1

Yellin' at Yellen

Hostile Sexism in the Federal Reserve Congressional Hearings

James Bisbee,[1‡] Nicolò Fraccaroli,[2] Andreas Kern,[3]

[1]Political Science Department, Vanderbilt University

[2] William R. Rhodes Center for International Economics and Finance, Brown University

[3] McCourt School of Public Policy, Georgetown Univeristy

[‡]To whom correspondence should be addressed: james.bisbee@nyu.edu

Abstract

How prevalent is hostile sexism among U.S. politicians? We analyze the transcripts of every congressional hearing attended by the chair of the U.S. Federal Reserve from 2001 to 2020 using Janet Yellen as a bundled treatment to provide carefully identified evidence of sexism. We find that legislators who interacted with both Yellen and at least one other male Fed chair over this period interrupt Yellen more and interact with her using more aggressive language. Furthermore, we show that having a daughter reduces a legislator's hostility toward Yellen. These results are not a function of differences in either the content Yellen discusses in these hearings (measured with topic models) or her tone of communication. Our results provide carefully identified evidence of hostile sexism among both male and female legislators from both sides of the aisle when confronted with a female in a novel position of power, supporting a rich literature on gender roles and forms of sexism.

11,146 Words
20,627 Appendix

*We would like to thank Isha Aora for outstanding research assistance and support. All errors are ours.

Introduction

In the United States in 2021, women comprised 40% of city councils, 33% of Supreme Court justices, 32% of mayors, 31% of state legislators, 30% of statewide elective offices, 27% of Congress, and 22% of governors.[1] A rich body of work examines underrepresentation at different stages of career advancement, starting with the decision to participate in public service (Fox and Lawless, 2010, 2014; Kanthak and Woon, 2015), charting additional challenges confronted by women at different stages in election campaigns (Sanbonmatsu, 2006; Barber et al., 2016; Holman, Merolla and Zechmeister, 2016; Van der Pas and Aaldering, 2020), and ending with the experiences of office among those who do attain it (Ban et al., 2022; Payson, Fouirnaies and Hall, 2021).

While different mechanisms bind at each stage of this "leaky pipeline", in the background is a pervasive gender bias against women in positions of power, in which sexism serves to maintain traditional gender roles in society and takes one of two forms depending on the context (Glick and Fiske, 1996).[2] The first – hostile sexism – captures hostile behavior toward women who violate traditionally male positions of status and power. The second – benevolent sexism – refers to behaviors that are substantively non-hostile but nevertheless reinforce traditional gender roles. Both forms of sexism combine to maintain the status quo in which men enjoy a privileged status in society and contribute to the "leaky pipeline" that produces imbalances in gender representation in American politics.

Yet, directly measuring the hostile ideology of sexism in politics is challenging. Existing evidence on sexism focuses on the decision on whether to run for office (Fox and Lawless, 2010, 2014; Kanthak and Woon, 2015), differences in campaign contributions (Barber et al., 2016; Sorensen and Chen, 2021), media coverage of campaigns (Heldman, Carroll and Olson, 2005; Conroy et al., 2015; Van der Pas and Aaldering, 2020), constituent prejudices (Huddy and Terkildsen, 1993; Lawless, 2009; Bock, Byrd-Craven and

[1] City council and mayoral statistics in the 100 largest cities in the country. All data from the Center for American Women and Politics: https://cawp.rutgers.edu/women-elective-office-2021.

[2] See Leaper and Ayres (2007*a*) for a review of the literature.

Burkley, 2017; Teele, Kalla and Rosenbluth, 2018), and the demands placed on women who attain office despite these barriers (Payson, Fouirnaies and Hall, 2021). Causally identified evidence of the latent trait of sexism is scant (Vera and Vidal, 2020).

We exploit the tenure of Janet Yellen as the first female chair of the Federal Reserve to provide causal evidence of a behavioral proxy for hostile sexism in the U.S. Congress: interruptions. Using a difference-in-differences specification, we show that legislators who interacted with both Yellen and at least one other male Fed chair (Alan Greenspan, Ben Bernanke, or Jerome Powell) were significantly more likely to interrupt Yellen. We test and disconfirm alternative mechanisms that arise due to the bundled nature of our treatment, including which topics Yellen discussed that might differ from her male counterparts and the tone with which she spoke at these hearings. Additionally, we show that the "Yellen effect" differs systematically by whether the legislator has any daughters, an as-good-as-random life experience that has been demonstrated to change the parent's gender bias in other settings (Washington, 2008; Shafer and Malhotra, 2011; Glynn and Sen, 2014; Borrell-Porta, Costa-Font and Philipp, 2019*a*). Finally, we demonstrate that the tone of interactions with Yellen are more hostile than those with her male counterparts.

Our findings contribute to several strands of the existing literature. First, we speak to a rich body of research on sexism in the U.S. Congress.[3] Existing work in this field has concentrated on voting behavior (Washington, 2008; Volden, Wiseman and Wittmer, 2016; Swers, 2020; Rolfes-Haase and Swers, 2021), verbal interactions in committee hearings (Kathlene, 1994; Mattei, 1998; Karpowitz and Mendelberg, 2014; Ban et al., 2022), wage gaps among staffers (Calcagno and Montgomery, 2020), and ongoing debates on women in Congressional politics (Krook, 2020). Our findings suggest that legislators reacted to the first female chair of the Fed with hostile sexism, providing further evidence of an existing gender bias in Congressional hearings (Kathlene, 1994; Mattei, 1998; Karpowitz and Mendelberg, 2014; Och, 2020; Ban et al., 2022).

[3]For example, see Kathlene (1994); Mattei (1998); Karpowitz and Mendelberg (2014); Och (2020); Vera and Vidal (2020).

Second, we complement a broader literature on women in central banks. Expanding on existing research on women's career advancement within central banks (Hospido, Laeven and Lamo, 2022; Carney, 2019; Comunale et al., 2023), female board representation (Diouf and Pépin, 2017; Charléty, Romelli and Santacreu-Vasut, 2017; Masciandaro, Profeta and Romelli, 2018), and its effect on monetary policy decision-making (Bodea and Kerner, 2021; Bodea et al., 2021),[4] our contribution shows that women in leadership roles are subject to greater scrutiny independent of their professional performance, uncovering a neglected deficiency in legislative oversight mechanisms of independent central banks.[5]

Finally, our paper makes a carefully identified contribution to the broader literature on gender roles, speaking to a number of theories found across the disparate fields of sociology, psychology, communications, linguistics, and anthropology (Glick and Fiske, 1996; Eagly and Karau, 2002; Leaper and Ayres, 2007*a*; Brescoll, 2016; Och, 2020; Vera and Vidal, 2020; Rudman and Glick, 2021; Goldin, 2023). We construe our results as evidence of hostile sexism in a natural experimental setting and disconfirm alternative interpretations of our headline finding that legislators interrupted Janet Yellen disproportionately more than either preceding or subsequent chairs of the Federal Reserve, all of whom were male. We show that this difference in interruptions occurs within individual legislators who interacted with both Yellen and at least one other (male) Fed chair; that Yellen's choice of topics and tone of speaking did not differ meaningfully from her male counterparts; and that male legislators with daughters are less likely to interrupt Yellen. Importantly, while the effect is strongest among Republicans and men, it also obtains for Democrats and women, underscoring the non-partisan nature of sexism in United States politics.

[4]Female FOMC members tend to have more dovish stances on monetary policy (Chappell Jr, Havrilesky and McGregor, 1993; Bodea and Kerner, 2021). For instance, Chappell Jr, Havrilesky and McGregor (1993) show that out of the seven women that served in the FOMC between 1966 and 1996, six of them are ranked among the most dovish members.

[5]This finding is also relevant to research on how elected representatives interact with central banks for accountability purposes. Previous work found that the governance, macroeconomic changes, and ideology influence the way legislators discuss with the central bank when holding it accountable (Fraccaroli et al., 2023, 2022*b*; Ferrara et al., 2021; Schonhardt-Bailey, Dann and Chapman, 2022; Maricut-Akbik, 2020). In this work, we identify that the gender of the central banker under scrutiny plays a key role.

1 Empirical Context

Based on "*stereotypically male qualities,*"[6] central banking is a policy domain where the implicit threat to traditionally male sources of power is salient (Hospido and Sanz, 2020; Comunale et al., 2023). In 2020, only 14 central banks in the world were headed by a woman, and one-fifth of central banks had no women in senior positions (Masciandaro, Profeta and Romelli, 2018). Only recently were women appointed to head major central banks such as the Federal Reserve (with Janet Yellen in 2014) and the European Central Bank (with Christine Lagarde in 2019).

In theory, these contexts embody the types of situations where hostile sexism should be most pronounced (Glick and Fiske, 1996; Goldin and Rouse, 2000). Specifically, women are still the overwhelming minority, exaggerating the salience of both descriptive and injunctive norms, which combine to provoke sexist reactions as women attempt to threat the needle of the contradicting demands of these norms (Eagly and Karau, 2002).[7] Furthermore, as discussed in Glick and Fiske (1996), path-breaking women who represent a threat to traditionally male positions of status and privilege further exacerbate hostile sexist responses. We concentrate on the tenure of Janet Yellen, the first female chair in the history of the Federal Reserve (Fed).[8]

Even prior to her appointment in 2014, Ezra Klein of The Washington Post considered what he called the "whispering campaign", and in some cases the "shouting campaign" against Yellen becoming the next Fed chair as a result of sexism.[9] In an Op-Ed for the New York Times, Paul Krugman echoed these claims, arguing that the two "sexism campaigns underway" were in the form of an implicit whisper campaign and the other involving "raw misogyny."[10] Although each Fed chair encountered some forms of political

[6] "Favorite for the Fed." Pittsburgh Post Gazette, September 19, 2013.

[7] We provide a detailed discussion of role congruity theory in the SI Section 2.

[8] We do not find similar evidence for Christine Lagarde in the context of ECB oversight hearings, although we posit that this more structured setting is less appropriate as a laboratory to empirically investigate our research question, consistent with work by (Fraccaroli et al., 2023).

[9] "The subtle, sexist whispering campaign against Janet Yellen." The Washington Post, June 19, 2013.

[10] "Sex, Money and Gravitas", New York Times, August 1, 2013

pressure or resistance (Binder and Spindel, 2019), Yellen's tenure stands out and has been described as a time of "pretty much constant harassment from politicians on all sides."[11]

To analyze to what extent this 'harassment' is an expression of hostile sexism toward the first female chair of the Fed, we examine the conversations between Yellen and members of the House and Senate who attended the bi-annual oversight committee hearings.

1.1 Construct Validity

We focus on interruptions as our primary outcome of interest. We defend the construct validity of our measure in two steps. First, we argue that interruptions are valid measures of hostile behavior in our setting. Second, we argue that this hostile behavior is a reflection of the latent quantity of interest: hostile sexism.

To support our first claim, we refer to a rich literature on communication, which recognizes that not all interruptions are hostile and divides interruptions into a "positive" and "negative" binary. Positive interruptions are a form of expressed support, which has been shown to strengthen norms of "niceness" in groups (Leaper and Ayres, 2007*b*; Karpowitz and Mendelberg, 2014). Conversely, negative interruptions are a power play in which the interrupter expresses opposition or deprecation toward the interrupted (Mendelberg, Karpowitz and Oliphant, 2014). These types of interruptions are used to silence speakers and suppress communication, both specifically – i.e., by talking over and disrupting a particular statement – and generally – i.e., by fostering a hostile communication setting that dissuades others from speaking up. These patterns can be particularly detrimental to women's participation: Beck (2001) shows that female council members are more likely than their male counterparts to "hold back" their views in response to hostility; Anderson and Leaper (1998) demonstrate that women are three times more likely to retreat from a conversation when interrupted; Mendelberg, Karpowitz and Oliphant (2014) show that these patterns are heightened in traditionally masculine domains; and Och (2020) argues

[11] "Toughest job in Washington?" Politico, December 1, 2014.

that the traditionally masculine setting of politics is where the assertion that interruptions are an expression of "symbolic violence against women" (pg. 390) is highly plausible.

In addition to this literature, we exploit advances in natural language processing in two ways. First, we score conversations on the basis of their aggressive tone. We demonstrate that interruptions in our setting are substantially more aggressive than other interactions (see SI section 10). Second, in our SI section 8.8, we rely on machine-assisted labeling via large language models (LLMs) to show that *any* interaction between Yellen and hearing attendees is less polite on average that a randomly selected interaction between one of the male Fed chairs at attendees.

1.2 Yellen as a bundled treatment

Our empirical setting is an attractive laboratory to investigate hostile sexism among elected officials for several reasons. First, by only examining hearings for the Federal Reserve, we hold relatively constant the domain of topics that might be discussed, making it unlikely that our results are driven by topics which might be both differentially inflammatory and differentially discussed by women.[12] Second, we focus on interactions with a single individual who occupies the same position throughout: the chair of the Federal Reserve, within which our treatment of interest varies. Finally, ensures both a strong treatment dosage with which to identify the quantity of interest, but also constrains variation in the outcome measure which – we believe – makes our evidence more convincing. On the former point, Yellen's path-breaking tenure as the first woman in a position of enormous power that had exclusively been male dominated means our treatment dosage is strong. On the latter point, the recurring, codified nature of these hearings makes them a relatively challenging test of hostile sexism insofar as this structure limits the possibility of free-flowing conversations where hostility is more likely to arise (Karpowitz

[12]In addition, we train a topic model on the data and add the utterance-level topic loadings as additional controls to soak up the remaining variation in what the Fed chairs talk about within the broader topic of monetary policy.

and Mendelberg, 2014). By maximizing the strength of the predictor of interest, while limiting variation in the outcome of interest, we argue our setting is ideally suited to providing carefully identified evidence of hostile sexism in the halls of power.

However, we acknowledge a major challenge with our setting as well: Yellen is the only Fed chair appointed by a Democratic president in our data. In this sense, her gender is fundamentally bundled with her partisan association. We point to the fact that both Bernanke and Powell were re-appointed by Democrats; that Yellen's topics of conversation do not differ from topics discussed by other Fed chairs; that she is interrupted less by legislators with daughters; that she is interrupted more by men than women (although she is interrupted more by both relative to her male counterparts); and that – despite being appointed by a Democrat – legislators from both sides of the aisle interrupted her more than her male counterparts.

We argue that these patterns are more plausibly explained by hostile sexism than by partisan antipathy. Taken together, we feel confident in our interpretation of interruptions as a type of social aggression designed to undermine a speaker's power and make the plausible assertion that this empirical proxy is a suitable reflection of hostile sexism.

2 Data and Research Design

Pursuant to the Federal Reserve Reform Act,[13] the chair of the Board of Governors of the Federal Reserve appears twice each year before the Senate Committee on Banking, Housing and Urban Affairs, and twice before the House Committee on Financial Services. Hearings begin with an opening remark by the Fed chair, followed by a Q&A session in which members of Congress debate monetary policy with the Fed chair.[14] After these

[13]The Federal Reserve Reform Act is accessible at the following link (last accessed in November 2021): https://www.govinfo.gov/content/pkg/STATUTE-91/pdf/STATUTE-91-Pg1387.pdf.

[14]Each committee is composed of a chairman, who is generally the majority party member with the greatest seniority; a Vice-chairman; and a Ranking Member. We identify variation within interlocutors, holding constant these potential confounders by position. For more details, see: https://www.senate.gov/artandhistory/history/common/briefing/Committees.htm.

lengthier opening statements have finished, the proper conversation begins, with the chair of the committee beginning the sequence of questions for the Fed chair, followed by other members of the committee. The Fed chair replies to these questions, which may be followed up on by the original interlocutor or by a different committee member. Each legislator is allocated a budget of time to engage the Fed chair, typically five minutes.[15]

These hearings are a matter of public record, and their transcripts are stored on *govinfo*, a website maintained by the US Government Publishing Office.[16] The transcripts capture the sequential conversational flow, indicating changes in who is speaking with a new paragraph that begins with the title and last name of the current speaker. We treat each segment of text as the core unit of analysis, which we refer to as an utterance.[17]

We process the textual data into a single data frame with rows indexing the speaker-utterance-hearing. We identify interruptions based on the notation used in the original transcripts: a series of two or more hyphens (--) to end an utterance. We rely on the sequential structure of the data to not only identify who is being interrupted, but also who is doing the interrupting. Our dataset covers the period from February 13th, 2001 to June 17th, 2020, consisting of 23,119 total utterances, 79 total hearings (40 in the House of Representatives and 39 in the Senate) attended by 242 unique legislators, 8 speaking experts, and 4 chairs of the Federal Reserve (Greenspan, Bernanke, Yellen, and Powell).[18]

Figure 1 plots the total utterances and total interrupted utterances by each speaker in our data who spoke more than 100 times.[19] As illustrated, the chairs of the Fed have the most utterances in our data, followed by the acting committee chairs. Janet Yellen

[15]The available data does not formally report on the amount of time allowed, although a careful search for relevant terms in the raw transcripts never indicates any budget other than 5 minutes for the legislators.

[16]The data can be accessed at the following link (last accessed on September 2021): https://www.govinfo.gov/app/collection/chrg.

[17]A snapshot of one of these transcripts as well as the distribution of utterances in the data, is provided in the Supporting Information Section 1.1.

[18]The Fed chairs in the period under analysis experienced different financial eras, each characterized by different challenges and policy stances. For a recent and detailed overview of the chairmanship and their policies, see Bauer and Faseruk (2020); Bernanke (2022). We describe the empirical setting and these specific pressures in greater detail in the Supporting Information, Section **??**.

[19]We reproduce this figure with a lower threshold for inclusion in the Supporting Information.

is among the speakers who is interrupted the most, measured as the proportion of her utterances that are interrupted and is interrupted, almost twice as frequently as the next.

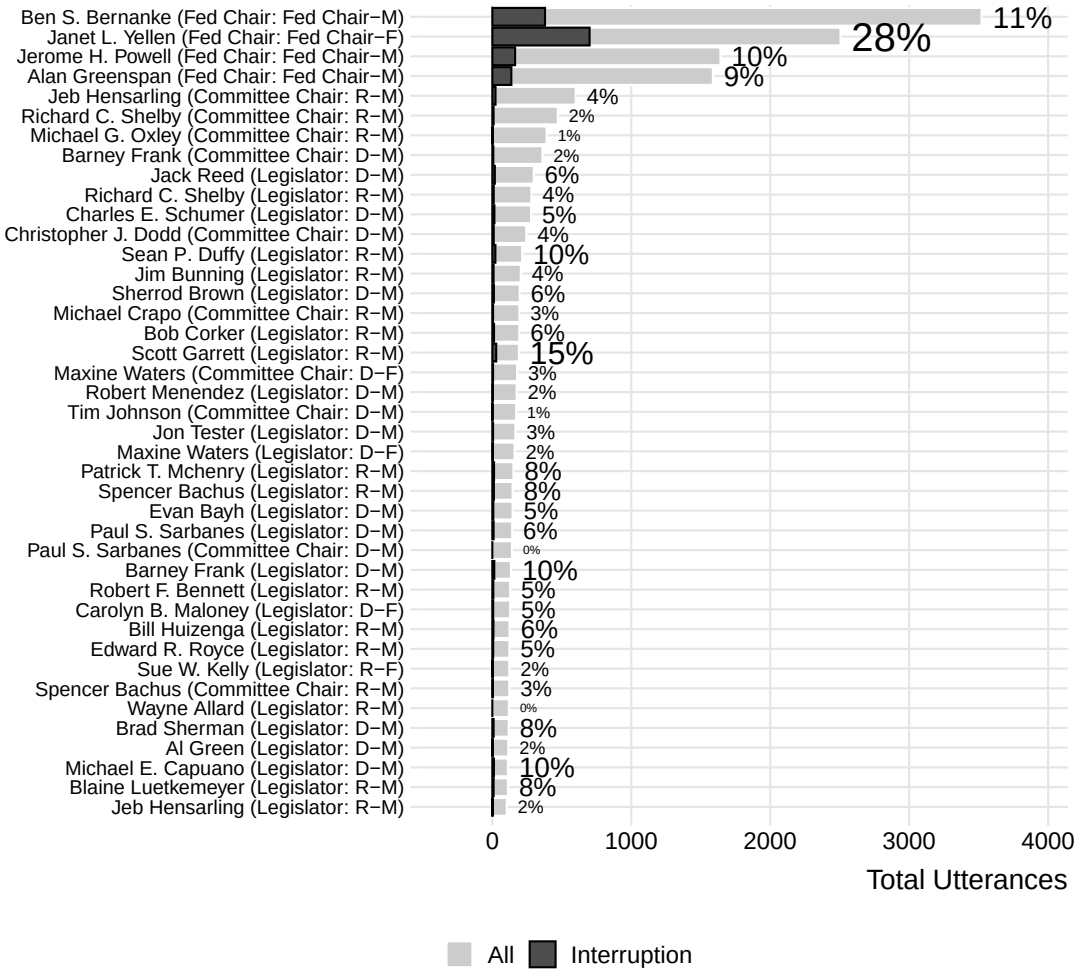

Figure 1: Total utterances by most talkative speakers (light grey bars), along with total interruptions (dark grey bars in the left panel) and total utterances that interrupt another speaker (dark grey bars in the right panel).

Figure 2 focuses on overtime changes in the House and the Senate, revealing a discontinuous increase in the number of times the Fed chair was interrupted that corresponds to Yellen's tenure. Notably, the increase in interruptions is most pronounced in the House of Representatives, where it is a stark discontinuity. The descriptive evidence in the Senate is more muddied. While Yellen's tenure captures the period of greatest interruptions of a Fed chair in the data, it caps a secular increase over the preceding 13 years.

We posit that the greater evidence of a clear difference in the way legislators in the House interact with Yellen and her male counterparts reflects different levels of professionalism in the two chambers, although are unable to test this theory with our data.[20]

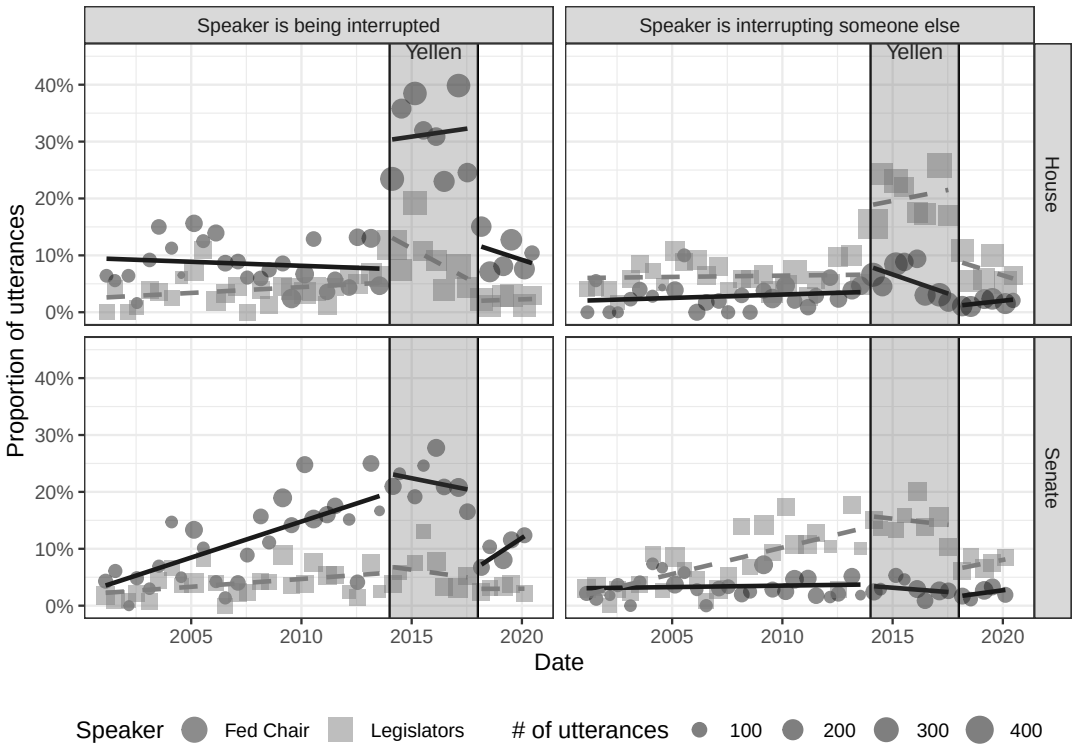

Figure 2: Proportion of utterances that are interrupted (left column) and that interrupt someone else (right column) by House (top row) and Senate (bottom row), broken out by whether the speaker is the Chair of the Fed (black circles) or a legislator (gray squares). The shaded region denotes the mandate of Janet Yellen as chair of the Federal Reserve (February 3, 2014 - February 3, 2018).

Consistent with previous research, we posit that interruptions are used to undermine female authority in a manner consistent with the concepts of role incongruity and hostile sexism, in which the tension between descriptive norms of femininity and the injunctive norms of leadership create an impossibly narrow needle to thread (Glick and Fiske, 1996;

[20]We also highlight the right column of Figure 2, which suggests that Yellen interrupted others more, especially toward the start of her tenure. In our Supporting Information Section 4, we demonstrate that much of this spike, particularly the February 2015 hearing in which Yellen interrupted others more than 20% of the time, was driven primarily by her continuing to speak through others' interruptions of her. When we remove these types of interruptions, the February 2015 outlier disappears.

Furthermore, we demonstrate that Yellen's ratio of interrupting others relative to being interrupted is far lower than that of her male counterparts.

Eagly and Karau, 2002; Och, 2020).[21] We support this interpretation using Natural Language Processing (NLP) methods to calculate the tone of each interaction, enabling us to score conversations on the basis of their aggressive tone. In SI Section 10 we demonstrate that interruptions in our setting are substantially more aggressive than other interactions. In addition, we use AI-assisted annotation to demonstrate that *any* interaction between Yellen and hearing attendees is more aggressive on average than a randomly selected interaction between one of the male Fed chairs and attendees (see SI Section 8.8).

2.1 Specifications

Our analysis relies on a variety of estimation strategies, all of which are built on a simple regression model in which we predict whether a Fed chair is interrupted (or interrupts) as a function of their gender, along with a battery of controls. Formally,

$$\text{Interrupted/ing}_{u,i,t} = \alpha_i + \delta_t + \gamma \mathbf{U}_u + \lambda \mathbf{X}_{i,t} + \varepsilon_{u,i,t} \tag{1}$$

where $Interrupted/ing_{u,i,t}$ is a binary measure of whether utterance u by individual i in time t is interrupted (or interrupts someone else); α_i are fixed effects for each individual who is being interrupted, relative to Bernanke; δ_t are date fixed effects (equivalent to hearing fixed effects, since there is never an instance in which two hearings are held on the same day); and $\mathbf{X}_{i,t}$ are controls measured at the speaker level (total utterances by that speaker, incumbent vote share, gender, seniority, age, party affiliation, and ideology).[22]

Of particular importance is $\mathbf{U}_u$, which is a vector of controls measured at the utterance level, including the logged number of characters, an indicator for whether the utterance itself interrupts/is interrupted by the preceding, and two vectors of NLP measures. The first NLP measure predicts the tone of each utterance based on Google's Perspective

[21] In the Supporting Information Section 4 we provide an example of text in which we detect interruptions, and discuss our substantive interpretation of this excerpt as evidence of hostile sexism.

[22] We rely on Poole and Rosenthal (2017)'s DW-NOMINATE measure for ideology. Summary statistics are reported in the Supporting Information, Section 1.

API,[23] which was designed to measure the tone of online comments or posts along several dimensions. This tool combines a powerful transformer machine learning algorithm with huge amounts of annotated training data to predict whether a small chunk of text is toxic, threatening, aggressive, attacking, and a number of other dimensions of unpleasant speech. We control for these predicted probabilities in our main specifications to hold constant the alternative pathway that Yellen's experiences are driven by *how* she speaks.[24]

Our second NLP-based measure estimates the topics each speaker talks about. Specifically, we estimate a topic model via Latent Dirichlet Allocation (LDA, Blei, Ng and Jordan 2003) with 100 topics on the utterances, treating each utterance as its own document.[25] The resulting model predicts the probability that each utterance is about each of the 100 topics. We include these predicted probabilities in our main specifications to control for the alternative pathway that Yellen's interruptions are driven by *what she talks about*.

Our main specification organizes the data dyadically and predicts whether a speaker is disproportionately more likely to interrupt Yellen than others. Formally, for ego i responding to alter j during a hearing held in time t, we estimate:

$$\text{Interrupting}_{u,i,j,t} = \alpha_i + \rho_j + \delta_t + \gamma \mathbf{U}_u + \varepsilon_{i,j,t} \tag{2}$$

We are interested in exploring whether speakers are more likely to interrupt Yellen than they are to interrupt others, which is achieved with the twin fixed effects (α_i and ρ_j). Here, the identifying variation comes from within a member of Congress who interacts

[23] See the SI section 8.1 for a discussion of this algorithm and its applicability to our empirical setting.

[24] Note that this approach is distinct from research by Dietrich, Hayes and Obrien (2019), who analyzes the audio recordings of speeches on the floor of Congress with specific attention to pitch to measure emotional states. Our method relies on the choice of words and their arrangement to predict the probability that an utterance would be interpreted as toxic, aggressive, flirtatious, and a host of other labels by a human interlocutor. For more information, please refer to the SI section 8.1.

[25] We examine the coherence of different choices of the number of topics k, and settle on a 100-topic model. We also test the robustness of our findings to other methods of topic modeling, including an approach that concatenates subsequent utterances by the same speaker to reduce the skew in document length inherent in our baseline measure. Our results are robust all alternative approaches, strengthening our confidence that these patterns are not reducible to what is being discussed at the hearings. See SI Section 8.6 more details.

with both Yellen and at least one other Fed chair.

Finally, our third specification uses a difference-in-differences design in which we compare the difference in how often a Fed chair is interrupted relative to a non-Fed chair before and after Yellen's tenure. Formally:

$$\begin{aligned} \text{Interrupting}_{u,i,\text{Fed},t} = \beta_1 \mathbb{I}\text{Fed}_t + \beta_2 \mathbb{I}\text{Yellen Tenure}_{i,t} \\ + \beta_3 \mathbb{I}\text{Fed}_t * \mathbb{I}\text{Yellen Tenure}_{i,t} \\ + \alpha_i + \gamma \mathbf{U}_u + \varepsilon_{i,j,t} \end{aligned} \tag{3}$$

Under the parallel trends assumption that the difference between utterances that respond to a Fed chair and those that respond to another legislator hold across periods, we can treat the difference between interruptions of a male Fed chair and another legislator as a valid counterfactual against which we compare the difference between interruptions of Yellen and another legislator. [26]

3 Results

We begin by confirming the descriptive results summarized above in Figure 1 – namely that Yellen is interrupted disproportionately more than others. Figure 3 plots the fixed effects for a subset of speakers who spoke 100 or more times in the data. Positive estimates indicate that these speakers were interrupted more than Bernanke, while negative coefficients indicate they were interrupted less. As illustrated, Yellen was interrupted significantly more frequently than *any* other speaker in our dataset who spoke 100 or more

[26] Note that the parallel trends assumption does not require us to believe that the pattern of interruptions experienced by the Fed Chair must be the same as the pattern of interruptions experienced by the legislator. Given the nature of these hearings, where the Fed chair is effectively the star of the show, such a claim is hardly credible. Instead, this assumption requires that the *difference* in interruptions experienced by the Fed chair and the legislators should be the same, regardless of who the Fed chair is. To the extent that this difference varies during Yellen's tenure, we can causally attribute this difference-in-differences to her presence as long as we believe the gap would have otherwise been the same had Bernanke continued as Fed chair, had Powell taken the post earlier or had another, male, Fed chair been appointed in 2014 instead of Janet Yellen. All standard errors are clustered at the speaker-hearing level.

times, even after controlling for what she talked about and how she talked.[27]

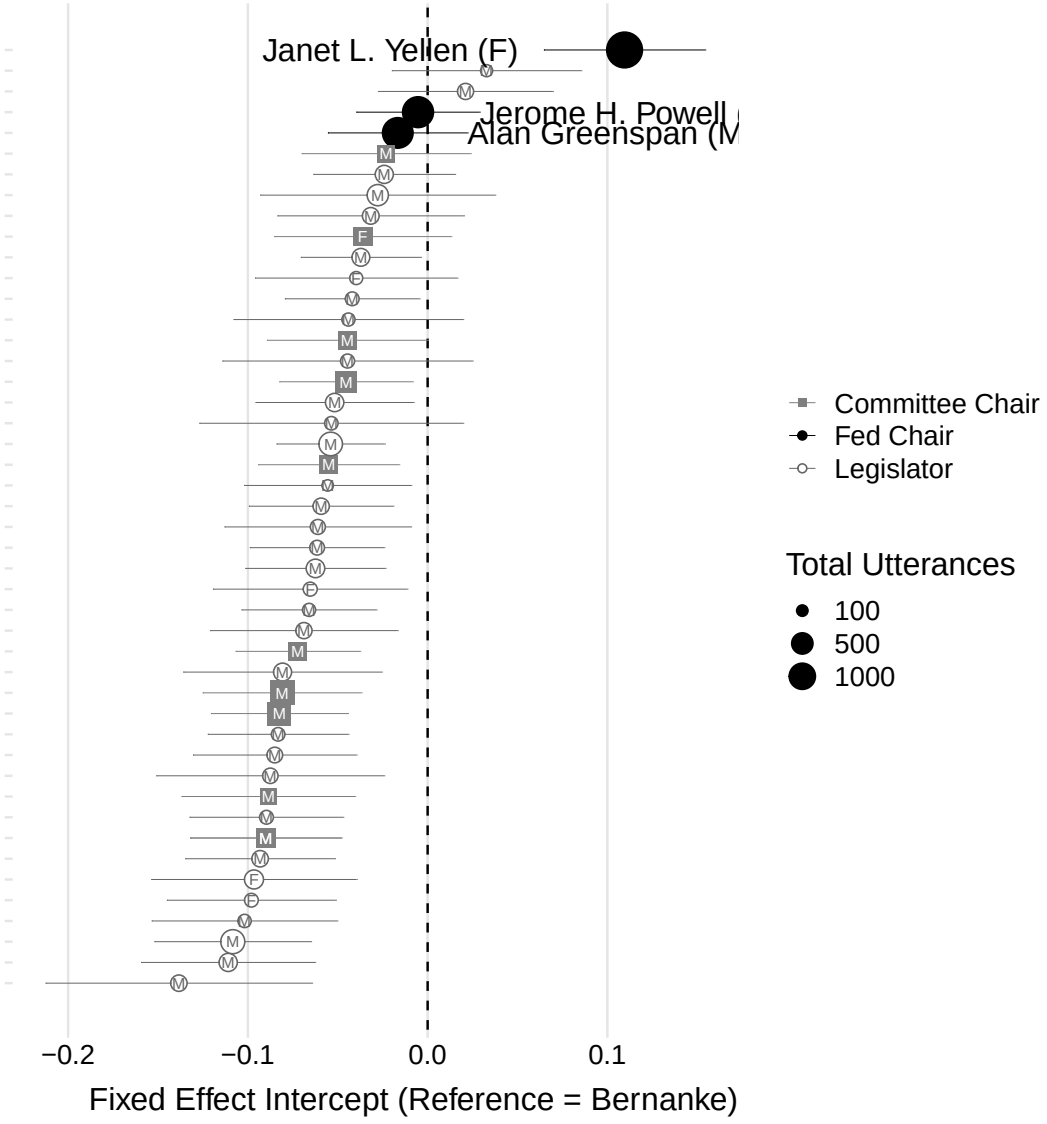

Figure 3: Coefficient plot summarizing the difference between Bernanke's propensity to be interrupted and a subset of other speakers (filtered to those who have 100 or more utterances in total). Points are sized by the total number of utterances for each speaker, colored to indicate whether the speaker is a committee chair (gray squares), Fed chair (black circles), or a legislator (white circles), and labeled according to the speaker's gender ("M" for male, "F" for female).

We can look at heterogeneities in these results in several ways.[28] First, we subset our

[27]We lower the threshold for inclusion to 30 or more utterances in the Supporting Information 5, confirming these results.

[28]We present a battery of heterogeneous effects by interlocutor characteristics, including gender, party, seniority, and ideology in the Supporting Information.

data to the House and the Senate to see if these patterns are stronger among one group of politicians than the other. As illustrated in the left panel of Figure 4, our findings are significantly stronger in the House than in the Senate, potentially reflecting the stronger norms of decorum in the latter chamber due to longer-term limits (Kousser, 2005).[29]

Second, we investigate whether Fed chairs themselves are interrupting others and whether Yellen differs systematically from her peers in this regard. As illustrated in the right panel of Figure 4, Yellen is no more likely to interrupt others than Ben Bernanke. If anything, she was *less* likely to be the interrupter in House hearings compared to Bernanke, whereas none of the other male Fed chairs differed from each other.

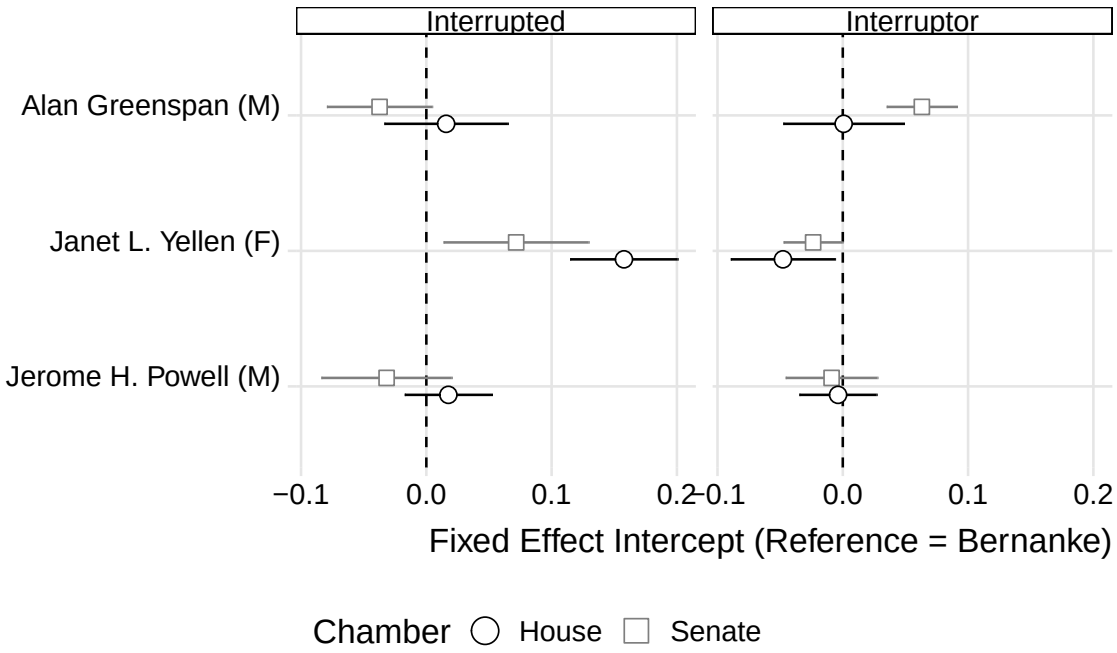

Figure 4: Coefficient plot summarizing the difference between Bernanke's propensity to be interrupted (left panel) and to interrupt someone else (right panel), and the other Fed Chairs, disaggregating to hearings in front of the House Committee on Financial Services (black circles) and hearings in front of the Senate Finance Committee (grey squares).

These results confirm the descriptive evidence that Janet Yellen alone experienced a spike in interruptions, disconfirming the concern that our results simply reflect a secular

[29]We believe this difference in chambers of Congress is striking and worthy of further investigation, beyond the scope of this paper.

increase in hostility over time. Nevertheless, it is possible that Yellen's tenure happened to co-occur precisely with an increase in hostility that is unrelated to her. To test for this possibility, we look for changes in behavior within legislators. Figure 5 indicates that those who interacted with Yellen and other chairs of the Fed were consistently more likely to interrupt Yellen. While this does not preclude the possibility that speakers simply became more hostile in 2014, it does mean that this shift would have had to occur *within* legislators, instead of reflecting an incoming class of more hostile interlocutors.

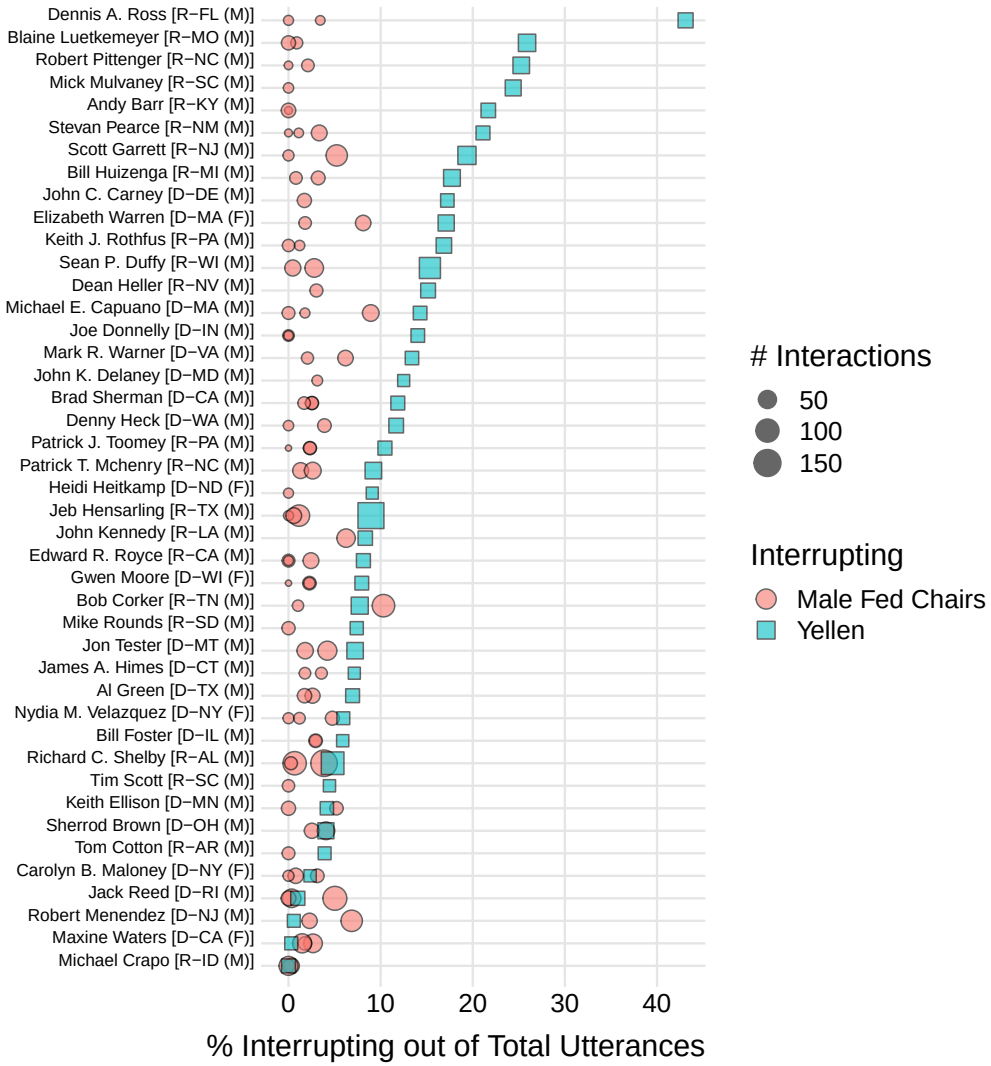

Figure 5: Points represent the share of total possible interactions (size) between a politician (x-axis) and one of the four Fed chairs that were interruptions. Squares indicate Yellen while circles indicate the three other male Fed chairs. Interlocutors are a subset of those who interacted with Yellen and at least one other Fed chair at least 20 times.

Second, we support this descriptive evidence with the dyadic regression described in Equation 2. Table 1 summarizes these results. Yellen is indeed consistently more likely to be interrupted than Bernanke. These results are robust to the inclusion of controls measured at both the speaker and hearing level. The most rigorous specification is presented in column 5, where we add speaker-fixed effects, meaning that the estimate captures the degree to which an individual legislator interrupted Yellen more than they interrupted Bernanke. Substantively, the estimate of 0.16 corresponds to a given legislator being 16 percentage points more likely to interrupt Janet Yellen than they were to interrupt Bernanke, who had a base rate of roughly 1 out of every 10 utterances being interrupted.

To further confirm our causal interpretation of these patterns, we turn to the difference-in-differences specification described in Equation 3. We plot the marginal effect of interrupting a Fed chair across periods, isolating only the within-legislator changes via legislator fixed effects. As displayed in Figure 6, we again find evidence consistent with a Yellen-specific effect of approximately 0.16, or 16 additional interruptions per 100 utterances compared to the male Fed chairs. Importantly, while we do find evidence that these increases are weaker among Democrats and female legislators (in both cases, the interaction terms are statistically significant at the 95% level of confidence), we underscore that the interaction estimates are still positive and – in the case of Democrats – statistically significantly so. Put plainly, neither progressivism nor gender is enough to totally suppress the sexist backlash we theorize is driving the increased interruptions of the first female chair of the Federal Reserve.

Finally, to investigate the degree to which these findings are driven by a secular increase in interruptions, we subset the data to the union of all hearings attended by Yellen and either Bernanke or Powell, the hearings specific to a single presidential administration (Obama's terms for Bernanke to Yellen, Trump's term for Yellen to Powell), and to the final year of one chair and the first year of the following. While narrowing the window of comparison in this fashion reduces the number of observations, it gives us a tighter

Table 1: Dyadic Interruptions

		Controls:		FEs:		Dropping
	Vanilla	Speaker	Utterance	Hearing	Speaker	Voters
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Yellen (ref. Bernanke)	0.169***	0.165***	0.132***	0.099***	0.135***	0.146***
	(0.011)	(0.014)	(0.016)	(0.019)	(0.028)	(0.030)
Powell (ref. Bernanke)	-0.013*	-0.011	-0.023*	-0.007	-0.008	0.005
	(0.006)	(0.008)	(0.011)	(0.015)	(0.029)	(0.030)
Greenspan (ref. Bernanke)	-0.026	-0.023	-0.026	-0.008	-0.006	-0.002
	(0.019)	(0.021)	(0.017)	(0.019)	(0.024)	(0.026)
Age (scaled)		0.010†	0.006	0.009†		
		(0.005)	(0.004)	(0.005)		
Vote Share (scaled)		-0.002	0.001	0.003		
		(0.005)	(0.008)	(0.008)		
Ideology (scaled)		-0.005	-0.007	-0.010		
		(0.010)	(0.011)	(0.012)		
Republican (ref. Democrat)		0.010	0.022	0.024		
		(0.031)	(0.028)	(0.028)		
Senate (ref. House)		-0.003	-0.006			
		(0.038)	(0.030)			
Male (ref. Female)		-0.010	-0.0002	0.017		
		(0.016)	(0.014)	(0.013)		
Seniority (scaled)		-0.012	-0.0008	-0.005		
		(0.009)	(0.006)	(0.007)		
Fed Oversight Sponsor		0.016†	0.012†	0.013†		
		(0.008)	(0.007)	(0.007)		
Oppose Yellen Conf.		-0.017	-0.024	-0.018		
		(0.016)	(0.017)	(0.018)		
Interrupted			0.084**	0.072**	0.066**	0.062**
			(0.026)	(0.025)	(0.023)	(0.020)
100 LDA Topics	No	No	Yes	Yes	Yes	Yes
Tone Probabilities	No	No	Yes	Yes	Yes	Yes
Fixed-effects						
Hearing				Yes	Yes	Yes
Speaker					Yes	Yes
Fit statistics						
Observations	20,151	20,151	20,151	20,151	20,151	18,981
R^2	0.078	0.087	0.218	0.228	0.243	0.238
Within R^2				0.192	0.166	0.160

Clustered (Speaker & respondingTo) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, †: 0.1*

Notes: Dyadic regressions predicting whether Greenspan, Powell, and Yellen were interrupted more than Bernanke. Column 2 includes speaker covariates only, including the total number of utterances of the interrupting politician during the hearing. Column 3 controls for the total number of characters (logged) in the interrupted utterance, utterance-topic probabilities from a 100-topic model, predicted probabilities of tone, and an indicator of whether the interruption itself was interrupted. Column 4 adds hearing fixed effects. Column 5 adds interrupter fixed effects. Column 6 drops all senators who either voted against Yellen's confirmation or abstained. Two-way standard errors are clustered at the speaker-hearing level to account for correlated errors within interlocutors as well as correlated errors within each hearing. † $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

comparison between Yellen and a male Fed chair in both the beginning and the end of her tenure.[30] Figure 7 plots the estimates from a fixed effects regression run on data collapsed

[30]Subsequent analyses that investigate marginal effects and mechanisms do not implement this nar-

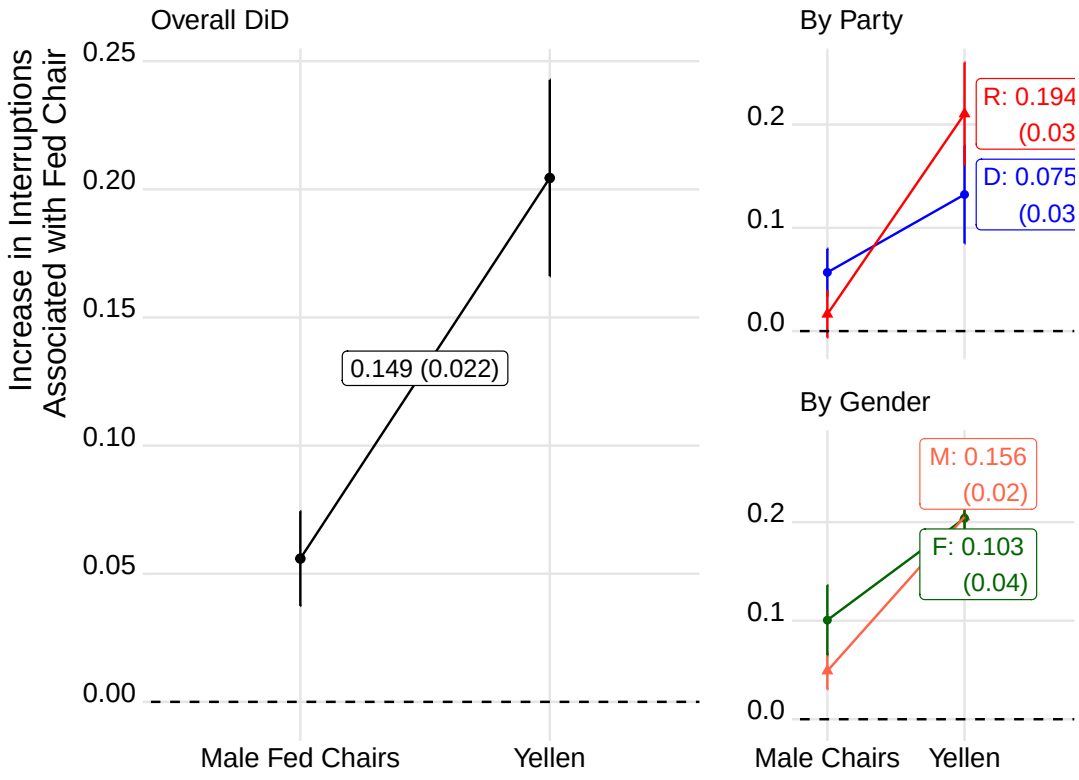

Figure 6: Difference-in-differences marginal effects comparing increase in interruptions of the Fed chair during Yellen's tenure among those who attended hearings with her and at least one other male Fed chair. Right facets illustrate marginal effects broken out by party (top-right facet) and by gender (bottom-right facet).

to the speaker-hearing level. As illustrated, the move from Bernanke to Yellen in 2015 saw a statistically significant increase in interruptions in the House of Representatives but a null result in the Senate (left panel). Conversely, the transition from Yellen to Powell saw a statistically significant decline in interruptions in both chambers.[31] As a placebo test, we re-run the above specifications but instead compare the male Fed chairs to each other, comparing the final term of Greenspan to the first term of Bernanke and Bernanke's

rowed window of analysis in order to retain sufficient statistical power. We defend this choice based on the lack of evidence of meaningful differences in the substantive conclusions across the windows displayed in Figure 7.

[31]The tightest identification subset – the comparison between the final year of an outgoing chair and the first year of an incoming chair – is also the one we have substantive reason to suspect should work against our findings. Specifically, the final hearings for an outgoing chair and the initial hearings for an incoming chair are those most likely to be more polite, as the outgoing chair is thanked for their service, and the incoming chair is welcomed.

final term to Powell's first. While the latter comparison skips the four years of Yellen's tenure, if anything, this should exaggerate the alternative explanation of growing partisan animosity and polarization that just happened to coincide with Yellen's tenure. That we find no evidence of increasing interruptions in any of these placebo tests disconfirms this alternative story, at least insofar as we expect this hostility to be increasing monotonically with time. See SI Section 7 for a detailed presentation of these results.

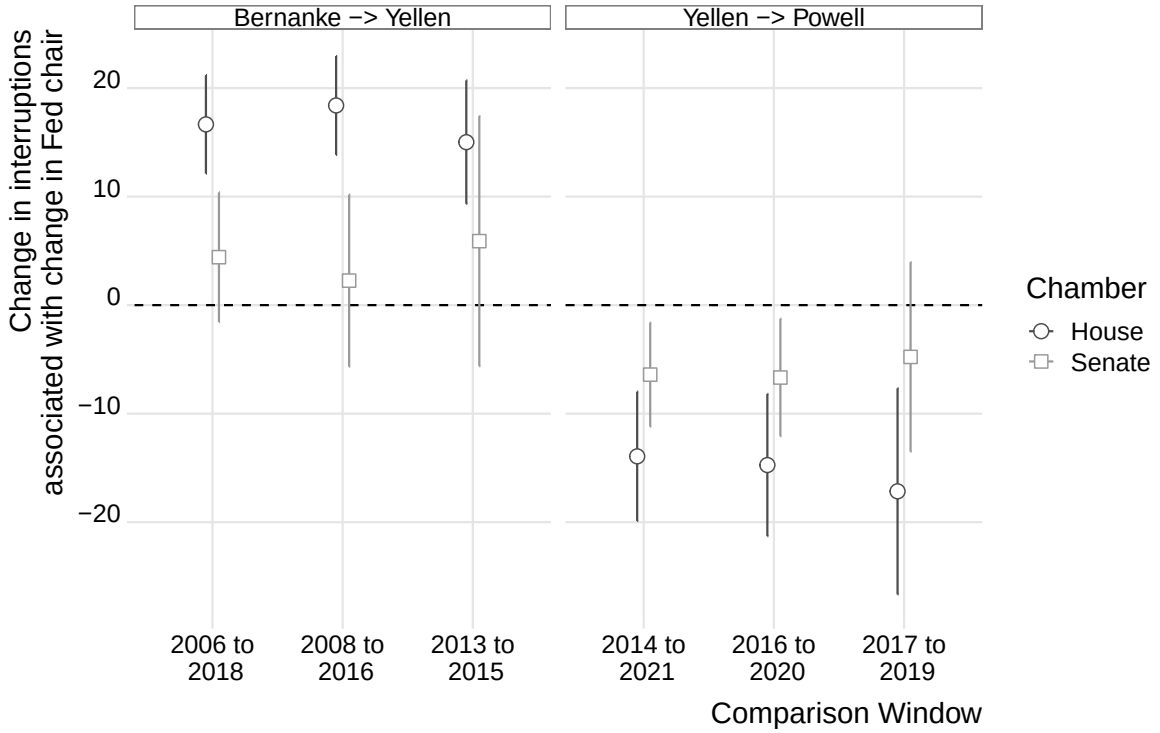

Figure 7: Coefficient estimates capturing the increase in the probability the speaker interrupts the Fed chair when the role shifts from Bernanke to Yellen (left panel) and when it shifts from Yellen to Powell (right panel). Points are shaded by whether the regression is run on hearings in the House of Representatives or in the Senate. The x-axis describes the subset of the data that is being used.

3.1 Mechanisms: Sexism and Daughters

The preceding results confirm that Janet Yellen was interrupted significantly more than the male chairs of the Fed. However, these findings do not confirm that we are capturing a sexist response. In our empirical setting, Janet Yellen is a bundled treatment whose

gender is only one dimension among many characteristics that might explain why she is interrupted more than other Fed chairs during Congressional hearings. While we have attempted to isolate this dimension in the preceding tests, here we turn to a heterogeneous effects analysis which further supports our interpretation of these patterns.

To identify whether sexism is the driver of interruptions, we turn to a modified interacted specification, in which we predict interruptions as a result of the interaction between Yellen's tenure (vis-à-vis the tenure of either Bernanke or Powell) and whether the interacting legislator has any daughters.[32] Specifically, we predict whether an utterance was interrupted as a function of whether the previous utterance was spoken by a Fed chair, interacted with an indicator for Yellen's tenure, and interacted again with an indicator for whether the interrupter had any daughters.

We are particularly interested in the coefficient estimate which reflects whether the increase in interruptions of the Fed chair associated with Yellen's tenure was attenuated by the speaker having daughters. We argue that evidence of this attenuation bolsters our conclusion that the increased interruptions faced by Yellen are due to sexism on the part of her interlocutors in Congress, and not due to other mechanisms. This interpretation rests on a well-developed body of empirical work that argues the experience of having a daughter makes the parent less sexist. Importantly, this theory posits that the effects are more notable among men, among whom the baseline empathy for a women's experience in a patriarchal society is low and for whom the experience of having a female child is therefore more influential (Washington, 2008; Shafer and Malhotra, 2011; Glynn and Sen, 2014; Borrell-Porta, Costa-Font and Philipp, 2019*b*). As such, we add a fourth interaction with the gender of the interlocutor, and visualize our findings with marginal effect plots.

Our results, summarized in Figure 8, indicate that having one or more daughters reduced the increased interruptions of the Fed chair associated with Yellen's tenure. Importantly, however, having daughters did not totally suppress the increase in interruptions

[32]Our results are robust to using the proportion of children that are daughters, as well as whether the first child was female.

that was observed during her time as Fed chair. Furthermore, this suppressing effect is only observed among male legislators. Among female legislators, there is no significant difference in the increased interruptions among those with and without daughters. These patterns are consistent with the existing literature that argues the experience of raising a female child reduces the parent's – especially the male parent's – attachment to traditional gender roles. In addition, these patterns further bolster our interpretation of the evidence of increased interruptions of the Fed chair during Yellen's tenure as reflecting sexism on the part of her interlocutors.

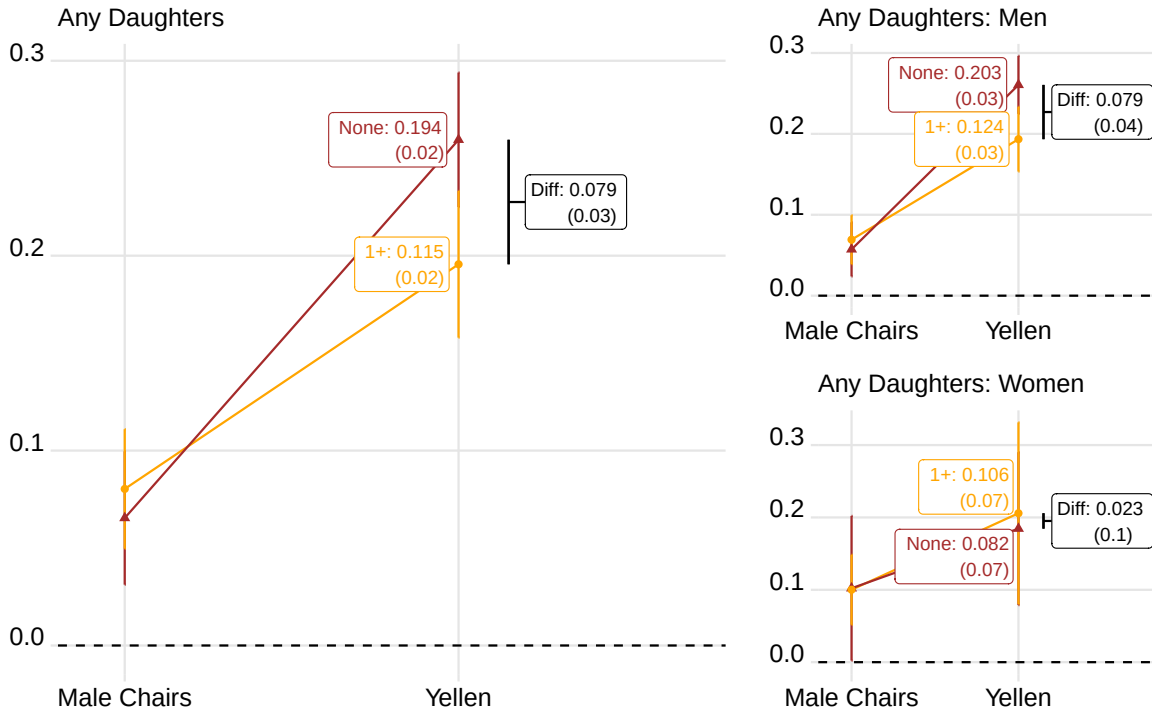

Figure 8: Marginal effect of Yellen taking over as chair of the Fed on whether legislators attending hearings interrupted the chair of the Fed, interacted with whether the legislators had any daughters. Left panel summarizes overall patterns, right column subsets data to men and women.

3.2 Alternative Mechanisms: Topic and Tone

Evidence of the attenuating effect of daughters notwithstanding, it may be that Yellen is simply a more interruptable interlocutor. She might speak more slowly, might be more cagey or unsatisfying in her responses, or might simply do more to avoid answering difficult questions. While it is unlikely that these styles of speech bias the previous results by triggering stronger responses among those without daughters, we nevertheless perform several tests to disconfirm this alternative mechanism.

First, we exploit the topic models that were previously used as a control to examine whether Yellen differs systematically from other Fed chairs in what she speaks about. We investigate this question in two ways. For one, we calculate the probability of a topic conditional on its utterance being interrupted or not, and then investigate whether Yellen chooses topics that are more heavily associated with interruptions than other Fed chairs.[33] Figure 9 shows these quantities of interest. It displays the difference in topic used by Yellen relative to the other male Fed chairs on the x-axis, and the difference in the same topics associated with interrupted and uninterrupted utterances on the y-axis.

As illustrated, there is strong evidence that some topics are more likely to be associated with an interrupted utterance (toward the top of the plot) and others are less likely to be associated with an interrupted utterance (toward the bottom of the plot). The highest-scoring terms associated with these topics provide an intuitive understanding of which types of speech are more or less likely to be interrupted. Specifically, speculative or imprecise language ("think", "look", "thing") is most likely to be interrupted, while polite formalities ("thank", "gentleman", "recognize") are least likely to be interrupted. However, we show that Yellen does not discuss interruptable topics more than other Fed chairs, as indicated by a local linear smoother.[34]

[33]When calculating this per-topic probability, we drop all of Yellen's own utterances. We do so to purge the measure of potential confounding factors driven by the fact that Yellen is more likely to be interrupted due to factors unrelated to the semantic content of the topic itself (i.e., sexism).

[34]In the Supporting Information section 8.4, we recreate these plots using a structural topic model (STM, Roberts et al. (2014)) to calculate the difference in topic use by Yellen relative to each other Fed

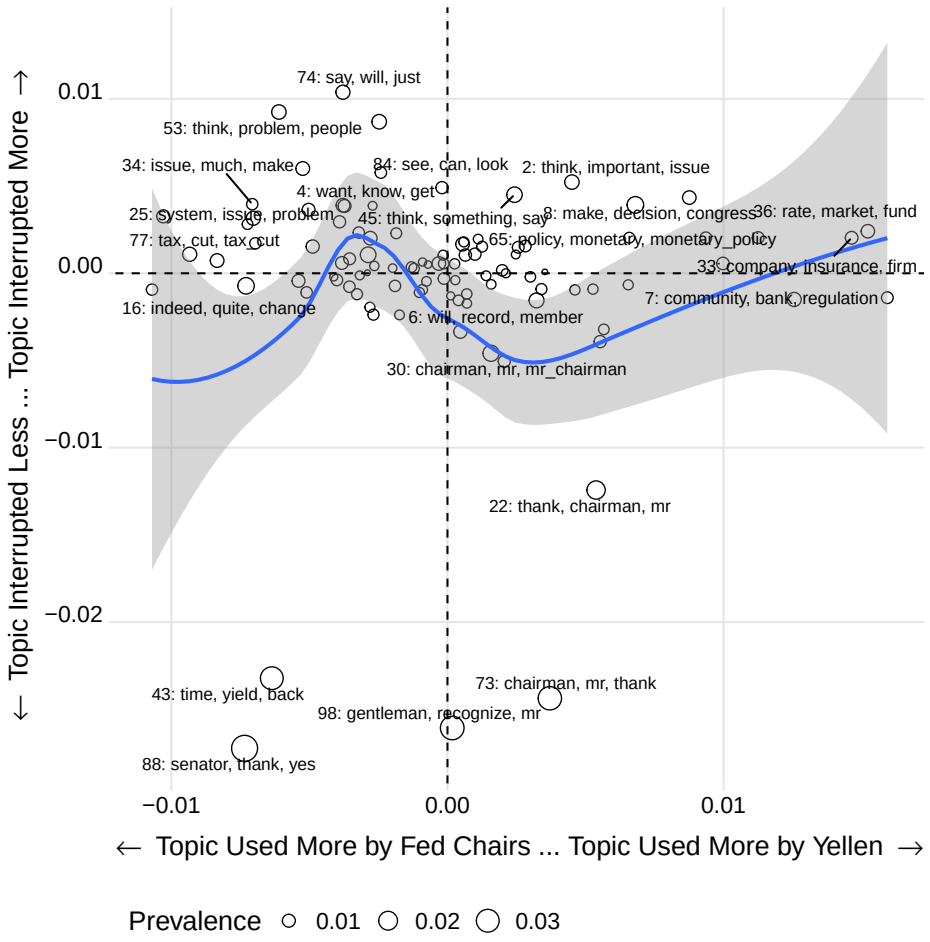

Figure 9: Scatter plot of topics by whether they were used more by Yellen or male Fed Chairs (x-axis) and whether they were more or less interrupted (y-axis).

Alternatively, we assign the highest scoring topic to every utterance, aggregate them by speaker, and inspect whether the utterance was interrupted. Figure 10 plots the topics on the y-axis, and the proportion of utterances associated with this topic for each group of speakers that were interrupted on the x-axis. As illustrated, Yellen is more likely to be interrupted across almost all topics compared to the other male chairs of the Fed. These results indicate that the systematic evidence of an interruption bias against Janet Yellen is unlikely to be driven by her focus on specific topics compared to other Fed chairs.

chair (x-axes) and compare these coefficients with topic-specific intercepts (y-axes) from a model that predicts whether an utterance is interrupted, conditional on speaker fixed effects, hearing fixed effects, and continuous measures of rhetorical hostility. Our results confirm the lack of a systematic relationship between the topics differentially used by Yellen and the likelihood she is interrupted.

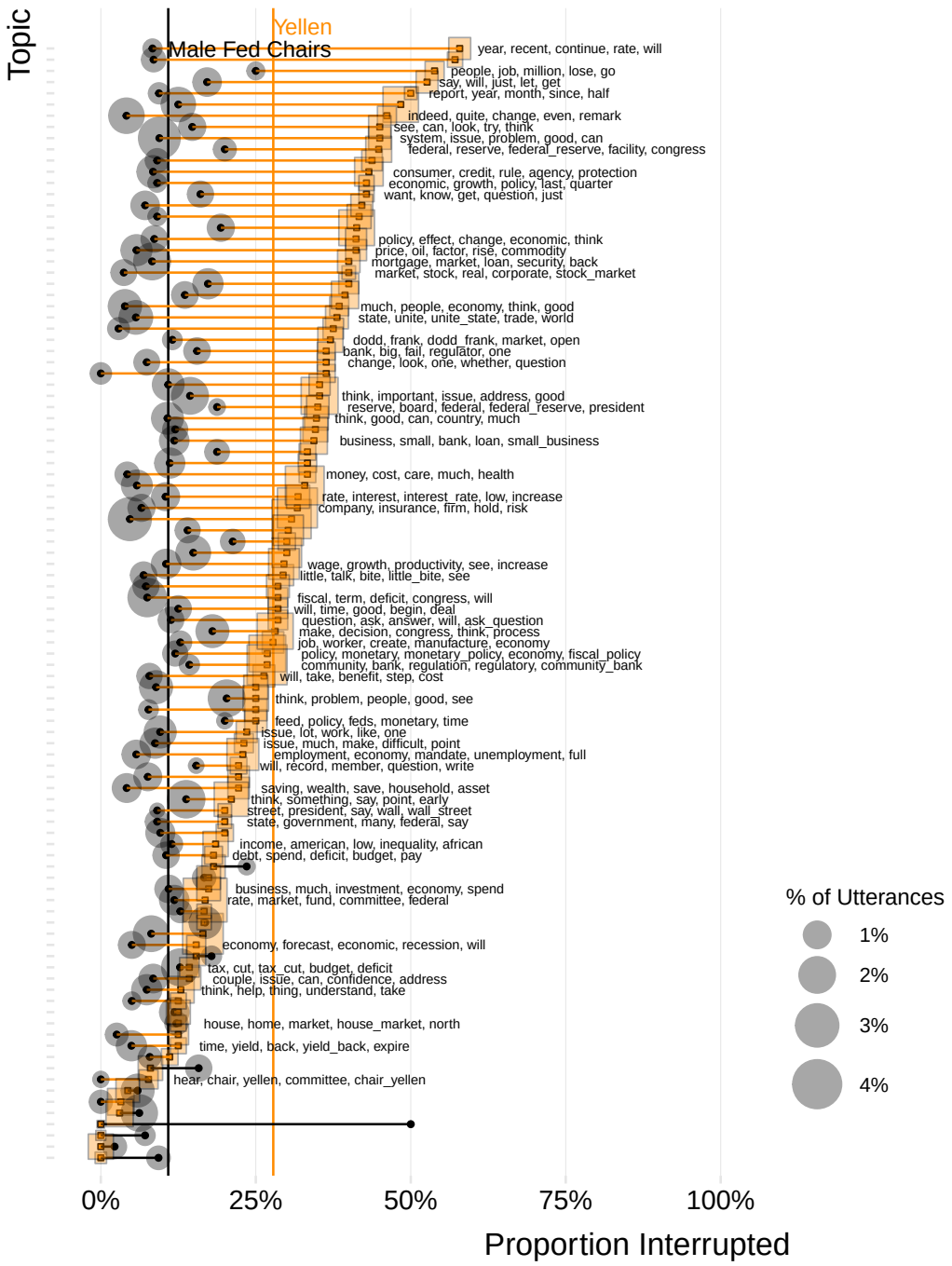

Figure 10: Each utterance is assigned to its highest-scored topic (y-axis) and then aggregated to the speaker. The x-axis indicate the proportion of each speaker's utterances that is interrupted, broken out by the utterance's highest scoring topic. Points are sized by the proportion of all utterances that are assigned to each topic.

However, topic models can only capture *what* was said, not *how* it was communi-

cated.[35] If Yellen talks about the same things as other Fed chairs, but does so in a more interruptable way, the causal effect we attribute to sexism might instead be due to the way she speaks.[36] To address this issue, we run a regression similar to those described above in equation 1, except that we replace the indicator for whether an utterance is interrupted with the continuous measure of the utterance's toxicity, aggression, or incoherence. These measures are produced by a sophisticated machine learning algorithm developed by Google for the purpose of online content moderation (*Using machine learning to reduce toxicity online*, 2021). The algorithm is trained on vector representations of chunks of text and generates a predicted score for a number of different dimensions of tone, including attack, identity attack, threat, insult, unsubstantial, incoherent, inflammatory, toxicity, sexually explicit, profanity, obscene, or flirtation. The algorithm's resulting score can be understood as the predicted probabilities that a human would label that text as belonging to one of the aforementioned dimensions.[37] We refer to these measures as capturing the "tone" of each utterance.[38]

In addition, we subset the data to only those utterances made by one of the chairs of the Federal Reserve, and focus attention on an indicator for Janet Yellen. Formally:

$$\text{Tone}_{u,i,j,t} = \mathbb{I}\text{Yellen}_t + \rho_j + \text{poly}_{t,3} + \gamma \mathbf{U}_u + \varepsilon_{i,j,t} \tag{4}$$

where $\text{poly}_{t,3}$ is a cubic polynomial for time, ρ_j are fixed effects for who the speaker is responding to, and $\mathbf{U}_u$ are utterance level controls including the length of the preceding

[35] In addition, topic models by their nature mix the substantive content of speech with the tone. In the SI section 8.5, we demonstrate that our conclusions persist if we instead rely on a dictionary-based measure of 6 topics specific to central banking, based on Fraccaroli et al. (2022*a*).

[36] Note that the concept of "in a more interruptable way" is itself a potentially gendered concept, making it post-treatment from the perspective of causal identification theory.

[37] We acknowledge that the use-case for this algorithm – content moderation for online platforms – differs from our application of it to Congressional hearings. Three human coders manually labelled a random sample of chunks and found validating support for our use of this algorithm. In addition, we further validated our analysis via ChatGPT 3.5 which allowed us to scale up the human labelling effort, again providing reassuring support for our conclusion that Yellen's experience was more hostile than that of her male counterparts. An extensive discussion of these results can be found in the SI Section 8.

[38] Our use of the term "tone" differs from related work by Dietrich, Hayes and Obrien (2019), which relies on audio data to measure divergences in pitch to capture emotional tones.

utterance (logged), the total number of utterances of the speaker (logged), and indicators for whether the utterance is interrupted and if it is itself interrupting someone else. The results are summarized in the left panel of Figure 11, illustrating that Janet Yellen's tone is either no different from her male counterparts or, if anything, is *less* offensive along the dimensions of toxicity, profanity, and her use of sexually explicit language (the latter of which is a very low probability in our context of congressional hearings writ large). In sum, we find little evidence to support the alternative hypothesis that Yellen is interrupted more because of what she talks about, or how she talks.

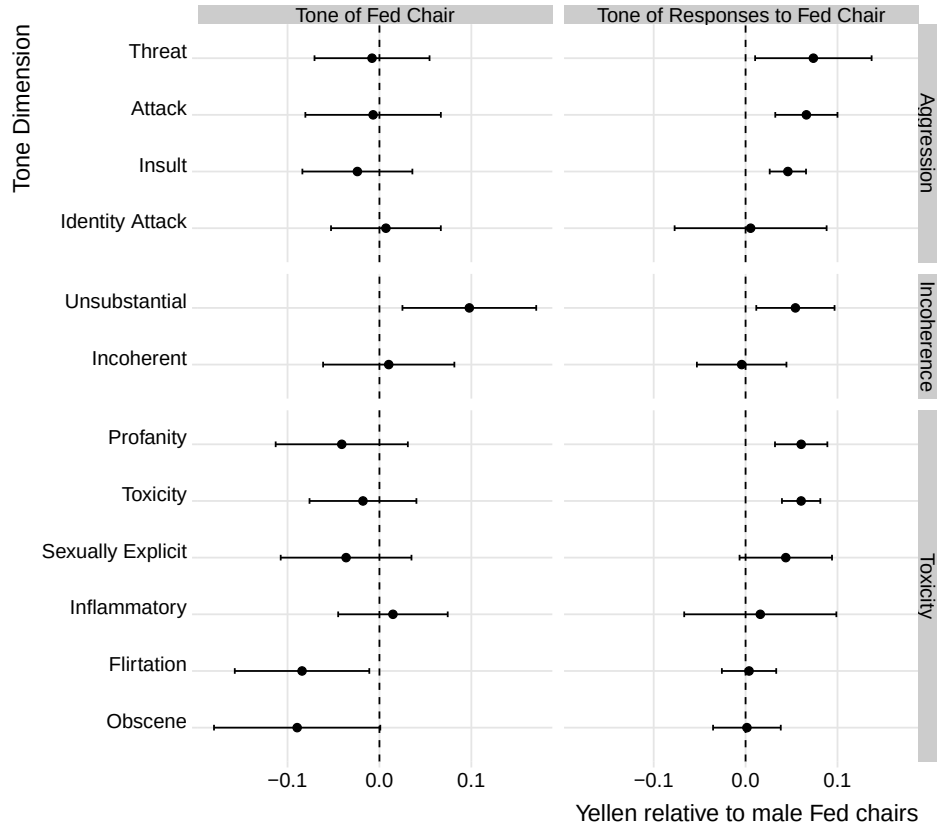

Figure 11: Coefficients capturing the difference in tone between utterances by Janet Yellen and the other three male Fed chairs.

As a final test, we reverse the analysis, looking at the utterances that are themselves responding to the Fed chair, and again focusing on whether these differ systematically by Yellen's tenure. As illustrated in the right panel of Figure 11, there is systematic

evidence suggesting that Yellen's experience was more hostile overall compare to her male counterparts. Utterances responding to her were more likely to be insulting, attacking, toxic, or profane.[39] In sum, the data indicates that Janet Yellen was both a more polite chair of the Federal Reserve compared to the men who came before and followed her, and yet was treated with greater hostility than them, patterns of behavior we argue reflect the latent trait of hostile sexism among members of Congress attending these hearings.

4 Discussion

Legislators attending congressional oversight hearings with the Chair of the Federal Reserve interrupted Janet Yellen – the first female chair – more than they interrupted her male predecessors and male successor. Yellen did not interrupt others more than her male counterparts, nor did she talk about different topics or use different language in her communication. Furthermore, the increase in interruptions persists after controlling for the gender, seniority, ideology, age, partisanship, chamber, and incumbent vote margin, as well as indicators for whether the legislator sponsored legislation to augment Congress's oversight on the Fed and for whether the legislator opposed or abstained Yellen's confirmation. Finally, we show that male legislators without daughters increased their interruptions more than those with daughters.

We argue that these patterns reflect an increase in hostile sexism against Janet Yellen, who threatened traditionally male positions of power with her tenure as the first female chair of the U.S. Federal Reserve. We support our interpretation of these patterns with two additional pieces of evidence: first, interactions with Yellen were more hostile than those with other (male) Fed chairs, where hostility is captured using a machine learning based estimate of toxicity. Second, interactions with Yellen were less polite than interactions with other (male) Fed chairs, where politeness is captured using unsupervised

[39] We examine these results in more detail in the SI Section 10, finding that, while the tone of the interruptions themselves did not differ by Fed chair, the overall tone was more hostile toward Yellen, particularly if we use quantile regressions to characterize more extreme outbursts.

zero-shot classification via OpenAI's GPT 3.5 model (see SI Section 8.8). Our results are consistent with a rich theoretical literature on sexism, spanning the fields of sociology, anthropology, education, feminist studies, and political science. By demonstrating the presence of hostile sexism in a setting where traditionally male positions of status are being threatened, we provide carefully identified empirical evidence in support of Ambivalent Sexism Theory (AST, Glick and Fiske 1996) and role congruity theory (Eagly and Karau, 2002). And by showing the attenuating influence of parenting daughters, we contribute to a growing literature on how socially and politically relevant attitudes can be shaped by events occurring later in life (Shafer and Malhotra, 2011).

Despite these comprehensive results, one might remain skeptical that Yellen's experience had nothing to do with her gender but instead reflected a secular increase in the hostility of Congressional hearings over time. In the SI Section 7, we drop Yellen entirely and compare the experiences of male Fed chairs to each other, finding systematically null results. Since these chairs served both prior to and following Yellen's tenure, these findings are inconsistent with this alternative explanation of a secular increase in hostility.

Another concern might be that Yellen was the only Fed chair in our data who was appointed by a Democratic president. While we cannot "unbundle" this treatment, we can point to two patterns that suggest the increase in interruptions is not merely a spurious byproduct of partisanship. First, Bernanke was re-appointed by Barack Obama in 2009, yet the difference in hostility between these two chairs was substantial. Second, we demonstrate that, while Republicans increased their interruptions of the Fed chair much more significantly than did Democrats during Yellen's tenure, legislators from both parties began interrupting Yellen more.

Nevertheless, we acknowledge that our findings only provide evidence consistent with our theoretically motivated story that hostile sexism manifests in the halls of leadership. Our substantive interpretation rests on an Occam's Razor intuition in which alternative stories are both empirically less consistent with our evidence and conceptually more post

hoc. In addition, many of these alternative stories – such as the concern that Yellen might simply be a less pleasant or more adversarial interlocutor – are potentially "post-treatment" in the sense that her speaking style is itself informed by her experience as a path-breaking woman in a traditionally male environment.

Given the focus on a single Fed chair, it is natural to wonder whether our results generalize beyond Janet Yellen. In particular, one might suspect that Yellen is something of an easy case for documenting these behaviors, given the extremely male-dominated context of central banking and the enormous power associated with it.

We do not disagree with this interpretation, and we suspect that evidence of similar hostile sexism elsewhere is likely prevalent as women make further inroads into the traditionally male halls of political power. Statistically, we emphasize that in our empirical framework, Janet Yellen is better understood as a treatment rather than an outcome. Her tenure as chair of the Fed is the variation in gender that we exploit to document systematic evidence of sexism among elected officials. Our results suggest, at minimum, that 242 democratically elected officials in the U.S. Congress exhibited significant evidence of hostile sexism. Substantively, we argue that Janet Yellen's 'grilling' in Congressional hearings is a pertinent example how gender bias and sexism can creep into political oversight and threaten the credibility of vital democratic accountability mechanisms. Even if we limit our contribution to the sexism experienced by Janet Yellen, we argue that this is – on its own – an important finding, as sexism undermines the purpose of these hearings.

From a practical or policy perspective, our contribution is an urgent call to recognize and remove the subtle layers of a political glass ceiling on women to create a more equitable policy environment benefiting the greater public good.

References

Anderson, Kristin J and Campbell Leaper. 1998. "Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how." *Sex roles* 39(3-4):225–252.

Ban, Pamela, Justin Grimmer, Jaclyn Kaslovsky, Emily West et al. 2022. "How Does the Rising Number of Women in the US Congress Change Deliberation? Evidence from House Committee Hearings." *Quarterly Journal of Political Science* 17(3):355–387.

Barber, Michael, Daniel M Butler, Jessica Preece et al. 2016. "Gender inequalities in campaign finance." *Quarterly Journal of Political Science* 11(2):219–48.

Bauer, Lawrence and Alex Faseruk. 2020. "Understanding Political Pressures. Monetary Policy, and the Independence of the Federal Reserve in the United States From 1960-2019." *Journal of Management Policy and Practice* 21(3):41–63.

Beck, Susan Abrams. 2001. Acting as Women: The Effects and Limitations of Gender in Local Governance. In *The impact of women in public office*, ed. Susan J. Carroll. Indiana University Press pp. 49–68.

Bernanke, Ben. 2022. *21st Century Monetary Policy The Federal Reserve from the Great Inflation to COVID-19*. W.W. Norton & Company.

Binder, Sarah and Mark Spindel. 2019. *The myth of independence: How Congress governs the Federal Reserve*. Princeton University Press.

Blei, David M, Andrew Y Ng and Michael I Jordan. 2003. "Latent dirichlet allocation." *the Journal of machine learning research* 3:993–1022.

Bock, Jarrod, Jennifer Byrd-Craven and Melissa Burkley. 2017. "The role of sexism in voting in the 2016 presidential election." *Personality and Individual Differences* 119:189–193.

Bodea, Cristina and Andrew Kerner. 2021. "The Gender Credibility Gap: All-Male Boards and Substantive Gender Representation in Central Banking." SSRN Working Paper No. 3780220.
URL: *https://doi.org/10.2139/ssrn.3780220*

Bodea, Cristina, Federico Maria Ferrara, Andrew Kerner and Thomas Sattler. 2021. "Gender and Economic Policy: When Do Women Speak with Authority on Economic Issues? Evidence from the Euro Area." SSRN Working Paper No. 3881069.
URL: *https://doi.org/10.2139/ssrn.3881069*

Borrell-Porta, Mireia, Joan Costa-Font and Julia Philipp. 2019*a*. "The mighty girleffect: does parenting daughters alter attitudes towards gender norms?" *Oxford Economic Papers* 71(1):25–46.

Borrell-Porta, Mireia, Joan Costa-Font and Julia Philipp. 2019*b*. "The mighty girleffect: does parenting daughters alter attitudes towards gender norms?" *Oxford Economic Papers* 71(1):25–46.

Brescoll, Victoria L. 2016. "Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders." *The Leadership Quarterly* 27(3):415–428.

Calcagno, Peter T. and Meg M. Montgomery. 2020. "The gender wage gap: an analysis of US congressional staff members." *Public Choice* 188(1-2):183–201.

Carney, Mark. 2019. "Finance by all, for all." Remarks at Women in Banking and Finance 22nd Annual Awards, London. June 14, 2019.
URL: *https://www.bankofengland.co.uk/-/media/boe/files/speech/2019/finance-by-all-for-all-speech-by-mark-carney.pdf*

Chappell Jr, Henry W, Thomas M Havrilesky and Rob Roy McGregor. 1993. "Partisan monetary policies: Presidential influence through the power of appointment." *The Quarterly Journal of Economics* 108(1):185–218.

Charléty, Patricia, Davide Romelli and Estefania Santacreu-Vasut. 2017. "Appointments to central bank boards: Does gender matter?" *Economics Letters* 155:59–61.

Comunale, Mariarosaria, Petra de Bruxelles, Kalpana Kochhar, Juliette Raskauskas and Filiz D. Unsal. 2023. "Who are Central Banks? Gender, Human Resources, and Central Banking." *IMF Working Papers* 2023(091):29.

Conroy, Meredith, Sarah Oliver, Ian Breckenridge-Jackson and Caroline Heldman. 2015. "From Ferraro to Palin: Sexism in coverage of vice presidential candidates in old and new media." *Politics, Groups, and Identities* 3(4):573–591.

Dietrich, Bryce J, Matthew Hayes and Diana Z Obrien. 2019. "Pitch perfect: Vocal pitch and the emotional intensity of congressional speech." *American Political Science Review* 113(4):941–962.

Diouf, Ibrahima and Dominique Pépin. 2017. "Gender and central banking." *Economic Modelling* 61:193–206.

Eagly, Alice H and Steven J Karau. 2002. "Role congruity theory of prejudice toward female leaders." *Psychological review* 109(3):573.

Ferrara, Federico M., Donato Masciandaro, Manuela Moschella and Davide Romelli. 2021. "Political voice on monetary policy: Evidence from the parliamentary hearings of the European Central Bank." *European Journal of Political Economy* p. 102143.
URL: *https://doi.org/10.1016/j.ejpoleco.2021.102143*

Fox, Richard L and Jennifer L Lawless. 2010. "If only theyd ask: Gender, recruitment, and political ambition." *The Journal of Politics* 72(2):310–326.

Fox, Richard L and Jennifer L Lawless. 2014. "Uncovering the origins of the gender gap in political ambition." *American Political Science Review* 108(3):499–519.

Fraccaroli, Nicolò, Alessandro Giovannini, Jean-François Jamet and Eric Persson. 2022*a*. "Does the European Central Bank speak differently when in parliament?" *The Journal of Legislative Studies* 28(3):421–447.

Fraccaroli, Nicolò, Alessandro Giovannini, Jean-François Jamet and Eric Persson. 2022*b*. "Ideology and monetary policy. The role of political parties' stances in the European Central Bank's parliamentary hearings." *European Journal of Political Economy* 74:102207.

Fraccaroli, Nicolò, Alessandro Giovannini, Jean-François Jamet and Eric Persson. 2023. "Central banks in Parliaments: A Text Analysis of the Parliamentary Hearings of the Bank of England, the European Central Bank and the Federal Reserve." *International Journal of Central Banking* 19(2):543–600.

Glick, Peter and Susan T Fiske. 1996. "The ambivalent sexism inventory: Differentiating hostile and benevolent sexism." *Journal of personality and social psychology* 70(3):491.

Glynn, Adam N. and Maya Sen. 2014. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?" *American Journal of Political Science* 59(1):37–54.

Goldin, Claudia. 2023. WHY WOMEN WON. Technical Report Working Paper 31762 National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w31762*

Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *American Economic Review* 90(4):715–741.

Heldman, Caroline, Susan Carroll and Stephanie Olson. 2005. ""She brought only a skirt": Print media coverage of Elizabeth Dole's bid for the Republican presidential nomination." *Political communication* 22(3):315–335.

Holman, Mirya R, Jennifer L Merolla and Elizabeth J Zechmeister. 2016. "Terrorist threat, male stereotypes, and candidate evaluations." *Political Research Quarterly* 69(1):134–147.

Hospido, Laura and Carlos Sanz. 2020. "Gender Gaps in the Evaluation of Research: Evidence from Submissions to Economics Conferences." *Oxford Bulletin of Economics and Statistics* 83(3):590–618.

Hospido, Laura, Luc Laeven and Ana Lamo. 2022. "The Gender Promotion Gap: Evidence from Central Banking." *The Review of Economics and Statistics* pp. 1–45.

Huddy, Leonie and Nayda Terkildsen. 1993. "The consequences of gender stereotypes for women candidates at different levels and types of office." *Political Research Quarterly* 46(3):503–525.

Kanthak, Kristin and Jonathan Woon. 2015. "Women don't run? Election aversion and candidate entry." *American Journal of Political Science* 59(3):595–612.

Karpowitz, Christopher F. and Tali Mendelberg. 2014. *The silent sex*. Princeton 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(3):560–576.

Kousser, Thad. 2005. *Term limits and the dismantling of state legislative professionalism*. Cambridge University Press.

Krook, Mona Lena. 2020. *Violence against Women in Politics*. Oxford University.

Lawless, Jennifer L. 2009. "Sexism and gender bias in election 2008: A more complex path for women in politics." *Politics & Gender* 5(1):70–80.

Leaper, Campbell and Melanie M Ayres. 2007*a*. "A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech." *Personality and Social Psychology Review* 11(4):328–363.

Leaper, Campbell and Melanie M. Ayres. 2007*b*. "A Meta-Analytic Review of Gender Variations in Adults' Language Use: Talkativeness, Affiliative Speech, and Assertive Speech." *Personality and Social Psychology Review* 11(4):328–363.

Maricut-Akbik, Adina. 2020. "Contesting the European Central Bank in Banking Supervision: Accountability in Practice at the European Parliament." *JCMS: Journal of Common Market Studies* 58(5):1199–1214.

Masciandaro, Donato, Paola Profeta and Davide Romelli. 2018. Do Women Matter in Monetary Policymaking? Working Paper 88 Baffi CAREFIN Centre.

Mattei, Laura R Winsky. 1998. "Gender and power in American legislative discourse." *The Journal of Politics* 60(2):440–461.

Mendelberg, Tali, Christopher F Karpowitz and J Baxter Oliphant. 2014. "Gender inequality in deliberation: Unpacking the black box of interaction." *Perspectives on Politics* 12(1):18–44.

Och, Malliga. 2020. "Manterrupting in the German Bundestag: Gendered Opposition to Female Members of Parliament?" *Politics & Gender* 16(2):388–408.

Payson, Julia, Alexander Fouirnaies and Andrew B Hall. 2021. "Decomposing the source of the gender gap in legislative committee service: evidence from US states." *Political Science Research and Methods* pp. 1–7.

Poole, Keith T and Howard Rosenthal. 2017. *Ideology & congress: A political economic history of roll call voting*. Routledge.

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.

Rolfes-Haase, Kelly L. and Michele L. Swers. 2021. "Understanding the Gender and Partisan Dynamics of Abortion Voting in the House of Representatives." *Politics & Gender* pp. 1–35.

Rudman, Laurie A and Peter Glick. 2021. *The social psychology of gender: How power and intimacy shape gender relations*. Guilford Publications.

Sanbonmatsu, Kira. 2006. "Do parties know that women win? Party leader beliefs about women's electoral chances." *Politics & Gender* 2(4):431–450.

Schonhardt-Bailey, Cheryl, Christopher Dann and Jacob Chapman. 2022. "The accountability gap: Deliberation on monetary policy in Britain and America during the financial crisis." *European Journal of Political Economy* 74:102209.

Shafer, Emily Fitzgibbons and Neil Malhotra. 2011. "The effect of a child's sex on support for traditional gender roles." *Social Forces* 90(1):209–222.

Sorensen, Ashley and Philip Chen. 2021. "Identity in Campaign Finance and Elections: The Impact of Gender and Race on Money Raised in 2010–2018 US House Elections." *Political Research Quarterly* p. 10659129211022846.

Swers, Michele L. 2020. *The difference women make*. University of Chicago Press.

Teele, Dawn Langan, Joshua Kalla and Frances Rosenbluth. 2018. "The ties that double bind: social roles and women's underrepresentation in politics." *American Political Science Review* 112(3):525–541.

Using machine learning to reduce toxicity online. 2021.
URL: *https://www.perspectiveapi.com/*

Van der Pas, Daphne Joanna and Loes Aaldering. 2020. "Gender differences in political media coverage: A meta-analysis." *Journal of Communication* 70(1):114–143.

Vera, Sebastian Vallejo and Analía Gómez Vidal. 2020. "The Politics of Interruptions: Gendered Disruptions of Legislative Speeches." *Journal of Politics* .

Volden, Craig, Alan E. Wiseman and Dana E. Wittmer. 2016. "Women's Issues and Their Fates in the US Congress." *Political Science Research and Methods* 6(4):679–696.

Washington, Ebonya L. 2008. "Female socialization: how daughters affect their legislator fathers." *American Economic Review* 98(1):311–32.