

Supplemental Online Material

Contents

A Measuring Advocate Gender	2
B Gendered Issues	3
C Confidence Intervals	5
D Justice-Level Estimates	6
E Heatedness of Discussions	7
F Back Channel Cues	9

A Measuring Advocate Gender

We operationalize the *gender signal* of advocates through the following two-stage process. In the first stage, we create an automated heuristic to extract the formal title of an advocate—Mr. or Ms.—stated by the Chief Justice when introducing the advocate in the utterance prior to the advocate’s first utterance. For example, in *Bowsher v. Synar*, Chief Justice Burger introduced the first advocate, Lloyd Cutler, by stating “Mr. Cutler, you may proceed whenever you are ready,” and later turned the floor over to opposing counsel Lois Williams by, after the conclusion of a lengthy statement by another advocate, stating “Very well. Ms. Williams?” Our heuristic extracts “Ms.” and assigns it to female and “Mr.” and assigns it to male. We move to the second stage when the first stage fails to extract a gender signal—which happens for just 0.75% of the unique advocates in our dataset—as when for example if the advocate is introduced with another title such as “General” or similar. In those cases, we look up the first name of the advocate in a first-name gender dictionary. We use the World Gender Name Dictionary (Raffo and Lax-Martinez 2018) which aggregates statistics on gender of names from administrative data such as the Social Security Administration. One limitation of this source is that it assigns binary gender to names that could be ambiguous, e.g. “Alex”, based on the observed most likely gender. However, this is an extremely small proportion of our overall data. In all, our two-stage approach assigns gender for approximately 99.8% of advocates, with the vast majority assigned based on pronouns used by the justices themselves.

B Gendered Issues

The substantive issue of the case stands as an additional, potent confounder. Specifically, prior work has established that women are interrupted less frequently across settings when they are perceived to be speaking from a position of authority (e.g., Miller and Sutherland 2022). In the context of oral argument, the attorneys are generally understood to be operating at a deficit with respect to the justices. However, when female advocates before the Court are addressing a “women’s issue”, they are likely to be perceived as operating from a position of authority; we might therefore expect female advocates to be interrupted less frequently than male advocates when the issue of the case is a “women’s issue” (Miller and Sutherland 2022; Patton and Smith 2017).

In our data, we find just such a gendered issue dynamic. Specifically, we separate the chunks into sets based on whether or not the case addressed a women’s issue. Applying the approach taken by Szmer, Sarver and Kaheny (2010), we identify as relating to women’s issues those cases coded by the Supreme Court Database as being about the issues of sex discrimination, abortion, and privacy. Out of 65,768 valid chunks, we identify 1,591 chunks in cases about women’s issues (approximately 2.4% of chunks), a small though meaningful subset of the Court’s work.

In Table A1, we present the conditional means of interruption rate (Y) of male and female advocates (T) for women’s issues and other issues (C). Among all other issues (i.e., cases not coded as women’s issues), male advocates are interrupted approximately 10.9 times per 1,000 tokens, whereas female advocates are interrupted approximately 12.4 times per 1,000 tokens. In stark contrast, in chunks from cases about women’s issues the dynamics are almost precisely opposite; male advocates are interrupted approximately 12.5 times per 1,000 tokens, whereas female advocates are interrupted approximately 10.8 times per 1,000 tokens.

Thus, in women’s issues cases—argument settings where female advocates are understood to be speaking from a position of authority—we find that they are significantly less likely to

$E[Y C = \text{Other issue}, T = M]$	10.90
$E[Y C = \text{Gender issue}, T = M]$	12.47
$E[Y C = \text{Other issue}, T = F]$	12.35
$E[Y C = \text{Gender issue}, T = F]$	10.82

Table A1: Conditional means of interruption, Y given whether the case topic is about gender issues or not (C). Here the number of chunks in which the case topic is a $C = \text{Gender issue}$ is only 1,591.

be interrupted than male advocates. Beyond the difference in interruption behavior by the justices, women are also more likely to represent parties in cases about women’s issues, with the odds ratio between T and C equal to 1.989. Considering these dynamics in conjunction with the very small number of chunks (and cases) that deal with women’s issues, we exclude women’s issues cases from subsequent analyses. Nevertheless, we hasten to emphasize that the difference we observe in gendered interruption rates between women’s issues and other issues highlights the ability of justices to behave differently towards advocates of different genders.

C Confidence Intervals

We use a non-parametric bootstrap method (Wasserman 2004, ch. 8) to infer confidence intervals for the per-justice interruption rates, effect of gender, and the effect of ideological alignment. We first subset chunks by justice, and then we utilize 10^3 bootstrap samples of chunks for each justice and report 95% bootstrap confidence intervals (CI) via the normal interval method (Wasserman 2004, ch. 8.3). To obtain the confidence intervals via this normal method for the results aggregated across all justices (or without loss of generality for the subset of male or female justices), we assume each of the n justices is independent and calculate $\frac{1}{n}\sqrt{\sum_{j=1}^n \sigma_j^2}$ where σ_j^2 is the variance of the bootstrap distribution for justice j .

D Justice-Level Estimates

Here we present a table corresponding to the estimates presented graphically in Figure 4 and Figure 5. The results are based on the formulas and details presented in the Research Design section of the manuscript. As we note there, our estimates are intentionally non-parametric and simplify to the conditional means from the data.

Justice	Interruption Rate	Gender Effect	Ideological Alignment Effect	N (Chunks)
Sandra Day O'Connor	10.15 (0.18)	2.18 (0.73)	-0.35 (0.35)	4847
John G. Roberts Jr.	9.71 (0.21)	2.11 (0.64)	0.64 (0.43)	2996
Anthony M. Kennedy	12.27 (0.18)	2.00 (0.60)	-0.72 (0.36)	5383
Antonin Scalia	13.07 (0.15)	1.65 (0.47)	0.67 (0.31)	11031
David H. Souter	12.76 (0.22)	1.5 (0.73)	-1.36 (0.47)	3633
Ruth Bader Ginsburg	9.22 (0.15)	1.16 (0.51)	0.18 (0.30)	5338
William H. Rehnquist	5.8 (0.12)	1.09 (0.47)	-0.48 (0.24)	5871
Stephen G. Breyer	16.8 (0.23)	1.08 (0.72)	0.18 (0.47)	5196
Samuel A. Alito Jr.	8.67 (0.26)	1.07 (0.64)	0.49 (0.6)	1807
Byron R. White	10.28 (0.16)	0.8 (0.65)	-0.44 (0.34)	4764
John Paul Stevens	10.47 (0.14)	0.58 (0.44)	-0.64 (0.28)	7041
Warren E. Burger	5.96 (0.28)	-0.02 (1.06)	-0.94 (0.56)	1005
Thurgood Marshall	11.2 (0.33)	-0.17 (1.13)	0.42 (0.66)	1300
Sonia Sotomayor	13.17 (0.25)	-0.57 (0.67)	-1.2 (0.51)	2496
Elena Kagan	8.98 (0.29)	-1.04 (0.92)	-0.18 (0.61)	1456

Table A2: *Table of Justice Interruption Rates, Effect of Gender, and Effect of Ideological Alignment.* Standard errors appear in parentheses, and are derived from a nonparametric bootstrap method, described in detail on Page 27.

E Heatedness of Discussions

While we focus on the justice interruptions of attorneys, the opposite does occur at the Court. As such, one potential concern is that male advocates interrupting female justices more frequently may lead to, on average, fewer opportunities for female justices to interrupt male advocates, or may actually lead to the reverse—a more heated discussion in which interruptions escalate. To examine this dynamic further, we look at the heatedness of conversations, or the extent to which justices and advocates are interrupting one another within chunks. The underlying idea is that if male advocate interruptions of female justices are decreasing opportunities for interruptions, then we might expect a negative relationship between advocate interruptions and just interruptions. On the other hand, one might also be concerned that the interruption of the female justice escalates interruption behavior, leading to more heated discussions in which the interruptions (by justice of advocate, and by advocate of justice) escalate. In either case, we would expect strong correlations between advocate and justice interruptions, with particular concern for differences across gender pairings.

We plot the results in Figure A1. Across gender groupings, we find no evidence that advocate interruptions lead to fewer opportunities for justice’s interrupting advocates; indeed, we primarily find only that as interruptions very loosely go hand-in-hand, with correlation coefficients consistently around 0.2, suggesting increases in advocate interruptions of justices are very slightly positively correlated with increases in justice interruptions of advocates across the different combinations of gender by judge and advocate.

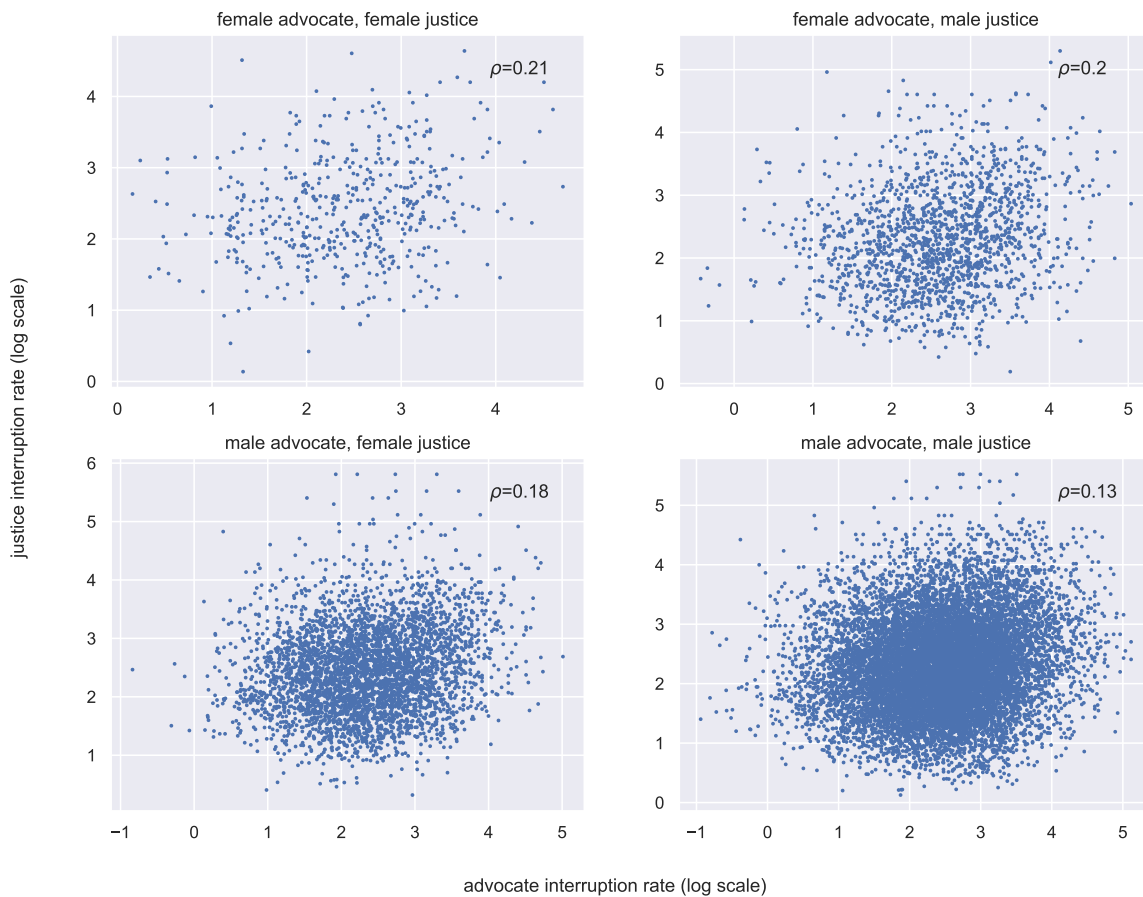


Figure A1: Heatedness: the relationship between advocates being interrupted by justices (x-axis) versus justices being interrupted by advocates (y-axis). Each blue dot is a chunk from a case. We put both axes in log scale and report the pearson correlation coefficient ρ .

F Back Channel Cues

Previous research (Gravano and Hirschberg 2009; Ruede, Müller, Stüker and Waibel 2017; Wang, Cheung, Zhang, Yang, Chen, Fu and Ngai 2023) suggests that certain short, phrasal *back channel cues*, such as “Right,” “Okay,” “I see”, may be a form a conversational maintenance and provide feedback from one speaker to the other. In the context of oral arguments, utterances in which a justice interrupts and advocate using a back channel cue may have a substantively different meaning than other interrupting utterances.

For a robustness check, we re-ran the entire analysis pipeline with the following change in our definition of a *valid chunk* (unbolded steps existed previously and changes are bolded). We operationalize a *valid chunk* as:

1. Four or more contiguous utterances in which there are exactly two speakers—a single justice and a single advocate, and
2. The advocate makes the first utterance, and
3. Each speaker has two or more **valid utterances**.
4. ***Invalid utterances:*** We discard an utterance from either a justice or an advocate if the entire utterance (after being lower-cased and stripped of ending punctuation) matches any of the following 18 phrasal back channel cues:
 - (a) right
 - (b) yes
 - (c) yeah
 - (d) okay
 - (e) ok
 - (f) i know

- (g) sure
- (h) uh-huh
- (i) uh huh
- (j) huh
- (k) really
- (l) true
- (m) that's true
- (n) that's right
- (o) all right
- (p) i see
- (q) go on
- (r) absolutely

After this change, many of our valid chunks remained intact. For example, if previously a valid chunk had six utterances and we discarded one utterance as invalid due to the presence of a back channel cue, the chunk would still meet our minimum utterance threshold and be a valid chunk.

Our approach to discard back channel cues is high precision and low recall. It is high precision because we believe each of the 18 words/phrases above to be a true positive back channel cue. However, it is low recall because there may be back channel cues we are missing. However, a different approach (such as discarding all utterances shorter than a particular length) may result in discarding substantively relevant utterances. In exploratory manual analysis, we found many short justice utterances (in which they interrupted an advocate) to be substantively meaningful. Take the following three examples of a justice (J) interrupting and advocate (A):

A: It is obviously interconnecting separately owned or managed buildings by means of a piece of cable will be cheaper in all instances for any–

J: Virtually costless.

A: –There is not a procedural bar–

J: Why is that?

A: –You have to agree–

J: Pick one.

Using the description of invalid chunks above, we discarded 19,407 utterances that matched the back channel cue list. This results in 63,472 valid chunks, which contrasts with the 64,164 valid chunks in the pipeline that does not remove back channel cues. Here are the aggregate results before and after discarding utterances that match the back channel cues:

Justices	θ_{Gender}	$\theta_{\text{Ideological Alignment}}$	$\frac{\theta_{\text{Gender}}}{\theta_{\text{Ideological Alignment}}}$
All	0.89±0.36	-0.25±0.23	3.59
Male	1.06±0.43	-0.20±0.26	5.34
Female	0.43±0.71	-0.39±0.45	1.12

Table A3: *Effects on advocate interruption rate, aggregated by justice gender including* utterances with back channel cues (same as in main paper).

Justices	θ_{Gender}	$\theta_{\text{Ideological Alignment}}$	$\frac{\theta_{\text{Gender}}}{\theta_{\text{Ideological Alignment}}}$
All	0.96±0.36	-0.24±0.23	3.99
Male	1.12±0.43	-0.21±0.26	5.22
Female	0.54±0.68	-0.32±0.46	1.70

Table A4: *Effects on advocate interruption rate, aggregated by justice gender excluding* utterances with back channel cues.

Comparing Tables A3 and A4, we see there are small changes in the point estimates of effect magnitudes. However, the direction and ordering of the all effects remains consistent before and after the back channel cue change, reassuring us of the robustness of our results.

References

- Gravano, Agustín and Julia Hirschberg. 2009. Backchannel-inviting cues in task-oriented dialogue. In *Tenth Annual Conference of the International Speech Communication Association*.
- Miller, Michael G. and Joseph L. Sutherland. 2022. “The Effect of Gender on Interruptions at Congressional Hearings.” *American Political Science Review* .
- Patton, Dana and Joseph Smith. 2017. “Lawyer, Interrupted: Gender Bias in Oral Arguments at the U.S. Supreme Court”.” *Journal of Law and Courts* pp. 337–361.
- Raffo, Julio and Gema Lax-Martinez. 2018. “WGND 1.0.”
URL: <https://doi.org/10.7910/DVN/YPRQH8>
- Ruede, Robin, Markus Müller, Sebastian Stüker and Alex Waibel. 2017. Enhancing Backchannel Prediction Using Word Embeddings. In *Interspeech*. pp. 879–883.
- Szmer, John J., Tammy A. Sarver and Erin B. Kaheny. 2010. “Have We Come a Long Way, Baby? The Influence of Lawyer Gender on Supreme Court Decision Making.” *Politics and Gender* 6:1–36.
- Wang, Kangzhong, MK Michael Cheung, Youqian Zhang, Chunxi Yang, Peter Q Chen, Eugene Yujun Fu and Grace Ngai. 2023. Unveiling Subtle Cues: Backchannel Detection Using Temporal Multimodal Attention Networks. In *Proceedings of the 31st ACM International Conference on Multimedia*. pp. 9586–9590.
- Wasserman, Larry. 2004. *All of statistics: a concise course in statistical inference*. Vol. 26 Springer.