Gender Attitudes in the Judiciary: Evidence from U.S. Circuit Courts

Elliott Ash, ETH Zurich
Daniel L. Chen, Toulouse School of Economics
Arianna Ornaghi, Hertie School*

April 21, 2022

Abstract

Do gender attitudes influence interactions with female judges in U.S. Circuit Courts? In this paper, we propose a judge-specific measure of gender attitudes based on use of gender-stereotyped language in the judge's authored opinions. Exploiting quasi-random assignment of judges to cases and conditioning on judges' characteristics, we validate the measure showing that higher-slant judges vote more conservatively in gender-related cases. Higher-slant judges interact differently with female colleagues: they are more likely to reverse lower-court decisions if the lower-court judge is a woman than a man, are less likely to assign opinions to female judges, and cite fewer female-authored opinions.

JEL Codes: J70, J16

Keywords: Gender attitudes, judiciary, stereotypes, NLP

*Elliott Ash, ETH Zurich, ashe@ethz.ch; Daniel Chen, Toulouse School of Economics, daniel.chen@iast.fr; Arianna Ornaghi, University of Warwick, ornaghi@hertie-school.org (corresponding author). We thank Jacopo Bregolin, David Cai, Christoph Goessmann, Ornelie Manzambi, and Seraina Zuber for helpful research assistance. We thank Ryan Copus, Arthur Dyevre, Mirko Draca, Patrick Kline, Emily Leslie, David Mimno, Amanda Pallais, and Evan Soltas as well as conference and seminar participants for helpful comments and suggestions. Funding from the Russell Sage Foundation is gratefully acknowledged.

1 Introduction

Women are underrepresented at the top of the legal profession. Although since the 1990s close to 45% of law school graduates have been female (Croft, 2016), women still compose only 20% of equity partners in large law firms (NAWL, 2019) and 30% of state and federal judgeships (George and Yoon, 2019; Root, 2019). In this paper, we focus on U.S. Circuits Courts of Appeal and explore a possible explanation for this disparity: differential treatment of female judges on the part of their colleagues due to variation in gender attitudes.

Attitudes toward social groups—most notably women and racial minorities—are highly predictive of judgments and choices (Bertrand, Chugh and Mullainathan, 2005). Attitudes matter even in high-stakes settings such as physician treatments (Green et al., 2007), voting (Friese, Bluemke and Wänke, 2007), hiring decisions (Rooth, 2010), employer-employee interactions (Glover, Pallais and Pariente, 2017), and teacher effectiveness (Carlana, 2019). If gender attitudes imply differential treatment of female judges as well, they might play a role in explaining the representation gap of women in the judiciary.

The major challenge to investigating these issues is that the measures of gender attitudes traditionally used in the social sciences, such as Implicit Association Tests, are not available for circuit court judges.[1] We address this challenge by proposing a novel measure of gender attitudes that exploits a unique feature of our setting—the large corpus of written text that is available for appellate judges—and the idea that text can provide important insights into human social psychology (Jakiela and Ozier, 2019). In particular, we draw on recent developments in natural language processing (NLP) and proxy judges' attitudes toward gender by measuring use of gender-stereotyped language in their writing (Pennington, Socher and Manning, 2014; Caliskan, Bryson and Narayanan, 2017; Kozlowski, Taddy and Evans, 2019; Antoniak and Mimno, 2018). That is, we develop a measure of *gender slant* based on how strongly judges associate men with careers and women with families in the opinions that they write.

The key NLP technology powering our approach, word embeddings, is an algorithm that distributes words in a vector space based on their co-occurrence in a corpus (Mikolov et al., 2013; Pennington, Socher and Manning, 2014). We use word embeddings because

[1] To the best of our knowledge, there exist only two papers that have collected IAT scores for judges, and neither of them links the scores to the real-world behavior of these judges. Rachlinski et al. (2009) measure implicit bias against African-Americans for 133 state and local trial judges, finding some evidence that more biased judges prefer harsher sentences when primed with hypothetical cases about African-Americans in a vignette experiment. Levinson, Bennett and Hioki (2017) collect implicit bias scores against Asians and Jews for 239 federal and state judges, and find that judges biased against Jews display different hypothetical sentencing behavior.

they are a language representation that preserves semantic relationships. First, words with similar meaning have similar representations: their vectors are close together in the space. But in addition to this, the relative position of word vectors also conveys meaning. For example, male and female words (e.g. $\overrightarrow{man}, \overrightarrow{woman}$ or $\overrightarrow{king}, \overrightarrow{queen}$) tend to hold similar relative positions to each other, and as a result we can identify a gender dimension in the space (equivalent to taking a step in the "male" direction) by taking the vector difference between male and female words. More generally, dimensions induced by word vector differences can be used to identify cultural concepts (Kozlowski, Taddy and Evans, 2019).

We operationalize these ideas to identify gender attitudes in language. If men are associated with career and women are associated with family, the relative position of male-to-female words should be similar to the relative position of career-to-family words. That is, the gender dimension ($\overrightarrow{male} - \overrightarrow{female}$) should be similar to the dimension representing stereotypical gender roles ($\overrightarrow{career} - \overrightarrow{family}$). More precisely, we measure the intensity of gender attitudes by looking at the cosine similarity between the gender dimension ($\overrightarrow{male} - \overrightarrow{female}$) and the stereotypical dimension ($\overrightarrow{career} - \overrightarrow{family}$). If the cosine similarity between the two dimensions is high, the corpus uses stereotyped language: men are associated with career, while women are associated with family. If the correlation is around zero, stereotyped language is not present in the text.

In the empirical analysis, we exploit the relative strength of the association to proxy for differential gender attitudes of judges. To construct a judge-specific gender slant measure, we consider the majority opinions authored by a given judge as a separate corpus. We then train embeddings for each judge, which allows us to calculate a judge-specific gender slant measure. To ensure that we obtain a high-quality representation notwithstanding the potentially small size of a judge's corpus, we train embeddings on different bootstrap samples following Antoniak and Mimno (2018) and define gender slant as the median value of the measure across the samples. Using this method, we calculate the gender slant of 139 circuit judges.

As a measure of stereotyped gender attitudes, gender slant behaves convincingly in a number of validation checks. Descriptively, we find that female and younger judges display lower gender slant and that having a daughter reduces gender slant. Getting into the language, we find that lower gender slant is associated with more frequent use of gender-neutral pronoun constructions, such as "he or she". Finally, in a human annotation exercise performed by a reader with legal training, we find that judges with higher slant tend to express less empathy towards women in their writing.

This approach to measuring gender attitudes is related to a growing literature using word embeddings to analyze bias in text. Bolukbasi et al. (2016) and Caliskan, Bryson and Narayanan (2017) demonstrate biases in general web corpora in terms of gender and a number of other dimensions. Garg et al. (2018) train separate embeddings by decade using the Google Books corpus and show that gender associations track demographic and occupational shifts. Two recent papers apply word embeddings to the judicial setting: Rice, Rhodes and Nteta (2019) detect racial slant in a corpus that includes U.S. circuit court opinions, and Ash, Chen and Galletta (2022) document sentiment towards social groups in a similar setting.[2] We build on this literature to construct author-specific measures of gender slant that we link to real-world behaviors.

There exist of course a number of alternative measures we could have used to proxy for these judges' gender attitudes. First, we could have asked human evaluators to qualitatively score the writing of judges. In addition to potential concerns regarding the subjectivity of the coding, the main problem with this approach is its prohibitive cost, considering that our corpus includes over 14 million sentences. Second, we could have proxied for gender attitudes by looking at judges' rulings or votes. While in principle this is indeed a reasonable approach, the small number of gender-related votes for which this information is available means that voting rates in gender decisions are unlikely to produce an empirically useful measure of gender attitudes for circuit judges.

The central research question of the paper is whether judges with different gender slant interact with female judges differently. Our empirical strategy relies on the fact that, within each circuit and year, judges are quasi-randomly assigned to cases. This ensures that higher-slant judges do not self-select into cases systematically based on the expected case outcome: cases assigned to higher- and lower-slant judges are comparable. In addition to this, we condition on detailed judges' characteristics to provide supportive evidence that our strategy identifies the effect of judges having different levels of slant, and not judges differing along some other characteristic such as gender or conservative ideology.

[2] Ash, Chen and Galletta (2022) measure text-based sentiment (positive/warm vs negative/cold) toward nineteen social groups (e.g. Black, White, women, Catholics, businesses, etc.) in U.S. circuit court opinions. The sentiment measure is based on embedding models broadly related to the ones used in this paper, although the implementation is different. Specifically, the authors use Doc2Vec (Le and Mikolov, 2014), an algorithm to represent documents in an embedding space, and apply it to each sentence appearing in the opinions. For each sentence, they compute the proximity of the vector representing the sentence to a vector representing positive/negative sentiment and the proximity to vectors representing the different social groups. They obtain each opinion's sentiment towards each social group by taking the average of the sentence-level sentiment weighted by the sentence's proximity to the group. This methodology differs from ours in the use of different embedding models (a single embedding model that allows for paragraph representation, versus different embedding models for each judge as we do here) and in the objective of creating a case-level sentiment measure rather than understanding judge-level attitudes.

We study whether higher-slant judges treat female judges differently focusing on three sets of interactions: reversals of district court decisions, opinion assignment, and citations. We find evidence supporting this hypothesis. First, we show using a differences-in-differences design that higher-slant judges are more likely to vote to reverse lower-court decisions authored by female district judges with respect to those authored by male district judges. The magnitude of the effect is sizable: being assigned a judge with a one standard deviation higher gender slant increases the probability that a female relative to a male district judge is reversed by 1 percentage point, which corresponds to around 5% of the outcome mean. Second, we consider the decision of a senior judge tasked with assigning the writing of the majority opinion to a panel member. Assigning judges with a one standard deviation higher gender slant are less likely to assign opinion authorship to a female judge by 1.7 percentage points, or 4.5% of the outcome mean. Finally, higher-slant judges are also less likely to cite opinions of female judges, although identification is weaker and the result is less robust. To the extent that these outcomes are relevant for future opportunities, interacting with judges with different gender attitudes as measured by gender slant has the potential to hinder the career progression of female relative to male judges.

To strengthen the interpretation that gender slant is a proxy for gender attitudes, we validate the measure by studying whether higher-slant judges take different decisions in gender-related cases. Consistent with the proposed interpretation, we find that judges with higher slant tend to vote more conservatively in gender-related issues (that is, against expanding women's rights). The magnitude of the effect is large: being assigned a judge with a one standard deviation higher gender slant increases the probability of voting against expanding women's rights by 4.2 percentage points, which corresponds to a 7% increase over the outcome mean. In addition to providing a validation of the measure, this is an interesting result per se, as it shows that by affecting how precedent is set, gender attitudes have the potential to impact real-word outcomes even outside the judiciary (Chen and Sethi, 2018; Chen and Yeh, 2014*c*,*a*).

Finally, we investigate whether higher-slant judges also respond differently to non-gender characteristics. While we do find that higher-slant judges tend to vote more conservatively in non-gender-related but ideologically divisive cases, the effect of being assigned a higher-slant judge is smaller than in gender-related cases. In addition, we find that gender slant has little to no effect on how judges interact with colleagues who were appointed by a Democratic President, who are minorities, or who are in a different age group. Overall, these results support the view that gender slant captures attitudes that are specific to gender.

If we think of courts as a workplace, this paper speaks to the literature on how gender shapes the labor market outcomes of women and why (see among others Bohren, Imas and Rosenberg, 2019; Bordalo et al. (2019); Card et al., 2019; Hengel, Forthcoming; Sarsons, 2019), in particular for women employed at the top end of the earnings distribution (Bursztyn, Fujiwara and Pallais, 2017; Bertrand, 2013; Bertrand, Goldin and Katz, 2010). Despite the richness of the data, the setting we study is quite novel: there is a scarcity of existing work that takes this approach toward the courts, in particular to study the potential for gender discrimination toward female judges.[3] In addition, we contribute to this literature by providing evidence that gender attitudes might play a direct role in determining differential labor market outcomes for men and women, even in high-stakes environments such as appellate courts.

More generally, this paper contributes to the growing literature demonstrating the importance of attitudes in decision making. For example, Glover, Pallais and Pariente (2017) show that attitudes regarding minorities influence manager-employee interactions in a way that impacts performance, and Carlana (2019) shows that teachers with stereotypical views of gender negatively impact the test scores and future scholastic careers of female students. Our novel text-based measure of gender attitudes allows us to study the role these attitudes play in determining the behavior of high-skilled professionals, who might otherwise be hard to reach through surveys or tests traditionally used in the literature.

In addition, by studying the effect of being assigned judges with different levels of slant on voting, this paper builds on the literature on the determinants of judicial decisions, which broadly shows that demographic and ideological characteristics of judges matter (Boyd and Spriggs II, 2009; Kastellec, 2013; Sunstein et al., 2006; Cohen and Yang, 2019). Recent work has shown that in criminal cases, judges systematically demonstrate racial (Arnold, Dobbie and Yang, 2018; Rehavi and Starr, 2014) and gender (Starr, 2014) disparities in their sentencing decisions. Instead of inferring bias from the decisions themselves, we take a different approach by considering how gender attitudes impact voting. The existing evidence on gender attitudes in the judiciary is limited to studies that look at the correlation between attitudes measured by Implicit Association Tests and decisions in hypothetical scenarios (Rachlinski et al., 2009; Levinson, Bennett and Hioki, 2017). Instead, our measure allows us to link gender attitudes to real world outcomes.

The remainder of the paper is organized as follows. Section 2 describes our measure of gender slant, and Section 3 discusses the empirical strategy and presents additional

[3] The one exception is contemporaneous work by Battaglini, Harris and Patacchini (Forthcoming), who show that random exposure to female judges on panels increases the probability of hiring a female clerk.

data sources. Section 4 provides descriptive statistics. Section 5 shows the relationship between higher-slant judges and decisions, while Section 6 describes the relationship between higher-slant judges and interactions with female judges. Finally, Section 7 concludes.

2 Gender Slant: Measuring Gender Attitudes in Text

We begin by constructing a measure of how gender is characterized in the language used by judges. Specifically, our measure aims to capture the strength of the association between gender identifiers and stereotypical gender roles, with men being more career-oriented and women being more strongly associated to family.

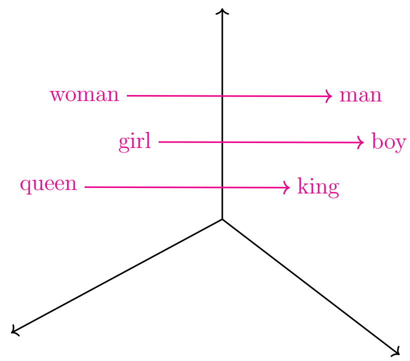
2.1 Word Embeddings

In order to construct our measure, we use word embeddings, a language modeling technique from NLP that relies on word co-occurrence to create a representation in a low dimensional Euclidean space in such a way as to preserve semantic meaning (Mikolov et al., 2013; Pennington, Socher and Manning, 2014).

Consider the simplest way of representing language. For a given vocabulary V, one possibility is to represent words as one-hot-encoded vectors, which are vectors with all values equal to 0 except the one entry corresponding to the word itself. This approach presents two issues. First, the dimensionality of the vector space grows linearly in the size of the vocabulary, as the one-hot-encoded vectors are V-dimensional by construction. Second, it is impossible to infer anything about the relationship between words in the resulting space: all word vectors are orthogonal to each other. Word embeddings offer a solution to both issues. The word representations are low dimensional—in our case, 300 dimensional—dense vectors which can accommodate large vocabularies and corpora without increasing dimensionality. In addition, the positions of word vectors in the space encode relations between words.

Word embeddings encode semantic meaning in two principal ways. First, distance in the word embedding space conveys semantic similarity between words. This is because the position of a word's representation in the vector space is assigned based on the context the word appears in: words that appear frequently in the similar context have representations close to each other in the space, while words that rarely do have representations that are far apart. Importantly, given that a vector's position is defined based on appearance in given contexts, word embedding distance can identify that two words are similar even if they do not necessarily often appear together, as long as the neighboring words tend to be similar.

Figure 1: Identifying a Gender Dimension in a Vector Space



Notes: The figure exemplifies how the gender dimension can be identified in the vector space based on the difference between vectors representing male and female words.

Second, the relative position of word vectors in the space also conveys meanings. For example, as Figure 1 illustrates, male and female words tend to be in similar relative positions to each other. This means that taking the difference between word vectors representing male and female words can identify a gender dimension, equivalent to taking a step in the male direction.

2.2 GloVe Embeddings Implementation

The specific model we use is Global Vectors for Word Representation (Pennington, Socher and Manning, 2014). GloVe is a weighted least squares model that trains word vectors on global co-occurrence counts. GloVe first computes a global co-occurrence matrix, which reports the number of times two words have occurred within a given context window. It then obtains word vectors $w_i \in \boldsymbol{w}$ to minimize the following objective function:

$$J(\boldsymbol{w}) = \sum_{i,j} f\left(X_{ij}\right)\left(w_i^T w_j - \log\left(X_{ij}\right)\right)^2$$

where X_{ij} is the co-occurrence count between words i and j and $f(\cdot)$ is a weighting function that serves to down-weight particularly frequent words.[4] The objective function $J(\cdot)$ effectively trains the word vectors to minimize the squared difference between the dot product of the vectors representing two words and their empirical co-occurrence in the corpus. Our GloVe implementation minimizes $J(\cdot)$ by stochastic gradient descent.

The two key hyperparameters for GloVe are the dimensionality of the vectors and the

[4] Following the original GloVe paper (Pennington, Socher and Manning, 2014), we use the following weighting function:

$$f(x) = \begin{cases} (x/x_{max})^a & \text{if } x < x_{max} \\ 1 & \text{otherwise} \end{cases}$$

where $a = 3/4$.

window size for computing co-occurence statistics. Previous experiments by NLP researchers suggest that increasing dimensionality beyond 300 has negligible improvements for downstream tasks (Pennington, Socher and Manning, 2014; Rodriguez and Spirling, 2022), so we follow that literature and train 300-dimensional vectors. In turn, we choose a standard window size of 10, a middle ground between shorter windows, which would tend to capture syntactic/functional relations between words, and longer windows, which tend to capture topical relations between words.[5]

A practical feature of GloVe is that the algorithm goes through the full corpus only once in order to build the initial co-occurrence matrix. This feature accounts for the considerable improvements in training time compared to other popular word embeddings algorithms such as word2vec (Mikolov et al., 2013), while obtaining embeddings of comparable quality (Pennington, Socher and Manning, 2014). Given that our approach requires the training of a large number of separate embeddings, this is a particularly attractive feature for our application.

2.3 Word Vectors and Gender Slant

We use word embeddings to identify cultural dimensions in language (Kozlowski, Taddy and Evans, 2019). As mentioned in the previous sub-section, a key feature of word embeddings is that the direction of the difference between word vectors in the space conveys meaning. Consider the vector representing the word $\overrightarrow{man}$ and the vector representing the word $\overrightarrow{woman}$. The vector difference between the two, i.e. the vector identified by $\overrightarrow{man} - \overrightarrow{woman}$, identifies a dimension in the space that corresponds to a step in the male direction.

In practice, this is true for $\overrightarrow{boy} - \overrightarrow{girl}$, $\overrightarrow{he} - \overrightarrow{she}$, and so on: we can identify a gender dimension in the space by taking the difference between the average normalized vector across a set of male words and the average normalized vector across a set of female words:

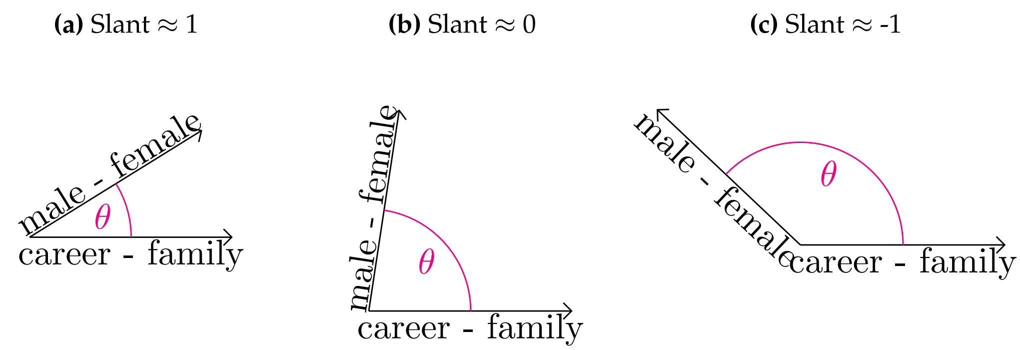
$$\overrightarrow{male} - \overrightarrow{female} = \frac{\sum_n \overrightarrow{male\,word_n}}{|N_{male}|} - \frac{\sum_n \overrightarrow{female\,word_n}}{|N_{female}|},$$

where $|N_{male}|$ is the number of words used to identify the male dimension and $|N_{female}|$ is similarly defined.

A desirable feature of the Euclidean geometry of the vector space and the gender dimension is that other words meaningfully project onto it. It is then possible to understand the connotation of other words along the gender dimension by looking at the cosine of the

[5] Appendix Figure 1 shows that 100- and 300-dimensional embeddings produce highly correlated measures of gender slant. The same is true for embeddings trained using 5, 10, or 15 word windows.

Figure 2: Measuring Gender Attitudes using Cosine Similarity



Notes: The figure exemplifies how gender slant varies depending on the relative position of the gender and the career-family dimensions in the vector space.

angle between the vector representing the word and the dimension itself. Formally, we use the cosine similarity, defined as:

$$sim\left(\vec{x}, \vec{y}\right) = cos\left(\theta\right) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \, \|\vec{y}\|} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}},$$

where $\vec{x}$ and $\vec{y}$ are non-zero vectors, θ is the associated angle, and $\|\cdot\|$ is the 2-norm. Note that $sim\left(\vec{x}, \vec{y}\right)$ varies between -1 and +1. Continuing with the example, words with male (female) connotations—e.g., male (female) first names—are going to be positively (negatively) correlated with the gender dimension defined by $\overrightarrow{male} - \overrightarrow{female}$.

We put these dimensions in service of constructing a gender slant measure. The goal is to capture the strength of the association between gender and stereotypical attitudes identifying men more closely with career, and women with family. Word embeddings provide an intuitive metric. If men are associated with career and women are associated with family in the corpus, the relative positions of male-to-female words should be similar to the relative positions of career-to-family words. Specifically, we use the cosine similarity between the vector representing the gender dimension, defined by $\overrightarrow{male} - \overrightarrow{female}$, and the vector representing the career-family dimension, defined by $\overrightarrow{career} - \overrightarrow{family}$. The statistical similarity summarizes how closely related the gender and stereotypical dimension are in the space, as illustrated in Figure 2. When the two concepts are strongly associated in a corpus, the two vectors are close together ($\theta \approx 0$), and the slant measure is close to 1 (Figure 2 (a)). If there is no association between the two, then $\theta \approx 90°$, and the slant measure is 0 (Figure 2 (b)). Finally, if the concepts are negatively associated in a corpus (e.g. male is associated to family and female is associated to career), the two vectors are far apart ($\theta \approx 180°$), and the slant measure will tend to -1 (Figure 2 (c)).

For this task, there are many potential combinations of male, female, career, and family

Table 1: Word Sets

Male	his, he, him, mr, himself, man, men, king, male, fellow
Female	her, she, ms, women, woman, female, herself, girl, girls, queen
Career	company, inc, work, business, service, pay, corp, employee, employment, benefits
Family	family, wife, husband, mother, father, parents, son, brother, parent, brothers

words that could be used to identify the gender and career-family dimension. We select the word sets to identify the dimensions using the following procedure. First, we identify potential word sets using Linguistic Inquiry and Word Count Dictionaries, which provide a human-validated list of words and word stems that correspond to certain concepts. We use the word sets for male, female, work and family. From these word sets, we eliminate words that could be ambiguous or have specific legal meanings in our setting (e.g. tribe, tribes for family; line, situation, trade for work). From each list, we then select the ten most frequent words in the full judicial corpus.[6] Table 1 reports the resulting word sets. In addition, we show how the results change when selecting the top five to top fifteen most frequent words and when dropping one word at a time in the career versus family dimension.

We focus on the stereotypical association between men and career versus women and family as opposed to other associations typically studied in the literature—for example, the association between men and science versus women and arts, and between men and positive attributes versus women and negative attributes—for the following reasons. First, words related to sciences and arts do not frequently appear in the corpus, which makes it difficult to identify the science/art dimension in the space (see Appendix Table 1). Second, we only find limited evidence that the full judicial corpus presents a stereotypical association between men and positive attributes versus women and negative attributes, which suggests limited scope for this measure to capture judge-specific variation in gender attitudes. In fact, Appendix Figure 2 shows that in the full judicial corpus, the cosine similarity between the gender and the positive-negative dimension is generally smaller and closer to zero with respect to the cosine similarity between the gender and the career-family dimension. When we perform a permutation test following Caliskan, Bryson and Narayanan (2017), we indeed find p-values lower than 0.10 in only three of the twenty-four bootstrapped embeddings trained on the full judicial corpus.

[6] We select words using these procedures as opposed to the word sets usually used in the gender-career IAT because we want to ensure that we are using words that meaningfully define gender in judicial language. For example, first names are rarely used in legal language as opposed to gender pronouns, which would introduce substantial noise in the slant measure.

2.4 Measuring Judge Gender Slant

Our goal is to produce measures of gender attitudes in the writing of circuit court judges. Our starting corpus is the universe of published opinions in circuit courts for the years 1890-2013, which consists of 380,000 opinions in thirteen courts from Bloomberg Law.

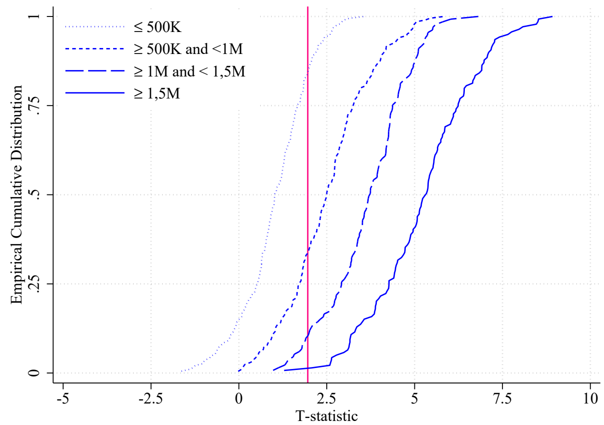
As a pre-processing step, we clean that text and exclude punctuation and numbers, although we retain hyphenated words. To avoid the word vectors being case sensitive, we transform all words to be lower cased. We then retain only the most common 50,000 words in all judicial opinions. Opinions are separated into sentences using punctuation, and each sentence is further tokenized into words. These tokenized sentences are the starting point of the model.

To obtain judge-specific gender slant measures, we take the set of majority opinions authored by each judge as a separate corpus. We train separate GloVe embeddings on each judge's corpus, and we used the resulting vectors to compute the gender slant measure as described in Subsection 2.3. To ensure convergence, we train vectors for 20 iterations with a learning rate of 0.05.

Creating judge-specific embeddings implies the use of relatively small corpora, which is potentially a problem given that word embeddings perform best when they are trained on large collections. To address this issue, we follow the approach suggested by Antoniak and Mimno (2018) and train embedding models on twenty-five bootstrap samples of each judge corpus. Specifically, we consider each sentence written by a judge as a document, and then create a corpus by sampling with replacement from all sentences. The number of sentences contained in the bootstrapped sample is the same as the total number of sentences in the original judge corpus. We then calculate our measure for all bootstrap samples and assign to each judge the median value of the measure across the samples. Given that embeddings trained on small corpora tend to be sensitive to the inclusion of specific documents, the bootstrap procedure produces more stable results. In addition, bootstrapping ensures stability with respect to the initialization of the word vectors, a potential concern given that GloVe presents a non-convex objective function (Rodriguez and Spirling, 2022). Importantly, this constraint on corpus size is why we are not able to construct time-varying measures of gender slant.

Even following the bootstrapping procedure, we might still worry about the quality of the judge-specific embeddings, and in particular, whether they are able to capture meaningful information about gender. To ensure that the judge-specific embeddings are able to do so, we compute the cosine similarity between the gender dimension and each of the vectors representing the most common 25 male and female names according to the 1990 census

Figure 3: Judge-Specific Word Embeddings Capture Gender Information



Notes: The graph shows that the gender dimension in judge-specific word embeddings captures gender information when the corpus of the judge is sufficiently large. In particular, we test whether male first names have a higher cosine similarity with the gender dimension than female names in the judge-specific embeddings. The graph reports the cumulative distribution of the t-statistics resulting from a series of regressions of an indicator variable equal to 1 if the name is male on the median cosine similarity between the vector representing the name and the gender dimension across bootstrap samples for each judge, by number of tokens included in the judge's corpus. We focus on the most common 25 male and female names according to the 1990 Population Census.

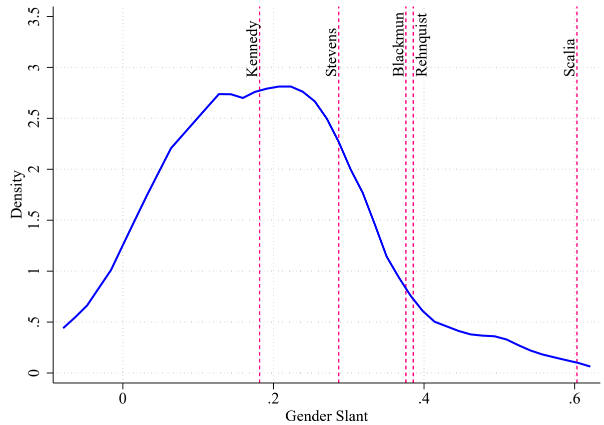
for each judge and bootstrap sample. We then regress a dummy for whether the name is male on the median cosine similarity between the vector representing the name and the gender dimension across bootstrap samples, separately for each judge.

Figure 3 shows the cumulative distribution of the t-statistics resulting from these regressions for sets of judges with different number of tokens. The figure shows that, for judges with a small number of tokens, the t-statistics are rarely above conventional significance level and are at times even negative. As the number of tokens increases, the t-statistics start to become significant and are never lower than zero, showing that the gender dimension identified in the embeddings does indeed contain meaningful gender information. Based on these experiments, all of our main results focus on the 139 judges whose corpus includes at least 1,500,000 tokens.[7] Out of the 139 judges with 1.5 million tokens, 12% are women (17 judges) and 42% were appointed by a Democratic President (58 judges).

The relatively small sample of judges for which we are able to define the measure might raise concerns related to the external validity of the findings. However, the judges in our sample do not appear to be highly selected as compared to other circuit judges: they are not significantly different in terms of party of appointing President, gender, and race, although they are more likely to be born after 1920 (see Appendix Table 4).

[7]For comparison, around 1,000 judges served in circuit courts from 1890-2013. Online Appendix C discusses in detail the robustness of our results to decreasing or increasing the tokens threshold. Consistent with the idea that our measure does not do a good job at capturing gender additudes for corpora that are too small, our results are not consistently robust to decreasing the token threshold below 1,500,000 tokens.

Figure 4: Distribution of Gender Slant



Notes: The graph shows the distribution of the gender slant measure for the 139 judges in the sample. The vertical lines indicate the gender slant of five Supreme Court justices who were appointed after 1970 and have a corpus of at least 1,000,000 tokens. The gender slant of Supreme Court justices is measured by training embeddings on the Supreme Court majority opinions that were authored by them. Gender slant is the cosine similarity between the gender and the career-family dimensions.

2.5 Variation in Judge Gender Slant

Figure 4 shows the distribution of gender slant across judges in the sample. Gender slant is almost always positive, reflecting that most judges display a stereotypical association between men and career and between women and family. However, there is significant variation in how strong this association is. Our empirical analysis leverages this variation to relate judges with higher or lower gender slant to judicial outcomes.[8]

To put the measure in perspective, the vertical lines on the same graph show the degree of gender slant for a sample of recent U.S. Supreme Court justices for which we were able to measure gender slant. Interestingly, the conservative Justice Antonin Scalia has the highest slant of the group, while more liberal judges such as Justice Anthony Kennedy have relatively lower slant.[9]

Where is the variation in gender slant across judges coming from? If judges are writing a legal opinion to support the decision in a case, do they still have latitude in their writing? Judges and legal scholars who have studied this topic have documented qualitative and quantitative differences in judge writing (see Posner, 1995, 2008). In the U.S. court

[8]Because we cannot ground our gender slant measure in an external measure of gender attitudes, we are unable to identify an objectively "correct" gender slant measure in this context. Nonetheless, we believe that our results can still provide important insights on the role of gender attitudes on judicial behavior exploiting the variation across judges in the relative strength of the stereotypical association. Because the units of the measure are not easily interpretable, we standardize it in the empirical analysis.

[9]The gender slant of Supreme Court justices is measured by training embeddings on the U.S. Supreme Court majority opinions that were authored by these judges. Finding judges with corpora of sufficient size is even more difficult in this setting: the figure shows the gender slant of justices with at least 1,000,000 tokens in their corpus. Unfortunately, none of the female justices (e.g. Ginsburg and O'Connor) reached the minimum token threshold.

system, especially, judges tend to express their individual personalities and attitudes in their opinions. These attitudes come out not just in the ruling and the legal arguments in support, but also, importantly, in how the facts and moral and policy issues are framed.

In particular, conditional on the facts and legal boundaries, a judge's choice of words can be expressive of attitudes towards gender. A familar example would be the choice of pronouns when discussing hypotheticals. A judge could either default to the masculine pronoun ("he"), as is the historical convention, or they could use a feminine pronoun ("she"), or they could use a gender-neutral expression (e.g. "he or she"). The latter, especially, tends to express commitment to gender-equality norms. Consistent with this intuition, Appendix Figure 3 shows that higher-slant judges are less likely to use gender-neutral pronoun constructions.

To add texture to these quantitative metrics, we asked a law student to read a selection of judicial opinions and annotate examples of passages demonstrating empathy for women. Appendix Table 2 provides a selection of these annotated snippets. They are interesting to read and illustrate how the personalities of judges are expressed in their opinions. The passages are notably expressive and not just bland technical writing, offering some qualitative evidence in our setting for judicial discretion in writing style.

2.6 Discussion of Alternative Measures

There are (perhaps many) other potential approaches to measure gender attitudes in judicial language. First, we could have had human evaluators qualitatively score the writing of judges. These coders could have done a deep reading of the text or a subjective coding of important themes (Glaser and Strauss, 2017). This qualitative approach is somewhat subjective and therefore lacks a rigorous method of replication (Ricoeur, 1981; DiMaggio, 1997). The binding constraint to the annotation approach, however, is the cost: it would be prohibitively expensive to implement on a large scale such as we have in our corpus, which contains over 14 million sentences. In this sense, our automated approach can be seen as a way of efficiently extracting information regarding how gender is talked about in large corpora.

To compare our approach to human coding, we tasked a law student with the following annotation exercise. The annotator read two randomly paired judicial opinions and assessed which opinion demonstrated more empathy towards women. To ensure that gender was a relevant element of the case, and to hold the direction of the ruling constant, the pairs were selected among gender-related cases that were decided in favor of expanding women's rights.[10] The pairs contained a higher-slant judge's opinion and a lower-slant judge's opinion (the top/bottom 10 percent of the slant distribution, to maximize power). Consistent with the idea that gender slant captures gender attitudes that are expressed in text, Appendix Figure 4 shows that opinions authored by higher-slant judges are less likely to be identified as having more empathic language. In addition, Appendix Figure 5 shows that they are less likely to include at least one empathic statement. Even among cases that are on a close topic (i.e. gender-related cases) and that are decided in the same direction (i.e. in favor of expanding women's rights), there is significant variation in the language that judges with higher and lower slant use.

While this annotation exercise provides some reassurance about our embedding-based measure, it also illustrates the infeasibility of hand-coding a comparable dataset. The law student needed about 30 minutes to annotate each case. At $30 per hour, it would be prohibitively expensive to annotate a sufficiently large corpus to do a comparable empirical analysis.

A second alternative to our embedding-based measure of gender attitudes would be to construct a count-based measure based on word or phrase frequencies. A count-based measure has intuitive appeal given that the language representation provided by word embeddings is itself built on proximate co-occurrence of words. Thus, co-occurrence statistics of male/female words with career/family words can be thought of as the building blocks of the gender slant measure. Consistent with this intuition, the relative co-occurrence of male words with career words relative to the co-occurrence of female words with family words is positively associated with the gender slant measure (Appendix Figure 6). In Appendix Table 3, we report ten randomly selected sentences that present these co-occurrences for reference.

Looking at the raw counts, however, we see that these explicit word co-occurrences are rare. For example, the median judge writes only 38 sentences where female and family words co-occur. This sparsity in the count-based measure is why a word embedding approach is preferrable, as embeddings do not require words to appear directly next to each other to register an association, and they are not strictly bound to the specified lexicon. As a result, embeddings are able to take into account more information contained in judges' writing. Instead, the count-based approach misses implicit and nuanced gender-stereotyped language and as a result provides a less precise measure.

On a different tack, it is also worth asking whether a linguistic analysis is needed at all. Why not just measure gender attitudes from the judge's rulings or votes? After all, in

[10]Section 3 and 5 provide more information on gender-related cases.

Section 5 below, we ask whether higher-slant judges vote differently in gender-related cases as a way of validating of our measure. In principle, one could measure gender attitudes by directly computing the average decision direction (progressive or conservative) in gender-related cases. However this approach does not work well in our setting in practice. As discussed further below, we are constrained by our judges showing up in datasets hand-coded for vote valence. The average judge has only 27 gender-decision votes, and many judges have fewer than 10. This sparsity of data points means that voting rates in gender decisions are unlikely to produce an empirically useful measure of gender attitudes.

3 Empirical Strategy and Additional Data Sources

3.1 Empirical Strategy

The main objective of the paper is to study whether judges displaying higher or lower levels of gender slant interact differently with female judges. The empirical strategy relies primarily on the quasi-random assignment of judges to cases, which means that higher- and lower-slant judges do not self-select into cases systematically based on the expected outcome. In addition, we condition on detailed judges' biographic characteristics to ensure that we are not confounding the effect of the judge having higher slant with other judge characteristics such as gender or conservative ideology.

Quasi-Random Assignment. A major concern for identification is the endogeneous selection of judges to cases, as we might worry that higher-slant judges might decide which panels to sit on based on the expected outcome of the case. We are helped in this respect by the fact that in circuit courts, cases are quasi-randomly assigned to panels formed by three judges, which ensures comparability of cases seen by judges with higher and lower gender slant.

Qualitative research based on court documents and interviews with court officials describe the process as follows (Chen and Sethi (2018) and Bowie, Songer and Szmer (2014)). Judges are chosen from a pool of 8-40 judges, depending on the specific circuit. Before oral arguments, available judges (including visiting judges) are assigned to cases following a random process, in recent years using computer programs. In some cases, randomization might be constrained by organizational considerations, such as accommodating judges' schedules. Importantly, in case of conflict of interest, judges might recuse themselves, but they would generally do so before random assignment.

Evidence supporting the quasi-random assignment of judges to cases has been provided by Chen and Sethi (2018), who show for a sample of gender-discrimination cases that pre-trial characteristics are uncorrelated with the demographic composition of the panel, and by Chen, Levonyan and Yeh (2014), who show that the panel composition does not appear to be autocorrelated over time. Chilton and Levy (2015) test whether the ideological composition of panels is indeed random by comparing a simulated panel formation process with actual panels sitting together in different sessions of the court, and they find evidence of non-random assigment for four circuits (namely, the 2nd, 8th, 9th, and D.C. Circuit). Because running their test requires detailed information of court calendars that are not available for the full period we study, we cannot replicate their test looking at the slant composition of panel. Nonetheless, we show that our results are robust to dropping the circuits where non-random assignment is suspected. Most importantly, based on extensive qualitative research, Chilton and Levy (2015) see the lack of random assignment as driven by organizational constraints, such as the need to ensure spacing of judicial assignments or to accommodate vacation schedules, and not by judges specifically selecting into panels, which would be more problematic in our setting.

Here, we provide two pieces of additional evidence supporting quasi-random assignment of judges to cases. First, we test whether higher-slant judges are systematically assigned to cases with different topics (including whether the case is gender-related) or that were originally decided by district judges with different characteristics. We do so by regressing case characteristics on gender slant, judge demographic controls, and circuit-year fixed effects. Appendix Table 5 shows that gender slant does not appear to be systematically correlated with the case characteristics we consider. Although slant is negatively correlated with the district judge being female, this is in line with statistical chance. Consistent with this interpretation, a joint test that the effect of slanted judges fails to reject the null hypothesis (p-value: 0.198).

Second, we test whether the variation in case characteristics across judges that we see in our data is consistent with quasi-random assignment. Following Abrams, Bertrand and Mullainathan (2012), we do this by comparing the variation we see in the data with the one resulting from simulating random assignment. Appendix Figure 7 shows, for each characteristic we consider, the distribution of the interquartile range of the mean characteristic across judges, for 1000 assignment simulations.[11] The vertical line shows

[11] For each characteristic, we estimate the actual interquartile range by taking the average value of the characteristic for each judge across all the cases that the judge is assigned to. Then, we simulate the process of quasi-random assignment by randomly assigning case characteristics to each judge. Because case characteristics vary substantially across circuits and years, we simulate the data to match the mean characteristics in each circuit and year. For each simulated dataset, we compute the average value of the characteristic for each judge and estimate the interquartile range of this value across the judges. We perform this exercise focusing on the 139 judges that are part of our main sample.

Table 2: Descriptive Statistics

	N	Mean	SD	N	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Judges					
Gender Slant	139	0.186	0.132			
Democrat	139	0.417	0.495			
Female	139	0.122	0.329			
Minority	139	0.079	0.271			
Born in 1920s	139	0.252	0.436			
Born in 1930s	139	0.295	0.458			
Born after 1940	139	0.230	0.422			
	Panel B: Cases					
	Analysis Sample			All Observations		
Conservative Vote, Gender-Related Cases	3086	0.605	0.489	5185	0.591	0.492
Conservative Vote, Non Gender-Related Cases	54777	0.569	0.495	9318	0.546	0.498
Voted to Reverse	145862	0.177	0.382	357920	0.203	0.402
Female District Judge	145862	0.120	0.325	357923	0.100	0.300
Author is Female	32052	0.383	0.486	51183	0.389	0.488
Cites at Least One Female Judge	107923	0.383	0.486	257923	0.322	0.467

Notes: This table reports descriptive statistics for judges' characteristics (Panel A) and for the main variables considered in the analysis (Panel B). In Panel B, columns (1) to (3) restrict the sample to observations included in the main analysis, while columns (4) to (6) include all observations.

the true value we observe in the data. Overall, the graphs show that the cross-judge variation we observe in the data is in line with the one we should expect given quasi-random assignment.

We interpret the evidence as supporting the fact that judges are quasi-randomly assigned to cases.

Conditioning on Observable Characteristics. Random assignment of judges to cases implies that the effect of being assigned a higher- or lower-slant judge is well-identified, but we might still worry of this effect being confounded by other judge characteristics such as judge gender or conservative ideology. We address this issue by exploiting detailed information on the demographic characteristics of these judges, which allows us to directly control for other characteristics of the judge that might correlate both with gender slant and with behavior. In Appendix Table 6, we assess this assumption using the method from Oster (2019).

3.2 Additional Data Sources

This paper combines four principal data sources, in addition to the text data used to construct the gender slant measure for each judge described in the previous section. Table 2 shows descriptive statistics for the data used in the analysis.

Judges' Demographic Characteristics. The data on judge characteristics are from the Appeals Court Attribute Data, the Federal Judicial Center, and previous data collection from Chen and Yeh (2014*b*). The final dataset has information on gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to an appellate court. It also includes information on the full history of federal court appointments held by the judge.

Judicial Decisions. To study judicial decisions, we use two existing legal datasets with hand-coded vote direction and topic, created by Epstein, Landes and Posner (2013) and Glynn and Sen (2015). The datasets were originally constructed by searching for cases related to a given set of topics, and then having research assistants read through the opinions to code whether each judge's vote was liberal or conservative.[12,13] For gender-related issues—namely, reproductive rights, gender discrimination, and sexual harassment—a liberal vote corresponds to a vote in support of extending women's rights. The analysis pools the two datasets.

In addition, we have information on circuit court cases from the U.S. Court of Appeals datasets (Songer, 2008; Kuersten and Haire, 2011). These data include a 5% random sample of circuit cases that were hand-coded for vote valence (liberal, conservative, or neutral/hard to code).

Circuit Court Cases. To study interactions with female judges, we exploit detailed records from all 380,000 circuit court cases for the years 1890-2013, which we obtained from Bloomberg Law. For each case, the records include information on year, circuit, and topic,

[12] The need to code the direction of the vote (in favor or against expanding women's rights) is why we are constrained to using these pre-existing datasets, as we do not know the 'directionality' of votes in the larger circuit court dataset that we use in the analysis regarding the treatment of female judges.

[13] In particular, the Epstein, Landes and Posner (2013) dataset contains all published opinions related to abortion, the Americans with Disabilities Act, affirmative action, campaign finance, capital punishment, the contracts clause, criminal appeals, environmental regulation, federalism, piercing the corporate veil, race discrimination, sex discrimination, sexual harassment, and takings. The dataset is updated until 2008, but the starting years of the dataset vary by issue, ranging from 1982 for abortion to 1995 to capital punishment. The original dataset was constructed by searching for cases related to each issue backwards from the present, and stopping when a sufficient number of cases was reached for that issue. The Glynn and Sen (2015) data contain all published and unpublished opinions from 1996 to 2002 that contain the words "gender", "pregnancy" or "sex" in the case headings. When the two datasets are pooled, we drop duplicate cases present in both datasets.

as well as the panel of judges assigned to decide on it. For each assigned judge, we have information on whether they voted to affirm or reverse the district court decision, whether they authored the majority opinion, and whether they dissented or concurred.

For a subsect of the cases, we are also able to match the case to the identity of the district judge who decided the corresponding lower-court case. The judge's name is obtained either directly from the district case's metadata or by parsing the name from the circuit opinion's case history.[14] We then match the name of the district judge to their biographic information from the Federal Judicial Center. We are able to assign a unique district judge (and associated gender) to 121,944 circuit cases (32% of all cases).

Citations. We use the full text of the opinions to extract information on which cases are cited in the same opinion. We use this dataset to define a citation network that includes both backward and forward citations for each case.

Clerks. Information on circuit court clerks are from Katz and Stafford (2010) and additional original data collection from Chen (2019).

4 Gender Slant and Judge Demographics

This section explores descriptively how the gender slant measure varies based on judge characteristics. We begin by correlating gender slant and judge characteristics using separate univariate regressions with judge-level data. Note that here and in the rest of the paper, the gender slant measure is standardized for ease of interpretation.

Table 3 reports estimates from these regressions. Column (1) shows that judges nominated by presidents of different political parties do not display different levels of gender slant. Column (2) shows that female judges display gender slant that is 0.5 standard deviations lower than male judges on average. The difference is statistically significant at the 10% level. Column (3) shows that there is no difference depending on judge race, possibly an artifact of there being very few minority judges included in the sample. As one would expect, older judges tend to have significantly higher gender slant (column (4)): judges that were born before 1920 display between 0.5 and 0.765 standard deviation higher slant than judges born between 1930 and 1939 and after 1940. While this is consistent with older judges holding more socially conservative views, this variation might reflect differences in the cases that were tried by the judges, as older judges served in court in periods

[14]More precisely, the algorithm starts with all district judges, then checks for every case/case history for the corresponding court. Out of the judges within that specific district court, it then narrows it down to the judges that were active during the time of the case, plus two years to allow for appeal proceedings. Finally, based on a name similarity measure (Levenshtein distance), district judges are assigned if the score is above 70 (out of a maximum 100).

Table 3: Correlates of Gender Slant

Dependent Variable	Gender Slant					
	(1)	(2)	(3)	(4)	(5)	(6)
Democrat	-0.027				-0.003	0.083
	(0.172)				(0.178)	(0.269)
Female		-0.502*			-0.592***	-0.713**
		(0.288)			(0.202)	(0.276)
Minority			-0.098		-0.164	0.453
			(0.329)		(0.194)	(0.283)
Born in 1920s				-0.069	0.080	0.152
				(0.191)	(0.208)	(0.299)
Born in 1930s				-0.765***	-0.740***	-0.606*
				(0.203)	(0.234)	(0.336)
Born after 1940				-0.537**	-0.558**	-0.381
				(0.229)	(0.258)	(0.338)
Daughter						-0.490*
						(0.275)
Observations	139	139	139	139	139	98
Outcome Mean	0.000	0.000	0.000	0.000	0.000	-0.085
Adjusted R2	-0.007	0.020	-0.007	0.087	0.440	0.529
Circuit FE					X	X
Additional Controls					X	X
Number of Children FE						X

Notes: The table shows the correlation between demographic characteristics and gender slant. We regress gender slant on demographic characteristics of the judge in separate regressions (columns (1) to (4)) and in a multivariate regression that includes additional controls and circuit fixed effects (column (5)). The additional controls are region of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. The omitted category for judge cohort are judges born before 1920. In column (6) we additionally include an indicator variable for having at least one daughter, and number of children fixed effects. Standard errors are robust. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. Data on judges' family composition is from Glynn and Sen (2015). *** p<0.01, ** p<0.05, * p<0.1.

with lower female labor force participation. Interestingly, judge cohort appears to have the strongest explanatory power across all different demographic characteristics.

Column (5) includes all characteristics in the same regression and additionally controls for judge religion, law school attended, whether the judge had federal experience prior to being appointed to circuit courts, and circuit fixed effects. The previously discussed correlations remain.[15] Overall, female judges and younger judges display the lowest gender slant.

[15]There is no difference in the gender slant of judges born in different regions. Jewish judges appear to be slightly less slanted than Protestant judges. As far as law schools are concerned, judges who received their J.D.s from Yale display lower slant than judges who received their J.D.s from Harvard, while Stanford judges have higher slant. Finally, judges from the 3rd (PA, DE, MD), 6th (MI, OH, KT, TN), and Federal Circuits have higher slant than judges from the 1st Circuit (ME, MA, RI, CT), while judges from the 7th Circuit (WI, IL, IN) display lower gender slant.

What explains the variation in gender slant across judges? Gender slant might be partially representational and reflect variation in the facts of the cases tried by the judges. Here, we explore a different possibility: exposure to women. In particular, we ask whether judges that have daughters display different levels of slanted language.[16] Given that, conditional on total number of children, gender should be as good as randomly assigned, we can estimate the causal effect of having daughters on slant (Washington, 2008; Glynn and Sen, 2015).

To perform the analysis, we combine our measure of slant with information of judges' family composition from Glynn and Sen (2015). We estimate the following specification:

$$Gender\ Slant_j \quad = \quad \beta Daughter_j + X_j'\gamma + \delta_{c(j)} + \delta_{n(j)} + \epsilon_j \tag{1}$$

where $Gender\ Slant_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of judge j, $Daughter_j$ is an indicator variable equal to 1 if judge j has at least one daughter, X_j are demographic characteristics of judge j (gender, party of nominating president, race, religion, region of birth, cohort of birth, law school attended, and whether the judge had federal experience prior to appointment), $\delta_{c(j)}$ are circuit fixed effects, and $\delta_{n(j)}$ are number of children fixed effects. Standard errors are robust.

Table 3 column (6) reports the estimates. We find that, conditional on the number of children, having a daughter lowers gender slant by 0.490 standard deviations. In comparison, female judges tend to have about 0.713 lower gender slant than male judges in this sample. The effect is only significant at the 10% level. While these estimates should be interpreted carefully, it is interesting to note that they are potentially consistent with the view that gender exposure may be important for gender attitudes, in line with the recent literature on the effect of direct contact on attitudes toward specific groups (Alesina et al., 2018; Corno, La Ferrara and Burns, 2019; Lowe, 2021).

5 Slanted Judges and Decisions in Gender-Related Cases

This section asks whether higher-slant judges take different decisions in gender-related cases, as a way of validating our interpretation that gender slant can be seen as a proxy for gender attitudes. We estimate the following specification:

$$Conservative\ Vote_{ij} \quad = \quad \beta Gender\ Slant_j + X_j'\gamma + \delta_{c(i)t(i)} + \varepsilon_{ij} \tag{2}$$

[16] The two explanations are not mutually exclusive. Exposure to women, and in particular having daughters, might matter for slant for potentially different reasons, including learning, empathy, or preference realignment (Glynn and Sen, 2015).

where $Conservative\ Vote_{ij}$ is an indicator variable equal to 1 if judge j voted conservatively (against expanding women's rights) in case i, $Gender\ Slant_j$ is the gender slant (standardized cosine similarity between the gender and the career-family dimensions) of judge j, X_j are demographic characteristics of judge j (gender, party of nominating president, race, religion, region of birth, cohort of birth, law school attended, and whether the judge had federal experience prior to appointment), and $\delta_{c(i)t(i)}$ are circuit-year fixed effects. The circuit-year fixed effects ensure that we are using within circuit-year variation, for which cases are quasi-randomly assigned to panels. Instead, controlling for demographic characteristics ensures that the effect of being assigned higher- or lower-slant judges is not confounded by other features of the judge. The dataset is at the vote level, i.e. for each case it includes one observation for the vote of each judge. Standard errors are clustered at the judge level.

Table 4 shows that judges with higher slant are more likely to vote conservatively in gender-related cases. In particular, column (2), that reports estimates from the baseline specification, shows that judges with a one standard deviation higher slant are less likely to vote in favor of expanding women's rights in gender-related cases by 4.2 percentage points, which is significant at the 1% level.[17,18] The magnitude of the effect is sizable, corresponding to 7% of the outcome mean. To put this in perspective, the effect of being assigned a judge with a one standard deviation higher slant has around one-third of the effect of being assigned a judge that was nominated by a Democratic President. Given that gender slant is measured with error, meanwhile, the estimates are likely to be attenuated toward zero. In addition to providing an important validation for our interpretation of the measure, this result is interesting per se in that it shows that slanted judges have the potential to influence real-world outcomes (Chen and Sethi, 2018; Chen and Yeh, 2014*c*,*a*)). Gender slant is policy-relevant.

Even if we control for detailed demographic characteristics of the judge, we might still worry that the results are picking up exposure to different types of cases. In fact, while random assignment of judges to cases ensures that judges are exposed to similar cases in a given circuit and year, it is still possible that the cases tried across circuits and years might differ quite significantly. We address this concern in two ways. First, in column (3),

[17] Appendix Table 6 displays the result of the test proposed in Oster (2019) for bounding selection of unobservables based on selection on observables. The test assesses the amount of selection on unobservables that is plausible in a setting by looking at the change in the coefficient of interest from an uncontrolled regression to a regression that includes all controls, scaled by the change in R^2. Overall, selection seems to be limited in this setting: selection on unobservable characteristics would need to be twice as large as selection on observables for the treatment effect to equal 0.

[18] Appendix Figure 8 shows a binned scatterplot of the main relationship between gender slant and decisions in gender-related cases, conditional on demographic controls and circuit-year fixed effects.

Table 4: Slanted Judges and Decisions in Gender-Related Cases

Dependent Variable	Conservative Vote				
	(1)	(2)	(3)	(4)	(5)
Gender Slant	0.041***	0.042***	0.042***	0.052***	0.046***
	(0.016)	(0.013)	(0.012)	(0.014)	(0.012)
Democrat		-0.144***	-0.141***	-0.136***	-0.148***
		(0.025)	(0.025)	(0.024)	(0.025)
Female		-0.033	-0.041	-0.016	-0.033
		(0.033)	(0.032)	(0.025)	(0.034)
Observations	3086	3086	3086	3086	3086
Clusters	113	113	113	113	113
Outcome Mean	0.605	0.605	0.605	0.605	0.605
Circuit-Year FE	X	X	X	X	X
Additional Controls		X	X	X	X
Year of Appointment			X		
Exposure FE				X	
No Gender-Related Cases					X

Notes: The table shows the effect of slanted judges on decisions in gender-related cases, i.e. cases related to reproductive rights, gender discrimination, and sexual harassment. We regress an indicator variable equal to 1 if the judge voted conservatively in a gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, race (i.e. whether minority), region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (1) does not include demographic controls. Column (2) estimates the baseline specification. Column (3) controls for year of first appointment of the judge to a circuit court. Column (4) includes exposure fixed effects, which are indicator variables equal to 1 if the judge sat on at least one panel in a given circuit over a given 25-year period. In column (5), gender slant is calculated using embeddings trained excluding gender-related cases. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Data on votes on gender-related cases are from Epstein et al. (2013)'s update of Sunstein's (2006) data and Glynn and Sen (2015). Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

we show that controlling for year of first appointment of the judge to an appellate court does not impact the results. Second, in column (4), we include exposure fixed effects, i.e. indicator variables equal to 1 if the judge sat on at least one case in a given circuit over a 25-year period, which ensures that we are only comparing judges who served in similar circuits and years. Again, the result is not affected.[19,20]

Robustness Checks. A potential concern with the gender slant measure is that the corpus of authored opinions might offer a limited reflection of the judges' preferences, if judicial clerks are the ones responsible for drafting them. However, Appendix Table 8 column (1) shows that controlling for the share of the clerks that are women does not impact the results, suggesting that gender slant is not just proxying for clerk characteristics. In

[19] Exposure fixed effects are different from circuit-year fixed effects: they are a biographic characteristic that refers to the career of the judge, and not a characteristic of the specific case. We are constrained in how detailed the exposure fixed effects can be by the relatively small number of judges in our sample.

[20] Appendix Table 7 shows the results are the same if we separately estimate the regression for the Epstein, Landes and Posner (2013) and Glynn and Sen (2015) datasets.

addition, column (2) shows that the result is robust to dropping cases decided in circuits where quasi-random assignment of judges to cases was contested by Chilton and Levy (2015) (namely, the 2nd, 8th, 9th, and D.C.).

Finally, the result is robust to two procedures that take into account the precision with which the slant of each judge is estimated. First, Appendix Figure 9 shows that the point estimate is virtually unchanged if we shrink gender slant using Empirical Bayes techniques.[21] Second, Appendix Table 8 column (3) shows that the point estimate is also not affected if we weight the regression by the inverse of the variance of gender slant across bootstrap samples, thus giving higher weight to judges whose slant is more precisely estimated. However, the fact that the result is robust to controlling for log number of tokens (column (4)) shows that the effect we estimate is not driven by differences in corpus size.

Robustness to Word Set Choice. The result does not depend on the specific choice of words used for constructing the gender and career-family dimensions. We experiment with expanding or restricting the word sets, or dropping single words at a time, and present the results in Appendix Figure 10. In particular, the graph to the left shows the coefficients and 95% confidence interval from separate regressions where slant is identified using the top five to top fifteen most frequent male, female, career, and family words from LIWC. The graph to the right shows coefficients and 95% confidence intervals from separate regressions where slant is measured by dropping one attribute word at the time. Smaller word sets give larger confidence intervals and weaker explanatory power of decisions in gender-related cases, but the result is otherwise robust. At the same time, no single word is driving the result: the coefficient is remarkably stable across all the regressions displayed in the graph to the right.

Gender Slant Excluding Gender-Related Cases. A potential concern with these results is that the gender slant measure could itself be determined by the text of opinions from gender-related cases. Under this argument, cases involving gender-normative situations (women at home and men at work) could be systematically correlated with more conservative decisions, and judges with higher slant might be more exposed to such cases in their careers. We believe this to be unlikely for the following reasons. First, as highlighted before, quasi-random assignment of judges to cases implies that cases are comparable within a given circuit and year. Still, while judges serving across circuits or time might be exposed to different types of cases, our results are robust to including judge cohort fixed effects, year of first appointment to a circuit court, and exposure fixed effects, which work

[21]See Appendix C for a detailed discussion of the procedure we use to implement the EB-adjustment, and why it does not seem to help us lower the tokens threshold.

to control for the types of cases judges have been exposed to. Second, gender-related cases constitute a small proportion of the texts in the full corpus: only 17,523 cases out of the 114,702 circuit court cases on which we train the judge-specific word embeddings are gender related.[22] Third, the bootstrap procedure we employ to train the word embeddings ensures that the measure is not driven by any specific opinions. Nonetheless, we also show directly that the results are not driven by writing in gender-related cases. As shown in Table 4 column (5), the result is unchanged if the gender slant measure is computed on embeddings trained excluding gender-related cases.[23]

Non-Gender-Related Cases. Finally, if gender slant proxies for gender preferences, we should expect it to have larger effects on gender-related cases as opposed to non-gender-related cases. Two separate datasets allow us to explore this question. First, we use the Epstein, Landes and Posner (2013) dataset, which also includes decisions in ideologically divisive but non-gender-related issues such as age discrimination or campaign finance, to study the effect of being assigned a slanted judge on decisions in these types of cases. Appendix Table 9 shows that higher-slant judges are more likely to vote conservatively in non-gender-related cases, but the effect is around two thirds as big as the effect in gender-related cases. Two additional pieces of evidence are consistent with a gender-focused effect. If we estimate the baseline specification controlling for the share of conservative votes of the judge in non-gender-related cases, the effect of being assigned a slanted judge on conservative votes in gender-related cases is unchanged (Appendix Table 8 column (5)). In addition, if we estimate a differences-in-differences specification in which we compare gender- and non-gender-related cases that are assigned to judges with different levels of gender slant, we find that judges with higher slant are more likely to vote conservatively in gender-related as opposed to non-gender-related cases (Appendix Table 10).

Second, we use the U.S. Court of Appeals datasets (Songer, 2008; Kuersten and Haire, 2011), which includes a 5% random sample of circuit cases that were hand-coded for vote

[22]We identify gender-related cases using a simple pattern-based sequence-classification method. The method identifies a case to be gender-related if it contains a word that is much more likely to appear in gender-related cases as identified by the Epstein et al. (2013) and Glynn and Senn (2015) datasets than in an equally sized random sample of cases not identified to be gender-related according to these two datasets. First, we match the cases identified to the related opinions string matching on citations and party names. We are able to match 86% of the cases. Second, we define a word to be gender-related if it is twenty-five times more likely to appear in gender-related cases versus non gender-related cases, and if it appears at least 500 times in gender-related cases.

[23]It is still possible, although unintuitive, for gender slant to be a proxy of non-gender related cases whose facts imply a closer association between men and career and women and family. In this case, we should interpret the effect of slanted judges as the effect of being exposed to such cases, in such a way that affects not only decision in other gender-related cases but also interaction with female judges.

valence (liberal, conservative, or neutral/hard to code). Appendix Table 11 shows that higher-slant judges are not differentially likely to cast a conservative vote across all specifications. Taken together, these results show that while gender slant may be correlated with holding liberal views, the measure does indeed capture attitudes that are specific to gender.

6 Slanted Judges and Interactions with Female Judges

In this section we ask whether gender attitudes, as proxied by our measure of slant, manifest themselves in differential treatment of female judges on part of their colleagues. We focus on dimensions that are relevant to a judge's career. In particular, we explore the following outcomes: whether lower-court decisions by female judges are more likely to be reversed with respect to lower-court decisions by male judges, whether female judges are assigned the writing of majority opinions, and forward citations by future judges.

6.1 Slanted Judges and Disparities in Reversals

District court trials are presided by a single judge and cases are assigned to district judges quasi-randomly within each district-year. Up to 40% of district cases are appealed and are therefore considered by circuit courts (Eisenberg, 2004). Importantly, reversals matter for career outcomes: as shown in Appendix Figure 11, district judges that have a higher share of their decisions reversed on appeal are less likely to be promoted to circuit courts.

Here, we ask whether judges with higher gender slant are differentially likely to reverse decisions authored by female relative to male district court judges. The empirical strategy we employ in this section is slightly different than the rest of the paper, although it also builds on quasi-random assignment of circuit judges to cases and on conditioning on observable characteristics. In particular, we identify the effect of being assigned a higher-slant judge on reversals using a differences-in-differences design that compares appealed cases decided by female and male district judges that are assigned to circuit judges with different levels of slant. Identification requires that cases originally decided by a male district judge and assigned to circuit judges with different level of slant provide a good control group for cases that were originally decided by a female district judge. The quasi-random assignment of cases to panels at the circuit level helps us in this respect: cases assigned to judges with higher or lower slant are comparable. Importantly, the identification strategy allows for cases decided by female and male judges to be different, for example because they are appealed at different rates, as long as there is no systematic assignment of cases to higher and lower slant judges based on the likely reversal outcome.

We estimate the following baseline specification:

$$\begin{aligned} \textit{Voted to Reverse}_{ij} = &\pi \textit{Female District Judge}_{k(i)} * \textit{Gender Slant}_j \\ &+ \textit{Female District Judge}_{k(i)} * X_j^{'}\gamma + \delta_{c(i)t(i)} + \delta_{k(i)} + \delta_j + \varepsilon_{ij} \end{aligned} \tag{3}$$

where $\textit{Voted to Reverse}_{ij}$ is an indicator variable equal to 1 if circuit judge j voted to reverse the district court decision in case i, $\textit{Gender Slant}_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of circuit judge j, $\textit{Female District Judge}_{k(i)}$ is a dummy equal to 1 if the district judge who originally decided the case (judge $k(i)$) is female, X_j are demographic controls for the circuit court judge (gender, party of nominating president, race, religion, region of birth, cohort of birth, law school attended, and whether the judge had federal experience prior to appointment), $\delta_{c(i)t(i)}$ are circuit-year fixed effects, $\delta_{k(i)}$ are district judge fixed effects, and δ_j are circuit judge fixed effects. The dataset is at the vote level. Standard errors are clustered at the circuit judge level.

The inclusion of the circuit-year fixed effects $\delta_{c(i)t(i)}$ ensures comparability of cases assigned to higher and lower slant judges because of quasi-random assignment of cases to judges. As before, controlling for the demographic characteristics X_j shows that the effect of higher-slant judges is not confounded by other judge characteristics. District judge fixed effects $\delta_{k(i)}$ ensure that we are comparing what happens when cases decided at the district court level by the same judge are assigned to circuit judges with higher and lower slant. Finally, the circuit judge fixed effects δ_j allow for circuit judges to differ in their baseline probability of reversing a district court decision.

Table 5 column (1) reports estimates from a specification that only includes the fixed effects, while column (2) estimates the baseline specification. Circuit judges with a one standard deviation higher slant are 1 percentage point (5.6% of the baseline mean) more likely to vote to reverse a district court decision if the district judge is female, relative to when the district judge is male. Other characteristics of the circuit judge do not make a difference: being appointed by a Democratic President or being female is unrelated to disparately voting to reverse female district judges. Female district judges are 2.3 percentage points less likely to be reversed, and going from a judge with the median slant to a judge with a one standard deviation higher slant decreases this gap almost by half.[24]

[24]Appendix Figure 12 shows a binned scatterplot of the main relationship between gender slant and reversals separately by gender of the district judge, conditional on demographic controls and circuit-year fixed effects. The figure shows that the reversal gap between male and female district judges is almost entirely driven by lower slanted circuit judges. Instead, the gap almost disappears for higher slanted circuit judges, who reverse male and female district judges at similar rates.

This pattern is consistent with two interpretations. If the reversal gap between male and female district judges for low slant circuit judges reflects what should be the "true" reversal rate based on decision qual-

Table 5: Slanted Judges and Reversals

Dependent Variable	Voted to Reverse District Decision				
	(1)	(2)	(3)	(4)	(5)
Gender Slant * Female District Judge	0.006*	0.010***	0.010***	0.010***	0.009**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Democrat * Female District Judge		-0.010	-0.010*	-0.010	-0.011*
		(0.006)	(0.006)	(0.006)	(0.006)
Female * Female District Judge		-0.003	-0.002	-0.003	-0.005
		(0.010)	(0.010)	(0.010)	(0.010)
Observations	145862	145862	145862	145862	145862
Clusters	133	133	133	133	133
Outcome Mean, Male District Judge	0.180	0.180	0.180	0.180	0.180
Outcome Mean, Female District Judge	0.157	0.157	0.157	0.157	0.157
Circuit-Year FE	X	X	X	X	X
Circuit Judge FE	X	X	X	X	X
District Judge FE	X	X	X	X	X
Additional Controls		X	X	X	X
Year of Appointment			X		
Exposure FE				X	
No Gender-Related Cases					X

Notes: The table shows the differential effect of slanted judges on the reversal probability of cases originally decided by male and female district judges using a differences-in-differences design. We regress an indicator variable equal to 1 if the judge voted to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge is female, demographic controls interacted with an indicator variable for whether the district judge is female, circuit judge fixed effects, district judge fixed effects and circuit-year fixed effects (equation (3)). Demographic controls are gender, party of appointing President, race (i.e. whether minority), region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and refer to the circuit judge. Column (1) does not include demographic controls. Column (2) estimates the baseline specification. Column (3) controls for year of first appointment of the judge to a circuit court interacted with an indicator variable for whether the district judge is female. Column (4) includes exposure fixed effects, which are indicator variables equal to 1 if the judge sat on at least one panel in a given circuit over a given 25-year period, interacted with an indicator variable for whether the district judge is female. In column (5), gender slant is calculated using embeddings trained excluding gender-related cases. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. The sample is restricted to cases for which we were able to determine the identity of the district judge. Standard errors are clustered at the circuit judge level. *** p<0.01, ** p<0.05, * p<0.1.

The result is not explained by judges being exposed to different cases over the course of their careers: the coefficient on slant is unchanged if we control for year of first appointment to a circuit court (column (3)), or if we include exposure fixed effects that ensure we are comparing judges who served in similar periods and areas (column (4)). Finally, using a gender slant measure calculated based on embeddings trained excluding gender-related cases barely affects the estimate (column (5)).

ity, then higher-slant judges over-reverse female judges. Alternatively, it is possible that male and female district judges should be reversed at the same rate, and that lower slant judges reverse decisions of male district judges more. Unfortunately, without a measure of decision quality that is independent of reversals, we cannot tease apart these two interpretations.

It is also interesting to explore whether the effect of assignment to a higher-slant judge changes over time. In particular, we might think that those who, in more recent years, still express stereotyped views of gender in their writings might display especially discriminatory behavior when dealing with female colleagues. Consistently with this hypothesis, in Appendix Figure 13 we show that the effect we estimate seems to be almost entirely driven by cases from after 2000.

Are higher-slant judges reacting specifically to the gender of the district judge? Not necessarily: it could be that higher-slant judges are reacting to a characteristic of appealed cases that correlates with the gender of the district judge. For example, female district judges might be systematically more lenient in criminal cases, and higher-slant judges might be responding along that margin. Because we don't observe such granular features of the district judge's decisions, we cannot fully rule it out. However, we note that with the differences-in-differences design, the pattern is not driven by level differences across cases. For example, if female district judges systematically make higher (or lower) quality decisions, or are appealed in different types of cases, that should be addressed by the fixed effects. These characteristics matter only to the extent that they interact with gender slant. Thus, the analysis is still informative of how gender preferences affect gendered incidence of reversals, even if it is driven by female judges making decisions that higher-slant judges tend to find fault with.

Robustness Checks. Appendix Table 12 shows that the main effect on reversals is robust to a number of additional robustness checks. Including district-year fixed effects, which are not needed for identification but might increase precision, does not impact the result (column (1)). Controlling for the share of female clerks (column (2)) and controlling for the circuit judge's vote record in ideologically divisive cases (column (3)) do not make a difference. In addition, the main effect is stable to restricting the sample to the post-1980 period when there were likely to be more female district judges (column (4)) or to circuits in which we are confident cases were quasi-randomly assigned to judges (column (5)). The effect on reversals is also not explained by some slant measures being estimated more precisely than others. The point estimate is, if anything, slightly larger when we use an EB-adjusted version of slant (Appendix Figure 14) or when we weight the regression by the inverse of the variance of slant across bootstrap samples (column (6)). Controlling for the size of the corpus also does not make a difference (column (7)). Finally, Appendix Figure 15 shows that the results are robust to using different word sets to identify the gender and career-family dimensions.

Back-of-the-Envelope Calculation. Given that reversals have a negative effect on district judges' promotion, higher-slant judges have the potential to hinder the career progression of female district judges with respect to male district judges. Using estimates of the effect of reversals on the probability that district judges get elevated to circuit courts, we can try to estimate the magnitude of this effect in a back-of-the-envelope calculation. In particular, we find that a female judge whose appealed decisions were assigned to circuit judges with on average a one standard deviation higher slant would be 6.3% less likely to be elevated than a male judge faced with similarly slanted circuit judges.[25]

6.2 Slanted Judges and Disparities in Opinion Authorship

In circuit courts, decisions are generally taken in conference by the three judges on the panel after oral arguments. The decision with the most votes becomes the majority position, and the judges in the majority then have to decide who is going to author the associated opinion. By custom, the most senior acting judge in the majority assigns the responsibility of writing, taking into consideration expertise, work load, and other factors (Bowie, Songer and Szmer, 2014). The majority opinion articulates the principles behind the decision, which are binding law for lower courts (Rohde and Spaeth, 1976): the authoring of the opinion itself is an important task. Given the policy stakes of opinion assignment, a relevant question is whether the preferences of the senior judges affect this procedure. In particular, we investigate whether senior judges with higher gender slant are differentially likely to assign majority opinions to female judges.

We estimate the following specification:

$$Female\ Author_i = \beta Gender\ Slant_{j(i)}^{SENIOR} + X_{j(i)}^{SENIOR'}\gamma + \delta_{c(i)t(i)} + \varepsilon_i \tag{4}$$

where $Female\ Author_i$ is a dummy equal to 1 if the author of the majority opinion of case i, $Gender\ Slant_{j(i)}^{SENIOR}$ is the gender slant (standardized cosine similarity between the gender and the career-family dimensions) of the most senior judge on the panel of case i, $X_{j(i)}^{SENIOR}$ are demographic characteristics of the most senior judge (gender, party of nominating president, race, religion, region of birth, cohort of birth, law school attended, and whether the judge had federal experience prior to appointment), and $\delta_{c(i)t(i)}$ are circuit-year fixed effects. The dataset is at the case level. Standard errors are clustered at the senior judge level.[26] For opinion assignment to a female or a male judge to be a meaningful decision, we drop per curiam (unsigned) opinions and restrict the sample to

[25] In fact, Appendix Table 13 shows that an increase from 0 to 1 in the share of votes to reverse the district court decision on appeal implies a 38 percentage points decrease in the probability of being elevated. The calculation follows from the fact that female district judges have around a 6.8% baseline probability of being elevated. Interestingly, the relationship between reversals and promotions is not differential by gender of the district judge.

[26] We determine the most senior judge on the panel using information on the career of appellate judges,

Table 6: Slanted Judges and Opinion Assignment

Dependent Variable	Author is Female				
	(1)	(2)	(3)	(4)	(5)
Gender Slant	-0.028**	-0.017**	-0.017**	-0.018*	-0.016**
	(0.011)	(0.008)	(0.008)	(0.010)	(0.008)
Democrat		-0.001	-0.001	-0.015	0.002
		(0.014)	(0.014)	(0.017)	(0.014)
Female		0.134***	0.133***	0.158***	0.137***
		(0.016)	(0.016)	(0.017)	(0.016)
Observations	32052	32052	32052	32052	32052
Clusters	125	125	125	125	125
Outcome Mean	0.383	0.383	0.383	0.383	0.383
Circuit-Year FE	X	X	X	X	X
Additional Controls		X	X	X	X
Year of Appointment			X		
Exposure FE				X	
No Gender-Related Cases					X

Notes: The table shows the effect of slanted judges on the probability of the majority opinion being assigned to a female judge. We regress an indicator variable equal to 1 if the authoring judge is female on the gender slant of the most senior judge on the panel, demographic controls, and circuit-year fixed effects (equation (4)). The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, race (i.e. whether minority), region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. Column (1) does not include demographic controls. Column (2) estimates the baseline specification. Column (3) controls for year of first appointment of the most senior judge of the panel to a circuit court. Column (4) includes exposure fixed effects, which are indicator variables equal to 1 if the most senior judge on the panel sat on at least one panel in a given circuit over a given 25-year period. In column (5), gender slant is calculated using embeddings trained excluding gender-related cases. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously. Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

cases that have at least one female judge on the panel. Since the decision to dissent or concur is possibly endogeneous, we also exclude cases that contain either a dissent or a concurrence.[27]

Table 6 reports the estimated effect of the most senior judge having higher or lower gender slant on whether the authoring judge is female. Column (1) reports the estimates from a specification that only controls for the circuit-year fixed effects, while column (2) estimates the baseline specification. When the most senior judge on the panel has a one standard deviation higher slant, the majority opinion is less likely to be authored by one of the female judges by 1.7 percentage points, corresponding to a 4.5% decrease over the

and exclude from the analysis cases for which we could not precisely determine who the identity of the most senior judge on the case, mainly because of there being multiple judges appointed in the same year.

[27] In Appendix Table 14 we check whether the gender slant of the panel's senior judge affects the probability of having a specific author (columns (1) and (2)) or having a per curiam opinion (columns (3) and (4)), and find no effect. We also show that the slant of the most senior judge on the panel does not impact the probability of unanimous decisions, that is, having dissents or concurrences (columns (5) and (6)).

outcome mean.[28,29] If the most senior judge is a woman instead, there is a 13.4 percentage points (35%) higher probability that the author of the opinion is a woman. While slant has a non-trivial effect on the gender of the authoring judge, the magnitude of the effect is substantially smaller than that of being female.[30]

Columns (3) and (4) show that the result is unchanged when we control for exposure to different types of cases. In addition, if we use a gender slant measure calculated based on embeddings trained excluding gender-related cases, the result is the same (column (5)). As was the case for reversals, Appendix Figure 17 shows that the effect of the most senior judge on the panel having higher slant is larger for more recent cases (i.e., those after 2000). Overall, these results show that judges with more conservative attitudes toward gender are less likely to assign an important career-relevant task to female judges.

Robustness Checks. Appendix Table 15 shows that the result is robust to a number of additional robustness checks. As before, in column (1) we control for the share of clerks that are female and find that it limitedly affects the main result: while the coefficient is no longer statistically significant—possibly because we have information on clerks only for a subset of the judges—it is very similar in magnitude. Reassuringly, column (2) shows that the result is not driven by confounded ideology of judges: if we control for a measure of how conservative a judge's voting record is, the coefficient on slant is unchanged.

Given the large effect of the senior judge being female on the probability that the authoring judge is a woman and that female judges have lower gender slant, we might worry that self-assignments are the driver behind the main effect. Column (3) shows that this is not the case: if we re-estimate the main specification excluding cases where the most senior judge on the panel is a woman, we find the same effect. Column (4) shows instead that the main result is, if anything, larger if we include cases that had dissents or concurrences. Columns (5) and (6) shows that the results are robust to restricting the sample to the post-1980 period and to dropping circuits for which Chilton and Levy (2015) doubted quasi-random assignment.

[28]Female judges are on average assigned opinions at the same rate as male judges. The baseline probability shown here is higher than 0.33 because we include panels with one, two or three female judges.

[29]Appendix Figure 16 shows a binned scatterplot of the main relationship between gender slant and opinion assignment, conditional on demographic controls and circuit-year fixed effects.

[30]To further put this number in perspective, it is helpful to compare the dynamics we see here with what would have been the opinion assignment process had it been random. If we exclude panels where the most senior judge is female to abstract from possible self-assignment issues (see 15 column (5)), we find that with random assignment the share of opinions authored by a female judge would have been 0.359. In the data we find that the observed proportion is lower at 0.347. The gap is larger for assigning judges with slant above the median (0.344) than for those with slant below the median (0.349). In other words, having a senior judge in the bottom half rather than in the top half of the slant distribution closes 40 percent of the gap.

The result is also robust to controlling for differences in how precisely slant is estimated for different judges, for example by using an EB-adjusted measure of slant (Appendix Figure 18) or giving more weight to judges whose gender slant is more precisely estimated (column (7)). Controlling for the size of the corpus (column (8)) does not make a difference. Finally, Appendix Figure 19 shows that the results are robust to using different word sets to identify the gender and career-family dimensions.

Assignment of Different Types of Cases. This result raises the question of whether slanted judges also assign different types of cases to female judges. In Appendix Figure 20 and Appendix Table 16, we look but do not find evidence of any differences. The cases assigned to female judges by higher-slant senior judges are not concentrated in specific areas of the law and do not have different expected importance, as proxied by forward citations predicted based on case characteristics, which is potentially an interesting result in light of the literature on discrimination in task assignments.

6.3 Slanted Judges and Disparities in Citations

The last career-related interaction we examine are citations. Law depends on precedent, and deciding which cases to cite in a specific opinion is a non-trivial decision: "Judges [and] lawyers who brief and argue cases [...] could all be thought, with only slight exaggeration, to make their living in part by careful citation of judicial decisions" (Posner, 2000). Meanwhile, many judges admit to monitoring and caring about whether they are cited by other judges (Posner, 2008), and citations to cases are commonly understood as a measure of judge quality (Ash and MacLeod, 2015). Citations are therefore relevant to judicial careers, and differential treatment of male and female judges in citation choices presents another potential domain for high-stakes discrimination.

Identification once again relies on quasi-random assignment of cases to judge panels. Conditional on circuit-year fixed effects, cases assigned to different panels are comparable. However, choice of authorship is endogenous. Even if the fact that we control for a number of judge characteristics improves comparability across judges, it is possible that judges with higher gender slant are systematically assigned authorship of cases for which it would be optimal to differentially cite female judges in the first place: the results in this sections have to be interpreted especially carefully.

The specification we estimate is:

$$Cites\ Female\ Judge_i = \beta Gender\ Slant_{j(i)} + X^{'}_{j(i)}\gamma + \delta_{c(i)t(i)} + \varepsilon_i \tag{5}$$

where $Cites\ Female\ Judge_{ictj}$ is an indicator variable equal to 1 if the opinion of case i cites at least one opinion authored by a female judge, $Gender\ Slant_{j(i)}$ is the slant (standard-

Table 7: Slanted Judges and Citations

Dependent Variable	Cites at Least One Female Judge				
	(1)	(2)	(3)	(4)	(5)
Gender Slant	-0.024***	-0.010*	-0.009*	-0.014**	-0.005
	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)
Democrat		-0.012	-0.011	-0.020*	-0.011
		(0.011)	(0.011)	(0.010)	(0.011)
Female		0.128***	0.125***	0.142***	0.131***
		(0.016)	(0.016)	(0.015)	(0.016)
Observations	107923	107923	107923	107923	107923
Clusters	139	139	139	139	139
Outcome Mean	0.383	0.383	0.383	0.383	0.383
Circuit-Year FE	X	X	X	X	X
Additional Controls		X	X	X	X
Year of Appointment			X		
Exposure FE				X	
No Gender-Related Cases					X

Notes: The table shows the effect of slanted judges on the probability of citing at least one female judge. We regress a dummy equal to 1 if the majority opinion cites at least one case authored by a woman on the gender slant of the author of the majority opinion, demographic controls, and circuit-year fixed effects (equation (5)). Demographic controls are gender, party of appointing President, race (i.e. whether minority), region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (1) does not include demographic controls. Column (2) estimates the baseline specification. Column (3) controls for year of first appointment to a circuit court. Column (4) includes exposure fixed effects, which are indicator variables equal to 1 if the judge sat on at least one panel in a given circuit over a given 25-year period. In column (5), gender slant is calculated using embeddings trained excluding gender-related cases. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases in which the opinion was authored by a specific judge. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

ized cosine similarity between the gender and career-family dimensions) of judge j who authored the majority of case i, $X_{j(i)}$ are demographic characteristics of judge j (gender, party of nominating president, race, religion, region of birth, cohort of birth, law school attended, and whether the judge had federal experience prior to appointment), and $\delta_{c(i)t(i)}$ are circuit-year fixed effects. The dataset includes one observation for each case, and is restricted to cases in which the opinion was authored by a specific judge. Standard errors are clustered at the judge level.

Table 7 reports the estimates from equation (5). Column (1) only controls for the circuit-year fixed effects, while column (2) estimates the baseline specification. Judges with a one standard deviation higher gender slant are 1 percentage point less likely to cite any opinions authored by female judges. The effect is significant at the 10% level, and relatively small in magnitude (2.6% of the outcome mean), especially if compared to the effect of the author being a woman (12.8 percentage points or 34%). Controlling for year of appointment (column (3)) does not affect the result, and neither does including career fixed effects (column (4)). However, the result is not robust to using gender slant measured on embeddings trained excluding gender-related cases (column (5)). As before, it is interesting to note that the effect of higher-slant judges is only present for opinions authored after 2000 (Appendix Figure 22).[31]

Robustness Checks. Appendix Table 17 elaborates on the robustness of the main effect on citations with different specifications and sample restrictions. The effect on citations is our least robust result. The main effect is not robust to controlling for the share of female clerks (column (1)), perhaps because clerks are important when determining citations or simply because of the smaller sample size. Controlling for the circuit judge's vote record in ideologically divisive cases (column (2)) limitedly impacts the main coefficient, and the effect is stable to restricting the sample to the post-1980 period =(column (3)). However, the result is not robust to excluding the 2nd, 8th, 9th, and D.C. circuits (column (4)).

As with opinion assignment, the large effect of being a female judge on citations suggests that the result might be explained by self-citations. Even when we define the outcome excluding self-cites, however, Appendix Table 17 column (5) shows that gender slant has a negative and statistically significant at the 10% level on the probability of citing a female judge. Interestingly, however, when self-cites are excluded, female judges appear to be less likely to cite other female judges.

In addition, the effect is not an artifact of gender slant being more precisely estimated for some judges: Appendix Figure 23 shows that the effect is robust to using an EB-adjusted version of slant and Appendix Table 17 column (6) shows that the result is not impacted by weighting by the inverse of the variance of the slant measure across bootstrap sample. However, as shown in column (7), the result is not robust to controlling for log number of tokens. Finally, the effect on citations is robust to using different word sets (Appendix Figure 24).

Overall, judges with higher gender slant appear to be less likely to cite opinions authored by female judges, although the effect is less robust than other findings presented in the paper and should be interpreted with special caution because of potential endogeneity.

6.4 Other Judge Characteristics Besides Gender

To strengthen the argument that gender slant is indeed proxying for gender attitudes, we explore whether higher-slant judges also interact differently with judges with specific demographic characteristics: political leanings, minority status, and age.

[31] Appendix Figure 21 shows a binned scatterplot of the relationship between gender slant and citations, conditional on demographic controls and circuit-year fixed effects.

Reversals. First, Appendix Table 18 shows that higher-slant judges are not differentially likely to reverse cases of district judges that were appointed by a Democratic President than cases of district judges that were appointed by a Republican President (column (1)). Circuit judges with higher slant, however, appear more likely to reverse cases in which the district judge is a minority (column (2)).

Opinion Assignment. Second, Appendix Table 19 shows that judges with higher slant are not differentially likely to assign the opinion of the court to Democratic judges (column (1)), minority judges (column (2)), or based on the age of the judge (column (3)).

Citations. Finally, Appendix Table 20 shows that judges with higher gender slant are less likely to cite opinions authored by Democrat-appointed judges (column (1)). Instead, judges with higher gender slant do not differentially cite minority judges (column (2)) or judges of different ages (column (3)). Interestingly, judges with higher slant are more likely to cite opinions authored by judges with higher slant as well (column (4)), and this pattern is not due to judge gender or party.

Overall, these results suggest that gender is the salient characteristic to which judges with higher slant respond: gender slant specifically proxies for attitudes toward women.

7 Conclusions

This paper investigates the role of gender attitudes in circuit courts. We find that gender attitudes, at least as far as they are expressed in judicial writing, matter. Judges with higher gender slant vote more conservatively on women's rights cases, are more likely to reverse district courts decisions when the district judge is a woman, are less likely to assign opinions to women, and cite fewer opinions authored by female judges.

These findings add to the literature on gender attitudes by showing that they matter even for skilled professionals making high-stakes, public-oriented decisions. Our text-based metric is a proxy for a psychological factor, and so the policy implications of the results should be considered with caution. Although we estimate the causal effect of assigning higher-slanted judges, we do not have evidence, for example, that forcing judges to use less stereotypical language would causally shift their decisions in gender law or their behavior toward female colleagues.

This research can be extended in a number of directions. First, it would be important to know how well text-based measures of gender attitudes correlate with other measures, such as scores on the implicit association test. Second, the text-based metrics could be computed for other decision-makers such as politicians, journalists, and professors. In these domains, as with judges, there are no traditional measures of attitudes, but large corpora of text are available.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *The Journal of Legal Studies*, 41(2): 347 – 383.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti. 2018. "Revealing Stereotypes: Evidence from Immigrants in Schools." NBER Working Paper No. 25333.

Antoniak, Maria, and David Mimno. 2018. "Evaluating the Stability of Embedding-Based Word Similarities." *Transactions of the Association for Computational Linguistics*, 6: 107–119.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Ash, Elliott, and W. Bentley MacLeod. 2015. "Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts." *Journal of Law and Economics*, 58(4).

Ash, Elliott, Daniel L Chen, and Sergio Galletta. 2022. "Measuring Judicial Sentiment: Methods and Application to US Circuit Courts." *Economica*, 89(354): 362–376.

Battaglini, Marco, Jorgen M. Harris, and Eleonora Patacchini. Forthcoming. "Professional Interactions and Hiring Decisions: Evidence from the Federal Judiciary." *Journal of Labor Economics*.

Bertrand, Marianne. 2013. "Career, Family, and the Well-Being of College-Educated Women." *American Economic Review*, 103(3): 244–50.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics*, 2: 228–255.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit Discrimination." *American Economic Review*, 95(2): 94–98.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg. 2019. "The Dynamics of Discrimination: Theory and Evidence." *American Economic Review*, 109(10): 3395–3436.

Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings." 4349–4357.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. "Beliefs about Gender." *American Economic Review*, 109(3): 739–73.

Bowie, Jennifer Barnes, Donald R Songer, and John Szmer. 2014. *The View from the Bench and Chambers: Examining Judicial Process and Decision Making on the U.S. Courts of Appeals.*

University of Virginia Press.

Boyd, Christina L., and James F. Spriggs II. 2009. "An Examination of Strategic Anticipation of Appellate Court Preferences by Federal District Court Judges." *Journal of Law & Policy*, 29: 37–81.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review*, 107(11): 3288–3319.

Caliskan, Aylin, Joanna J Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases." *Science*, 356(6334): 183–186.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2019. "Are Referees and Editors in Economics Gender Neutral?" *Quarterly Journal of Economics*, 135: 269–327.

Carlana, Michela. 2019. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *The Quarterly Journal of Economics*, 134(3): 1163–1224.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review*, 106(8): 2110–44.

Chen, Daniel L. 2019. "Judicial Analytics and the Great Transformation of American Law." *Artificial Intelligence and Law*, 27(1): 15–42.

Chen, Daniel L., and Jasmin Sethi. 2018. "Insiders and Outsiders: Does Forbidding Sexual Harassment Exacerbate Gender Inequality?" TSE Working Paper No. 16-687.

Chen, Daniel L., and Susan Yeh. 2014*a*. "The Construction of Morals." *Journal of Economic Behavior and Organization*, 104: 84–105.

Chen, Daniel L., and Susan Yeh. 2014*b*. "Growth Under the Shadow of Expropriation? The Economic Impacts of Eminent Domain." ETH Zurich and George Mason University Working Paper.

Chen, Daniel L., and Susan Yeh. 2014*c*. "How Do Rights Revolutions Occur? Free Speech and the First Amendment." ETH Zurich Working Paper.

Chen, Daniel L., Vardges Levonyan, and Susan Yeh. 2014. "Do Policies Affect Preferences? Evidence from Random Variation in Abortion Jurisprudence." ETH Zurich Working Paper.

Chilton, Adam S, and Marin K Levy. 2015. "Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals." *Cornell L. Rev.*, 101: 1.

Cohen, Alma, and Crystal S Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy*, 11(1): 160–91.

Corno, Lucia, Eliana La Ferrara, and Justine Burns. 2019. "Interaction, Stereotypes and Performance: Evidence from South Africa." IFS working paper W19/03.

Croft, Jane. 2016. "Law School Admissions Collapse Continues: They are being Forced to Innovate or Face being Left Behind." Financial Times.

DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology*, 23(1): 263–287.

Eisenberg, Theodore. 2004. "Appeal Rates and Outcomes in Tried and Nontried Cases: Further Exploration of Anti-Plaintiff Appellate Outcomes." *Journal of Empirical Legal Studies*, 1(3): 659–688.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. *The Behavior of Federal Judges.* Harvard University Press.

Friese, Malte, Matthias Bluemke, and Michaela Wänke. 2007. "Predicting Voting Behavior with Implicit Attitude Measures: The 2002 German Parliamentary Election." *Experimental Psychology*, 54(4): 247–255.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *Proceedings of the National Academy of Sciences*, 115(16): E3635–E3644.

George, Tracey E., and Albert H. Yoon. 2019. "The Gavel Gap: Who Sits in Judgement on State Courts?" American Constitution Society for Law and Policy.

Glaser, Barney G, and Anselm L Strauss. 2017. *Discovery of Grounded Theory: Strategies for Qualitative Research.* Routledge.

Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *The Quarterly Journal of Economics*, 132(3): 1219–1260.

Glynn, Adam N, and Maya Sen. 2015. "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?" *American Journal of Political Science*, 59(1): 37–54.

Green, Alexander R, Dana R Carney, Daniel J Pallin, Long H Ngo, Kristal L Raymond, Lisa I Iezzoni, and Mahzarin R Banaji. 2007. "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients." *Journal of General Internal Medicine*, 22(9): 1231–1238.

Hengel, Erin. Forthcoming. "Publishing while Female. Are Women Held to Higher Standards? Evidence from Peer Review." *Economic Journal*.

Jakiela, Pamela, and Owen Ozier. 2019. "Gendered Language." CGD Working Paper No. 500.

Kastellec, Jonathan P. 2013. "Racial Diversity and Judicial Influence on Appellate

Courts." *American Journal of Political Science*, 57(1): 167–183.

Katz, Daniel M, and Derek K Stafford. 2010. "Hustle and Flow: a Social Network Analysis of the American Federal Judiciary." *Ohio St. LJ*, 71: 457.

Kozlowski, Austin C, Matt Taddy, and James A Evans. 2019. "The Geometry of Culture: Analyzing Meaning through Word Embeddings." *American Sociological Review*, 84(5).

Kuersten, Ashlyn K., and Susan B. Haire. 2011. "Update to the Appeals Courts Database 1997 - 2002." 'http://www.songerproject.org/data.html'.

Le, Quoc, and Tomas Mikolov. 2014. "Distributed Representations of Sentences and Documents." Vol. 32 of *Proceedings of Machine Learning Research*.

Levinson, Justin D, Mark W Bennett, and Koichi Hioki. 2017. "Judging Implicit Bias: A National Empirical Study of Judicial Sterotypes." *Fla. L. Rev.*, 69: 63.

Lowe, Matt. 2021. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review*, 111(6): 1807–44.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." 3111–3119.

NAWL. 2019. "Survey Report on the Promotion and Retention of Women in Law Firms."

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vectors for Word Representation." 1532–1543.

Posner, Richard. 2008. *How Judges Think.* Harvard University Press.

Posner, Richard A. 1995. "Judges' Writing Styles (And Do They Matter?)." *The University of Chicago Law Review*, 62(4): 1421–1449.

Posner, Richard A. 2000. "An Economic Analysis of the Use of Citations in the Law." *American Law and Economics Review*, 2(2): 381–406.

Rachlinski, Jeffrey J., Sheri Lynn Johnson, Andrew J. Wistrich, and Chris Guthrie. 2009. "Does Unconscious Bias Affect Trial Judges?" *Notre Dame Law Review*, 84(3): 1195–1246.

Rehavi, M Marit, and Sonja B Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy*, 122(6): 1320–1354.

Rice, Douglas, Jesse H Rhodes, and Tatishe Nteta. 2019. "Racial Bias in Legal Language." *Research & Politics*, 6(2): 2053168019848930.

Ricoeur, Paul. 1981. *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation.* Cambridge university press.

Rodriguez, Pedro L, and Arthur Spirling. 2022. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *The Journal of Politics*, 84(1): 000–000.

Rohde, David W, and Harold J Spaeth. 1976. *Supreme Court Decision Making.* WH Freeman.

Root, Danielle. 2019. "Women Judges in the Federal Judiciary – A Series of Fact Shhts on the Lack of Diversity on the Bench." Center for American Progress.

Rooth, Dan-Olof. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence." *Labour Economics*, 17(3): 523–534.

Sarsons, Heather. 2019. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals."

Songer, Donald R. 2008. "The United States Courts of Appeals Database." 'http://www.songerproject.org/data.html'.

Starr, Sonja B. 2014. "Estimating Gender Disparities in Federal Criminal Cases." *American Law and Economics Review*, 17(1): 127–159.

Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki. 2006. *Are Judges Political?: An Empirical Analysis of the Federal Judiciary.* Brookings Institution Press.

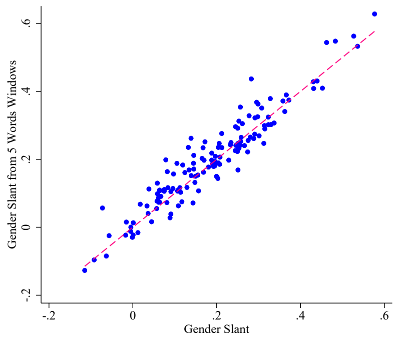
Washington, Ebonya L. 2008. "Female Socialization: How Daughters Affect Their Legislator Fathers." *American Economic Review*, 98(1): 311–32.

Online Appendix

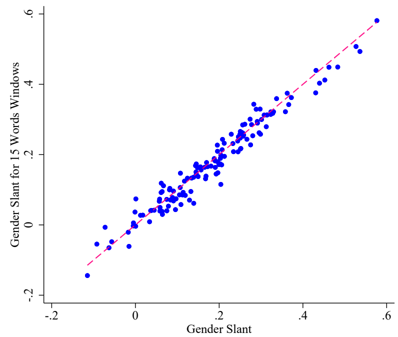
A Appendix Figures

Appendix Figure 1: Gender Slant for Different Window Sizes and Embedding Dimensions

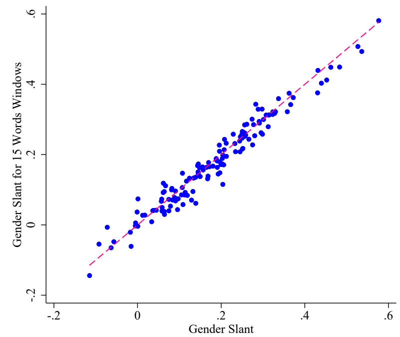
(a) 5 Words



(b) 15 Words

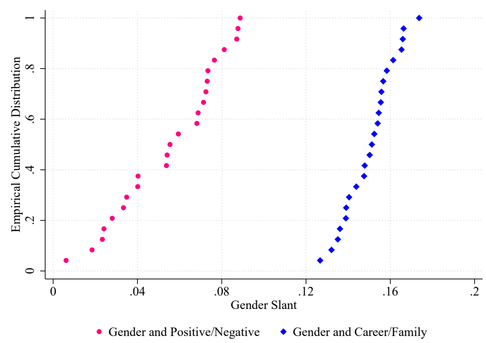


(c) 100-Dimensional



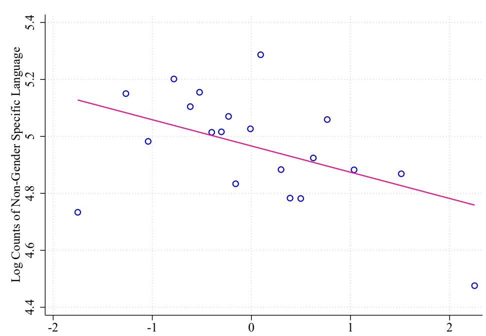
Notes: The graphs show a scatter plot of the gender slant measure obtained by training embeddings using different window sizes to construct the co-occurrence matrix (panel (a) and (b)) and embeddings with different dimensions (100 versus 300) (panel (c)).

Appendix Figure 2: Cumulative Distribution of Slant for Different Associations



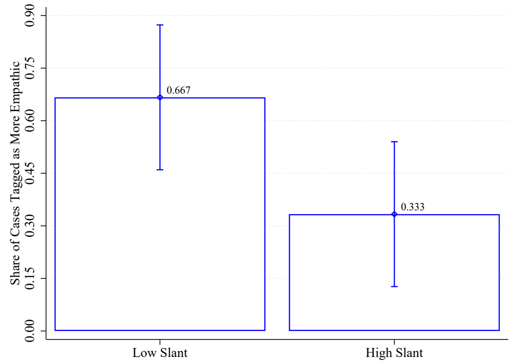
Notes: The graph shows the empirical cumulative distribution of gender slant measured using the stereotypical association between males and career versus female and family and using the stereotypical association between male and positive attributes versus female and negative attributes. The distibution comes for the 24 repetitions of bootstrapped embeddings for the full judicial corpus.

Appendix Figure 3: Gender Slant and Gender Neutral Language



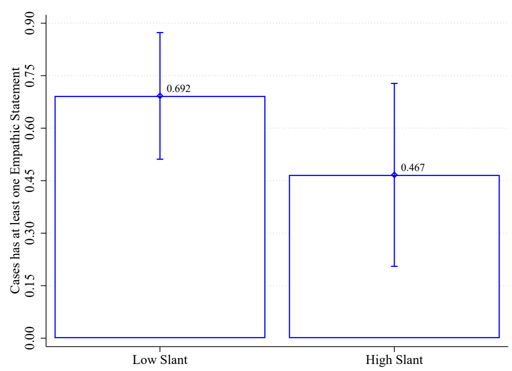
Notes: The graph shows a binned scatterplot of the relationship between the number of sentences in which the judge uses gender neutral language (e.g. 'he or she', 'he/she', 'he and she', etc.) and gender slant, conditional on the log number of tokens in the corpus.

Appendix Figure 4: Human Coding of Opinions and Gender Slant, Most Empathic



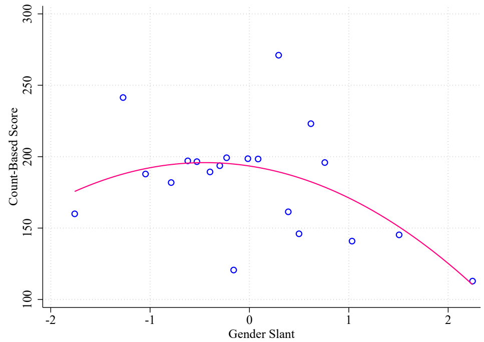
Notes: The graph shows the share of cases that were tagged as being the most empathic within a pair, by whether the opinion was authored by a judge with high or low slant. The sample includes 40 cases (20 pairs). The pairs were selected among gender-related cases that we were able to match with the opinion text from Bloomberg Law and that were decided in favor of expanding women' rights. The sample was further restricted to cases whose majority opinon was authored by a judge with the slant in the top/bottom 10% of the slant distribution. Cases were then randomly paired with replacement, with one case from the top and one from the bottom of the slant distribution. A reader, blind to the slant of each judge, was asked to assess which case presented the opinon that was most empathic towards women within each pair. In 13 out of the 20 pairs, the case selected to be the most empathic was the one with the lowest slant.

Appendix Figure 5: Human Coding of Opinions and Gender Slant, Has Empathic Statement



Notes: The graph shows the share of cases that were tagged as having at least one empathic statement, by whether the opinion was authored by a judge with high or low slant. The sample includes 38 cases. The pairs were selected among gender-related cases that we were able to match with the opinion text from Bloomberg Law and that were decided in favor of expanding women' rights. The sample was further restricted to cases whose majority opinon was authored by a judge with the slant in the top/bottom 10% of the slant distribution. A reader, blind to the slant of each judge, was asked to annotate each case.

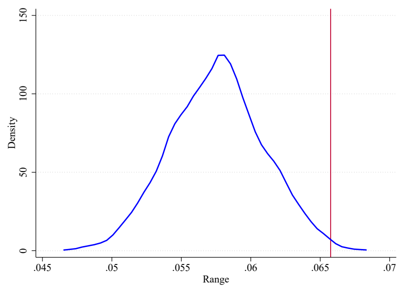
Appendix Figure 6: Gender Slant and Count-Based Score



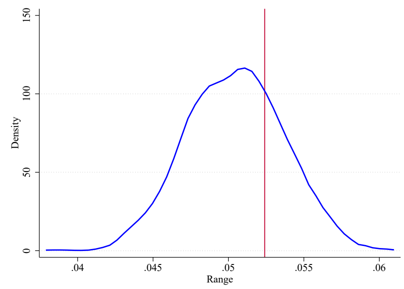
Notes: The graph shows a binned scatterplot of the relationship between the count based score and gender slant. The count-based score is defined as the ratio of the difference between the number of male/career snippets and the number of female/career snippets over the difference between the number of male/family snippets minus the number of female/family snippets, where male/career snippets are snippets in which a male and career word appear within a ten word window from each other and the other terms are similarly defined.

Appendix Figure 7: Randomization Check

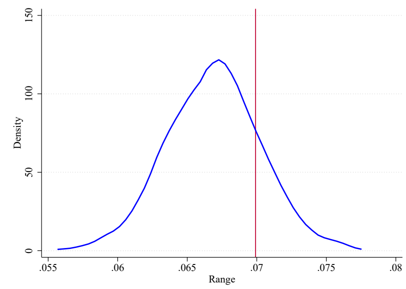
(a) Criminal



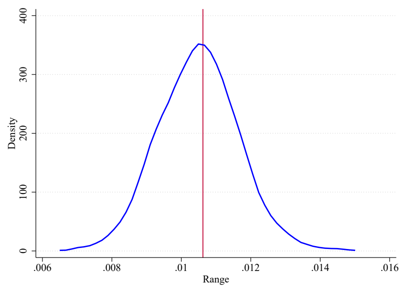
(b) Civil



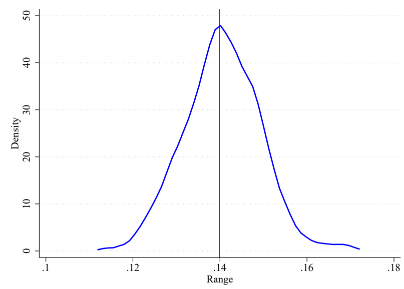
(c) Economic



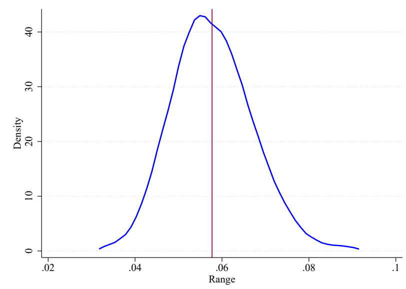
(d) Gender-Related



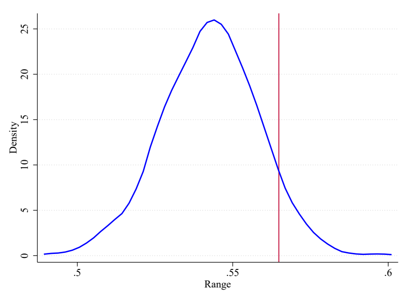
(e) Female District Judge



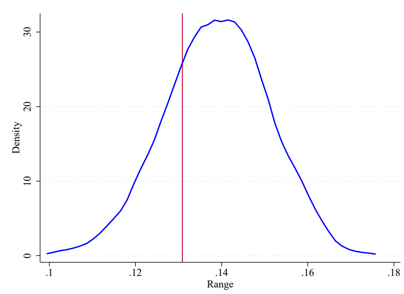
(f) Democrat District Judge



(g) District Judge Born Before 1934

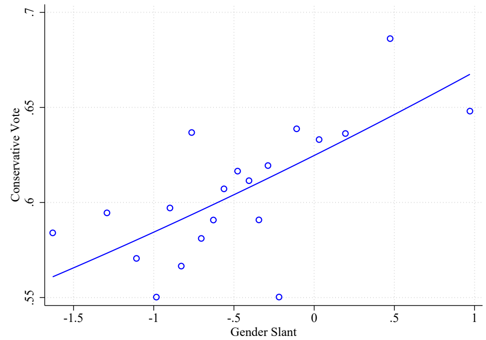


(h) Minority District Judge



Notes: These figures shows, for eight different judge characteristic, the distribution of the interquantile range of judge-specific averages resulting from 1000 simulated datasets. The vertical line represents the actual interquantile range.

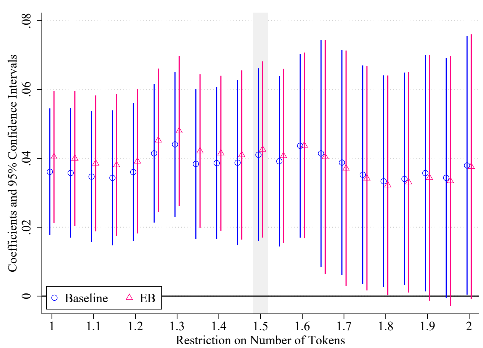
Appendix Figure 8: Slanted Judges and Decisions in Gender-Related Cases, Binned Scatterplot



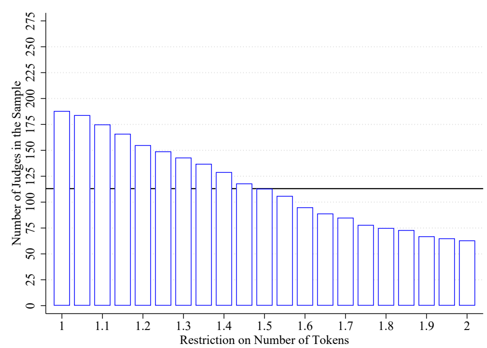
Notes: The graph shows a binned scatterplot of the relationship between gender slant and conservative votes in gender-related cases, conditional on demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court.

Appendix Figure 9: Slanted Judges and Decisions in Gender-Related Cases, Robustness to EB-Adjustment and Tokens Threshold

(a) Point Estimates and 95% CI

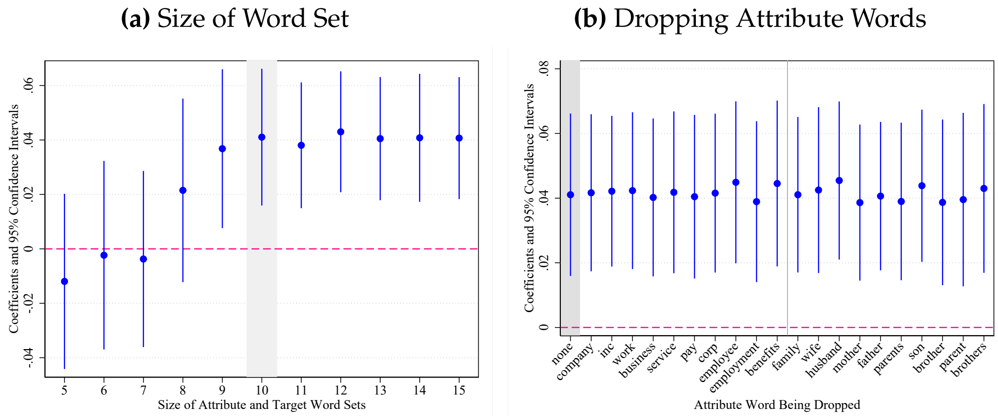


(b) Number of Judges



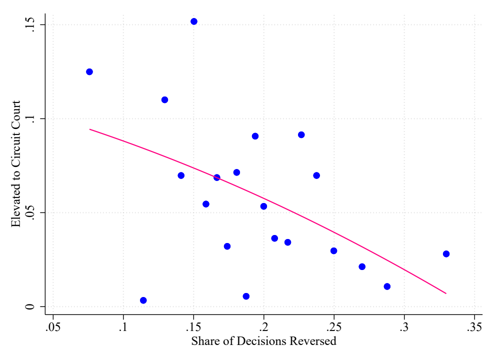
Notes: The graphs show how the effect of slanted judges on decisions in gender-related cases varies based on the tokens thresholds used to select the sample. The graph on the left shows both the point estimate and the 95% confidence interval for the baseline measure of gender slant and EB-adjusted gender slant, estimated including judges selected using different token tresholds. The graph on the right shows the number of judges included in the analysis for each token threshold. We regress an indicator variable equal to 1 if a judge voted conservatively in a gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Standard errors are clustered at the judge level.

Appendix Figure 10: Slanted Judges and Decisions in Gender-Related Cases, Robustness to Word Set Choice



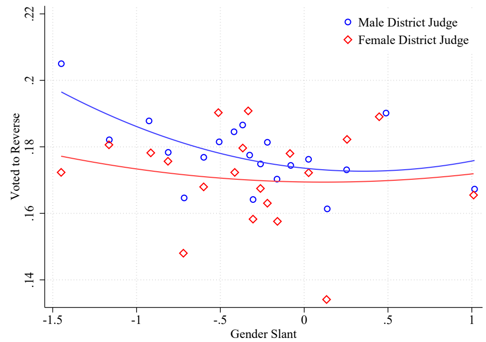
Notes: The graphs show how the effect of slanted judges on decisions in gender-related cases varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the right shows robustness to dropping one attribute word at the time. The graphs show the coefficient on gender slant, together with 95% confidence intervals. We regress an indicator variable equal to 1 if a judge voted conservatively in a gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. Standard errors are clustered at the judge level.

Appendix Figure 11: Reversals and Promotions from District to Circuit Courts



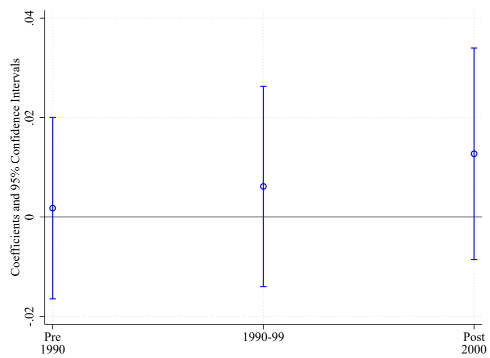
Notes: The graph shows the relationship between the probability of being elevated from a district to a circuit court and the share of decisions that were reversed on appeal, conditional on demographic controls and circuit fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. The sample is restricted to district judges for which we observe at least 50 cases.

Appendix Figure 12: Slanted Judges and Reversals, Binned Scatter Plot



Notes: The graph shows a binned scatterplot of the relationship between gender slant and the probability of voting to reverse the district court decision by the gender of the district judge, conditional on demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. The sample is restricted to cases for which we were able to determine the identity of the district judge.

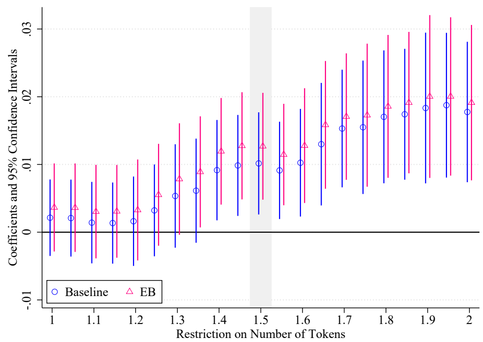
Appendix Figure 13: Slanted Judges and Reversals, by District Judge Gender



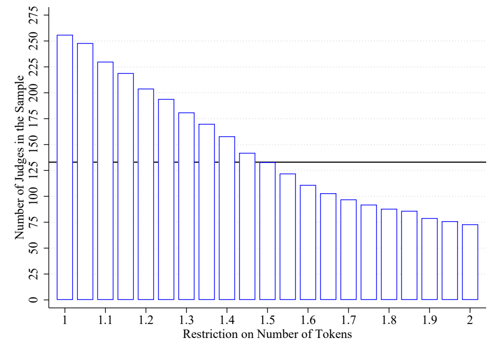
Notes: The graphs show how the differential effect of slanted judges on the reversal probability of cases originally decided by male and female district judges varies over time. The graph shows both the point estimate and the 95% confidence interval for three periods: before 1990, 1990-1999, and after 2000. We regress an indicator variable equal to 1 if the judge voted to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge is female and dummies for time period, demographic controls interacted with an indicator variable for whether the district judge is female, circuit judge fixed effects, district judge fixed effects and circuit-year fixed effects (equation (3)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. The sample is restricted to cases for which we were able to determine the gender of the district judge. Standard errors are clustered at the judge level.

Appendix Figure 14: Slanted Judges and Reversals, Robustness to EB-Adjustment and Token Threshold

(a) Point Estimates and 95% CI



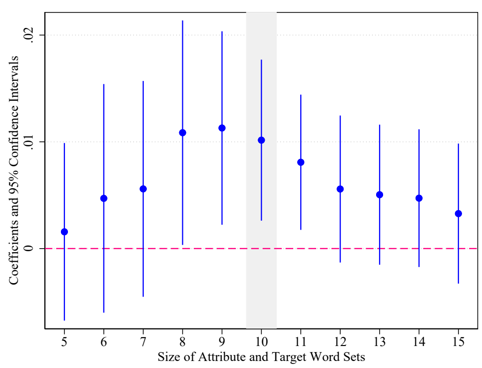
(b) Number of Judges



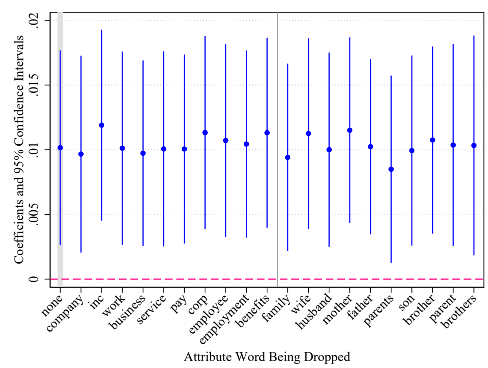
Notes: The graphs show how the differential effect of slanted judges on the reversal probability of cases originally decided by male and female district judges varies based on the token thresholds used to select the sample. The graph on the left shows both the point estimate and the 95% confidence interval for the baseline measure of gender slant and EB-adjusted gender slant, estimated including judges selected using different token thresholds. The graph on the right shows the number of judges included in the analysis for each token threshold. We regress an indicator variable equal to 1 if the judge voted to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge is female, demographic controls interacted with an indicator variable for whether the district judge is female, circuit judge fixed effects, district judge fixed effects and circuit-year fixed effects (equation (3)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. The sample is restricted to cases for which we were able to determine the gender of the district judge. Standard errors are clustered at the judge level.

Appendix Figure 15: Slanted Judges and Reversals, Robustness to Word Set Choice

(a) Size of Word Set

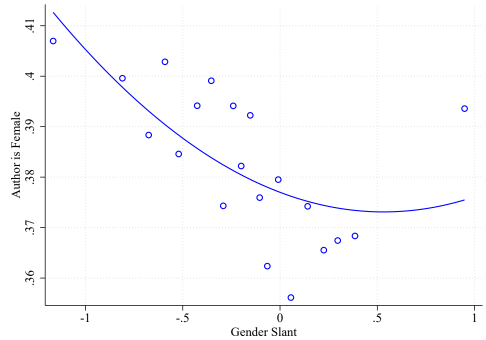


(b) Dropping Attribute Words



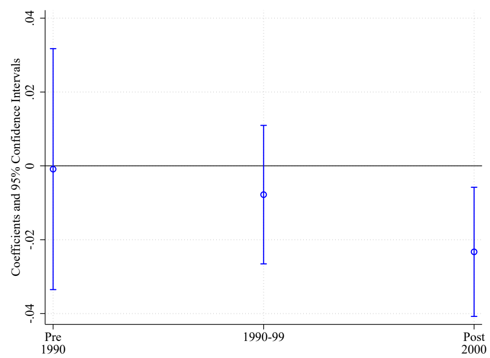
Notes: The graphs show how the differential effect of slanted judges on the reversal probability of cases originally decided by male and female district judges varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the right shows robustness to dropping one attribute word at the time. The graphs show the coefficient on gender slant interacted with an indicator variable for whether the district judge is female, together with 95% confidence intervals. We regress an indicator variable equal to 1 if the judge voted to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge is female, demographic controls interacted with an indicator variable for whether the district judge is female, circuit judge fixed effects, district judge fixed effects and circuit-year fixed effects (equation (3)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and refer to the circuit judge. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. The sample is restricted to cases for which we were able to determine the gender of the district judge. Standard errors are clustered at the circuit judge level.

Appendix Figure 16: Slanted Judges and Opinion Assignment, Binned Scatterplot



Notes: The graph shows a binned scatterplot of the relationship between gender slant and the probability of the majority opinion beign assigned to a female judge, conditional on demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously.

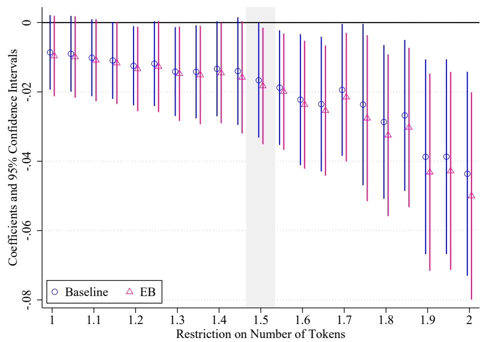
Appendix Figure 17: Slanted Judges and Opinion Assignment, Over Time



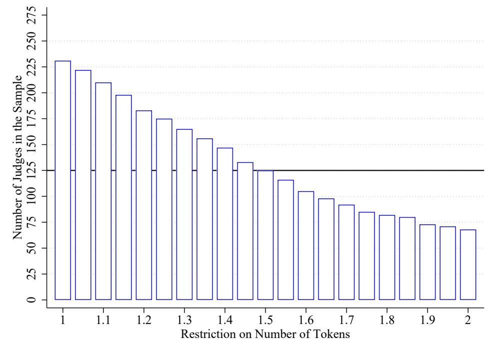
Notes: The graph shows how the effect of slanted judges on the probability of the majority opinion being assigned to a female judge varies over time. The graph shows both the point estimate and the 95% confidence interval for three periods: before 1990, 1990-1999, and after 2000. We regress an indicator variable equal to 1 if the authoring judge is female on the gender slant of the most senior judge on the panel interacted with dummies for time period, demographic controls, and circuit-year fixed effects (equation (4)). The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously. Standard errors are clustered at the judge level.

Appendix Figure 18: Slanted Judges and Opinion Assignment, Robustness to EB-Adjustment and token threshold

(a) Point Estimates and 95% CI



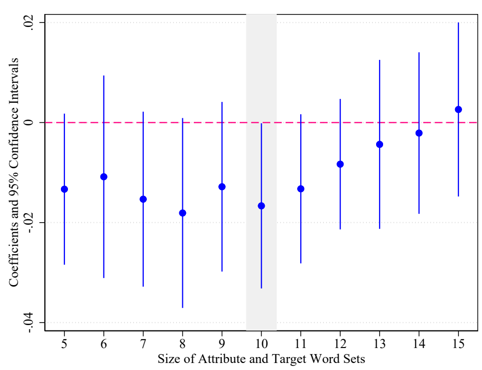
(b) Number of Judges



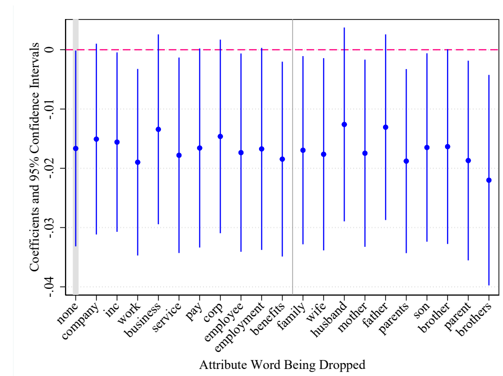
Notes: The graphs show how the effect of slanted judges on the probability of the majority opinion being assigned to a female judge varies based on the token thresholds used to select the sample. The graph on the left shows both the point estimate and the 95% confidence interval for the baseline measure of gender slant and EB-adjusted gender slant, estimated including judges selected using different token tresholds. The graph on the right shows the number of judges included in the analysis for each token threshold. We regress an indicator variable equal to 1 if the authoring judge is female on the gender slant of the most senior judge on the panel, demographic controls, and circuit-year fixed effects (equation (4)). The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously. Standard errors are clustered at the judge level.

Appendix Figure 19: Slanted Judges and Opinion Assignment, Robustness to Word Set Choice

(a) Size of Word Set

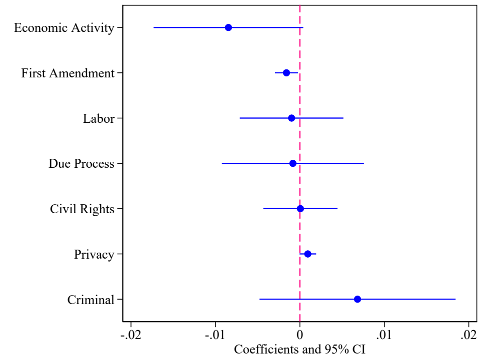


(b) Dropping Attribute Words



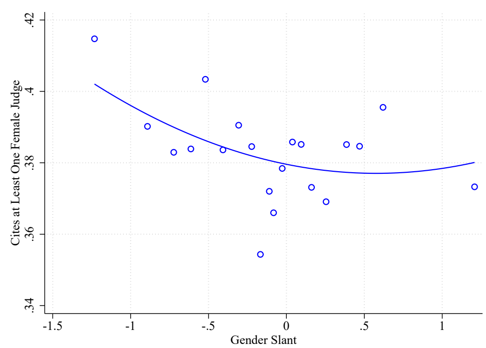
Notes: The graphs show how the effect of slanted judges on the probability of the majority opinion being assigned to a female judge varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the right shows robustness to dropping one attribute word at the time. The graphs show the coefficient on gender slant, together with 95% confidence intervals. We regress an indicator variable equal to 1 if the authoring judge is female on the gender slant of the most senior judge on the panel, demographic controls, and circuit-year fixed effects (equation (4)). The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and refer to the most senior judge on the panel. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously. Standard errors are clustered at the judge level.

Appendix Figure 20: Differential Effect of Slanted Judges on Case Topic by Author's Gender



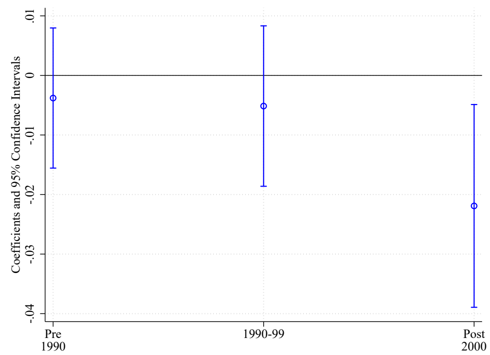
Notes: The graph explores whether slanted judges assign different cases to female judges. We regress an indicator variable equal to 1 if the case is on one of seven topics on an indicator variable for whether the opinion is assigned to a female judge, the gender slant of the most senior judge on the panel interacted with an indicator variable for whether the opinion is assigned to a female judge, demographic controls for the most senior judge interacted with an indicator variable for whether the opinion is assigned to a female judge, senior judge fixed, and circuit-year fixed effects. The graphs show the coefficient on gender slant interacted with an indicator variable for whether the opinion is assigned to a female judge, together with 95% confidence intervals. The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously. Standard errors are clustered at the senior judge level.

Appendix Figure 21: Slanted Judges and Citations, Binned Scatterplot



Notes: The graph shows a binned scatterplot of the relationship between gender slant and the probability of citing at least one female judge, conditional on demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. The sample is restricted to cases in which the opinion was authored by a specific judge.

Appendix Figure 22: Slanted Judges and Citations, Over Time

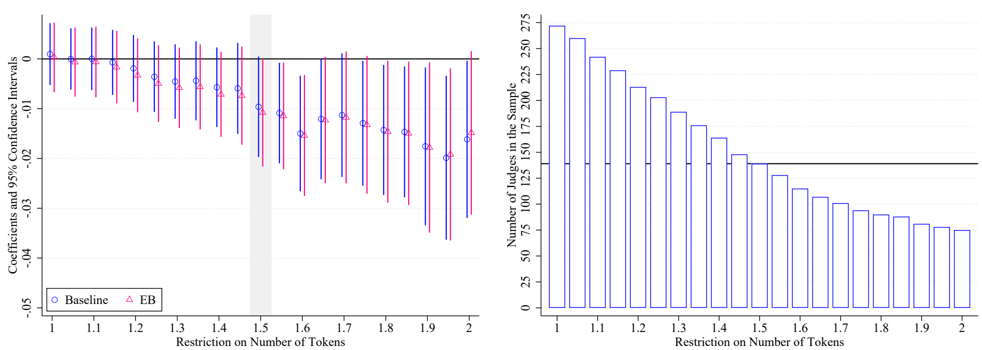


Notes: The graphs show how the effect of slanted judges on the probability of citing at least one female judge varies over time. The graph shows both the point estimate and the 95% confidence interval for three periods: before 1990, 1990-1999, and after 2000. We regress a dummy equal to 1 if the majority opinion cites at least one case authored by a woman on the gender slant of the author of the majority opinion interacted with dummies for time period, demographic controls, and circuit-year fixed effects (equation (5)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the case level. The sample is restricted to cases in which the opinion is authored by a specific judge. Standard errors are clustered at the judge level.

Appendix Figure 23: Slanted Judges and Citations, Robustness to EB-Adjustment and token threshold

(a) Point Estimates and 95% CI

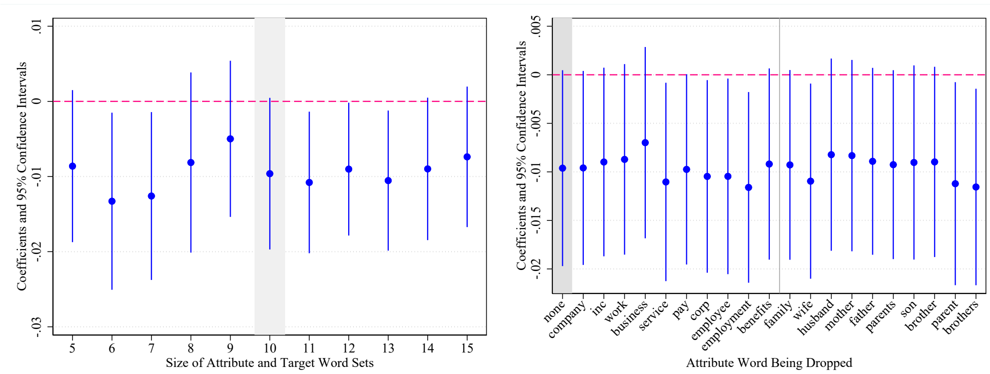
(b) Number of Judges



Notes: The graphs show how the effect of slanted judges on the probability of citing at least one female judge varies based on the token thresholds used to select the sample. The graph on the left shows both the point estimate and the 95% confidence interval for the baseline measure of gender slant and EB-adjusted gender slant, estimated including judges selected using different token tresholds. The graph on the right shows the number of judges included in the analysis for each token threshold. We regress a dummy equal to 1 if the majority opinion cites at least one case authored by a woman on the gender slant of the author of the majority opinion, demographic controls, and circuit-year fixed effects (equation (5)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the case level. The sample is restricted to cases in which the opinion is authored by a specific judge. Standard errors are clustered at the judge level.

Appendix Figure 24: Slanted Judges and Citations, Robustness to Word Set Choice

(a) Size of Word Set **(b)** Dropping Attribute Words



Notes: The graphs show how the effect of slanted judges on the probability of citing at least one female judge varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the right shows robustness to dropping one attribute word at the time. The graphs show the coefficient on gender slant, together with 95% confidence intervals. We regress a dummy equal to 1 if the majority opinion cites at least one case authored by a woman on the gender slant of the author of the majority opinion, demographic controls, and circuit-year fixed effects (equation (5)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases in which the opinion is authored by a specific judge. Standard errors are clustered at the judge level.

B Appendix Tables

Appendix Table 1: Median Word Count by Concept

Concept	Median Word Count
Male	1,114,493
Female	24,563
Career	359,683.5
Family	44,925
Positive	43,651
Negative	73,200
Art	12,399
Science	5,117.5

Notes: The table shows the median number of times that words used to define the gender, career-family, positive-negative, and art-science dimensions appear in the full judicial corpus.

Appendix Table 2: Examples of Empathic Passages from Human Coding of Opinions

It is obvious from Davis's distraught journal entries that these incidents upset her and made it more difficult for her to work.
The Sheriff's Department's response was an institutional shrug of the shoulders. It neither investigated further nor did it discipline Gamble. Instead, in response to Smith's request that further action be taken, one Investigator Sullivan made light of the incident and jokingly suggested that Smith should "kiss and make up" with Gamble.
DiTusa walked past Plaintiff, paced around the common area of the trailer, swearing loudly. He returned to the office and glared at Plaintiff. Plaintiff feared for her safety.
Valentine alleged that Tominello's harassment was very frequent: he rubbed his crotch in front of her nearly every day; asked her on twenty occasions to leave her fiance'; asked her on dates between 30 and 40 times; made repeated comments about her "tits" and "ass"; and on six occasions rubbed Valentine's arm or shoulder. Valentine also alleged that Tominello's behavior was humiliating.
On the other hand, by Hostetler's description (which at this point is undisputed), the new assignment brought with it a lengthy commute and a marathon work schedule.
Cynthia Stoll was sexually harassed, raped, and abused by supervisors and coworkers at the Sacramento Post Office. As a result of the defendant's plainly wrongful conduct, Stoll was severely psychiatrically impaired. She presented compelling direct evidence, which the district court failed to consider, that this impairment interfered with her relationship with her lawyer and rendered her unable to communicate with him or to protect her legal rights.
Rowe testified to a constant fear of Moore and to experiencing panic attacks variously characterized by nausea, headaches, sweating, and hyperventilation. She was so afraid of Moore that she moved to a different home, obtained a gun card, purchased mace, and since June of 2000 has been eating lunch and taking coffee breaks in the women's restroom to avoid any contact with Moore.

Appendix Table 3: Text Snippets

male and career	"there is no question here that neither the trustee nor mrs coggin executed *service* on coggin *himself*"
	"*he* then contracted with manhattan consolidated gold mines, *inc*"
	"kosereis was required to *work* in a particular building that *he* says lacked ventilation and was dirty"
	"eventually, *he* ordered white back to *work* in the infirmary, however"
	"after being joined by other officers, they cornered *mr* avery on a crowded street in the town's *business* district"
	"if *he* failed to make any payment, *he* forfeited the *business* plus any payments made before the default"
	"talbert left the management conference about noon and returned to *his* regular post of *work*"
	"at the same time, *he* was continuing full time secular *employment*"
	"in 1986, *he* had to stop *work* because of the back pain did you do any *work* on *his* appeal?"
female and career	"1291, this court affirms. adt hired harris as a customer *service* specialist in 1997, promoting *her* to team manager the next year"
	"from 1980 until shortly before the end of *her employment* on august 23, 1987, *she* worked as an "extra board" *employee*"
	"however, *her* condition remained of such severity as to preclude her from engaging in sedentary *work*"
	"*she* said *she* did not feel well enough to return to *work*"
	"neither of the decisions cited by *her* involved an *employee* who was receiving owcp *benefits*"
	"this effort was due in part to mrs arlinghaus' need for cash to *pay* the federal tax on *her* husband's estate"
	"metlife's letter outlined its reasoning for denying *her benefits* under the any occupation period"
	"an *employee* is deemed qualified only if she can perform all of the essential functions of *her* job, whether accommodated or not"
	"the resume contained *her* home and *work* addresses and telephone numbers"
male and family	"before trial, appellant's *husband* died and appellant, as administratrix of *his* estate, was substituted as plaintiff in *his* stead"
	"reynolds told defendant mcpheters that raymond lived with *his mother*"
	"sultan refused to bring *his son* to the police because the *family* was ashamed of the sexual abuse"
	"on may 1, 1942, delfino ferdinando cinelli died, leaving *his* estate of spannocchia to *his wife* and children"
	"in support of *his* claim, lambros refers to the government's agreement not to prosecute *his wife*"
	"at the last january term, as my learned *brother* informs me, *he* intimated that the case, in *his* opinion, was against the defendant"
	"syed reports that each time *he* returned to hyderabad *he* was told that *he* would be killed if *he* left his *wife* again"
	"holloman, *wife* of the plaintiff, and for *his* use"
	"*he* testified bolyard told abigando that they knew either *he* or *his wife* had "some connection" with the mustang"
female and family	"further, the alj did not believe cox, or *her husband* and neighbor, who both testified at cox's hearing on *her* behalf"
	"mrs willging and *her husband* were wheat farmers, owning community property, and reporting their income on the accrual basis"
	"*she* points to the incidents involving *her* father, *her mother* and *her* father's associates"
	"*she* does aver that some of the personnel in the entities sued by *her father* were high-ranking officials within the government"
	"no evidence was presented to show that mrs gordon intended to inculpate *her husband* falsely"
	"whitten testified, however, that tilley was not the *father* and claimed that *she* had lied on the birth certificate"
	"told *her mother* about problems at eagle's house"
	"ms johnson, accompanied by members of *her family* and jonathan young, went to a grocery store with a western union office"
	"second, during the assault, natasha had been subjected to physical abuse and death threats made against *her* and *her family*"

Notes: This table report randomly selected text snippets used to create the count-based measure as an example. Words that are part of the selected sets used to construct the gender slant measure are in italics.

Appendix Table 4: Correlates of Having a Sufficiently Large Corpus

Dependent Variable	Tokens >= 1.5m		
	(1)	(2)	(3)
Democrat	0.004	-0.011	-0.011
	(0.021)	(0.021)	(0.021)
Female	-0.027	0.003	0.021
	(0.048)	(0.046)	(0.045)
Minority	-0.049	0.000	0.030
	(0.050)	(0.049)	(0.047)
Born in 1920s	0.311***	0.253***	0.257***
	(0.050)	(0.051)	(0.050)
Born in 1930s	0.383***	0.315***	0.295***
	(0.052)	(0.055)	(0.054)
Born after 1940	0.121***	0.107***	0.099***
	(0.035)	(0.036)	(0.043)
Observations	951	951	951
Adjusted R2	0.138	0.190	0.231
Additional Controls		X	X
Circuit FE			X

Notes: The table shows what demographic characteristics correlate with the judge having a sufficiently large corpus to be included in the main sample. In column (1) we regress an indicator variable equal to one if the judge's corpus includes more than 1.5m tokens (i.e., the judge is included in the sample) on gender, party of appointing President, race (i.e. whether minority), region of birth, cohort of birth. Column (2) additionally controls for region of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (3) include circuit fixed effects. The sample is composed of 951 judges who served in circuit courts 1890-2013. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 5: Randomization Check

Dependent Variable	Topic				District Judge			
	Criminal	Civil	Economic	Gender-Related	Female	Democrat	Old	Minority
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Slant	-0.001	0.001	0.001	-0.000	-0.003**	-0.001	0.003	-0.003
	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)
Democrat	-0.003	-0.003	0.002	-0.001	-0.002	0.003	-0.002	0.005
	(0.003)	(0.002)	(0.003)	(0.001)	(0.003)	(0.004)	(0.004)	(0.005)
Female	0.007*	-0.001	-0.006*	0.001	0.000	-0.006	0.004	-0.006
	(0.004)	(0.003)	(0.004)	(0.001)	(0.004)	(0.005)	(0.005)	(0.005)
Observations	399141	399141	399141	147124	148881	148881	148881	148881
Clusters	139	139	139	118	133	133	133	133
Outcome Mean	0.245	0.320	0.277	0.017	0.119	0.398	0.547	0.124
Circuit-Year FE	X	X	X	X	X	X	X	X
Controls for Demographics	X	X	X	X	X	X	X	X

Notes: The table provides suggestive evidence supporting the hypothesis that cases are quasi-randomly assigned to judges within each circuit-year. We regress indicator variables equal to 1 if the case has a certain characteristic on the gender slant of the judge, demographic controls, and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the vote level. Column (4) restricts the sample to cases after 1995. Columns (5) to (8) restrict the sample to cases that were matched to a district judge. Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 6: Oster Test for Selection on Unobservables

β Uncontrolled	β Controlled	$\overline{R_{max}}$	δ for $\beta = 0$
(1)	(2)	(3)	(4)
Panel (a) Decisions in Gender Related Cases			
-0.041	-0.042	0.057	3.556
		0.076	1.893
		0.191	0.497
Panel (b) Disparities in Reversals			
0.006	0.010	0.000	43.772
		0.000	16.585
		0.001	4.529
Panel (c): Disparities in Opinion Authorship			
-0.028	-0.017	0.025	0.877
		0.033	0.445
		0.082	0.113
Panel (d): Disparities in Citations			
-0.024	-0.010	0.022	0.670
		0.030	0.338
		0.074	0.085

Notes: The table shows the results from applying the method proposed by Oster (2019) to assess bias from unobservables based on selection on observables. Column (1) shows the estimate of the coefficient on gender slant from the uncontrolled specification, in which we drop demographic controls. Column (2) shows the estimate of the coefficient from the baseline specification (equation (2)). Based on the recommendations in Oster (2019), we perform the test for three different levels of R_{max} (the $\tilde{R}^2$ of a regression that included all unobservable charactestics): 1.5, 2, and 5 times the of the controlled regression. For each value of R_max, we compute the degree of selection on unobservables as a proportion of the selection on observables that would be needed to obtain a bias-adjusted coefficient equal to 0 (δ). Column (3) shows the value of R_max for which δ (column (4)) is computed.

Appendix Table 7: Slanted Judges and Decisions in Gender-Related Cases, by Dataset

Dependent Variable	Conservative Vote	
Dataset	Epstein et al.	Glynn-Sen
	(1)	(2)
Gender Slant	0.037***	0.041*
	(0.014)	(0.024)
Democrat	-0.145***	-0.086
	(0.026)	(0.053)
Female	-0.054	0.033
	(0.033)	(0.056)
Observations	2335	738
Clusters	112	104
Outcome Mean	0.583	0.673
Circuit-Year FE	X	X
Additional Controls	X	X

Notes: The table shows the effect of slanted judges on decisions in gender-related cases, separately by dataset. We regress an indicator variable equal to 1 if the judge voted conservatively in a gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 8: Slanted Judges and Decisions in Gender-Related Cases, Additional Robustness Checks

Dependent Variable	Conservative Vote				
	(1)	(2)	(3)	(4)	(5)
Gender Slant	0.040***	0.033*	0.039***	0.033***	0.040***
	(0.013)	(0.017)	(0.012)	(0.014)	(0.013)
Democrat	-0.158***	-0.130***	-0.158***	-0.155***	-0.138***
	(0.031)	(0.032)	(0.026)	(0.026)	(0.027)
Female	-0.009	-0.021	-0.024	-0.032	-0.029
	(0.035)	(0.038)	(0.032)	(0.033)	(0.033)
Share Female Clerks	0.006				
	(0.088)				
Log Tokens				-0.068**	
				(0.031)	
Conservative Score					0.060
					(0.098)
Observations	2348	2348	3086	3086	3078
Clusters	72	91	113	113	111
Outcome Mean	0.612	0.626	0.606	0.606	0.606
Circuit-Year FE	X	X	X	X	X
Additional Controls	X	X	X	X	X
Drops 2nd, 8th, 9th, and D.C. Circuits		X			
Weights by Inverse of Slant Variance			X		

Notes: The table shows the robustness of the effect of slanted judges on decisions in gender-related cases. We regress an indicator variable equal to 1 if the judge voted conservatively in a gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (1) additionally controls for the share of clerks that are female. Column (2) drops the 2nd, 8th, 9th, and D.C. circuits. Column (3) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample. Column (4) additionally controls for the log number of tokens in a judge's corpus, while column (5) for the judge's share of conservative votes in non gender-related cases from the Epstein et al. (2013) data. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Data on votes on gender-related cases are from Epstein et al. (2013)'s update of Sunstein's (2006) data and Glynn and Sen (2015). Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 9: Slanted Judges and Decisions in Non-Gender-Related Cases

Dependent Variable	Conservative Vote			
	(1)	(2)	(3)	(4)
Gender Slant	0.027**	0.027***	0.009	0.018*
	(0.012)	(0.012)	(0.013)	(0.010)
Democrat	-0.070***	-0.075***	-0.059***	-0.070***
	(0.020)	(0.020)	(0.020)	(0.018)
Female	-0.060**	-0.046*	-0.075***	-0.067***
	(0.026)	(0.024)	(0.020)	(0.024)
Observations	5477	5477	5477	5477
Clusters	112	112	112	112
Outcome Mean	0.569	0.569	0.569	0.569
Circuit-Year FE	X	X	X	X
Additional Controls	X	X	X	X
Year of Appointment		X		
Exposure FE			X	
No Gender-Related Cases				X

Notes: The table shows the effect of slanted judges on decisions in non-gender-related cases. We regress an indicator variable equal to 1 if the judge voted conservatively in a gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (2) controls for year of first appointment of the judge to a circuit court. Column (3) includes exposure fixed effects, which are indicator variables equal to 1 if the judge sat on at least one panel in a given circuit over a given 25-year period. In column (4), gender slant is calculated using embeddings trained excluding gender-related cases. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 10: Slanted Judges and Decisions in Gender-Related and Non-Gender-Related Cases, Differences-in-Differences Specification

Dependent Variable	Conservative Vote
	(1)
Gender-Related	0.0087
	(0.080)
Gender Slant * Gender-Related	0.027**
	(0.013)
Democrat * Gender-Related	-0.084***
	(0.029)
Female * Gender-Related	0.031
	(0.039)
Observations	8565
Clusters	113
Outcome Mean	0.582
Circuit-Year FE	X
Judge FE	X

Notes: The table tests whether slanted judges are more likely to vote conservatively in gender-related rather than in non-gender-related cases. We regress an indicator variable equal to 1 if a judge voted conservatively in a gender-related case on the gender slant of the judge interacted with an indicator variable for the case being gender-related, demographic controls interacted with an indicator variable for the case being gender-related, judge fixed effects, and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 11: Slanted Judges and Decisions in All Cases, Songer Data

Dependent Variable	Conservative Vote			
	(1)	(2)	(3)	(4)
Gender Slant	0.003	0.004	0.007	-0.002
	(0.007)	(0.007)	(0.012)	(0.006)
Democrat	-0.027**	-0.024**	-0.010	-0.028**
	(0.012)	(0.012)	(0.016)	(0.012)
Female	0.025	0.021	0.003	0.023
	(0.020)	(0.020)	(0.018)	(0.020)
Observations	13420	13420	13420	13420
Clusters	136	136	136	136
Outcome Mean	0.620	0.620	0.620	0.620
Circuit-Year FE	X	X	X	X
Additional Controls		X	X	X
Controls for Year of Appointment		X		
Includes Exposure FEs			X	
No Gender-Related Cases				X

Notes: The table shows the effect of slanted judges on decisions in all cases. We regress an indicator variable equal to 1 if the judge voted conservatively in a non-gender-related case on the judge's gender slant, demographic controls, and circuit-year fixed effects (equation (2)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (2) controls for year of first appointment of the judge to a circuit court. Column (3) includes exposure fixed effects, which are indicator variables equal to 1 if the judge sat on at least one panel in a given circuit over a given 25-year period. In column (4), gender slant is calculated using embeddings trained excluding gender-related cases. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. Data on votes are from the U.S. Court of Appeal Dataset (Songer, 2008; Kuersten and Haire, 2011). Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 12: Slanted Judges and Reversals, Additional Robustness Checks

Dependent Variable	Votes to Reverse District Decision						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender Slant * F District Judge	0.012***	0.009**	0.012***	0.010**	0.015**	0.016***	0.011***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Democrat * F District Judge	-0.010	-0.023**	-0.010	-0.010	-0.023**	-0.012*	-0.008
	(0.006)	(0.009)	(0.007)	(0.006)	(0.009)	(0.007)	(0.006)
Female * F District Judge	-0.005	0.018	-0.000	-0.003	-0.004	-0.003	-0.004
	(0.011)	(0.012)	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)
Share Fem. Clerks * F District Judge		0.035					
		(0.039)					
Conservative Score * F District Judge			0.007				
			(0.026)				
Log Tokens * F District Judge							0.010
							(0.011)
Observations	145863	83751	129677	130381	96637	145862	145862
Clusters	133	68	106	119	110	133	133
Outcome Mean for Male Judges	0.180	0.163	0.167	0.168	0.158	0.180	0.180
Outcome Mean for Female Judges	0.157	0.151	0.157	0.157	0.135	0.157	0.157
Circuit-Year FE	X	X	X	X	X	X	
Circuit Judge FE	X	X	X	X	X	X	
District Judge FE	X	X	X	X	X	X	
Additional Controls	X	X	X	X	X	X	
District -Year FE	X						
After 1980				X			
Drops 2nd, 8th, 9th, and D.C. Circuits					X		
Weights by Inverse of Slant Variance						X	

Notes: The table shows the robustness of the differential effect of slanted judges on the reversal probability of cases originally decided by male and female district judges. We regress an indicator variable equal to 1 if the judge voted to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge is female, demographic controls interacted with an indicator variable for whether the district judge is female, circuit judge fixed effects, district judge fixed effects and circuit-year fixed effects (equation (3)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and refer to the circuit judge. Column (1) includes district-year fixed effects. Column (2) additionally controls for the share of clerks that are female, and column (3) for the share of conservative votes of the judge in non gender-related cases from the Epstein et al. (2013) data both interacted with an indicator variable for the district judge being female. Column (4) restricts the sample to cases decided after 1980, and column (5) drops cases decided in the 2nd, 8th, 9th, and D.C. circuits. Column (6) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample, while column (7) controls for the log number of tokens in a judge's corpus interacted with an indicator variable for the district judge being female. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. The sample is restricted to cases for which we were able to determine the identity of the district judge. Standard errors are clustered at the circuit judge level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 13: Reversals and Promotion from District to Circuit Courts

Dependent Variable	Elevated to Circuit Court	
	(1)	(2)
Share of Decisions Reversed on Appeal	-0.345**	
	(0.136)	
Share of Votes to Reverse on Appeal		-0.367***
		(0.116)
Female	0.030	0.031
	(0.028)	(0.028)
Democrat	-0.001	0.002
	(0.018)	(0.018)
Observations	862	862
Outcome Mean	0.058	0.058
Circuit FE	X	X
Additional Controls	X	X

Notes: The table shows the relationship between reversals and promotion of judges from district to circuit courts. We regress an indicator variable equal to 1 if the judge was elevated to a circuit court on the share of decisions that were reversed on appeal (column (1)) or the share of circuit judges that voted to reverse the decision (column (2)), demographic controls and circuit fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. The sample is restricted to district judges for which we observe at least 50 cases (this requires that the case was appealed, and that we were able to match the circuit court case to the respective district judge). Standard errors are clustered at the district judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 14: Slanted Judges and Whether the Opinion has Specific Author, or the Opinion is Per Curiam

Dependent Variable	Has Author		Per Curiam		Decided Unanimously	
	(1)	(2)	(3)	(4)	(5)	(6)
Gender Slant	0.002	0.003	-0.000	-0.001	0.001	0.001
	(0.005)	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)
Democrat	0.002	-0.010	-0.007	0.003	-0.017	-0.002
	(0.011)	(0.010)	(0.006)	(0.006)	(0.010)	(0.009)
Female	-0.001	0.013	0.005	-0.004	0.020*	0.009
	(0.011)	(0.010)	(0.005)	(0.004)	(0.010)	(0.010)
Observations	171441	43601	171441	43601	171441	43601
Clusters	139	125	139	125	139	125
Outcome Mean	0.803	0.847	0.092	0.045	0.887	0.874
Circuit-Year FE	X	X	X	X	X	X
Controls for Demographics	X	X	X	X	X	X
One Female Judge on Panel		X		X		X

Notes: The table shows the effect of slanted judges on whether the opinion has a specific author or is per curiam, and on whether the decision was unanimous. We regress an indicator variable equal to 1 if the opinion has a specific author (columns (1) and (2)), if the opinion is per curiam (columns (3) and (4)), or if the panel decided unanimously (columns (5) and (6)) on the gender slant of the most senior judge on the panel, demographic controls, and circuit-year fixed effects. The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. In columns (2), (4), and (6) the sample is restricted to cases with at least one female judge on the panel. Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 15: Slanted Judges and Opinion Assignment, Additional Robustness Checks

Dependent Variable	Author is Female							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Slant	-0.016	-0.018**	-0.012*	-0.020**	-0.017**	-0.028***	-0.029***	-0.020**
	(0.011)	(0.009)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)
Democrat	-0.016	-0.002	0.001	-0.001	-0.001	0.001	0.005	-0.006
	(0.025)	(0.017)	(0.014)	(0.013)	(0.014)	(0.021)	(0.015)	(0.015)
Female	0.172***	0.134***		0.133***	0.134***	0.171***	0.126***	0.140***
	(0.018)	(0.017)		(0.017)	(0.016)	(0.026)	(0.017)	(0.017)
Share Female Clerks	-0.021							
	(0.055)							
Conservative Score		0.023						
		(0.040)						
Log Tokens								-0.048***
								(0.023)
Observations	20543	30614	22828	36939	31998	21380	32052	32052
Clusters	72	111	108	125	124	80	125	125
Outcome Mean	0.396	0.387	0.347	0.383	0.383	0.383	0.383	0.383
Circuit-Year FE	X	X	X	X	X	X	X	X
Controls for Demographics	X	X	X	X	X	X	X	X
Excludes Female Senior Judges			X				X	
Includes Dissents/Concurrences				X				
After 1980					X			
Drops 2nd, 8th, 9th, and D.C. Circuits						X		
Weights by Inverse of Slant Variance							X	

Notes: The table shows robustness of the effect of slanted judges on the probability of assigning the majority opinion to a female judge. We regress an indicator variable equal to 1 if the authoring judge is female on the gender slant of the most senior judge on the panel, demographic controls, and circuit-year fixed effects (equation (4)). The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. Column (1) additionally controls for the share of clerks that are female, and column (2) for the share of conservative votes of the judge in non gender-related cases from the Epstein et al. (2013) data. Column (3) excludes panels in which the most senior judge is female. Column (4) does not restrict the sample to cases decided unanimously, but includes cases with dissents or concurrences. Column (5) restrict the sample to post-1980 cases, and column (6) drops cases decided in the 2nd, 8th, 9th, and D.C. circuits. Column (7) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample, while column (8) controls for the log number of tokens in a judge's corpus. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously (with the exception of column (5). Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 16: Differential Effect of Slanted Judges on Predicted Case Importance by Author's Gender

Dependent Variable	Predicted Forward Citations
	(1)
Gender Slant * Female Author	0.001
	(0.002)
Democrat * Female Author	0.005
	(0.009)
Female * Female Author	-0.010
	(0.009)
Observations	31616
Clusters	123
Outcome Mean	1.726
Circuit-Year FE	X
Judge FE	X
Additional Controls	X

Notes: The table tests whether slanted judges assign different types of cases to female judges. We regress predicted forward citations on an indicator variable for whether the opinion is assigned to a female judge, the gender slant of the most senior judge on the panel interacted with an indicator variable for whether the opinion is assigned to a female judge, demographic controls for the most senior judge interacted with an indicator variable for whether the opinion is assigned to a female judge, senior judge fixed, and circuit-year fixed effects. The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that were decided unanimously. Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 17: Slanted Judges and Citations, Additional Robustness Checks

Dependent Variable	Cites at Least One Female Judge						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender Slant	-0.006	-0.009	-0.010	-0.006	-0.009*	-0.008	-0.006
	(0.009)	(0.006)	(0.006)	(0.008)	(0.005)	(0.006)	(0.005)
Democrat	-0.038**	-0.020	-0.013	-0.017	-0.027***	-0.012	-0.007
	(0.018)	(0.013)	(0.012)	(0.012)	(0.010)	(0.011)	(0.010)
Female	0.157***	0.125***	0.128***	0.135***	-0.084**	0.138***	0.126***
	(0.022)	(0.017)	(0.017)	(0.022)	(0.018)	(0.017)	(0.017)
Share Female Clerks	-0.018						
	(0.042)						
Conservative Score		-0.039					
		(0.035)					
							0.034**
							(0.015)
Observations	54301	86910	83680	67497	107923	107923	107923
Clusters	73	112	125	114	139	139	139
Outcome Mean	0.536	0.452	0.487	0.413	0.383	0.383	0.383
Circuit-Year FE	X	X	X	X	X	X	X
Additional Controls	X	X	X	X	X	X	X
After 1980			X				
Drops 2nd, 8th, 9th, and D.C. Circuits				X			
Excludes Self-Citations					X		
Weights by Inverse of Slant Var.						X	X

Notes: The table shows the robustness of the effect of slanted judges on the probability of citing at least one female judge. We regress a dummy equal to 1 if the majority opinion cites at least one case authored by a woman on the gender slant of the author of the majority opinion, demographic controls, and circuit-year fixed effects (equation (5)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Column (1) additionally controls for the share of clerks that are female, and column (2) for the judge's share of conservative votes in non gender-related cases from the Epstein et al. (2013) data. Column (3) restricts the sample to cases decided after 1980, and column (4) drops cases decided in the 2nd, 8th, 9th, and D.C. circuits. Column (5) defines the outcome excluding self-citations. Column (6) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample, while column (8) controls for the log number of tokens in a judge's corpus. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases in which the opinion is authored by a specific judge. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 18: Slanted Judges and Reversals, Characteristics other than Gender

Dependent Variable	Votes to Reverse District Decision	
	(1)	(2)
Gender Slant * Democrat District Judge	0.005	
	(0.004)	
Democrat * Democrat District Judge	-0.006	
	(0.007)	
Female * Democrat District Judge	-0.003	
	(0.010)	
Gender Slant * Minority District Judge		0.0112**
		(0.005)
Democrat * Minority District Judge		0.002
		(0.007)
Female * Minority District Judge		0.016
		(0.011)
Observations	145862	145862
Clusters	133	133
Outcome Mean	0.177	0.177
Circuit-Year FE	X	X
Circuit Judge FE	X	X
District Judge FE	X	X
Additional Controls	X	X

Notes: The table shows the differential effect of slanted judges on the reversal probability of cases originally decided by district judges with different characteristics. We regress an indicator variable equal to 1 if the judge voted to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge was appointed by a Democratic President (column (1)) or is a minority (column (2)), demographic controls interacted with an indicator variable for whether the district judge was appointed by a Democratic President (column (1)) or is a minority (column (2)), circuit judge fixed effects, district judge fixed effects and circuit-year fixed effects (similar to equation (3)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and refer to the circuit judge. Gender slant is the standardized cosine similarity between the gender and the career-family dimensions. The dataset is at the vote level. The sample is restricted to cases for which we were able to determine the identity of the district judge. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 19: Slanted Judges and Opinion Assignment, Characteristics other than Gender

Dependent Variable	Author is Democrat	Author is Minority	Author Age
	(1)	(2)	(3)
Gender Slant	-0.007	0.005	0.041
	(0.006)	(0.008)	(0.175)
Democrat	0.224***	-0.002	0.081
	(0.011)	(0.013)	(0.382)
Female	0.030	0.027*	0.056
	(0.019)	(0.016)	(0.563)
Observations	92816	23436	120365
Clusters	139	126	139
Outcome Mean	0.616	0.340	63.030
Circuit-Year FE	X	X	X
Additional Controls	X	X	X
Panel Includes Democrat Judge	X		
Panel Includes Minority Judge		X	

Notes: The table shows the the effect of slanted judges on the probability of assigning the majority opinion to judges with different characteristics. We regress an indicator variable equal to 1 if the authoring judge was appointed by a Democratic President (column (1)), if the authoring judge is minority (column (2)) and age of the authoring judge (column (3)) on the gender slant of the most senior judge on the panel, demographic controls, and circuit-year fixed effects (equation (4)). The most senior judge on the panel is customarily in charge of assigning the majority opinion. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court and they refer to the most senior judge on the panel. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases with a specific author that were decided unanimously. Column (1) additionally restricts the sample to cases with one democratic judge on the panel and column (2) to cases with one minority judge on the panel. Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 20: Slanted Judges and Citations, Characteristics other than Gender

Dependent Variable	Cites Democrat	Cites Minority	Average Age	Average Bias
	(1)	(2)	(3)	(4)
Gender Slant	-0.008**	-0.006	-0.072	0.118***
	(0.004)	(0.005)	(0.082)	(0.011)
Democrat	0.008	-0.021*	-0.071	0.011
	(0.007)	(0.011)	(0.105)	(0.016)
Female	0.023**	0.058***	0.026	-0.022
	(0.009)	(0.011)	(0.173)	(0.019)
Observations	107923	107923	107923	98435
Clusters	139	139	139	139
Outcome Mean	0.875	0.336	61.407	0.052
Circuit-Year FE	X	X	X	X
Additional Controls	X	X	X	X

Notes: The table shows the the effect of slanted judges on the probability of citing judges with different characteristics. We regress a dummy equal to 1 if the majority opinion cites at least one case authored by a judge nominated by a Democratic President (column (1)), at least one case authored by a minority judge (column (2)), the average age of the authors of cited opinions (column (3)), and the average slant of the authors of the cited opinions (column (4)) on the gender slant of the author of the majority opinion, demographic controls, and circuit-year fixed effects (equation (5)). Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to a circuit court. Gender slant is the standardized cosine similarity between the gender and career-family dimensions. The dataset is at the case level. The sample is restricted to cases in which the opinion is authored by a specific judge. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

C Robustness to Empirical Bayes Adjustment and Sample Restrictions

This appendix shows that our results are robust to applying shrinkage techniques from the Empirical Bayes literature to our gender slant measure. Because these techniques are often used to deal with noisy estimates, they are especially attractive in our setting as they might help us expand the sample to include judges with smaller corpora.

We implement the EB-adjustment procedure described in Chandra et al. (2016). For each judge, we estimate a gender slant parameter as described in Section 2. As a measure of the precision of each estimate, we use the standard deviation of slant in our twenty-five bootstrap samples. We assume that the underlying mean of slant is a function of the biographical characteristics of the judges, and only run the EB procedure only on judges with more than 1 million tokens. Appendix Figure 25 shows that as one would expect, the EB-adjusted estimates tend to shrink toward the mean slant in the sample.

In Appendix Figures 9, 14, 18, and 23, we report the coefficient estimates and 95% confidence intervals from our baseline specification estimated using the gender slant measure we use throughout the paper and the EB-adjusted estimate, for different restrictions on the number of tokens for the corpus of each judge (from 1 million tokens to 2 million tokens). To put the restriction of tokens in perspective, we also report the number of judges that are included in the baseline estimation for each token threshold.

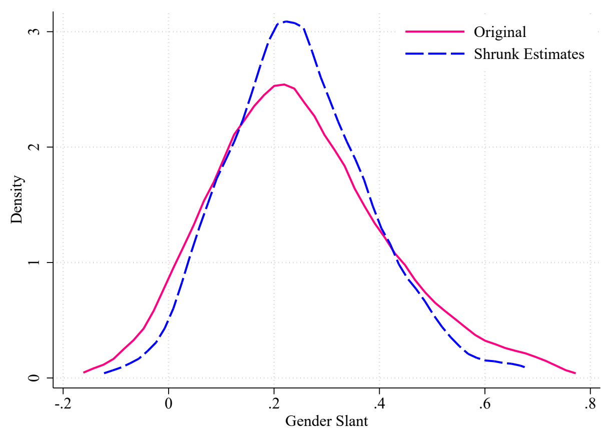
Our main results are always robust to using an EB approach to reduce noise in the estimation. However, the EB-adjustment approach is limitedly helpful in increasing our sample size. More precisley, the results on decisions and opinion assignment are robust to including all judges with more than 1 million tokens. However, the result on reversals is only robust to including all judges with more then 1.3 million tokens, and the one on citations is generally not robust to restricting the sample. Further restricting the sample to higher thresholds generally yields larger point estimates and larger confidence intervals.

Two points are worth noting. When we decrease the size of the corpus, we are not just introducing noise (which would be accounted for by the EB-adjustment), we are also introducing bias. As we reduce the size of the corpus, we lose predictive power to the point that, for judges with small corpora, the quality of the embeddings is not sufficient to even predict the gender of the very common first names (see Figure 3). It is perhaps not surprising that moving towards these lower quality embeddings gives us different results.

Second, even lowering the token threshold to 1.25 million tokens already implies a large increase in the number of judges included in the regression. While we present all estimates to readers so that they can come to an independent conclusion, we are confident that our result are not limited to a very specific group of judges.

Still, the figures suggest heterogeneity in the treatment effects, with judges with a larger number of tokens displaying larger effects of gender slant. These judges tend to be born in more recent cohorts (as Appendix Table 4) and thus are more likely to operate in recent years. Also, they tend to have lower slant (as shown in Table 3). In line with our findings when we estimate heterogeneous treatment effects over time, it is plausible that those who still express stereotyped views of gender in their writings in recent years might display especially discriminatory behavior when dealing with female colleagues. However, a potential concern is that the results are purely driven by the size of a judge's corpus. To check whether this is the case, we estimate specifications that control for the log number of tokens, a proxy for corpus size. With the exception of the effect on citations, our results are robust to this additional check (as shown in Appendix Tables 8, 12, 15, and 17).

Appendix Figure 25: Distribution of Gender Slant



Notes: The graph shows the distribution of the gender slant measure for 272 judges with corpora with more than 1,000,000 tokens, for the baseline measure of gender slant and EB-adjusted gender slant.