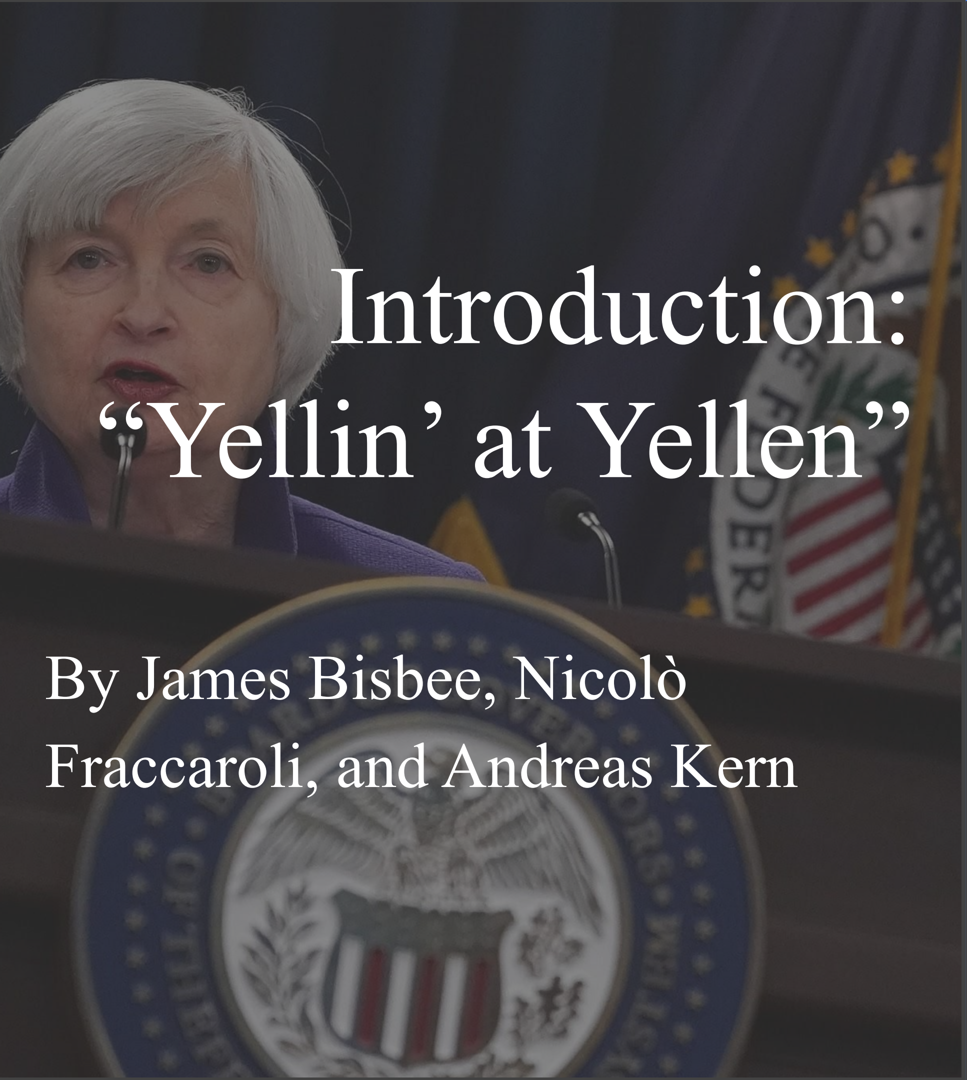
Replication by Jiani Zhao, Kaitlyn Vana

“Yellin’ at Yellen”: Hostile Sexism in the Federal Reserve Congressional Hearings

Overview

1. Introduction to “Yellin’ at Yellen”
2. Methods:
   - Text Processing
   - Data Exploration
3. Specifications, Results, Differences
4. Autopsy
5. Extension



Introduction: “Yellin’ at Yellen”

By James Bisbee, Nicolò Fraccaroli, and Andreas Kern

Q: *“How prevalent is hostile sexism among US politicians?”*

Data: corpus of transcripts, every **congressional hearing** attended by the chair of the US Federal Reserve 2001-2020

Method: measuring sexism, using Janet Yellen as a bundled treatment (Yellen understood as a treatment, not outcome)

Results: “...**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, the paper shows that having a daughter reduces a legislator’s hostility toward Yellen.**

Contribution: literature on forms of sexism, responses to women in novel positions of power. Acknowledging potential for non-generalizability given male-dominated field, argue example how gender bias, sexism exists in political oversight, threatens credibility of vital democratic accountability mechanisms

Methods: Text Processing

Senate Committee on Banking, Housing, and Urban Affairs (x2), House Committee on Financial Services (x2), legislators allocated budget of time to engage Fed chair, 2001-2020

Public record: transcripts stored on GovInfo

Interruptions identified based on the notation used in the original transcripts: a series of two or more hyphens (--)

Single data frame with rows indexing the speaker-utterance-hearing

Data structure: 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 Fed (Greenspan, Bernanke, Yellen, and Powell)

Methods: Data

Chairs of the Fed have the most utterances, followed by acting committee chairs

Janet Yellen among speakers who is interrupted the most:

measured as the proportion of her utterances that are interrupted and interrupts

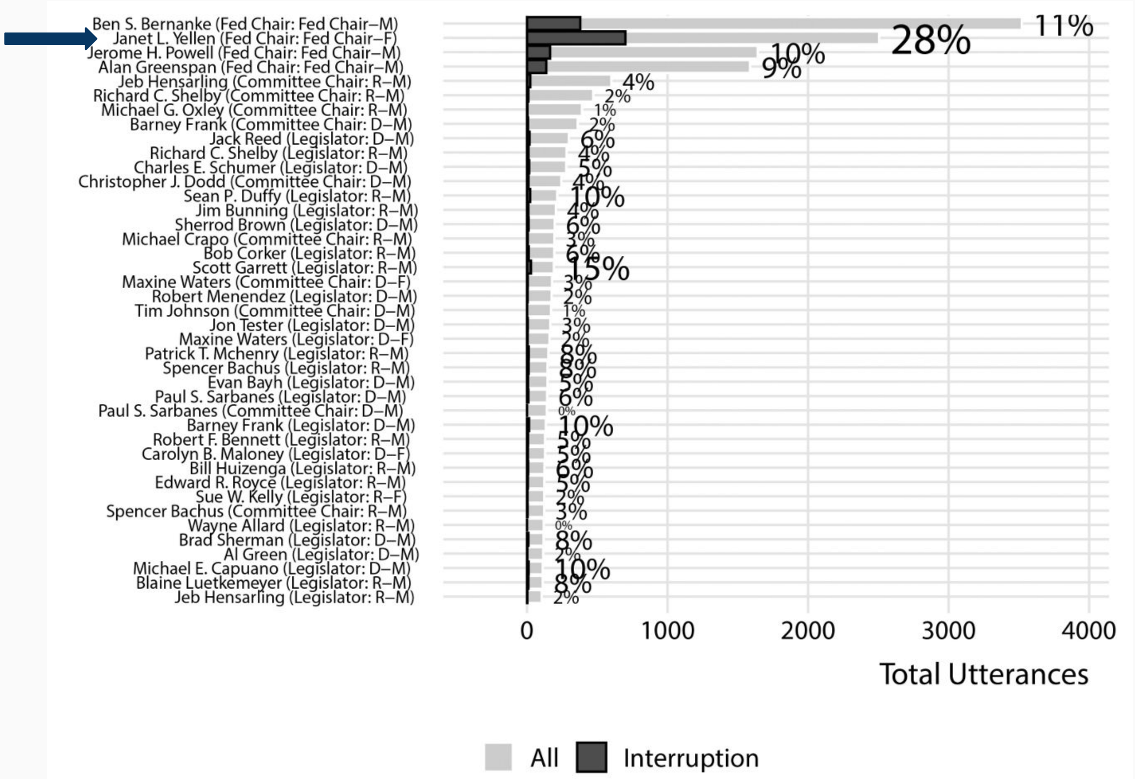


Figure 1. Total utterances by most talkative speakers (light gray bars), along with total interruptions (dark gray bars in the left panel) and total utterances that interrupt another speaker (dark gray bars in the right panel).

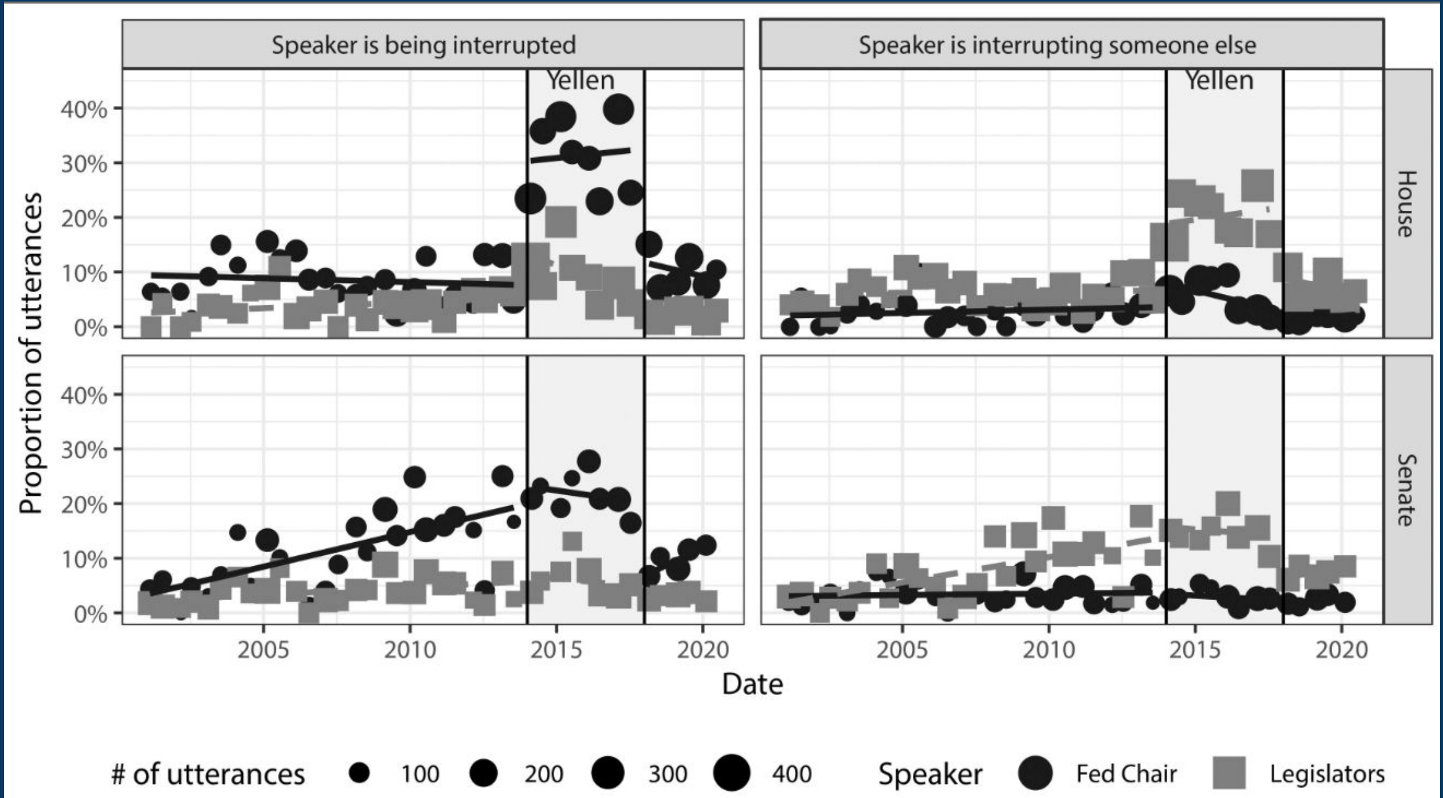


Figure 2. Proportion of utterances that are interrupted (left column) and that interrupt someone else (right column) by House (top row) and Senate (bottom row), broken out by whether the speaker is the chair of the Fed (black circles) or a legislator (gray squares). The shaded region denotes the mandate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

Specifications:

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

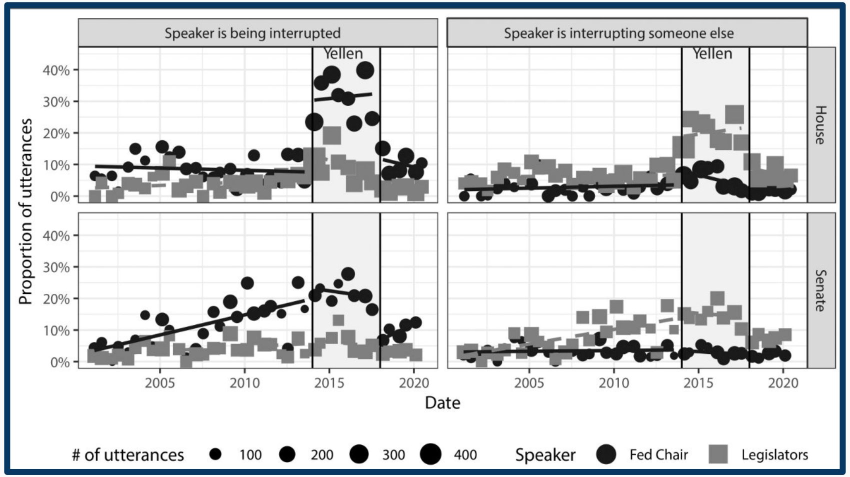
$$\text{Interrupting}_{u,i,j,t} = \alpha_i + \rho_j + \delta_t + \gamma \mathbf{U}_u + \varepsilon_{i,j,t}$$

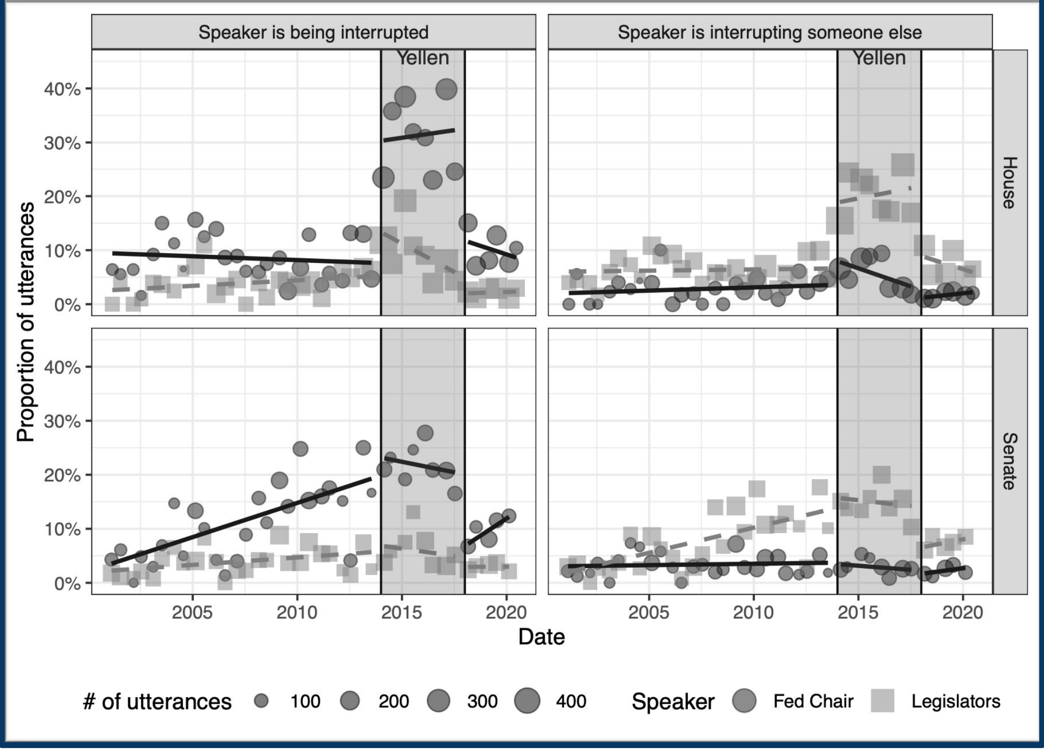
$$\begin{aligned}\text{Interrupting}_{u,i,\text{Fed},t} = {} & \beta_1 \mathbb{I}\text{Fed}_t + \beta_2 \mathbb{I}\text{Yellen Tenure}_{i,t} \\ & + \beta_3 \mathbb{I}\text{Fed}_t^* \mathbb{I}\text{Yellen Tenure}_{i,t} \\ & + \alpha_i + \gamma \mathbf{U}_u + \varepsilon_{i,j,t}\end{aligned}$$

Two NLPs measures: *first* NLP measure predicts the tone of each utterance based on Google's Perspective API
second NLP-based measure estimates the topics each speaker talks about by estimating a topic model via LDA with 100 topics on the utterances

Results, Differences

Proportion of utterances that are interrupted (left column) and that interrupt someone else (right column) by House (top row) and Senate (bottom row)

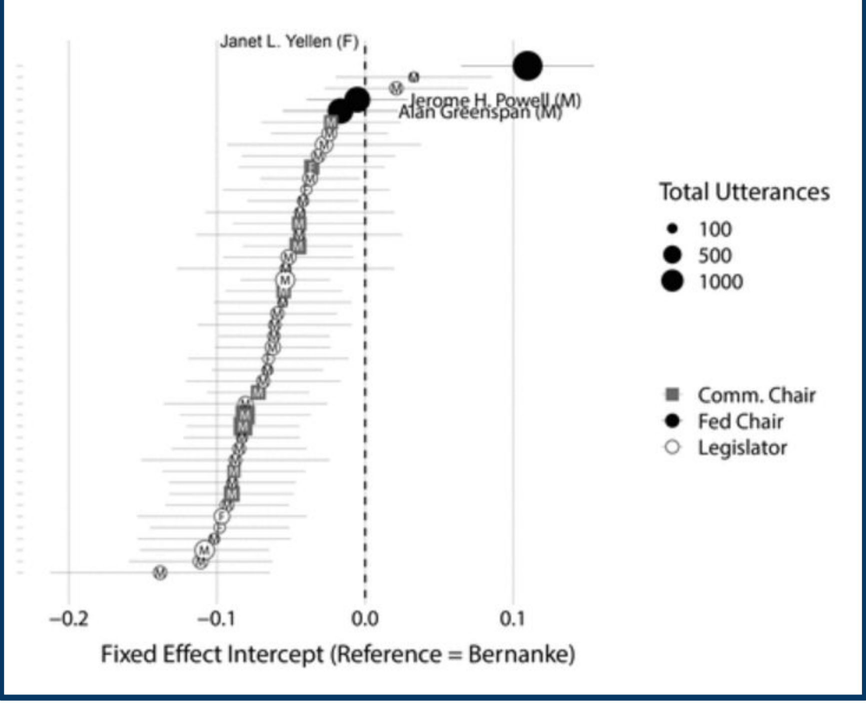


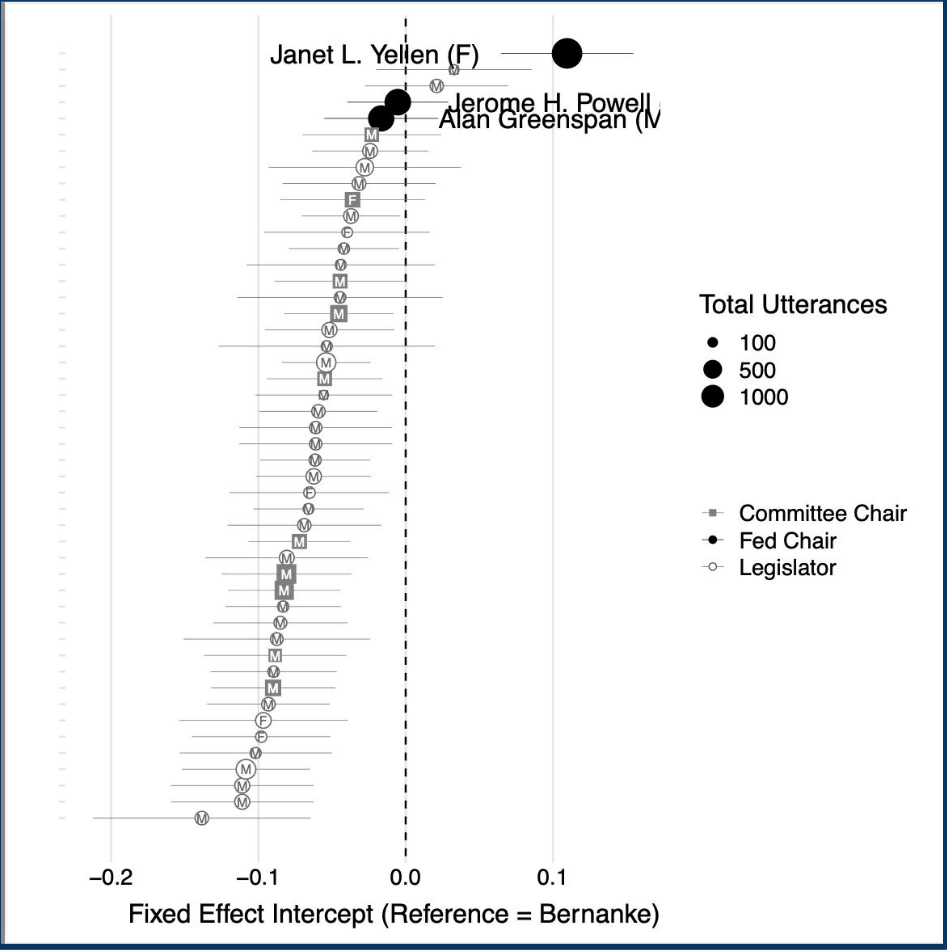


Left: from article pictured earlier in presentation, Right: generated in replication

Results, Differences

Coefficient plot summarizing the difference between Bernanke's propensity to be interrupted and a subset of other speakers

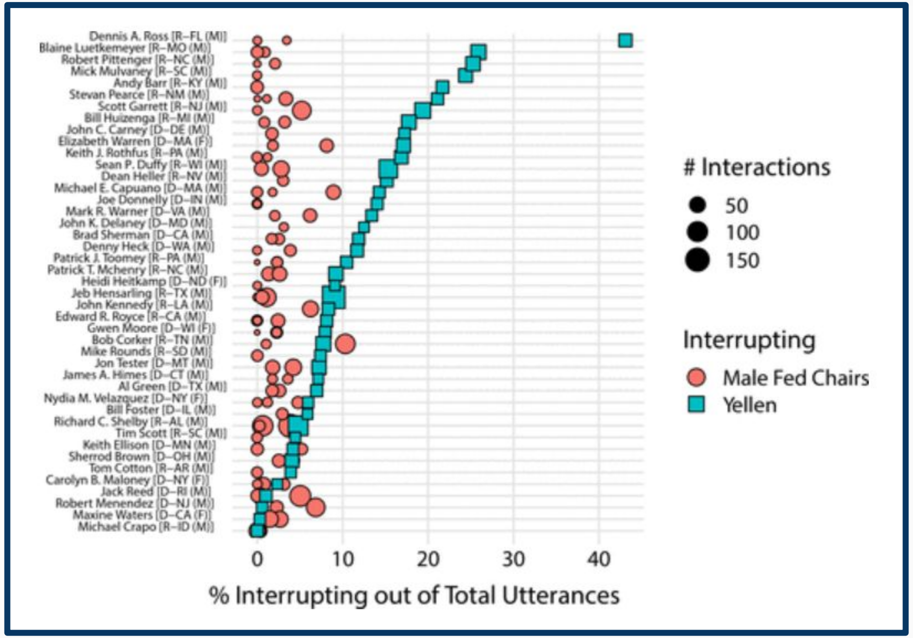


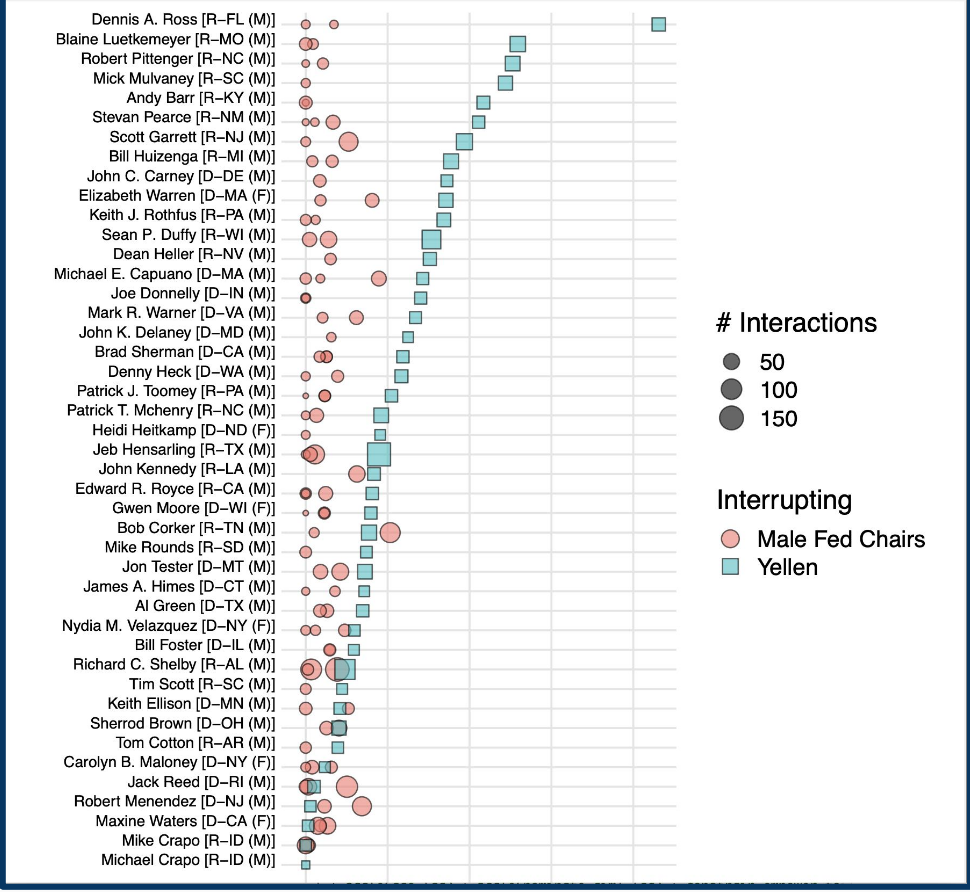


Left: from article, Right: generated in replication

Results, Differences

Points represent the share of total possible interactions (size) between a politician (x-axis) and one of the four Fed chairs that were interruptions.

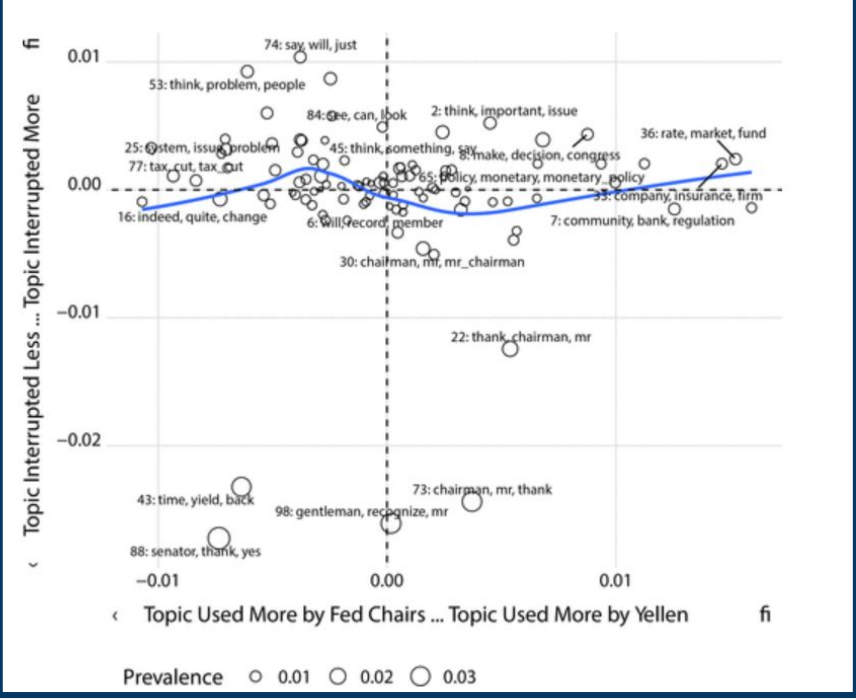


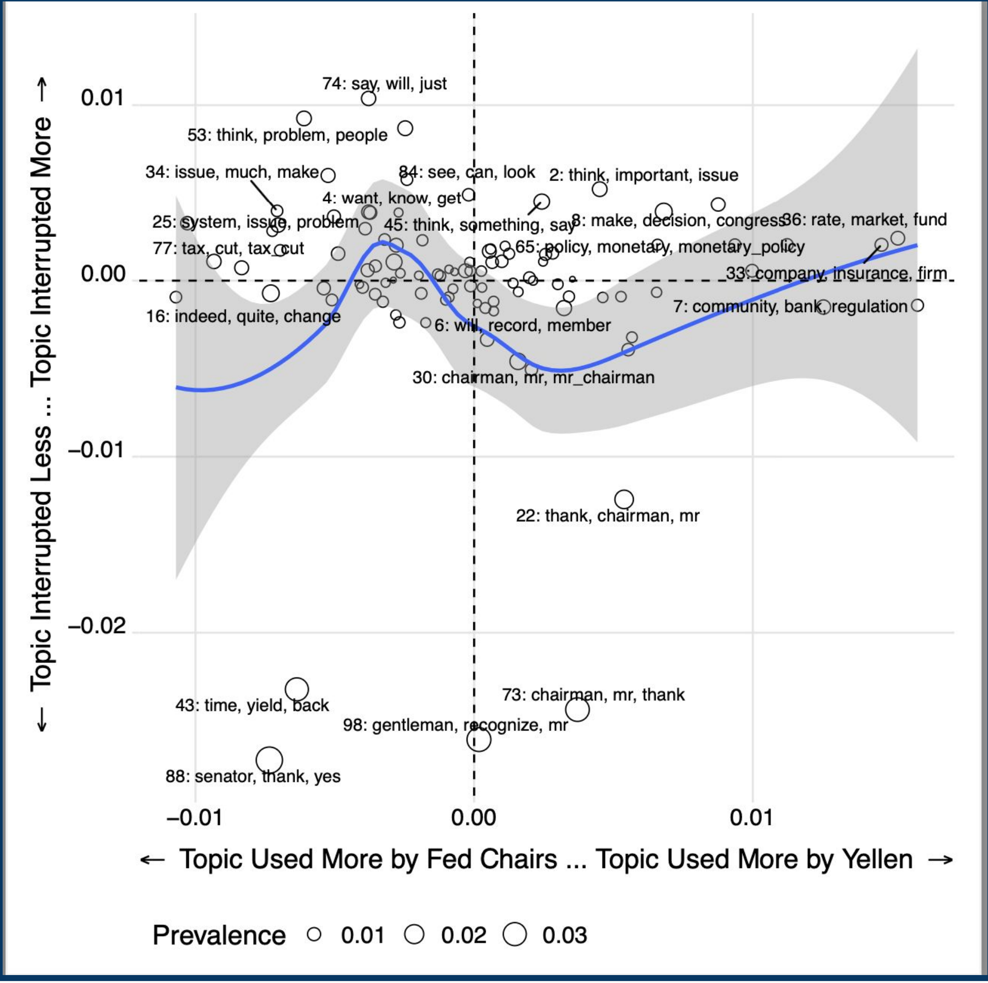


Left: from article, Right: generated in replication

Results, Differences

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).

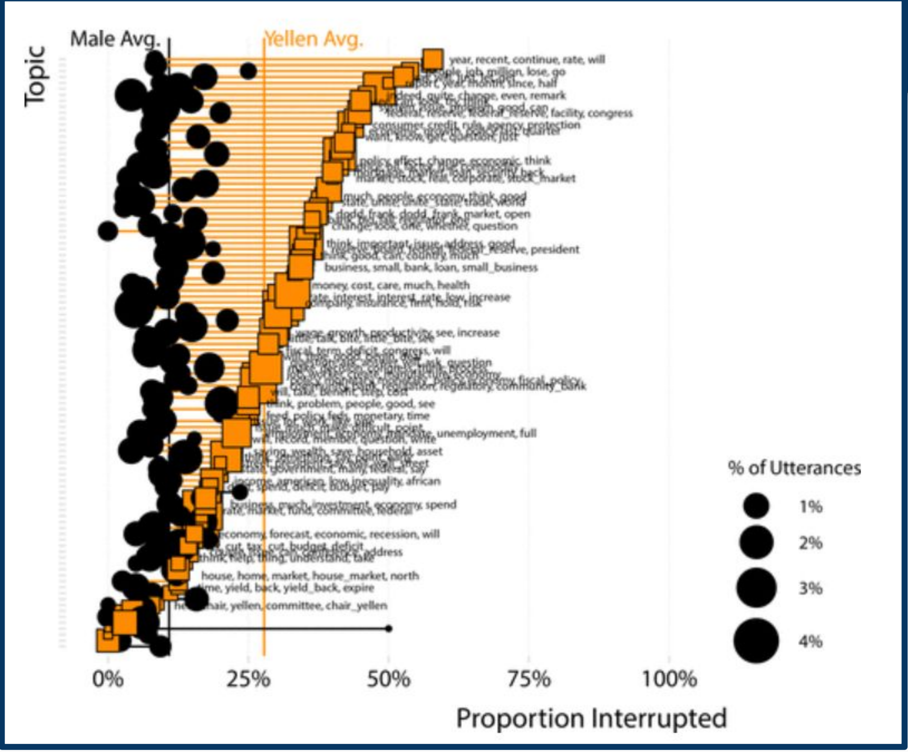


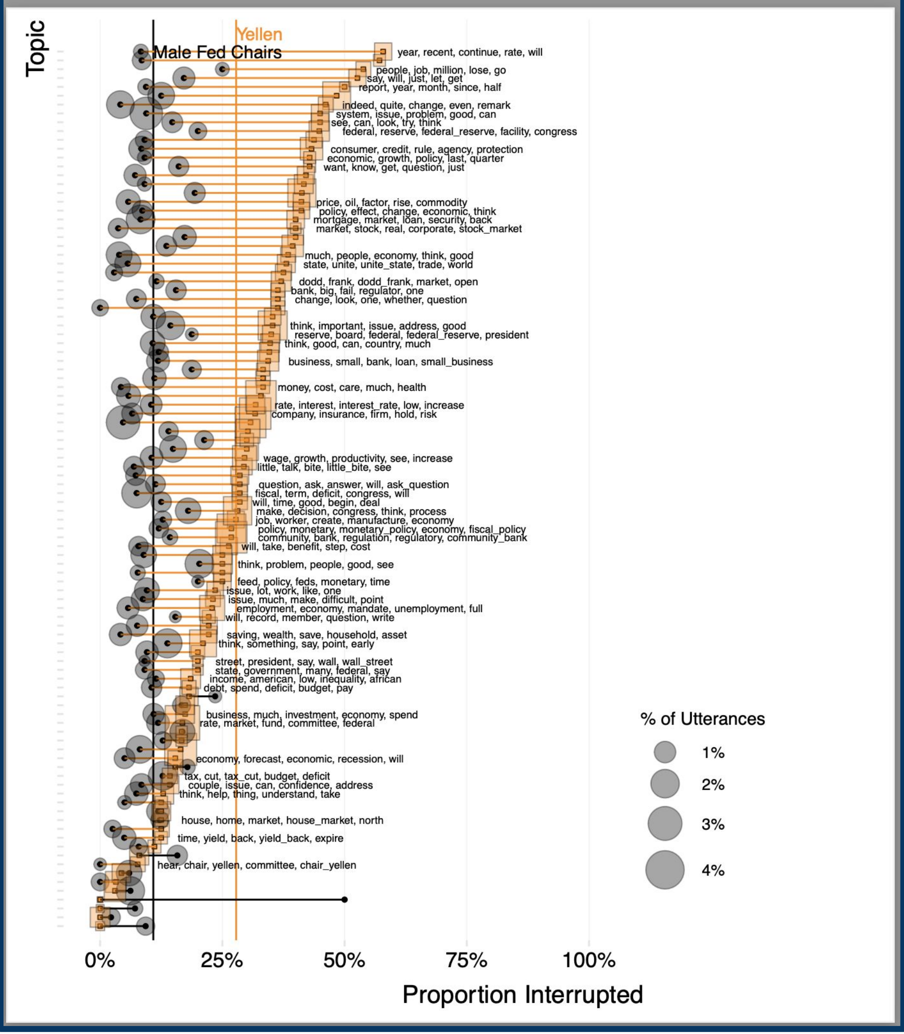


Left: from article, Right: generated in replication

Results, Differences

Each utterance is assigned to its highest-scored topic (y-axis) and then aggregated to the speaker.





Left: from article, Right: generated in replication

Autopsy

What Worked: Text Processing

Data Cleaning & Standardization

- Cleaned hearing transcripts (removed headers, fixed whitespace, standardized text).
- Harmonized speaker attributes (party, gender, demographics) and patched missing values.

Enriched Speaker Information

- Added electoral data, family variables, and biographical corrections for Fed officials/experts.
- Merged in bill-level data for each speaker (legislative activity controls).

What Worked: LDA Model (Baseline Model)

Preprocessing

- Clean raw text (tokenize, remove stopwords, stems, etc.)
- Split data into **train/test** for model selection

Build vocabulary + matrices

- **Create vocabulary** on *training* data. Compute:
  - **DTM (Document-Term Matrix):** how often each word appears in each document.
  - **TCM (Term Co-occurrence Matrix):** how often word-pairs co-occur within a sliding window.

Fit LDA + Model selection

- Train LDA on training documents using different K.
- Evaluate by:
  - **Perplexity** (predictive performance)
  - **Semantic coherence** (interpretability)
- Pick the best **K = 100**.
- Refit final LDA on the **entire dataset** to obtain final topic distributions.

What Worked: STM Model

Preprocessing

- Use **textProcessor()** and **prepDocuments()**
- No train/test split → STM is not evaluated via perplexity, not selecting K; focus is on covariate effects.

Model Setup

- Reuse the **same K = 100** as LDA to maintain interpretability across methods.
- Prevalence formula includes speaker & hearing attributes. (θ~gender+party+interrupted+…)

Estimating Covariate Effects

- *"Female speakers are 15% more likely to talk about financial stability (Topic 12)."*
- *Interrupted legislators shift toward defensive topics.*
- *Republicans emphasize inflation differently than Democrats.*

What Worked: RR1 Robustness Work

Re-run LDA under:

Chunk-Level Topic Model

1. Aggregate short utterances from the same speaker into large chunks (>1000 characters).
2. Split into **50/50 train–test**.
3. Search over many K using: perplexity and coherence
4. Best K ≈ **70** (different from 100 because text units changed)
5. Fit LDA on all grouped documents using K = 70.

Speaker-Level Topic Model

1. Aggregate all text by speaker.
2. Use 80/20 split.
3. Fit LDA with best K (again ≈70).

What Didn’t Work

1. cannot recreate all original figures
    a. plot_cme() no longer exists
    b. marginaleffects() no longer exists

Extension:

Different Context/Field

For generalizability: Secretary of Commerce, Secretary of Education, CFPB Director, CBO Director

Intra-Party, Family Life

Investigate/causality of family life, Democrat/Republican divisions

Alternative Methods

Revise pre-processing steps; Hard-encoding may help control for any overall hostility increases

Thank you!

Questions?