is a backlash when a large number of female MCs speak with emotional intensity about women. Past research has shown that as women become more prevalent in legislatures, male politicians act to minimize their influence in order to maintain dominance (Heath, Schwindt-Bayer and Taylor-Robinson 2005; Kanthak and Krause 2012; Krook 2015), including becoming more aggressive and controlling of deliberation (Kathlene 1994). For this reason, it is important to consider the possibility that male MCs may respond negatively to female MCs speaking with more intensity about women.

In the main text, we test whether there is any evidence of male backlash using the interaction between the total number of female speeches on women and their average vocal pitch. Using the standardized vote measure introduced in Section S6.1, we show in Table 7 that male MCs tend to vote more with female MCs when they collectively speak with more emotional intensity about women. The result is then replicated using an unstandardized vote measure in Section S6.2. We next re-estimated our main results including measures for male speech on women. These results are reported in Table S41. Regardless of the model, we find no evidence of male backlash. This suggests when female MCs speak with greater intensity about women, they do not seem to face any immediate detrimental effects.

Our dyadic models are meant to give additional support to this claim. We do not argue that a single speech can have a large influence on men's (or women's) behavior, nor do we make any strong claims about the persuasive effects of speech in general. Instead, we use the dyadic models outlined above—combined with the corresponding placebo tests—to demonstrate that their is no evidence of male backlash against female MCs. Across all of our models, the most consistent statistical relationship is between a female MC speaking about women with intensity and an increase in the subsequent male MC's likelihood of voting with her. Our placebo tests provide added confidence that this statistical relationship is not solely due to male speakers being more likely to vote with female speakers regardless of the content of their speeches. This should not, however, be misconstrued as determining that a single female MC's speech can be pivotal in persuading a male MC to vote in a particular direction. There are a multitude of factors that influence an MC's vote choice, originating both within and outside the legislative chamber, and we could not possibly hope to rule out all of these omitted variables. Instead, what we offer here is simply a test of whether our data provide evidence consistent or inconsistent with a male backlash against female MC's efforts to speak on behalf of women.

In sum, whether it is talking more about women (Pearson and Dancey 2011*b*) or "women's issues" (Gerrity, Osborn and Mendez 2007; Osborn and Mendez 2010), scholars have consistently shown that female representatives are more likely to elevate the voice of women both

Table S32: Female MCs More Likely to Talk About Women, with Greater Intensity (No Outliers)

	Dependent variable:			
	"Women" Mentioned		Standardized Vocal Pitch	
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	−2.442***	−2.195***	−0.051***	0.054**
	(0.036)	(0.186)	(0.005)	(0.026)
Female	0.865***	0.783***	0.069***	0.052***
	(0.080)	(0.082)	(0.011)	(0.012)
"Women" Mentioned			0.018	−0.051***
			(0.014)	(0.014)
Female × "Women" Mentioned			0.060**	0.086***
			(0.013)	(0.013)
Controls		✓		✓
N_1	68,150	68,150	68,150	68,150
N_2	613	613	613	613
Log Likelihood	−21,837.380	−20,817.400	−85,688.450	−84,612.430
AIC	43,680.760	41,656.790	171,388.900	169,252.900

Note: Re-estimated the models from Table 1 only including speeches which had a standardized vocal pitch ±2 standard deviations. Controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

within (Pearson and Dancey 2011*a*) and beyond the halls of government (Herrnson, Lay and Stokes 2003). Our work suggests that such efforts do not lead to male backlash, which should give some comfort to gender and politics scholars who emphasize the importance of women's speech.

S8.3 Potential Outliers: Reduced Models

To address concerns that our results are being driven by a handful of influential and extreme speeches, we re-estimated all models in the main text eliminating potential outliers. To do so, we use a conservative definition of what constitutes an outlier: any speech with a vocal pitch more than ±2 standard deviations away from a speaker's baseline. We find that the substantive results we report in the main text remain unchanged even after restricting our data in this way.

Table S32 re-estimates Models 1-4 in Table 1 in the main text. As this table shows, our initial results are robust even after eliminating any speeches with very high or low vocal pitch relative to a speaker's baseline. Similar results can be found in Table S33, which re-estimates the models from Table 5 in the main text excluding days in which the average vocal pitch of

Table S33: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women (No Outliers)

	Dependent variable:	
	"Women" Mentioned	
	(1)	(2)
Fixed Effects		
Constant	-2.691^{***}	-2.210^{***}
	(0.041)	(0.221)
Female Speeches	0.056^{***}	0.061^{***}
	(0.003)	(0.003)
Female Pitch	-0.174^{***}	-0.170^{***}
	(0.034)	(0.035)
Female Speeches × Female Pitch	0.009	0.012^{*}
	(0.006)	(0.006)
Controls		✓
N_1	49,598	49,598
N_2	509	509
Log Likelihood	−14,572.020	−13,793.910
AIC	29,154.050	27,617.810

Note: Re-estimated the models from Table 5 only including speeches which had a standardized vocal pitch ±2 standard deviations. Controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

female speeches on women is not within ±2 standard deviations of the mean. In the main text, the interaction between `Female Speeches` and `Female Pitch` was only statistically significant when additional controls were included in the model. We find the same in Table S33, suggesting the substantive interpretation is equivalent even when potential outliers are removed from the data.

We also find generally consistent results when the dependent variable is male vocal pitch. This is shown in Table S34, which re-estimates the models from Table 6 in the main text. Even though the interaction between `Female Speeches`, `Female Pitch`, and `''Women'' Mentioned` is not statistically significant at the 0.05-level when speeches are restricted to ±2 standard deviations (see Model 1), the effect is in the same direction as our main result in Table 6. Moreover, when we relax our definition of an outlier to include only speeches that are ±3 standard deviations from the mean (see Model 2), we find results that are nearly identical to our main findings.

Finally, when potential outliers are excluded, we still find almost identical results to those presented in Table 7. Indeed, regardless of whether controls are (Model 2) or are not (Model 1) included, the results in Table S35 are essentially the same as those found in the main text. This suggest that our male vote results are not being driven by extreme cases in which

Table S34: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch (No Outliers)

	Dependent variable:	
	Male Vocal Pitch	
	(1)	(2)
Constant	-0.020^{***}	-0.023^{***}
	(0.007)	(0.007)
"Women" Mentioned	0.020	0.023
	(0.022)	(0.022)
Female Speeches	0.001	0.004^{***}
	(0.001)	(0.001)
Female Pitch	0.086^{***}	0.080^{***}
	(0.010)	(0.009)
"Women" Mentioned × Female Speeches	-0.001	-0.003
	(0.003)	(0.003)
"Women" Mentioned × Female Pitch	0.034	0.025
	(0.033)	(0.031)
Female Speeches × Female Pitch	0.020^{***}	0.011^{***}
	(0.002)	(0.002)
"Women" Mentioned × Female Speeches × Female Pitch	0.006	0.014^{***}
	(0.005)	(0.005)
Random Effects		
MC	0.000	0.000
N_1	49,324	49,884
N_2	506	506
Log Likelihood	$-69{,}035.120$	$-70{,}357.910$
AIC	138,090.200	140,735.800

Note: Re-estimated the models from Table 6 only including speeches which had a standardized vocal pitch ±2 standard deviations. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S35: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (No Outliers)

	Dependent variable:	
	Male Votes Cast	
	(1)	(2)
Fixed Effects		
Constant	0.021	0.096*
	(0.015)	(0.051)
Female Speeches	0.001	−0.00002
	(0.001)	(0.001)
Female Pitch	−0.184***	−0.176***
	(0.014)	(0.014)
Female Speeches × Female Pitch	0.015***	0.012***
	(0.002)	(0.002)
Controls		✓
N_1	21,769	21,769
N_2	485	485
Log Likelihood	−27,885.740	−27,870.470
AIC	55,783.480	55,770.930

Note: Re-estimated the models from Table 7 only including speeches which had a standardized vocal pitch ±2 standard deviations. Controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: *p < 0.1; **p < 0.05; ***p < 0.01. Standard errors are reported in parentheses.

female vocal pitch exceeds ±2 standard deviations when talking about women. Altogether, these robustness checks give us greater confidence that our main results are not dependent on a handful of extreme speeches or extreme speaking days where vocal pitch is far outside of the range we might expect in normal legislative discourse.

S8.4 Potential Confounders: Interacting CQ Bills and Women Bills

To address concerns regarding the legislative activities on a given day, we re-estimated the models found in Tables 5–7 including an interaction between `CQ Bills` and `Women Bills`. Here, the logic is relatively simple – female MCs might be especially likely to be emotionally intense when speaking about women on days where important legislation dealing with women's issues is being debated on the House floor. If this is the case, both male and female behavior may be explained by the bills being debated (rather than women's speech on women). Tables S36 and S37 attempt to gain traction on this question by including `CQ Bills` × `Women Bills` as an additional control.

Table S36 replicates our result from Table 5 with the inclusion of this new interaction term. Our results including this new interaction term are identical without controls; `Female`

Table S36: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women (Women's Bills × CQ Bills)

	Dependent variable:	
	"Women" Mentioned	
	(1)	(2)
Fixed Effects		
Constant	-2.695^{***}	-2.219^{***}
	(0.041)	(0.220)
Female Speeches	0.056^{***}	0.059^{***}
	(0.003)	(0.003)
Female Pitch	-0.124^{***}	-0.124^{***}
	(0.031)	(0.032)
Women Bills		-0.042
		(0.031)
CQ Bills		-0.108^{***}
		(0.039)
Women Bills × CQ Bills		0.109^{*}
		(0.058)
Female Speeches × Female Pitch	0.009	0.009
	(0.005)	(0.006)
Additional Controls		✓
N_1	50,235	50,235
N_2	509	509
Log Likelihood	−14,735.510	−13,948.520
AIC	29,481.010	27,929.030

Note: Re-estimated the models from Table 5 including the interaction between `Women Bills` and `CQ Bills`. Additional controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S37: Effect of Quantity and Intensity of Women's Speech on Men's Voting Patterns (Women Bills × CQ Bills)

	Dependent variable:	
	Male Votes Cast	
	(1)	(2)
Fixed Effects		
Constant	0.019	0.109**
	(0.015)	(0.051)
Female Speeches	0.001	−0.001
	(0.001)	(0.001)
Female Pitch	−0.187***	−0.176***
	(0.013)	(0.013)
Women Bills		0.048***
		(0.010)
CQ Bills		−0.030**
		(0.015)
Women Bills × CQ Bills		0.071***
		(0.023)
Female Speeches × Female Pitch	0.015***	0.011***
	(0.002)	(0.002)
Additional Controls		✓
N_1	21,920	21,920
N_2	485	485
Log Likelihood	−28,122.730	−28,103.140
AIC	56,257.460	56,238.280

Note: Re-estimated the models from Table 7 including the interaction between `Women Bills` and `CQ Bills`. Additional controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

`Pitch` and `Female Speeches` remain significant predictors of men's speech about women. In Model 2, however, the coefficient associated with the interaction term loses significance. However, the substantive effect is almost identical to that reported in the main text: 0.009 versus 0.011. Given the similarity to our main results, we remain confident that we are not simply picking up on the effect of important women's bills being on the agenda.

We are also able to replicate our findings for men's voting patterns. As shown in Model 2 of Table S37, our findings are nearly identical to those presented in Table 7. With the additional control for `Women Bills` × `CQ Bills`, our substantive results remain the same: `Female Pitch` and `Female Speeches` × `Female Pitch` remain statistically significant predictors of the likelihood of men voting with female speakers, and the magnitude of these coefficients is nearly identical.

The only findings we were unable to replicate with the inclusion of this additional control were those from Table 6 in the main text. Regardless of the specification, when `CQ Bills` was interacted with `Women Bills` the multilevel model failed to converge. Consequently, we cannot rule out this competing explanation for the effect of female speech on male vocal pitch. Still, the results presented here do not suggest that our main findings are being driven solely by the issues on the agenda. Rather, most of our results appear robust even when accounting for the interaction between important bills and bills concerning women.

S8.5 Potential Confounders: Questions

Throughout our paper, we argue that subtle changes in vocal pitch are indicative of emotional intensity (or activation), and our validation exercises presented in Section S2 above support that argument. Still, it is important to note that changes in vocal pitch can also be indicative of other linguistic features. For example, English speakers typically increase their vocal pitch at the end of a sentence to denote a question. This phenomenon (rising pitch tail) could contribute to the increase in vocal pitch we observe among legislators. In this section, we control for changes in vocal pitch associated with questions by re-estimating our models from the main text including a control for whether the speech included a question. We identify questions in speeches using the text of the speeches – any speech that included a question mark was coded as 1 on the `Question` variable. As we present below, our results are robust to the inclusion of this control.

Table S38 replicates our original results from Table 1. Our results here are the same. In Model S38.2, we see that `Female` remains statistically significant with a comparable magnitude after the inclusion of our control for speeches with questions. In Model S38.4, we find that the interaction term between `Female` and `''Women'' Mentioned` remains statistically significant, and the magnitude of the replicated results (0.103) is substantively similar to our original results (0.112).

Our results for men's quantity of speeches about women are similarly robust to the inclusion of this new control. In Table S39, we re-estimate our models from Table 5 and are primarily interested in the interaction between `Female Speeches` and `Female Pitch`. As in our main results, this interaction term is positive and statistically significant, with a substantive magnitude (0.012) that is nearly identical to our original result (0.011). These findings together suggest that our main results for women's speeches cannot be attributed to the rising pitch tail associated with questions.

We next consider whether the results reported in Table 6 are robust to the inclusion of a control for speeches including a question. As shown in Table S40, all of our key independent variables remain essentially unchanged from the main text; male MCs' vocal pitch is significantly higher when referencing women on days in which a large number of female MCs gave emotionally intense speeches about women. With the inclusion of this control, the coefficient on this interaction term is 0.013, almost identical to the 0.016 reported in the main text. As with our preceding analyses, this suggest that questions are not driving the higher pitch we observe in male MCs' responses to female MCs' speech.

Table S38: Female MCs More Likely to Talk About Women, with Greater Intensity (Controlling for Questions)

	Dependent variable:			
	"Women" Mentioned		Standardized Vocal Pitch	
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	-2.427^{***}	-2.190^{***}	-0.002	0.063^{***}
	(0.035)	(0.183)	(0.004)	(0.024)
Female	0.866^{***}	0.787^{***}	-0.017	-0.022^{**}
	(0.078)	(0.081)	(0.011)	(0.011)
Question		-0.107^{***}		0.370^{***}
		(0.030)		(0.009)
"Women" Mentioned			0.020	-0.040^{***}
			(0.014)	(0.014)
Female × "Women" Mentioned			0.090^{***}	0.103^{***}
			(0.027)	(0.026)
Additional Controls		✓		✓
N_1	74,151	74,151	71,198	71,198
N_2	619	619	613	613
Log Likelihood	−23,909.700	−22,780.610	−100,720.100	−98,736.320
AIC	47,825.410	45,585.210	201,452.100	197,502.600

Note: Re-estimated models from Table 1 in the main text including a dummy variable for whether the speech included a question. Additional controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S39: The Effect of Quantity and Intensity of Women's Speech on Quantity of Men's Speeches About Women (Controlling for Questions)

	Dependent variable:	
	"Women" Mentioned	
	(1)	(2)
Fixed Effects		
Constant	-2.695^{***}	-2.216^{***}
	(0.041)	(0.219)
Female Speeches	0.056^{***}	0.060^{***}
	(0.003)	(0.003)
Female Pitch	-0.124^{***}	-0.128^{***}
	(0.031)	(0.032)
Question		-0.086^{**}
		(0.039)
Female Speeches × Female Pitch	0.009	0.012^{**}
	(0.005)	(0.006)
Additional Controls		✓
N_1	50,235	50,235
N_2	509	509
Log Likelihood	−14,735.510	−13,947.810
AIC	29,481.010	27,927.620

Note: Re-estimated models from Table 5 in the main text including a dummy variable for whether the speech included a question. Additional controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Table S40: The Effect of Quantity and Intensity of Women's Speech on Men's Vocal Pitch (Controlling for Questions)

	Dependent variable:	
	Male Vocal Pitch	
	(1)	(2)
Fixed Effects		
Constant	-0.022^{***}	0.029
	(0.007)	(0.030)
"Women" Mentioned	0.026	-0.028
	(0.022)	(0.022)
Female Speeches	0.003^{***}	0.004^{***}
	(0.001)	(0.001)
Female Pitch	0.077^{***}	0.072^{***}
	(0.009)	(0.009)
Question		0.354^{***}
		(0.010)
"Women" Mentioned × Female Speeches	-0.003	-0.003
	(0.003)	(0.003)
"Women" Mentioned × Female Pitch	0.004	0.004
	(0.031)	(0.030)
Female Speeches × Female Pitch	0.012^{***}	0.010^{***}
	(0.002)	(0.002)
"Women" Mentioned × Female Speeches × Female Pitch	0.016^{***}	0.013^{**}
	(0.005)	(0.005)
Additional Controls		✓
N_1	49,914	49,914
N_2	506	506
Log Likelihood	$-70{,}478.580$	$-69{,}255.900$
AIC	140,977.200	138,553.800

Note: Re-estimated models from Table 6 in the main text including a dummy variable for whether the speech included a question. Additional controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Altogether, these findings are consistent with vocal pitch capturing an important aspect of emotions in speech. We acknowledge that vocal pitch may indicate other linguistic features – such as questions – in ordinary speech. However, it is clear from these results that we are not simply detecting the effects of questioning sentences in speeches. And there is no reason to believe that increased vocal pitch over the entire duration of a speech has linguistic value in the same way that rising pitch tail conveys a query to an English speaker. For this reason, these robustness checks give us greater confidence that our measure of vocal pitch is detecting larger shifts in emotional content, rather than smaller linguistic features of sentences.

S8.6 Potential Confounders: Men Speaking About Women

On pages 26–27 in the main text, we argue that a large number of female MCs speaking intensely about women could affect Congressmen's behavior. Since we are relaying on observational data, we cannot establish a clear causal relationship, but we are able to present a series of models showing a clear relationship between female MCs' speaking behavior and male MCs' voting behavior. Specifically, we show that on legislative days when many female MCs give emotionally intense speeches about women, male MCs are more likely to cast votes in the same direction as those female MCs. In this section, we examine whether we see a similar relationship between when *male MCs* give emotionally intense speeches about women and men's voting behavior. As we show below, our results in the main text are robust to the inclusion of male MCs' speaking behavior, suggesting a unique relationship between female MCs' speeches about women and male MCs' voting behavior.

Table S41 shows our replication of Table 7 from the main text with the addition of the number and vocal pitch of male MCs' speeches about women. In both Models 1 and 2, our original findings are robust to the inclusion of men's speaking behavior as a predictor of men's votes. Specifically, our original coefficient for `Female Pitch` in Model 1 was -0.187, which increases to -0.207 after including controls for male MCs' speaking behavior. We similarly find robust results for the interaction between `Female Speeches` and `Female Pitch`. This suggests that the results reported in Table 7 in the main text are robust even after accounting for male speaking behavior, and indeed appear to be conservative estimates of the coefficients for women's speaking behavior.

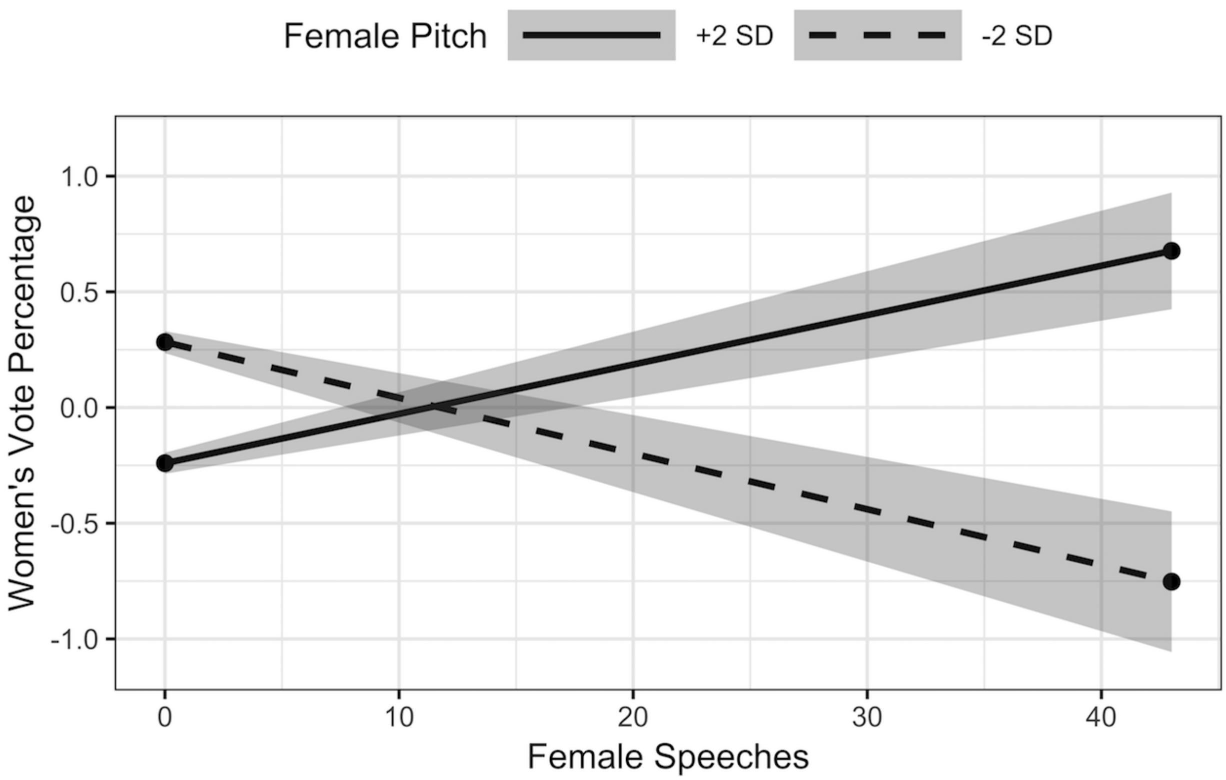
Also worth noting is that the interaction between *Male Speeches* and *Male Pitch* has a negative relationship with the degree to which male MCs vote in the same direction as female MCs. In both our original findings and the results presented here, we find that when many female MCs give emotionally intense speeches about women, male MCs are more likely to vote in the same direction as female MCs. Indeed, if we compare Figure S12 to Figure 3 in the main text, we see an essentially unchanged relationship between female MCs' speeches and male MCs' voting behavior. In Figure S13, however, we see a very different relationship between male MCs' speeches about women and male MCs' voting behavior. Here, the y-axis is the same as in Figure S12, but the x-axis is the number of male speeches referencing women. The solid line shows the predicted effect with male vocal pitch set at two standard

Table S41: Relationship between Women's Speech and Men's Voting Patterns (Controlling for Male Speeches)

	Dependent variable:	
	Male Votes Cast	
	(1)	(2)
Fixed Effects		
Constant	−0.011	0.107
	(0.016)	(0.079)
Female Speeches	−0.001	−0.002
	(0.002)	(0.002)
Female Pitch	-0.207^{***}	-0.195^{***}
	(0.013)	(0.014)
Male Speeches	0.004^{***}	0.004^{***}
	(0.001)	(0.001)
Male Pitch	−0.006	−0.022
	(0.019)	(0.019)
Female Speeches × Female Pitch	0.019^{***}	0.017^{***}
	(0.002)	(0.002)
Male Speeches × Male Pitch	-0.005^{***}	-0.005^{**}
	(0.002)	(0.002)
Additional Controls		✓
N_1	21,614	21,614
N_2	509	509
Log Likelihood	−27,702.840	−27,687.190
AIC	55,423.680	55,412.370

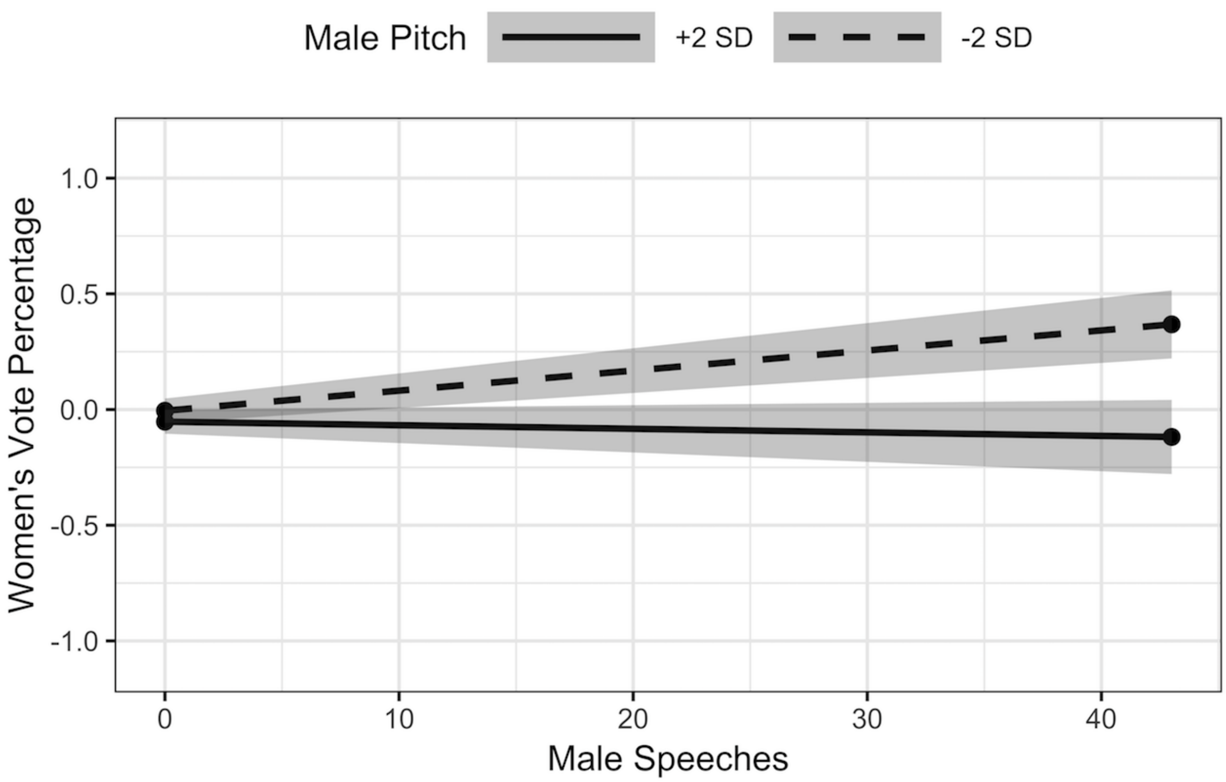
Note: Re-estimated models from Table 5 in the main text including the interaction between the number of male speeches mentioning women and the average vocal pitch of those speeches. Additional controls excluded to save space. Full models available upon request. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Figure S12: Relationship between Women's Speech and Men's Voting Patterns (Controlling for Male Speeches)



Note: Predicted male voting behavior from Model 2 in Table S41 holding all other variables constant. Solid and dashed lines indicate `Female Pitch` was set to two standard deviations above (1.41) and below (-1.28) the mean respectively. On the x-axis `Female Speeches` is allowed to vary from its minimum (0) to maximum (43). The y-axis has the percentage of time the male MC voted with women, as described on pages S36–S41. The gray ribbons represent 95 percent confidence intervals.

Figure S13: The Quantity and Intensity of Women's Speech Affects Men's Voting Patterns



Note: Predicted male voting behavior from Model 2 in Table S41 holding all other variables constant. Solid and dashed lines indicate `Male Pitch` was set to two standard deviations above (1.13) and below (-1.01) the mean respectively. On the x-axis `Male Speeches` is allowed to vary from its minimum (0) to maximum (55). The y-axis has the percentage of time the male MC voted with women, as described on pages S36–S41. The gray ribbons represent 95 percent confidence intervals.

deviations above the mean, and the dashed line shows the predicted effect with male vocal pitch set at two standard deviations below the mean. What the solid line shows is that as many male MCs give emotionally intense speeches about women, the percentage of votes cast in the same direction as female MCs is essentially unchanged (and perhaps slightly lower). This suggests that the positive relationship we see between emotionally intense speeches about women and male MCs' voting behavior is unique to female MCs' speeches. This is entirely consistent with our argument about the importance of women's presence in legislative discourse.

S8.7 Potential Confounders: Expertise

The intensity of female MCs' speech about women could be influenced by their level of expertise. That is, the changes we observe in female MCs' vocal pitch when speaking about women might represent their greater confidence in speaking, rather than an underlying emotional commitment to representing women. Although we think it is likely that female MCs who speak intensely about women are also likely to have expertise on women's issues, in this section we test whether changes in vocal pitch can be captured by variables that measure expertise, including the number of women's bills introduced and the average interest group rating.

Table S42: Female MCs More Likely to Talk with Greater Intensity About Women (Controlling for the Number of Women's Bills Introduced)

	Dependent variable:			
	Standardized Vocal Pitch		Standardized Vocal Pitch	
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	-2.427^{***}	-2.218^{***}	-0.002	0.151^{***}
	(0.035)	(0.183)	(0.004)	(0.024)
Female	-0.020^{*}	-0.032^{***}	-0.014	-0.028^{**}
	(0.011)	(0.011)	(0.011)	(0.011)
Women's Bills Introduced	0.006	0.002	0.011^{**}	0.005
	(0.004)	(0.004)	(0.005)	(0.005)
"Women" Mentioned	0.019	-0.056^{***}	0.016	-0.057^{***}
	(0.015)	(0.015)	(0.015)	(0.015)
Female × "Women" Mentioned	0.090^{***}	0.113^{***}	0.070^{**}	0.091^{***}
	(0.027)	(0.027)	(0.029)	(0.029)
Female × Women's Bills Introduced			-0.026^{***}	-0.019^{**}
			(0.010)	(0.010)
"Women" Mentioned × Women's Bills Introduced			-0.026^{***}	-0.019^{**}
			(0.010)	(0.010)
Female × "Women" Mentioned × Women's Bills Introduced			0.076^{***}	0.068^{***}
			(0.025)	(0.024)
Additional Controls		✓		✓
N_1	69,644	69,644	69,644	69,644
N_2	588	588	588	588
Log Likelihood	−98,531.450	−97,494.600	−98,535.470	−97,500.110
AIC	197,076.900	195,019.200	197,090.900	195,036.200

Note: Re-estimating Models 3 and 4 from Table 1 in the main text including the number of bills sponsored which deal with women's issues as defined by Volden, Wiseman and Wittmer (2018). Additional controls excluded to save space. Full models available upon request. `Women's Bills Introduced` is standardized using the mean and standard deviation from each Congress. More details can be found on pages S80–S81. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

We begin with Table S42. Here, we include a variable that captures the total number of women's bills introduced by each MC. We define "women's bills" using the issues outlined by Volden, Wiseman and Wittmer (2018) (see page 28 in the main text). We next determine the number of bills each MC sponsors in a given Congress that fall into the major topic areas defined by Volden, Wiseman and Wittmer (2018). We then divide this sum by the total number of bills the MC sponsored in the same Congress. For example, Rep. Rosa DeLauro

(D-CT) sponsored 207 bills in the 111th Congress of which 46 addressed women's issues. This means that 22.22 percent of the bills she sponsored in the 111th Congress fell into at least one of the major topic areas defined by Volden, Wiseman and Wittmer (2018).

We create a different measure for each Congress, as some terms are likely more conducive to the advancement of women's issues than others. We then standardize the percentage of women's bills using the mean and standard deviation for a given Congress. We call this variable `Women's Bills Introduced`. Here, positive values indicate MCs introduced more women's bills than we would expect given the percentage of women's bills introduced that Congress. Conversely, negative values suggest MCs were below average in terms of the percentage of their sponsored bills addressing women's issues.[11]

We present the results in Table S42. In Models 1 and 2, we include `Women's Bills Introduced` as a control variable. We find essentially the same results as those presented in the main text. The coefficient for the interaction between `Female` and `"Women" Mentioned` is unchanged when `Women's Bills Introduced` is included as a control. This provides strong evidence that vocal pitch and the number of women's bills introduced are not interchangeable, which suggests we are capturing something new with our measure of emotional intensity.

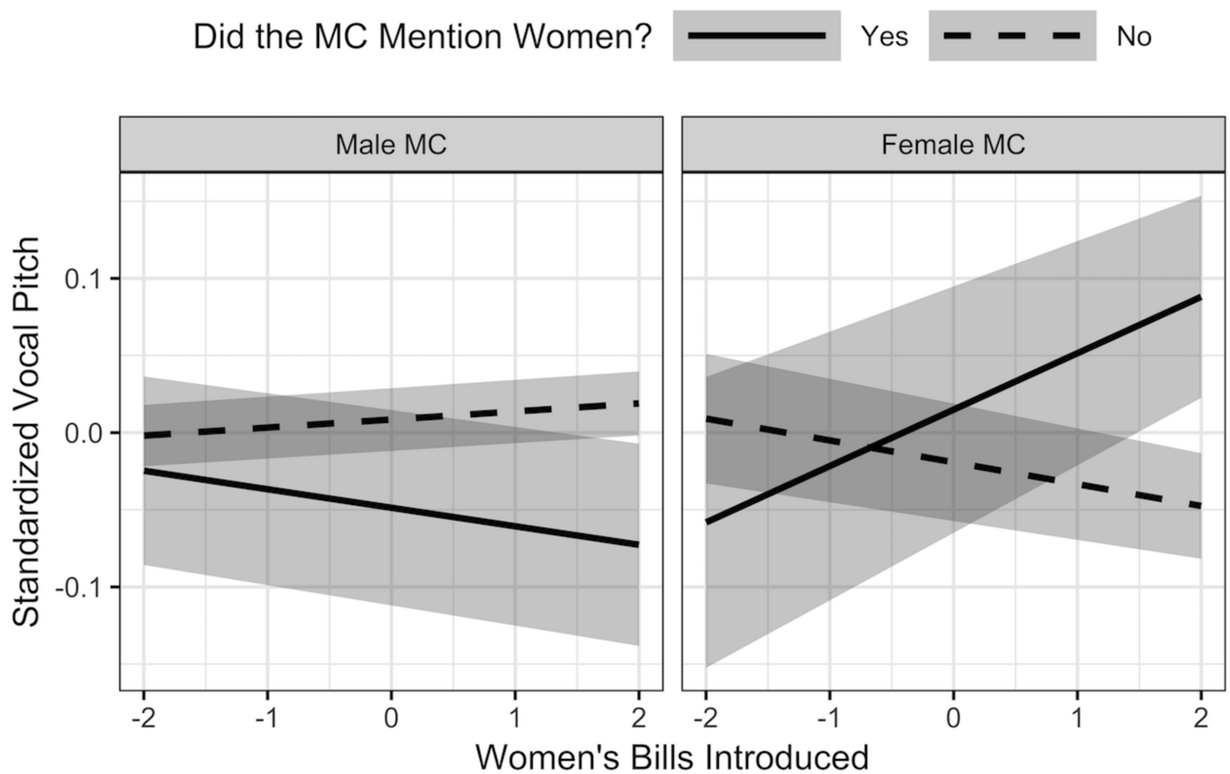
In Models 3 and 4 we interact our measure of emotional intensity with the number of women's bills introduced. These models test whether female MCs who introduce more women's bills also speak about women with greater emotional intensity. Although the interaction between `Female`, `"Women" Mentioned`, and `Women's Bills Introduce` is positive and statistically significant at the 0.001-level, it is difficult to directly interpret the coefficient. Figure S14 thus reports predicted values when `Women's Bills Introduced` is allowed to vary ± 2 standard deviations.

When both `Female` and `"Women" Mentioned` are set to 1, MCs' are predicted to speak at a higher vocal pitch as `Women's Bills Introduced` increases. Said differently, female MCs who introduce more women's bills tend to speak with *more* emotional intensity when speaking about women. The dashed line also shows the inverse is true when these female legislators do not mention women. That is, Congresswomen who introduce more women's bills tend to speak with *less* emotional intensity when `"Women" Mentioned` is set to 0. Not only is this result consistent with our broader argument, but it also demonstrates that vocal pitch may yield additional insights when used in conjunction with more traditional variables (such as the percentage of sponsored bills which deal with women's issues). Regardless, these results provide strong evidence that the number of women's bills introduced (i.e., confidence/expertise) is not a substitute for our vocal pitch measure.

In Table S43, we conduct a similar analysis, except instead of including `Women's Bills Introduced` as a control we use the women's interest group ratings we describe on page 19 in the main text. More specifically, for each MC we computed the average score from the 24 groups outlined in Table S4 for a given Congress. Similar to `Women's Bills Introduced`,

[11] This standardization means that the results outlined in Table S42 are on the same scale as those outlined in Table S43. As women's bills and women's interest group ratings are on different scales, this standardization is needed in order to make these two sets of results more comparable.

Figure S14: Intensity of Speeches about Women by Number of Women's Bills Introduced



Note: Predicted vocal pitch derived from Model 4 in Table S42 holding all other variables constant. Solid lines indicate the speech included at least one of the Pearson and Dancey (2011*b*) women's dictionary terms. Dashed lines indicate all other speeches. For a given MC, `Women's Bills Introduced` captures whether an MC (as compared to the rest of the Congress) tended to dedicate a greater percentage of his/her sponsored bills to women's issues. More details can be found on pages S80–S81. On the x-axis, `Women's Bills Introduced` is allowed to vary from ± 2 standard deviations. The y-axis reports the predicted standardized vocal pitch with positive values implying greater emotional intensity. The gray ribbons represent 95 percent confidence intervals.

Table S43: Female MCs More Likely to Talk with Greater Intensity About Women (Controlling for Women's Interest Group Ratings)

	Dependent variable:			
	Standardized Vocal Pitch		Standardized Vocal Pitch	
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	-2.427^{***}	-2.218^{***}	-0.002	0.151^{***}
	(0.035)	(0.183)	(0.004)	(0.024)
Female	-0.014	-0.031^{***}	-0.010	-0.025^{**}
	(0.011)	(0.011)	(0.011)	(0.011)
Women's Group Rating	-0.008^{**}	-0.019^{***}	-0.010^{**}	-0.021^{***}
	(0.004)	(0.006)	(0.004)	(0.007)
"Women" Mentioned	0.020	-0.054^{***}	0.021	-0.054^{***}
	(0.014)	(0.015)	(0.014)	(0.015)
Female × "Women" Mentioned	0.090^{***}	0.111^{***}	0.042	0.064^{**}
	(0.027)	(0.027)	(0.030)	(0.029)
Female × Women's Group Rating			-0.008	-0.014
			(0.011)	(0.011)
"Women" Mentioned × Women's Group Rating			0.009	0.012
			(0.015)	(0.014)
Female × "Women" Mentioned × Women's Group Rating			0.095^{***}	0.090^{***}
			(0.030)	(0.029)
Additional Controls		✓		✓
N_1	71,154	71,154	71,154	71,154
N_2	612	612	612	612
Log Likelihood	−100,650.600	−99,570.690	−100,651.700	−99,572.040
AIC	201,315.100	199,171.400	201,323.400	199,180.100

Note: Re-estimating Models 3 and 4 from Table 1 in the main text including the average women's interest group rating. Additional controls excluded to save space. Full models available upon request. `Women's Group Rating` is standardized using the mean and standard deviation from each Congress. More details can be found on page S81. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

we standardized this measure using the mean and standard deviation for a given Congress. We call this variable `Women's Group Rating`. Here, positive (negative) values imply MCs cast more (fewer) votes in the preferred direction of the 24 groups as compared to the average MC for that Congress. Again, this make the results outlined in Tables S42 and S43 more comparable.

We present the results in Table S43. In Models 1 and 2 we simply include `Women's Group Rating` as a control variable. When the coefficient for the interaction between `Female` and `"Women" Mentioned` is compared to the original coefficient we report in Table 1 in the main text, we again find essentially the same results. This once again provides strong evidence that vocal pitch and confidence/expertise (here, as captured by women's interest group ratings) are not interchangeable.

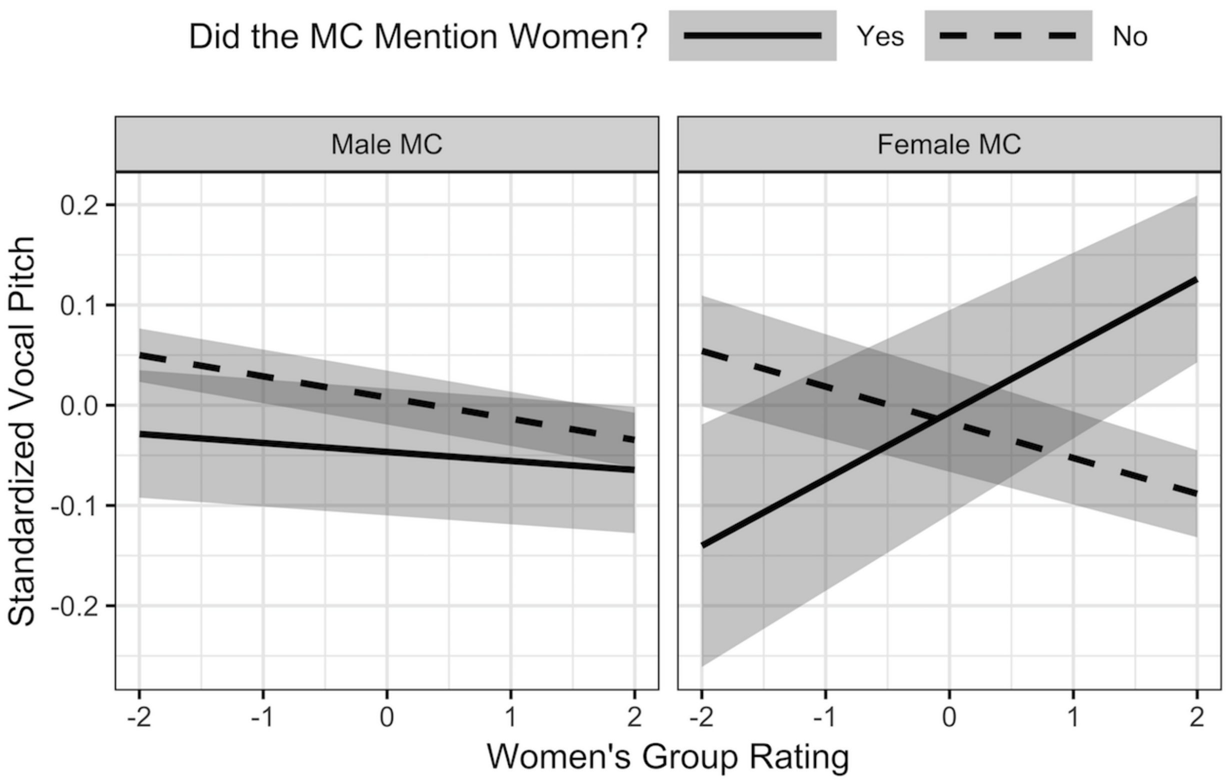
In Models 3 and 4 we interact our measure of emotional intensity with the average women's interest group rating. This tests whether female MCs who tend to vote in the preferred direction of women's interests groups also tend to speak with more intensity when referencing women. The positive and significant coefficient associated with the interaction between `Female`, `"Women" Mentioned`, and `Women's Group Rating` suggests that the interaction with `Women's Group Rating` likely functions similarly to the interaction with `Women's Bills Introduced`. We again plot predicted values to make the interaction more interpretable. These results are reported in Figure S15.

Beginning with the right panel, when both `Female` and `"Women" Mentioned` are set to 1, female MCs are predicted to speak at a higher vocal pitch as `Women's Group Rating` increases. Said differently, female MCs who tend to vote in the preferred direction of women's interest groups also speak with *more* emotional intensity when talking about women. The dashed line also shows the inverse is true when they do not mention women. That is, Congresswomen who vote with women's interest groups tend to speak with *less* emotional intensity when `"Women" Mentioned` is set to 0. Not only is this result consistent with what we found with `Women's Bills Introduced`, but it again demonstrates that vocal pitch may yield additional insights when used in conjunction with more traditional variables. Regardless, these results provide additional evidence consistent with our broader theoretical argument.

S8.8 Potential Confounders: Anxiety

As we explain in Section S2.1, emotions can be characterized as a mixture of two dimensions: a valence dimension and an arousal/activation/intensity dimension. As we argue in our paper, heightened vocal pitch is a useful indicator of this second dimension—arousal/activation/intensity—because when we are in this state our heart naturally begins to race and our muscles, including our vocal cords, tighten. The latter causes our vocal pitch to increase, which is why scholars use pitch to measure the intensity of emotional expressions. However, one may be concerned that we are detecting only emotional anxiety about

Figure S15: Intensity of Speeches about Women by Women's Interest Group Rating



Note: Predicted vocal pitch derived from Model 4 in Table S43 holding all other variables constant. Solid lines indicate the speech included at least one of the Pearson and Dancey (2011*b*) women's dictionary terms. Dashed lines indicate all other speeches. For a given MC, `Women's Group Rating` captures whether s/he was more likely than the average legislator to vote in the preferred direction of the 24 women's interest groups outlined in Table S4. More details can be found on page S81. On the x-axis, `Women's Group Rating` is allowed to vary from ± 2 standard deviations. The y-axis reports the predicted standardized vocal pitch with positive values implying greater emotional intensity. The gray ribbons represent 95 percent confidence intervals.

speaking on the House floor, rather than emotions related to the topic of the speech.[12]

We do not believe, however, that more emotionally intense speeches simply reflect an anxiety about speaking on the subject of the speech. In Table 4 we show that legislators speak with higher vocal pitch on issues owned by their party, and decreased pitch on issues owned by the opposing party. It is highly unlikely that this reflects lawmakers' greater anxiety when speaking about owned issues. Indeed, this would run counter to scholarship on issue ownership by Petrocik and others (Petrocik 1996; Petrocik, Benoit and Hansen 2003), which assumes that partisans advance party issues because they are thought to be better able to handle them.

We also address the concern that we might be detecting overall emotional anxiety about speaking on the House floor, rather than emotions related to the topic of the speech. It is important to note that if women MCs are simply anxious about speaking in an overwhelmingly masculine institution, then they should exhibit higher anxiety than men regardless of speech topic, and may in turn speak with a higher baseline vocal pitch. But since our measure takes the legislator's baseline pitch into account, what we are capturing is deviations from this (potentially) already-heightened baseline. Additionally, a legislator's first speech in a given Congress does not exhibit meaningfully higher vocal pitch than other speeches. In Table S44, we find that the interaction between `Female` and `''Women'' Mentioned` is still positive and statistically significant when a dummy variable is included for the first speech. This suggests that our measure of emotional intensity is not simply picking up a general anxiety about speaking on the floor of Congress.

References

Bänziger, Tanja and Klaus R Scherer. 2007. Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In *International conference on affective computing and intelligent interaction.* Springer pp. 476–487.

Boersma, Paul. 1993. "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound." *Proceedings of the Institute of Phonetic Sciences* 17:97–110.

Dietrich, Bryce J, Ryan D. Enos and Maya Sen. 2019. "Emotional arousal predicts voting on the US supreme court." *Political Analysis* 27(2):237–243.

Gerrity, Jessica C., Tracy Osborn and Jeanette Morehouse Mendez. 2007. "Women and Representation: A Different View of the District?" *Politics & Gender* 3(2):179–200.

Giannakopoulos, Theodoros and Aggelos Pikrakis. 2014. *Introduction to audio analysis: a MATLAB® approach.* Academic Press.

[12]We thank a helpful reviewer for pushing us on this point.

Table S44: Female MCs More Likely to Talk About Women, with Greater Intensity (Controlling for the First Speech)

	Dependent variable:			
	Standardized Vocal Pitch		Standardized Vocal Pitch	
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	−0.002	0.151***	0.152***	0.152***
	(0.004)	(0.024)	(0.024)	(0.024)
Female	−0.017	−0.032***	−0.033***	−0.033***
	(0.011)	(0.011)	(0.011)	(0.011)
First Speech	0.033	−0.015	−0.036	−0.036
	(0.040)	(0.040)	(0.047)	(0.047)
"Women" Mentioned	0.020	−0.054***	−0.054***	−0.054***
	(0.014)	(0.014)	(0.015)	(0.015)
Female × "Women" Mentioned	0.090***	0.112***	0.110***	0.110***
	(0.027)	(0.027)	(0.027)	(0.027)
Female × First Speech			0.096	0.096
			(0.115)	(0.115)
"Women" Mentioned × First Speech			−0.035	−0.035
			(0.140)	(0.140)
Female × "Women" Mentioned × First Speech			0.277	0.277
			(0.286)	(0.286)
Additional Controls		✓	✓	✓
N_1	71,198	71,198	71,198	71,198
N_2	613	613	613	613
Log Likelihood	−100,722.000	−99,647.330	−99,648.740	−99,648.740
AIC	201,458.000	199,324.700	199,333.500	199,333.500

Note: Re-estimating Models 3 and 4 from Table 1 in the main text including a control for whether the speech was the first the MC delivered. Additional controls excluded to save space. Full models available upon request. `First Speech` is dummy variable which equals 1 when the speech is the MC's first in a given Congress. More details can be found on pages S86–S86. Levels of significance are reported as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are reported in parentheses.

Goudbeek, Martijn and Klaus Scherer. 2010. "Beyond arousal: Valence and potency/control cues in the vocal expression of emotion." *The Journal of the Acoustical Society of America* 128(3):1322–1336.

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.

Heath, Roseanna, Leslie A Schwindt-Bayer and Michelle M Taylor-Robinson. 2005. "Women on the Sidelines: Women's Representation on Committees in Latin American Legislatures." *American Journal of Political Science* 49(2):420–436.

Heilman, Kenneth M., Robert T. Watson and Edward Valenstein. 2011. Neglect and Related Disorders. In *Clinical Neuropsychology*, ed. Kenneth M. Heilman and Edward Valenstein. New York, NY: Oxford University Press pp. 296–348.

Herrnson, Paul S, J Celeste Lay and Atiya Kai Stokes. 2003. "Women Running 'as Women': Candidate Gender, Campaign Issues, and Voter-Targeting Strategies." *Journal of Politics* 65(1):244–255.

Hess, Wolfgang J. 2007. Pitch and Voicing Determination of Speech with an Extension Toward Music Signal. In *Springer Handbook of Speech Processing*, ed. Jacob Benesty, M. Mohan Sondhi and Yiteng Huang. 1st ed. Springer chapter 10, pp. 181–211.

Jones, Barbara E. 2003. "Arousal Systems." *Frontiers in Bioscience* 8:438–451.

Kanthak, Kristin and George A Krause. 2012. *The Diversity Paradox: Political Parties, Legislatures, and the Organizational Foundations of Representation in America*. New York: Oxford University Press.

Kathlene, Lyn. 1994. "Power and Influence in State Legislative Policymaking: The Interaction of Gender and Position in Committee Hearing Debates." *American Political Science Review* 88(03):560–576.

Klofstad, Casey A. 2016. "Candidate Voice Pitch Influences Election Outcomes." *Political Psychology* 37(5):725–738.

Koch, Michael and Ulrich Ebert. 1993. "Enhancement of the Acoustic Startle Response by Stimulation of an Excitatory Pathway from the Central Amygdala/Basal Nucleus of Meynert to the Pontine Reticular Formation." *Experimental Brain Research* 93(2):231–241.

Krook, Mona Lena. 2015. "Empowerment versus backlash: gender quotas and critical mass theory." *Politics, Groups, and Identities* 3(1):184–188.

Lang, Peter J., Mark K. Greenwald, Margaret M. Bradley and Alfons O. Hamm. 1993. "Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions." *Psychophysiology* 30(3):261–273.

Laukka, Petri. 2004. Vocal expression of emotion: discrete-emotions and dimensional accounts PhD thesis Acta Universitatis Upsaliensis.

Mauss, Iris B. and Michael D. Robinson. 2009. "Measures of Emotion: A Review." *Cognition and Emotion* 23(2):209–237.

Osborn, Tracy and Jeanette Morehouse Mendez. 2010. "Speaking as Women: Women and Floor Speeches in the Senate." *Journal of Women, Politics & Policy* 31(1):1–21.

Pearson, Kathryn and Logan Dancey. 2011*a*. "Elevating Women's Voices in Congress Speech Participation in the House of Representatives." *Political Research Quarterly* 64(4):910–923.

Pearson, Kathryn and Logan Dancey. 2011*b*. "Speaking for the Underrepresented in the House of Representatives: Voicing Women's Interests in a Partisan Era." *Politics & Gender* 7(4):493–519.

Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40(3):825–850.

Petrocik, John R., William L. Benoit and Glenn J. Hansen. 2003. "Issue Ownership and Presidential Campaigning, 1952–2000." *Political Science Quarterly* 118(4):599–626.

Poole, Keith T. and Howard Rosenthal. 2001. "D-Nominate after 10 Years: A Comparative Update to Congress." *Legislative Studies Quarterly* 26(1):5–29.

Posner, Jonathan, James A. Russell and Bradley S. Peterson. 2005. "The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive development, and Psychopathology." *Development and Psychopathology* 17:715–734.

Re, Daniel E, Jillian JM O'Connor, Patrick J Bennett and David R Feinberg. 2012. "Preferences for very low and very high voice pitch in humans." *PLoS ONE* 7(3):e32719.

Roberts, Margaret E., Brandon M. Stewart and Dustin Tingley. 2014. "stm: R package for structural topic models." *R Package* 1:1–12.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley and Edoardo M. Airoldi. 2013. The Structural Topic Model and Applied Social Science. In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation.*

Rosen, Jeffery B., Janice M. Hitchcock, Catherine B. Sananes, Mindy J. D. Miserendino and Michael Davis. 1991. "A Direct Projection from the Central Nucleus of the Amygdala to the Acoustic Startle Pathway: Anterograde and Retrograde Tracing Studies." *Behavioral Neuroscience* 105(6):817–825.

Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 39(6):1161.

Russell, James A. 2003. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110:145–172.

Scherer, Klaus R. 2013. "Vocal markers of emotion: Comparing induction and acting elicitation." *Computer Speech & Language* 27(1):40–58.

Scherer, Klaus R. 2014. "Corpus design for studying the expression of emotion in speech." *Spoken Corpora and Linguistic Studies* 61:210.

Volden, Craig, Alan E Wiseman and Dana E Wittmer. 2018. "Women's Issues and Their Fates in the US Congress." *Political Science Research and Methods* 6(4):679–696.