CAGE

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

CAGE working paper no. 462

Mar 2020 (Revised June 2020)

Elliott Ash
Daniel L. Chen
Arianna Ornaghi

This paper has been revised June 2020 previous title:

**Stereotypes in High-Stakes Decisions: Evidence from U.S. Circuit Courts**

WARWICK
THE UNIVERSITY OF WARWICK

UKRI

Economic and Social Research Council

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

Elliott Ash, ETH Zurich

Daniel L. Chen, Toulouse School of Economics

Arianna Ornaghi, University of Warwick*

June 25, 2020

**Abstract**

Do gender attitudes influence interactions with female judges in U.S. Circuit Courts? In this paper, we propose a novel 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 slanted judges vote more conservatively in gender-related cases. Slant influences interactions with female colleagues: slanted judges 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.

*Elliott Ash, ETH Zurich, ashe@ethz.ch; Daniel Chen, Toulouse School of Economics, daniel.chen@iast.fr; Arianna Ornaghi, University of Warwick, a.ornaghi@warwick.ac.uk (corresponding author). We thank Jacopo Bregolin, David Cai, Christoph Goessmann, and Ornelie Manzambi for helpful research assistance. We thank Ryan Copus, Arthur Dyevre, Mirko Draca, Patrick Kline, Emily Leslie, David Mimno, 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 in the upper echelons 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 the U.S. Circuit Courts of Appeal, the focus of our study, 75 out of 290 sitting judges (25.9%) were women in 2019 (Root, 2019). In this paper, we 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 et al., 2005). Attitudes matter even in high-stakes settings such as physician treatments (Green et al., 2007), voting (Friese et al., 2007), hiring decisions (Rooth, 2010), employer-employee interactions (Glover et al., 2017), and teacher effectiveness (Carlana, 2019). If gender attitudes imply differential treatment of female judges as well, that 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 their use of gender-stereotyped language (Pennington et al., 2014; Caliskan et al., 2017; Kozlowski et al., 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 their writing.

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;

[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 of 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 et al. (2017) collected implicit bias against Asians and Jews for 239 federal and state judges, and they found that judges biased against Jews display different hypothetical sentencing behavior.

Pennington et al., 2014). We use word embeddings because 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. Second, the relative positions of words in the space also convey 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 differences can be used to identify cultural concepts (Kozlowski et al., 2019).

We operationalize these concepts to identify gender attitudes in language. If men are associated with career and women are associated with family in the corpus, 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 ($\overrightarrow{man} - \overrightarrow{woman}$) 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 ($\overrightarrow{man} - \overrightarrow{woman}$) 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 we obtain a high-quality representation notwithstanding the small size of judge corpora, we train embeddings on different bootstrap samples of each corpus 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. Descriptively, we find that female and younger judges display lower gender slant and that having a daughter reduces gender slant as well.

This approach is related to a growing literature using word embeddings to analyze bias in text. Bolukbasi et al. (2016) and Caliskan et al. (2017) demonstrate biases in general web corpora in terms of gender attitudes 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. The closest analysis to ours is Rice et al. (2019), who detect a racial slant in a corpus of U.S. state and federal court opinions. We build on this literature to construct author-specific measures of slant that can be linked to real-world behaviors.

The central research question of the paper is how assigning a judge with different gender slant impacts the treatment of female judges. The empirical strategy to estimate the causal effect of gender slant builds on the following two features of the setting. First, we rely on the quasi-random assignment of judges to cases to ensure that slanted judges do not self-select into cases systematically based on the outcome. This means that cases assigned to higher and lower slanted judges are comparable. Second, we condition on detailed judges' characteristics to show that gender slant has an effect that goes above and beyond the judge's other features and is not simply acting as a proxy for some other characteristic such as gender or conservative ideology.

We study whether assigning a judge with different slant impacts the treatment of female judges focusing on three sets of interactions: reversals of district court decisions, opinion assignment, and citations. We find that judges with different gender slant treat female judges differently. First, we show using a differences-in-differences design that more slanted 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, corresponding to around 5% of the baseline 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 higher gender slant are less likely to assign opinion authorship to a female judge by 1.7 percentage points, or 4.5% of the baseline mean. Finally, more slanted 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 promotion opportunities, the gender attitudes measured by gender slant have the potential to hinder the career progression of female relative to male judges.

To strengthen the interpretation that gender slant is a proxy of gender attitudes, we validate the measure by studying its effect on decisions in gender-related cases. Consistent with the proposed interpretation, we find that judges with higher gender slant tend to vote more conservatively in gender-related issues (that is, against expanding women's rights). The magnitude of the effect is large: a one standard deviation increase in the gender slant of the judge 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 judges vote in appellate cases and therefore impacting 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, 2014c,a).

Finally, we investigate whether slanted judges also respond differently to non-gender characteristics. While we do find that higher-slanted judges tend to vote more conservatively in non-gender-related but ideologically divisive cases, the effect of slant 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, the paper speaks to the literature on how gender shapes the labor market outcomes of women and why (see among others Bohren et al., 2019; Bordalo et al. (2019); Card et al., 2019; Hengel, 2019; Sarsons, 2019), in particular for women employed at the top end of the earnings distribution (Bursztyn et al., 2017; Bertrand, 2013; Bertrand et al., 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.[2] 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 et al. (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

[2]The one exception is contemporaneous work by Battaglini et al. (2019), who show that random exposure to female judges on panels increases the probability of hiring a female clerk.

used in the literature.

In addition, by studying the effect of gender 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 et al., 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.[3] 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 et al., 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 data sources. Section 4 provides descriptive statistics. Section 5 shows the relationship between gender slant and judicial decisions, while Section 6 describes the relationship between gender slant and interactions with female judges. Finally, Section 7 concludes.

## 2 Gender Slant: Measuring Gender Attitudes in Text

The starting point of the paper is to construct a measure of how gender is characterized in the language used by judges, in particular whether a judge's writing displays stereotypical views of gender. 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.

[3] We could potentially construct a measure of voting disparities in gender-related cases and use this measure to study how this impacts female judges. Our setting is not well suited for this approach. In particular, as we will discuss in detail below, we are constrained to use pre-coded datasets for vote valence. This means that we have at most 50 votes per judge, which would make calculating gender bias measures based on votes challenging.

## 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 (relatively) low dimensional Euclidean space in such a way as to preserve semantic meaning (Mikolov et al., 2013; Pennington et al., 2014).

Consider the simplest way of representing language. For a given vocabulary $V$, one possibility is to represent words as one-hot-encoded 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. 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 same context have representations close to each other in the space, while words that appear rarely together 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.

Second, differences between vectors that identify directions in the space also convey meaning. As Figure 1 illustrates, going from the vector representing the word $\overrightarrow{woman}$ to the vector representing the word $\overrightarrow{man}$ means taking a step in the 'maleness' direction, and taking the same step from the vector representing the word $\overrightarrow{queen}$ will bring us close to the vector representing the word $\overrightarrow{king}$ as well.

## 2.2 GloVe Embeddings Implementation

The specific model we use is Global Vectors for Word Representation (Pennington et al., 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. 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 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 et al., 2014; Spirling and Rodriguez, 2019), 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.[4]

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, namely word2vec (Mikolov et al., 2013), while obtaining embeddings of comparable quality (Pennington et al., 2014). Given that our approach requires the training of a large number of separate embeddings, this is a particularly attractive feature for our application.

[4] Appendix Figure 3 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 ( Appendix Figure 2).

### 2.3 Word Vectors and Gender Slant

We use word embeddings to identify cultural dimensions in language (Kozlowski et al., 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, since 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 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, $\theta$ 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 the 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 $\theta \approx 90^{\circ}$, 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 ($\theta \approx 180^{\circ}$), 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 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.[5] 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 several 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 language 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

[5] 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.

women and negative attributes, which suggests limited scope for this measure to capture judge-specific variation in gender attitudes. In fact, Appendix Figure 1 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 et al. (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.

### 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 (Spirling and Rodriguez, 2019). Importantly, this constraint on corpus size is why we are not able to construct time-varying measures of gender slant.

Figure 3 shows the distribution of gender slant across judges in the sample. Note that here and in the remainder of the paper, the measure is standardized across judges for ease of interpretation. The figure shows there is significant variation in how strongly judges associate men with career and women with family in their language. To put the measure in perspective, the vertical lines on the same graph show the degree of gender slant of 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.[6]

### 2.5 Validating Judge-Specific Embeddings

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 validate the judge-specific embeddings, 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 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 4 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

[6] 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 (see the following sub-section) 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.

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 judges with a minimum corpus size of 50,000 tokens: 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 2).

### 2.6 Discussion of Alternative Measures

There are (perhaps many) other approaches that we could have used to measure gender attitudes in judicial language. In particular, these two alternatives come to mind.

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). However, this qualitative approach is somewhat subjective and therefore lacks a rigorous method of replication (Ricoeur, 1981; DiMaggio, 1997). It would also be prohibitively expensive to implement on a large scale such as we have in our corpus, which contains over 14 million sentences.

Given that the language representation provided by word embeddings is inherently built on co-occurrence of words within ten word windows, co-occurrence statistics of male and female words with career and family words can be thought of as the building blocks of the gender slant measure. In Appendix A.4, we report ten randomly selected sentences that present these co-occurrences. Consistent with this intuition, the relative co-occurrence of male and female words with career words relative to the co-occurrence of male and female words with family words is positively associated with the gender slant measure (Appendix Figure 4).

However, looking at the raw counts also shows that these types of explicit co-occurrences are

[7]For comparison, around 1000 judges served in circuit courts from 1890-2013. Of these, 475 have at least 500,000 tokens in their corpus, while 272 have at least 1,000,000.

rare, as for example the median judge only writes 38 sentences where female and family words co-occur. This is why using word embeddings is preferrable to using this or other simple count measures, as embeddings do not require words to appear directly next to each other to register an association. As a result, embeddings are able to take into account more information contained in judges' writing. Instead, a count-based approach would likely miss implicit and nuanced gender-stereotyped language and therefore provide a less precise measure.

# 3 Empirical Strategy and Additional Data Sources

## 3.1 Empirical Strategy

The main objective of the paper is to study the effect of gender slant on the behavior of judges. The empirical analysis relies on the quasi-random assignment of judges to cases – which means that slanted judges do not self-select into cases systematically based on the outcome – and on conditioning on detailed judges' biographic characteristics, which ensures that gender slant is not acting as a proxy for some other of the judges' features.

**Quasi-Random Assignment.** A major concern for identification is the endogeneous selection of judges to cases, as we might worry that slanted judges decide which panels to sit on based on the expected outcome of the case. We address this issue by exploiting 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 slant.[8]

Evidence supporting the quasi-random assignment of panels to cases has been provided by Chen and Sethi (2018), who show for a sample of gender-discrimination cases that pre-trial characteristics of the case are uncorrelated with the demographic composition of the panel, and by Chen et al. (2014), who show that the panel composition does not appear to be autocorrelated over time. Here, we provide additional evidence by showing that higher-slanted judges do not appear to be systematically hearing cases on specific topics. In particular, we perform a series of

[8] Chen and Sethi (2018) and Bowie et al. (2014) report detailed information on how the random assignment of cases to panels work, based on qualitative interviews with officials of the courts. 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 following a quasi-random process. In recent years, assignment is often done using a computer program. The program might place some limits on judges serving on the same panel and on how many cases are assigned to senior judges. It is possible that judges recuse themselves at times, and in rare occasions randomization is not used for some specialized cases (e.g. cases involving the death penalty) but they would generally do so before random assignment.

regressions where an indicator variable for a given topic (one of 89 topics) is regressed on judge gender slant and circuit-year fixed effects, with standard errors clustered by judge. Appendix Figure 5 shows that just 15 of the 89 coefficients are significant at the 10 percent level, in line with what we would expect by chance.

**Conditioning on Observable Characteristics.** Random assignment of judges to cases implies that the effect of being assigned a slanted judge is well-identified, but the effect of gender slant itself could be confounded with other judge characteristics that affect decisions or behavior. For example, male judges might be more conservative and also have higher slant. 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 decisions. In other words, identification requires that conditional on the covariates, slant is not systematically correlated with other omitted factors related to voting or treatment of female judges. In Appendix A.8, 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.

**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 (2014b). 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 the effect of gender slant on decisions, we use two existing legal datasets with hand-coded vote direction and topic, created by Epstein et al. (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.[9,10] 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, 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/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.[11] 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 145,862 circuit cases (38% 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.

[9] 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.

[10] In particular, the Epstein et al. (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.

[11] 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).

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

## 4 Gender Slant and Demographic Characteristics

This section explores descriptively how the gender slant measure varies based on judge characteristics. We begin by correlating the gender slant measure and different judge characteristics using separate univariate regressions with robust standard errors.

Table 2 reports estimates from these regressions. Column (1) shows that judges nominated by Presidents of different parties do not appear to display different levels of gender slant. Looking at the entire distribution of gender slant separately by party (Figure 5 (a)) confirms the result, and according to a Kolmogorov-Smirnov test for equality of distribution functions, the difference is not statistically significant ($p = 0.952$).

Column (2) shows that female judges display on average gender slant that is 0.5 standard deviations lower than male judges. The difference is statistically significant at the 10% level. Figure 5 (b) shows that the distribution of slant for female judges is clearly shifted to the left relative to the one for male judges, and the difference is significant according to a Kolmogorov-Smirnov test for equality of distributions ($p = 0.012$).

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 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.[12] Overall, female judges and younger judges display the lowest 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.[13] 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 from information of judges' family composition from Glynn and Sen (2015). We estimate the following specification:

$$Gender\ Slant_j \quad = \quad \beta Daughter_j + X_j^{'}\gamma + \delta_c + \delta_n + \epsilon_j \tag{1}$$

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), $\delta_c$ are circuit fixed effects, and $\delta_n$ are number of children fixed effects. Standard errors are robust.

Table 2 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 (Lowe, 2019; Alesina et al., 2018; Corno et al., 2019).

[12] 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.

[13] 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).

# 5 Effect of Gender Slant on Judicial Decisions

This section asks whether gender slant influences how judges make 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:

$$Conservative\ Vote_{jcit} \quad = \quad \beta Gender\ Slant_j + X_j^{'}\gamma + \delta_{ct} + \varepsilon_{ictj} \tag{2}$$

where $Conservative\ Vote_{ictj}$ is an indicator variable equal to 1 if judge $j$ of circuit $c$ voted conservatively (against expanding women's rights) in case $i$ during year $t$, $Gender\ Slant_j$ 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 $\delta_{ct}$ are circuit-year fixed effects. The circuit-year fixed effects $\delta_{ct}$ ensure that we are using within-circuit-year variation, for which cases are quasi-randomly assigned to panels. Instead, controlling for demographic characteristics tests whether gender slant contains signal above and beyond that of 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 3 shows how slant affects decisions in gender-related cases. Column (1) shows that judges with higher slant are more likely to vote conservatively in gender-related cases. In particular, 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 4.1 percentage points. The coefficient is significant at the 1% level.[14,15]

The magnitude of the effect is sizable. In the baseline specification, a one standard deviation increase in the gender slant of the judge decreases the probability of voting in favor of expanding women's rights by 4.2 percentage points. This corresponds to a 7% decrease over the outcome

[14]Appendix Table 3 displays the result of the test proposed in Oster (2019) for bounding selection of unobservable based on selection on observables. Based on the test, 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.

[15]Appendix Figure 6 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.

mean. To put this in perspective, the effect of increasing the gender slant measure by one standard deviation has around one-third of the effect of the judge being 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 by affecting decisions in gender-related cases, attitudes reflected by gender slant have the potential to influence real-world outcomes that are directly impacted by them (Chen and Sethi, 2018; Chen and Yeh, 2014c,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 (2) we show that controlling for year of first appointment of the judge to an appellate court does not impact the results. Second, in column (3) 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.[16,17]

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

As a placebo test, we check whether the association between gender and positive/negative attributes, which does not appear to be a relevant measure of gender attitudes in this setting as discussed in Appendix A.1, holds any explanatory power in explaining decisions in gender-related cases. Reassuringly, Appendix Table 8 column (2) shows that this is not the case: the standardized cosine similarity between the gender and positive-negative dimensions is not correlated with

[16] 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.

[17] Appendix Table 7 shows the results are the same if we separetely estimate the regression for the Epstein et al. (2013) and Glynn and Sen (2015) datasets.

decisions in gender-related cases.

Finally, to address the concern that some slant measures might be estimated more precisely than others, in column (4) we weight the regression by the inverse of the variance of the gender slant measure across bootstrap samples. The effect is virtually unchanged.

**Robustness to Word Set Choice.** These results do 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 coefficient on slant from the baseline specification (equation (2)), together with 95% confidence intervals, from different regressions where slant is identified using the top five to top fifteen most frequent male, female, career, and family words from LIWC in the full judicial corpus. The graphs to the right show 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 to the choice of the word sets. At the same time, no single word is driving the result: the main coefficient is remarkably stable across all the regressions displayed in the graphs 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 texts of the gender-related case opinions. 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, 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 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.[18] Third, the bootstrap procedure we employ to train the word embeddings ensures

[18]We identify gender-related cases using a simple pattern-based sequence-classification method. The method iden-

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 3 column (4), the result is unchanged if the gender slant measure is computed on embeddings trained excluding gender-related cases.[19]

**Non-Gender-Related Cases.** Finally, to make the case that 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 et al. (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 gender slant on decisions in these types of cases. Appendix Table 9 shows that gender slant has a positive effect on voting conservatively in non-gender-related cases, but the effect is around two thirds as big as the effect of slant 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 slant on conservative votes in gender-related cases is unchanged (Appendix Table 8 column (3)). 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 valence (liberal, conservative, or neutral/hard to code). Appendix Table 11 shows that gender slant has no effect on the probability of casting a liberal vote across all specifications. Taken together, these results show that while gender slant may be correlated with holding liberal views, the measure does

tifies 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.

[19]It is still possible, although unintuitive, for gender slant to be measuring as 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 gender slant 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.

indeed also capture attitudes that are specific to gender.

## 6 Effect of Gender Slant on Treatment of Female Judges

In this section we ask whether gender attitudes, as proxied by our measure of slant, manifest themselves in differential treatment by judges toward their female 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 Effect of Gender Slant on Disparities in Reversals

District court trials are presided by a single judge and, similarly to the Appellate Level, 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 14, district judges that see 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 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 slant 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 gender 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:

$$\begin{aligned} \textit{Voted to Reverse}_{jictk} = & \pi \textit{Female District Judge}_k * \textit{Gender Slant}_j \\ & + \textit{Female District Judge}_k * X_j^{'}\gamma + \delta_{ct} + \delta_k + \delta_j + \varepsilon_{jictk} \end{aligned} \quad (3)$$

where $\textit{Voted to Reverse}_{id\tau ctj}$ is an indicator variable equal to 1 if circuit judge $j$ voted to reverse the district court decision in case $i$ in circuit $c$ in year $t$, originally decided by district judge $k$. On the right-hand side, $\textit{Gender Slant}_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of circuit judge $j$, $\textit{Female District Judge}_k$ is a dummy equal to 1 if the district judge 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), $\delta_{ct}$ are circuit-year fixed effects, $\delta_k$ are district judge fixed effects, and $\delta_j$ 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 $\delta_{ct}$ ensures comparability of cases assigned to higher and lower slanted judges because of quasi-random assignment of cases to judges, while as before, controlling for the demographic characteristics $X_j$ shows that gender slant matters independently of other judge characteristics. The inclusion of district judge fixed effects $\delta_k$ ensures 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 $\delta_j$ allow for circuit judges to differ in their baseline probability of reversing a district court decision.

Table 4 column (1) estimates the baseline specification. We find that 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. Interestingly, female district judges are 2.3 percentage points less likely to be reversed, and gender slant goes against this gradient.[20]

The result is not explained by judges being exposed to different cases over the course of their careers: the results are unchanged if we control for year of first appointment to a circuit court (column (2)), or if we include exposure fixed effects that ensure we are comparing judges who served in similar periods and areas (column (3)). Including district-year fixed effects, which are not needed for identification but might increase precision, also does not impact the result (column (4)). Finally, using a gender slant measure calculated based on embeddings trained excluding gender-related cases barely affects the coefficient (column (5)).

**Robustness Checks.** Appendix Table 12 shows that the main effect on reversals is robust to a number of additional checks. In particular, controlling for the share of female clerks (column (1)), including a placebo association between gender and positive/negative attributes (column (2)), and controlling for the circuit judge's vote record in ideologically divisive cases (column (3)) do not make a difference. Column (4) shows that the main coefficient is similar if slightly larger in magnitude if we weight by the inverse of the variance of the gender slant measure across bootstrap samples, which gives more weight to judges for which the measure is more precisely estimated. 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 (5)). Appendix Figure 11 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, slant has the potential to hinder the career progression of female district judges with respect to male district judges. Using estimates ofn 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.[21]

[20] Appendix Figure 8 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. It appears that low slant judges are less likely to reverse female judges with respect to males.

[21] 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.

### 6.2 Effect of Gender Slant on 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 et al., 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 slanted senior judges are differentially likely to assign majority opinions to female judges.

In particular, we estimate the following specification:

$$Female\ Author_{ictj} = \beta Gender\ Slant_j^{SENIOR} + X_j^{SENIOR'}\gamma + \delta_{ct} + \varepsilon_{ictj} \tag{4}$$

where $Female\ Author_{ictj}$ is a dummy equal to 1 if the author of the majority opinion of case $i$ in circuit $c$ in year $t$ decided by a panel with senior judge $j$ is a female judge, $Gender\ Slant_j^{SENIOR}$ is the gender slant (standardized cosine similarity between the gender and the career-family dimensions) of the most senior judge on the panel, $X_j^{SENIOR}$ are demographic characteristics of the 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 $\delta_{ct}$ are circuit-year fixed effects. The dataset is at the case level. Standard errors are clustered at the senior judge level.[22]

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 cases that have at least one female judge on the panel.[23] Since the decision to dissent or concur is possibly endogeneous, we also exclude cases

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

[23] 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.

that contain either a dissent or a concurrence.[24]

Table 5 reports the estimated effect of the gender slant of the assigning judge on whether the authoring judge is female. Column (1) estimates the baseline specification. When the most senior judge on the panel has higher slant by one standard deviation, 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 baseline probability.[25,26] If the senior judge is a woman instead, there is 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.

Columns (2) and (3) show that the result is unchanged when we control for exposure to different types of cases. Even if the coefficient is not significant when we include exposure fixed effects, the magnitude of the effect remains comparable to that of the baseline specification. In addition, if we use a gender slant measure calculated based on embeddings trained excluding gender-related cases, the result is the same (column (4)). 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 specification 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, in column (2) we show that the placebo association between gender and positive/negative attributes does not predict opinion assignment to a female judge. Column (3) 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. Finally, in column (4) we show that the effect is larger in magnitude if we give higher weight to judges whose gender slant measure is more precisely estimated.

[24] In Appendix Table 14, we 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)). We looked at the effect of gender slant on probability of dissenting against or concurring with female judges and did not find any effect.

[25] 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.

[26] Appendix Figure 8 shows a binned scatterplot of the main relationship between gender slant and opinion assignment, conditional on demographic controls and circuit-year fixed effects.

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 (5) 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 (6) shows instead that the main result is if anything larger if we include cases that had dissents or concurrences. Finally, column (7) shows that the results are robust to restricting the sample to the post-1980 period. In addition, Appendix Figure 12 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 A.12, we look but do not find evidence of any differences. The cases assigned to female judges by higher-slant judges are not concentrated in specific areas of the law and do not have different expected importance, as proxied by predicted forward citations (based on case characteristics), which is potentially an interesting result in light of the literature on discrimination in task assignments.

### 6.3 Effect of Gender Slant on Disparities in Citations

The last career-related outcome 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). This measure is therefore relevant to judicial careers, and differential treatment of male and female judges in citation choices presents another potential domain for high-stakes discrimination. In this section we analyze this type of differential treatment: are gender attitudes among judges reflected in their decisions about which precedents to cite in their opinions? In particular, we ask whether judges with higher gender slant are differentially likely to cite opinions authored by female judges.

Identification once again relies on 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 carefully.

The specification we estimate is:

$$Cites\ Female\ Judge_{ictj} = \beta Gender\ Slant_j + X_j'\gamma + \delta_{ct} + \varepsilon \tag{5}$$

where $Cites\ Female\ Judge_{ictj}$ is an indicator variable equal to 1 if the opinion of case $i$ authored by judge $j$ in circuit $c$ during year $t$ cites at least one opinion authored by a female judge, $Gender\ Slant_j$ is the slant (standardized cosine similarity between the gender and 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 $\delta_{ct}$ 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 6 shows the estimates from Equation (5) on how gender slant affects the probability of citing at least one female judge. Column (1) reports the estimates from 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%). Including interacted controls does not affect the result (column (2)), and neither does including career fixed effects (column (3)). However, the result is not robust to using gender slant measured on embeddings trained excluding gender-related cases (column (4)).[27]

**Robustness Checks.** Appendix Table 17 elaborates on the robustness of the main effect on citations with different specifications and sample restrictions. The main effect is not robust to

[27] Appendix Figure 9 shows a binned scatterplot of the relationship between gender slant and citations, conditional on demographic controls and circuit-year fixed effects.

controlling for the share of female clerks (column (1)), perhaps because clerks are important when determining decisions or simply because of the smaller sample size. Instead, including the placebo association between gender and positive/negative attributes (column (2)) and controlling for the circuit judge's vote record in ideologically divisive (column (3)) limitedly impact the main coefficient. Column (4) shows that the result is not impacted by weighting by the inverse of the variance of the slant measure across bootstrap sample. In addition, column (5) shows that the main effect is stable to restricting the sample to the post-1980 period when there were likely to be more female district judges.

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 (6) 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. Finally, the effect is robust to using different word sets (Appendix Figure 13). 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 gender slant also affects interactions with judges with different demographic characteristics – namely, political leanings, minority status, and age.

**Reversals.** First, we check whether judges with higher gender slant are also more likely to reverse decisions of district judges based on party of appointing President or minority status. Appendix Table 18 shows that circuit judges with higher slant 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 explores the effect of gender slant of the most senior judge on the panel on opinion assigment to whether the authoring judge is Democrat, whether the authoring judge is minority, and the age of the authoring judge. We find that judges with higher slant are not differentially likely to assign the opinion of the court to a Democratic judge (column (1)), which reassures us that the measure is indeed a meaningful proxy for attitudes toward gender and not overall conservativeness. Consistent with this idea, columns (2) and (3) show that slant does not impact the probability of the judge assigning the opinion to a minority judge, or the age of the judge.

**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, gender slant does not affect the probability of citing minority judges (column (2)) or the average age of the judges cited (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 less likely to assign opinions to women, are more likely to reverse district courts decisions when the district judge is a woman, 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. 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


**Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti**, "Revealing Stereotypes: Evidence from Immigrants in Schools," 2018. NBER Working Paper No. 25333. 4

**Antoniak, Maria and David Mimno**, "Evaluating the Stability of Embedding-Based Word Similarities," *Transactions of the Association for Computational Linguistics*, 2018, *6*, 107–119. 1, 2.4

**Arnold, David, Will Dobbie, and Crystal S Yang**, "Racial Bias in Bail Decisions," *The Quarterly Journal of Economics*, 2018, *133* (4), 1885–1932. 1

**Ash, Elliott and W. Bentley MacLeod**, "Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts," *Journal of Law and Economics*, 2015, *58* (4). 6.3

**Battaglini, Marco, Jorgen M. Harris, and Eleonora Patacchini**, "Professional Interactions and Hiring Decisions: Evidence from the Federal Judiciary," 2019. NBER Working Paper No. 26726. 2

**Bertrand, Marianne**, "Career, Family, and the Well-Being of College-Educated Women," *American Economic Review*, 2013, *103* (3), 244–50. 1

**__ , Claudia Goldin, and Lawrence F. Katz**, "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, July 2010, *2*, 228–255. 1

**__ , Dolly Chugh, and Sendhil Mullainathan**, "Implicit Discrimination," *American Economic Review*, 2005, *95* (2), 94–98. 1

**Bohren, J Aislinn, Alex Imas, and Michael Rosenberg**, "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 2019, *109* (10), 3395–3436. 1

**Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai**, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings," in "Advances in Neural Information Processing Systems" 2016, pp. 4349–4357. 1

**Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer**, "Beliefs about Gender," *American Economic Review*, 2019, *109* (3), 739–73. 1

**Bowie, Jennifer Barnes, Donald R Songer, and John Szmer**, *The View from the Bench and Chambers: Examining Judicial Process and Decision Making on the U.S. Courts of Appeals*, University of Virginia Press, 2014. 8, 6.2

**Boyd, Christina L. and James F. Spriggs II**, "An Examination of Strategic Anticipation of Appellate Court Preferences by Federal District Court Judges," *Journal of Law & Policy*, 2009, *29*, 37–81. 1

**Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais**, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, *107* (11), 3288–3319. 1

**Caliskan, Aylin, Joanna J Bryson, and Arvind Narayanan**, "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases," *Science*, 2017, *356* (6334), 183–186. 1, 2.3

**Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri**, "Are Referees and Editors in Economics Gender Neutral?," *Quarterly Journal of Economics*, February 2019, *135*, 269–327. 1

**Carlana, Michela**, "Implicit Stereotypes: Evidence from Teachers' Gender Bias," *The Quarterly Journal of Economics*, 2019, *134* (3), 1163–1224. 1

**Chen, Daniel L**, "Judicial Analytics and the Great Transformation of American Law," *Artificial Intelligence and Law*, 2019, *27* (1), 15–42. 3.2

**Chen, Daniel L. and Jasmin Sethi**, "Insiders and Outsiders: Does Forbidding Sexual Harassment Exacerbate Gender Inequality?," October 2018. TSE Working Paper No. 16-687. 1, 3.1, 8, 5

— **and Susan Yeh**, "The Construction of Morals," *Journal of Economic Behavior and Organization*, August 2014, *104*, 84–105. 1, 5

— **and** — , "Growth Under the Shadow of Expropriation? The Economic Impacts of Eminent Domain," Working Paper, ETH Zurich and George Mason University December 2014. 3.2

— **and** — , "How Do Rights Revolutions Occur? Free Speech and the First Amendment," Working Paper, ETH Zurich December 2014. 1, 5

— **, Vardges Levonyan, and Susan Yeh**, "Do Policies Affect Preferences? Evidence from Random Variation in Abortion Jurisprudence," Working Paper, ETH Zurich May 2014. 3.1

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

**Corno, Lucia, Eliana La Ferrara, and Justine Burns**, "Interaction, Stereotypes and Performance: Evidence from South Africa," 2019. IFS working paper W19/03. 4

**Croft, Jane**, "Law School Admissions Collapse Continues: They are being Forced to Innovate or Face being Left Behind," 2016. Financial Times. 1

**DiMaggio, Paul**, "Culture and Cognition," *Annual Review of Sociology*, 1997, *23* (1), 263–287. 2.6

**Eisenberg, Theodore**, "Appeal Rates and Outcomes in Tried and Nontried Cases: Further Exploration of Anti-Plaintiff Appellate Outcomes," *Journal of Empirical Legal Studies*, 2004, *1* (3), 659–688. 6.1

**Epstein, Lee, William M. Landes, and Richard A. Posner**, *The Behavior of Federal Judges*, Harvard University Press, 2013. 3.2, 10, 17, 5

**Friese, Malte, Matthias Bluemke, and Michaela Wänke**, "Predicting Voting Behavior with Implicit Attitude Measures: The 2002 German Parliamentary Election," *Experimental Psychology*, 2007, *54* (4), 247–255. 1

**Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou**, "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes," *Proceedings of the National Academy of Sciences*, 2018, *115* (16), E3635–E3644. 1

**George, Tracey E. and Albert H. Yoon**, "The Gavel Gap: Who Sits in Judgement on State Courts?," 2019. American Constitution Society for Law and Policy. 1

**Glaser, Barney G and Anselm L Strauss**, *Discovery of Grounded Theory: Strategies for Qualitative Research*, Routledge, 2017. 2.6

**Glover, Dylan, Amanda Pallais, and William Pariente**, "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *The Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260. 1

**Glynn, Adam N and Maya Sen**, "Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?," *American Journal of Political Science*, 2015, *59* (1), 37–54. 3.2, 10, 4, 13, 17

**Green, Alexander R, Dana R Carney, Daniel J Pallin, Long H Ngo, Kristal L Raymond, Lisa I Iezzoni, and Mahzarin R Banaji**, "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients," *Journal of General Internal Medicine*, 2007, *22* (9), 1231–1238. 1

**Hengel, Erin**, "Publishing while Female. Are Women Held to Higher Standards? Evidence from Peer Review," 2019. 1

**Jakiela, Pamela and Owen Ozier**, "Gendered Language," 2019. CGD Working Paper No. 500. 1

**Kastellec, Jonathan P.**, "Racial Diversity and Judicial Influence on Appellate Courts," *American*

*Journal of Political Science*, January 2013, *57* (1), 167–183. 1

**Katz, Daniel M and Derek K Stafford**, "Hustle and Flow: a Social Network Analysis of the American Federal Judiciary," *Ohio St. LJ*, 2010, *71*, 457. 3.2

**Kozlowski, Austin C, Matt Taddy, and James A Evans**, "The Geometry of Culture: Analyzing Meaning through Word Embeddings," *American Sociological Review*, 2019, *84* (5). 1, 2.3

**Kuersten, Ashlyn K. and Susan B. Haire**, "Update to the Appeals Courts Database 1997 - 2002," 2011. 'http://www.songerproject.org/data.html'. 3.2, 5

**Levinson, Justin D, Mark W Bennett, and Koichi Hioki**, "Judging Implicit Bias: A National Empirical Study of Judicial Sterotypes," *Fla. L. Rev.*, 2017, *69*, 63. 1, 1

**Lowe, Matt**, "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," 2019. Working Paper. 4

**Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean**, "Distributed Representations of Words and Phrases and their Compositionality," 2013, pp. 3111–3119. 1, 2.1, 2.2

**NAWL**, "Survey Report on the Promotion and Retention of Women in Law Firms," 2019. 1

**Oster, Emily**, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204. 3.1, 14, A.8

**Pennington, Jeffrey, Richard Socher, and Christopher Manning**, "Glove: Global Vectors for Word Representation," in "Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing" 2014, pp. 1532–1543. 1, 2.1, 2.2

**Posner, Richard**, *How Judges Think*, Harvard University Press, 2008. 6.3

**Posner, Richard A**, "An Economic Analysis of the Use of Citations in the Law," *American Law and Economics Review*, 2000, *2* (2), 381–406. 6.3

**Rachlinski, Jeffrey J., Sheri Lynn Johnson, Andrew J. Wistrich, and Chris Guthrie**, "Does Unconscious Bias Affect Trial Judges?," *Notre Dame Law Review*, 2009, *84* (3), 1195–1246. 1, 1

**Rehavi, M Marit and Sonja B Starr**, "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy*, 2014, *122* (6), 1320–1354. 1

**Rice, Douglas, Jesse H Rhodes, and Tatishe Nteta**, "Racial Bias in Legal Language," *Research & Politics*, 2019, *6* (2), 2053168019848930. 1

**Ricoeur, Paul**, *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*, Cambridge university press, 1981. 2.6

**Rohde, David W and Harold J Spaeth**, *Supreme Court Decision Making*, WH Freeman, 1976. 6.2

**Root, Danielle**, "Women Judges in the Federal Judiciary – A Series of Fact Shhts on the Lack of Diversity on the Bench," 2019. Center for American Progress. 1

**Rooth, Dan-Olof**, "Automatic Associations and Discrimination in Hiring: Real World Evidence," *Labour Economics*, 2010, *17* (3), 523–534. 1

**Sarsons, Heather**, "Interpreting Signals in the Labor Market: Evidence from Medical Referrals," 2019. 1

**Songer, Donald R.**, "The United States Courts of Appeals Database," 2008. 'http://www.songerproject.org/data.html'. 3.2, 5

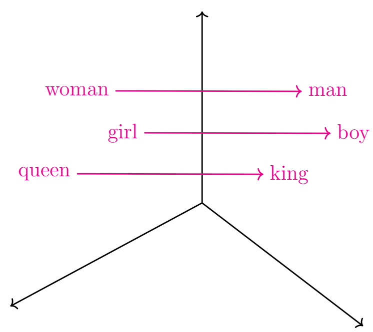
**Spirling, Arthur and P Rodriguez**, "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research," 2019. 2.2, 2.4

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

**Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki**, *Are Judges Political?: An Empirical Analysis of the Federal Judiciary*, Brookings Institution Press, 2006. 1

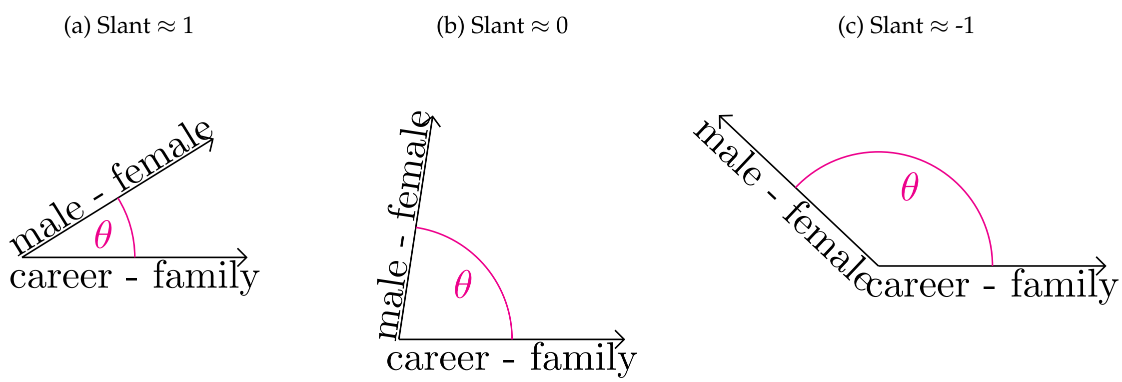
**Washington, Ebonya L**, "Female Socialization: How Daughters Affect Their Legislator Fathers," *American Economic Review*, 2008, *98* (1), 311–32. 4

Figure 1: Identifying a Gender Dimension in a Vector Space



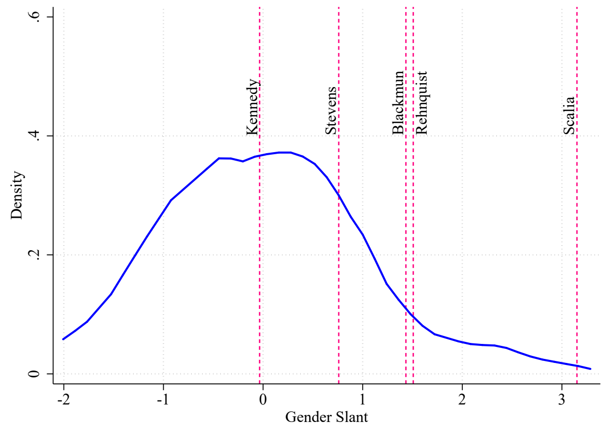

Notes: The figure exemplifies how the gender dimension can be identified in the vector space based on the difference between vectors representing male and female words.

Figure 2: Measuring Gender Attitudes using Cosine Similarity



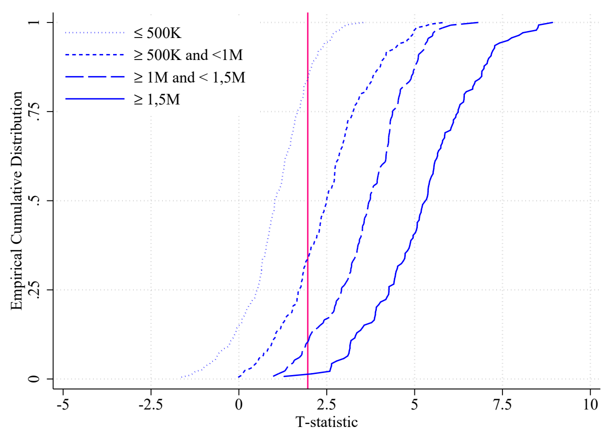

Notes: The figures exemplify how gender slant varies depending on the relative position of the gender and the career-family dimensions in the vector space.

Figure 3: Distribution of Gender Slant



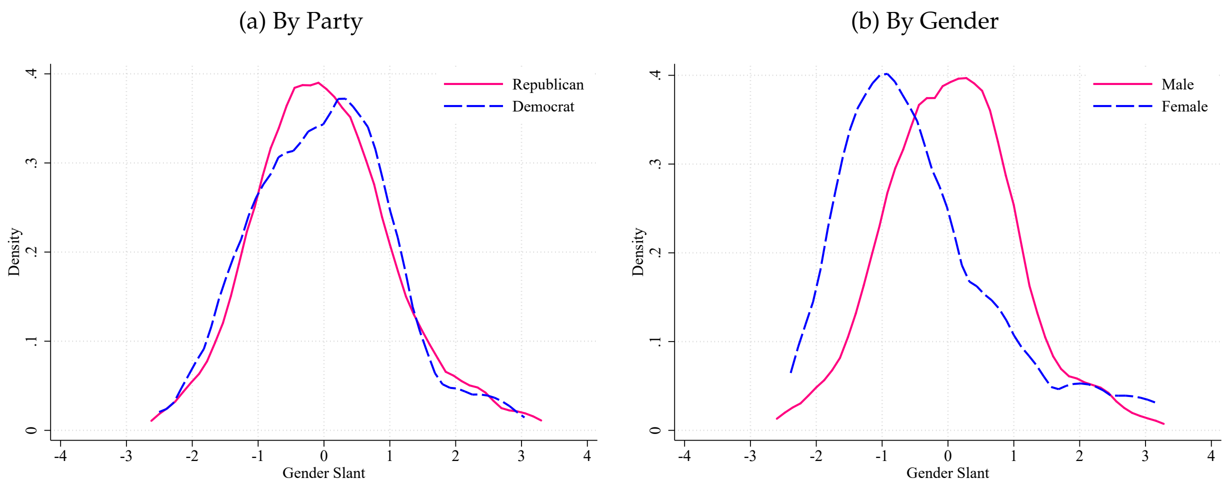

Notes: The graph shows the distribution of the gender slant measure for the 139 judges in the sample. The vertical lines indicate the gender slant of five Supreme Court justices who were appointed after 1970 and have a corpus of at least 1,000,000 tokens. The gender slant of Supreme Court justices is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

Figure 4: 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. The graph shows that for almost all judges with at least 1,500,000 tokens, there is a positive and significant at the 5% level correlation between the gender connotation of a name and a dummy indicating the correct gender, while this is not the case for judges with smaller corpora.

Figure 5: Gender Slant, by Party and Gender




Notes: The graphs show the distribution of the gender slant measure, by party of nominating President (graph to the left) and by judge gender (graph to the right). Gender slant is the standardized cosine similarity between the gender and the career-family dimensions.

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 |

Table 2: 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 Demographic 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.

Table 3: Effect of Gender Slant on Decisions in Gender-Related Cases

| Dependent Variable | Conservative Vote | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Gender Slant | 0.041*** | 0.041*** | 0.050*** | 0.046*** |
| | (0.013) | (0.012) | (0.014) | (0.012) |
| Democrat | -0.144*** | -0.141*** | -0.135*** | -0.148*** |
| | (0.025) | (0.025) | (0.023) | (0.025) |
| Female | -0.031 | -0.042 | -0.017 | -0.034 |
| | (0.032) | (0.032) | (0.025) | (0.034) |
| Observations | 3086 | 3086 | 3086 | 3086 |
| Clusters | 113 | 113 | 113 | 113 |
| Outcome Mean | 0.606 | 0.606 | 0.606 | 0.606 |
| Circuit-Year FE | X | X | X | X |
| Additional Demographic Controls | X | X | X | X |
| Year of Appointment | | X | | |
| Exposure FE | | | X | |
| Slant Excludes Gender-Related Cases | | | | X |

Notes: The table shows the effect of gender slant 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, 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 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.

Table 4: Differential Effect of Gender Slant on Reversals of District Court Cases by Gender of District Judge

| Dependent Variable | Voted to Reverse District Decision | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Gender Slant * Female District Judge | 0.010*** | 0.010*** | 0.010*** | 0.012*** | 0.009** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Democrat * Female District Judge | -0.010 | -0.010* | -0.010 | -0.010 | -0.011* |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| Female * Female District Judge | -0.003 | -0.002 | -0.003 | -0.005 | -0.005 |
| | (0.010) | (0.010) | (0.010) | (0.011) | (0.010) |
| Observations | 145862 | 145862 | 145862 | 145563 | 145862 |
| Clusters | 133 | 133 | 133 | 133 | 133 |
| Outcome Mean for Male District Judges | 0.180 | 0.180 | 0.180 | 0.180 | 0.180 |
| Outcome Mean for Female District Judges | 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 Demographic Controls | X | X | X | X | X |
| Year of Appointment | | X | | | |
| Exposure FE | | | X | | |
| District-Year FE | | | | X | |
| Slant Excludes Gender-Related Cases | | | | | X |

Notes: The table shows the differential effect of gender slant 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, 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 (2) 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 (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, interacted with an indicator variable for whether the district judge is female. Column (4) includes district-year fixed effects. 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.

Table 5: Effect of Assigning-Judge Gender Slant on Author Gender

| Dependent Variable | Author is Female | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Gender Slant | -0.017** | -0.017** | -0.014 | -0.016** |
| | (0.008) | (0.008) | (0.010) | (0.008) |
| Democrat | -0.001 | -0.001 | -0.012 | 0.002 |
| | (0.014) | (0.014) | (0.016) | (0.014) |
| Female | 0.134*** | 0.133*** | 0.158*** | 0.137*** |
| | (0.016) | (0.016) | (0.017) | (0.016) |
| Observations | 32052 | 32052 | 32052 | 32052 |
| Clusters | 125 | 125 | 125 | 125 |
| Outcome Mean | 0.383 | 0.383 | 0.383 | 0.383 |
| Circuit-Year FE | X | X | X | X |
| Additional Demographic Controls | X | X | X | X |
| Year of Appointment | | X | | |
| Exposure FE | | | X | |
| Slant Excludes Gender-Related Cases | | | | X |

Notes: The table shows the effect of the gender slant of the most senior judge on the panel, who is in charge of assigning opinion authorship, on the gender of the judge authoring the majority decision. 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)). 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 (2) controls for year of first appointment of the most senior judge of the panel to a circuit court. Column (3) 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 (4), 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.

Table 6: Effect of Gender Slant on the Probability of Citing a Female Judge

| Dependent Variable | Cites at Least One Female Judge | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Gender Slant | -0.010* | -0.009* | -0.013* | -0.005 |
| | (0.005) | (0.005) | (0.007) | (0.005) |
| Democrat | -0.012 | -0.011 | -0.018* | -0.011 |
| | (0.011) | (0.011) | (0.010) | (0.011) |
| Female | 0.128*** | 0.125*** | 0.139*** | 0.131*** |
| | (0.016) | (0.016) | (0.015) | (0.016) |
| Observations | 107923 | 107923 | 107923 | 107923 |
| Clusters | 139 | 139 | 139 | 139 |
| Outcome Mean | 0.383 | 0.383 | 0.383 | 0.383 |
| Circuit-Year FE | X | X | X | X |
| Additional Demographic Controls | X | X | X | X |
| Year of Appointment | | X | | |
| Exposure FE | | | X | |
| Slant Excludes Gender-Related Cases | | | | X |

Notes: The table shows the effect of the gender slant of the author of the majority opinion 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 (2) controls for year of first appointment 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 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. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# A Appendix

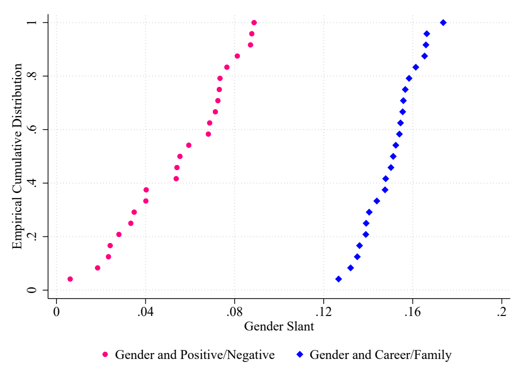
## A.1 Alternative Stereotypical Dimensions

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 Figure 1: Cumulative Distribution of Gender Slant based on the Association Considered




Notes: The graph shows the empirical cumulative distribution of gender slant measured using the stereotypical association between males and career versus female and family as opposed to 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.

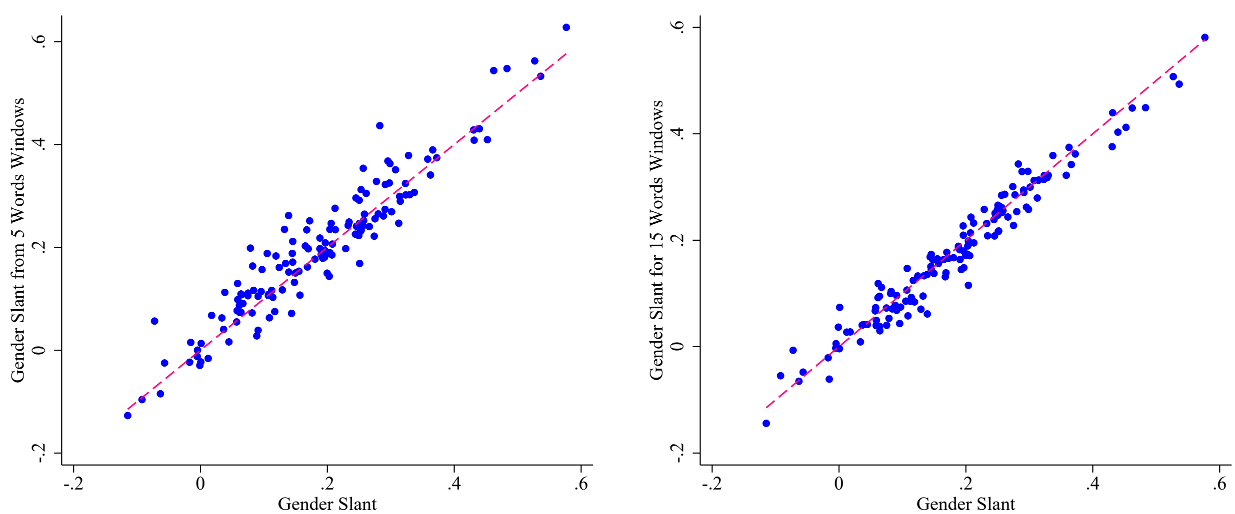
## A.2 Correlation of Gender Slant for Embeddings Trained using Different Parameters

A key hyperparameter for GloVe is the window size for computing co-occurence statistics. Here, we show that gender slant measures obtained from embeddings estimated using a 5, 10, or 15 window size are highly correlated. In particular, the two scatterplots below illustrate the correlation between the baseline gender slant measure we use in the paper, obtained using a 10 window size, and gender slant obtained using a 5 word window size (graph to the left) or a 15 word window size (graph to the right). Both scatterplots are clustered around the 45 degree line, which illustrates the strong correlation across the measures, although it is worth noting that the correlation is lower for smaller window sizes than for larger ones. Similarly, training lower (100- as opposed to 300-) dimensional embeddings leaves the gender slant measure almost unchanged.

Appendix Figure 2: Correlation of Gender Slant for Embeddings of Different Window Sizes

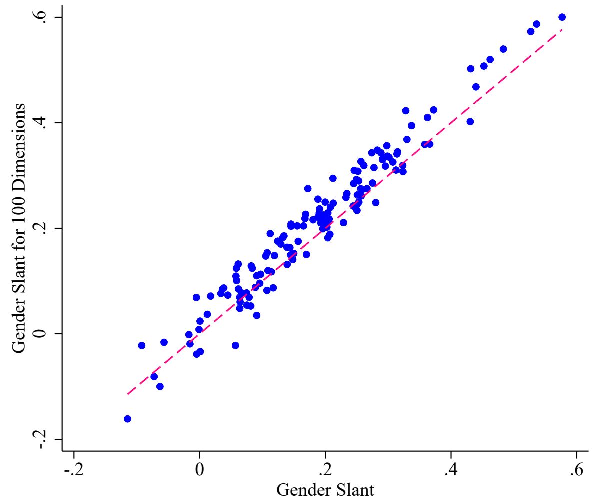
(a) Window of 5 Words

(b) Window of 15 Words




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.

Appendix Figure 3: Correlation of Gender Slant for Embeddings of Different Dimensions




Notes: The graph shows a scatter plot of the gender slant measure obtained by training 100-dimensional and 300-dimensional embeddings.

## A.3 Sample Selection

Appendix Table 2: Correlates of Having a Sufficiently Large Corpus

| Dependent Variable | Tokens >= 1.5m |
|---|---|
| | (1) |
| Democrat | -0.012 |
| | (0.030) |
| Female | 0.034 |
| | (0.055) |
| Minority | 0.035 |
| | (0.062) |
| Born in 1920s | 0.278*** |
| | (0.054) |
| Born in 1930s | 0.320*** |
| | (0.059) |
| Born after 1940 | 0.131*** |
| | (0.050) |
| Observations | 667 |
| Adjusted R2 | 0.181 |
| Circuit FE | X |
| Additional Demographic Controls | X |

Notes: The table shows what demographic characteristics correlate with the judge having a sufficiently large corpus to be included in the main sample. 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 demographic controls and circuit fixed effects. Demographic 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. The sample is restricted to 667 judges judges with a minimum corpus size of 50,000 tokens. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

## A.4 Example Sentences

Here, we report ten randomly selected text snippets that present the co-occurrences at the basis of the gender slant measure. Words that are part of the selected sets used to construct the gender slant measure are in italics.

1. Snippets that include *male* and *career* words within a ten word window

   - "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?"

2. Snippets that include *female* and *career* words within a ten word window

   - "1291, this court affirmsi.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"
- "by most accounts, *she* was a hard-working, competent, loyal *employee*"

3. <u>Snippets that include *male* and *family* words within a ten word window</u>

   - "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"
- "william mynd devised considerable estate to *his son* william mynd, charged with certain legacies to *his* daughters"

4. Snippets that include *female* and *family* words within a ten word window

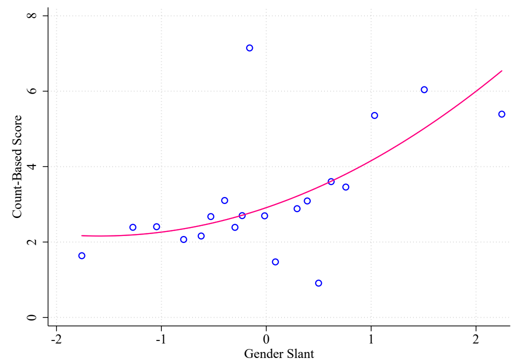
- "further, the alj did not believe cox, or *her husband* and neighbor, who both testified at cox's hearing on *her* behalf"
- "mrs willging and *her husband* were wheat farmers, owning community property, and reporting their income on the accrual basis"
- "*she* points to the incidents involving *her* 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"
- "eleni argues that the stipulation fails to illustrate *her* exercise of control over *her husband*"

We define a count-based score of gender slant as follows:

$$Count\ Based\ Score = \frac{(\#Male/Career\ Snippets - \#Female/Career\ Snippets)}{(\#Male/Family\ Snippets - \#Female/Family\ Snippets)}$$

The count-based score measures how much stronger the association between male and career words with respect to female is relative to the association between male and family words with respect to female. The figure show that the count based measure positively correlated with our gender slant measure.

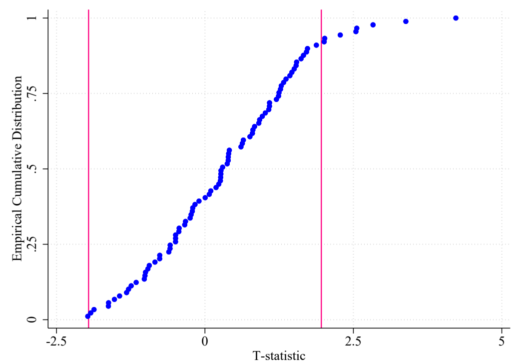
Appendix Figure 4: Count-Based Score and Gender Slant




Notes: The graph shows a binned scatterplot of the relationship between the count based score defined above and gender slant.

## A.5 Randomization

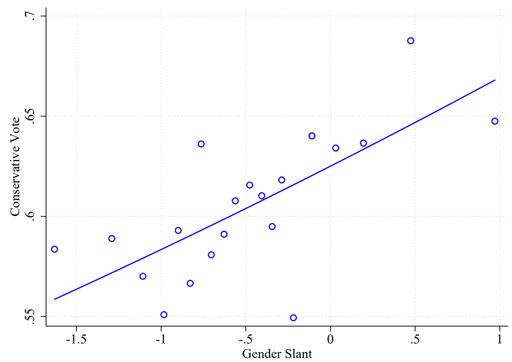
Appendix Figure 5: Randomization Check




Notes: The graph shows the cumulative distribution of the t-statistics from a series of regressions of an indicator variable for a given topic on the gender slant of the judge and circuit-year fixed effects. Standard errors are clustered are the judge level.

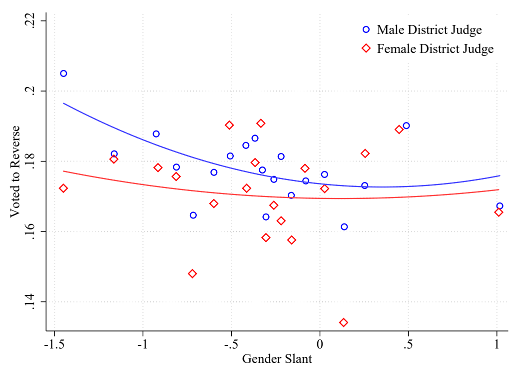
## A.6 Binscatter Plots

Appendix Figure 6: Gender Slant and Decisions in Gender-Related Cases



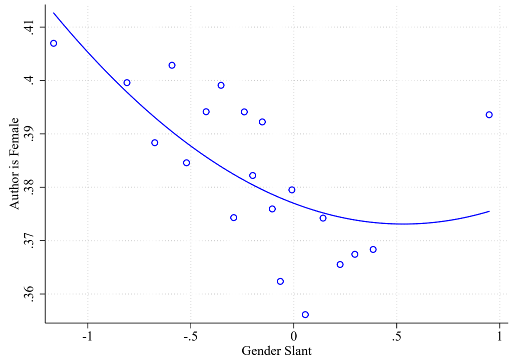

Notes: The graph shows a binned scatterplot of the relationship between gender slant and the probability that a judge voted conservatively in gender-related cases, conditional on demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region 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 7: Gender Slant and Reversals, by District Judge Gender



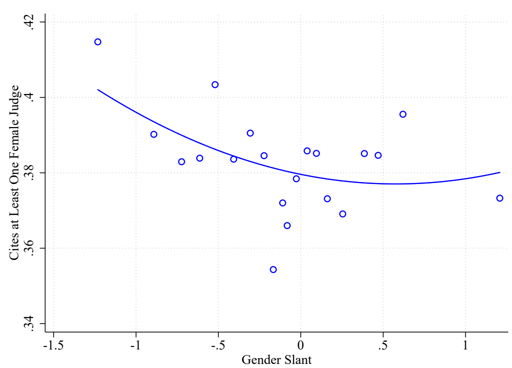

Notes: The graph shows a binned scatterplot of the relationship between gender slant and the probability of voting to reverse the district court decision by the gender of the district judge, conditional on demographic controls and circuit-year fixed effects. Demographic controls are gender, party of 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 8: Assigning-Judge Gender Slant and Author Gender




Notes: The graph shows a binned scatterplot of the relationship between the gender slant of the most senior judge on the panel, who is in charge of assigning opinion authorship, and the gender of the authoring 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 9: Gender Slant and Probability of Citing a Female Judge




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.

## A.7 Robustness to Different Word Choice

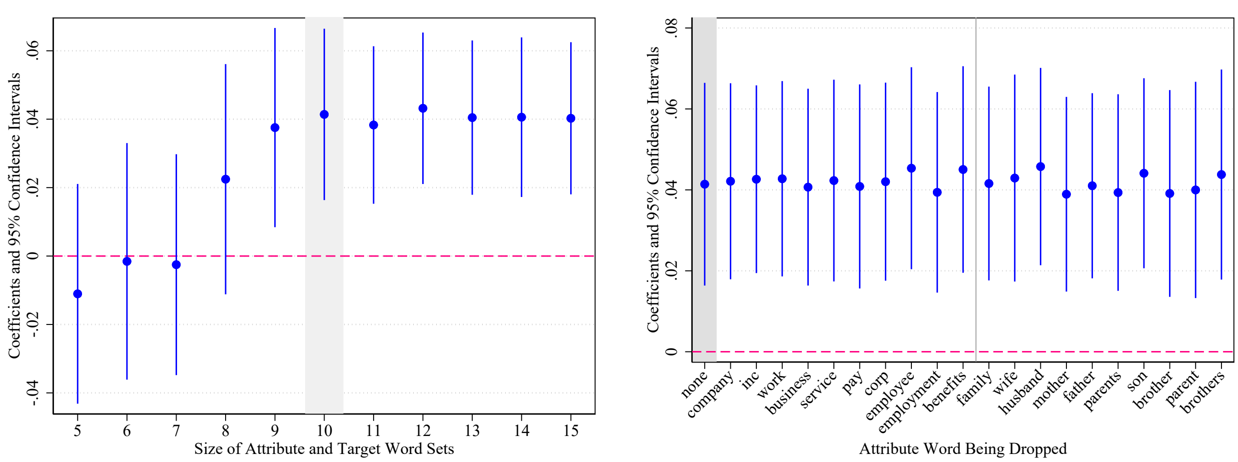
An important choice we made when constructing the lexical gender slant measure was which words to use to identify the gender and career-family dimension in the word embedding space. As we explain in Section 2.3, we followed a principled procedure which selected the words from the Linguistic Inquiry and Word Count Dictionary associated with a given concept that appeared more frequently in the judicial corpus.

In this section, we explore how the main results of the paper are affected by perturbation of the word sets used. First, we show how the results are affected if we choose the top 5 to top 15 most frequent words appearing in the judicial corpus. Second, we show how the results change if we drop one word used to identify the career-family dimension at the time. Overall, the results appear to be robust to the choice of words used to identify the two dimensions.

Appendix Figure 10: Effect of Gender Slant on Gender-Related Decisions, Robustness to Word Set Choice

(a) Robustness to Size of Word Set

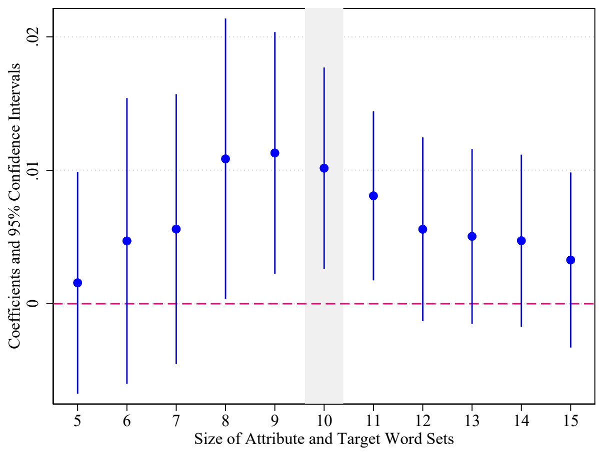
(b) Robustness to Dropping Attribute Words



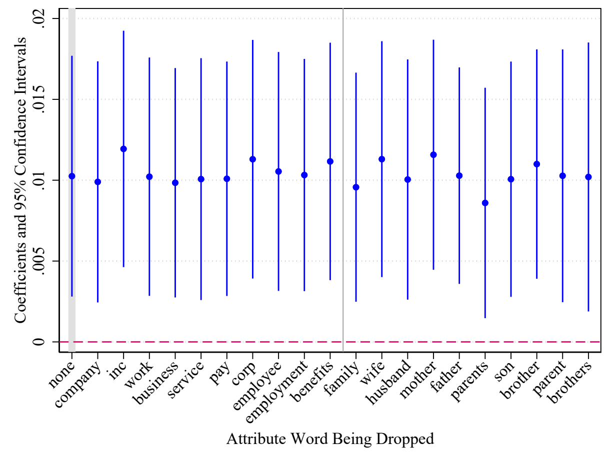

Notes: The graphs show how the effect of gender slant on decisions in gender-related cases varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the right shows robustness to dropping 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: Differential Effect of Gender Slant on Reversals in District Court Cases by Gender of District Judge, Robustness to Word Set Choice

(a) Robustness to Size of Word Set



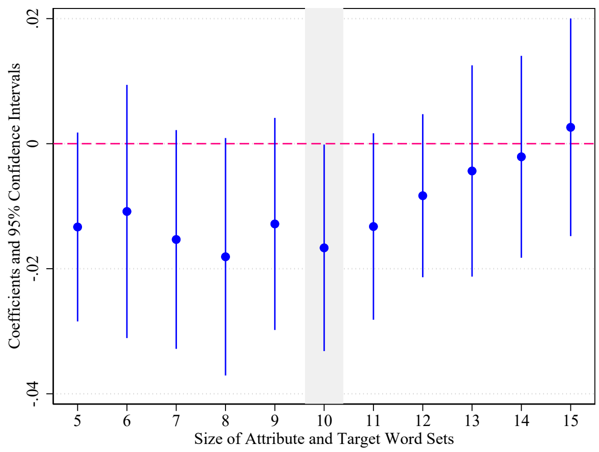

(b) Robustness to Dropping Attribute Words



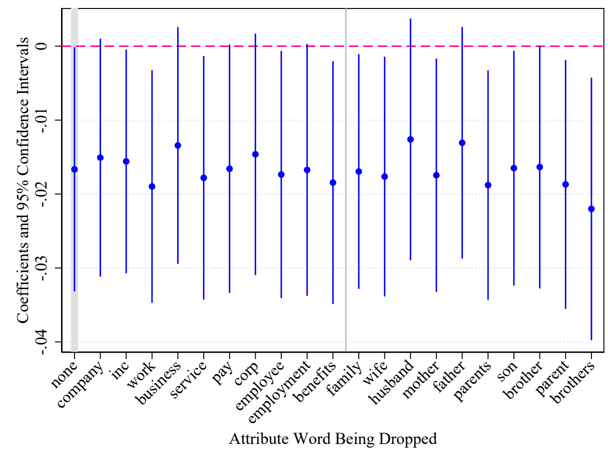

Notes: The graphs show how the differential effect of gender slant on the reversal probability of cases originally decided by male and female district judges varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of 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 12: Effect of Assigning-Judge Gender Slant on Author Gender, Robustness to Word Set Choice

(a) Robustness to Size of Word Set




(b) Robustness to Dropping Attribute Words

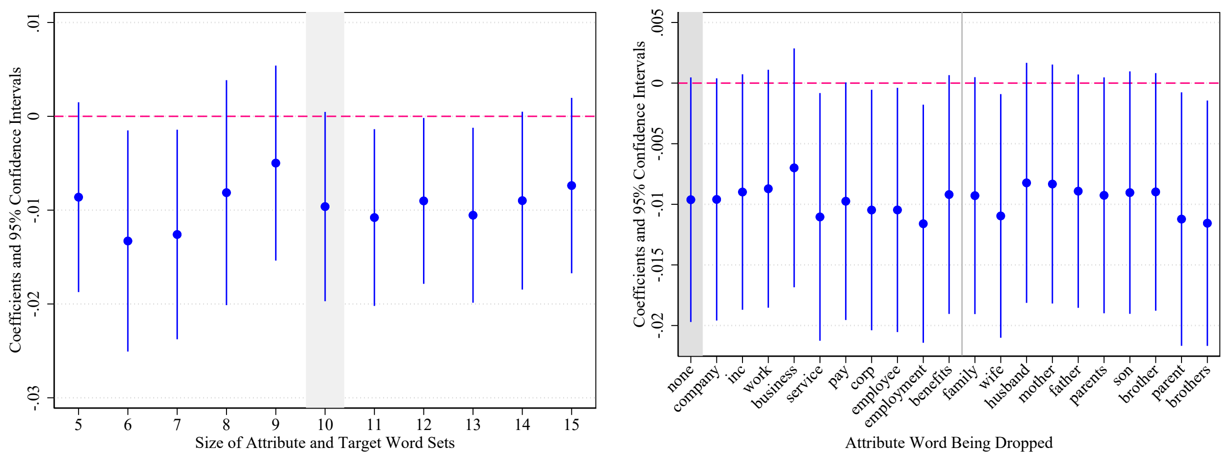



Notes: The graphs show how the effect of the gender slant of the most senior judge on the panel, who is in charge of assigning opinion authorship, on the gender of the judge authoring the majority decision varies based on the word sets used to identify the gender and attribute dimension. The graph on 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)). 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 13: Effect of Gender Slant on the Probability of Citing a Female Judge, Robustness to Word Set Choice

(a) Robustness to Size of Word Set

(b) Robustness to Dropping Attribute Words




Notes: The graphs show how the effect of the gender slant of the author of the majority opinion 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.

### A.8 Oster (2019) Test for Selection on Unobservables

Our identification strategy relies on two assumptions: first, random assignment of judges to cases – which ensures that gender slant is not systematically related to the outcome; and second, on conditioning on a detailed set of observable characteristics – which ensures that gender slant is not proxying for other characteristics of the judge (other than gender preferences). Of course, judges with different levels of gender slant might still be different according to some unobservable characteristic. In this section, we provide some suggestive evidence on the extent to which this is an issue by implementing the test proposed by Oster (2019). 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 $R^2$.

The tables below report the coefficient from the uncontrolled regression (column (1)), where we drop the demographic controls for the judge, and the coefficient from the controlled regression (column (2)). We perform the test for three different levels of $\overline{R_{max}}$, which is the $R^2$ of a regression that included all unobservable charactestics. Based on the recommendations in Oster (2019), we consider three different values of $\overline{R_{max}}$: 1.5, 2, and 5 times the $R^2$ of the controlled regression. For each value of $\overline{R_{max}}$, we then 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.

Overall, selection of unobservables does not appear to be a major concern. The exception is the effect of gender slant on citations, where the bias-adjusted treatment effect would be zero if selection on unobservable were only half as large as the amount of selection on observables.

Appendix Table 3: Oster Test: Effect of Gender Slant on Decisions in Gender-Related Cases

| $\beta$ Uncontrolled | $\beta$ Controlled | $\overline{R_{max}}$ | $\delta$ for $\beta = 0$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| -0.041 | -0.041 | 0.057 | 3.500 |
| | | 0.076 | 1.861 |
| | | 0.190 | 0.489 |

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)). Column (3) shows the value of $R_max$ for which $\delta$ (column (4)) is computed. $\delta$ is 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.

Appendix Table 4: Oster Test: Differential Effect of Gender Slant on Reversals of District Court Cases by Gender of District Judge

| $\beta$ Uncontrolled | $\beta$ Controlled | $\overline{R_{max}}$ | $\delta$ for $\beta = 0$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| 0.006 | 0.010 | 0.000 | 43.772 |
| | | 0.000 | 16.585 |
| | | 0.001 | 4.529 |

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 (3)). Column (3) shows the value of $R_max$ for which $\delta$ (column (4)) is computed. $R_max$ is set to be 1.5, 2, or 5 times the adjusted $R^2$ from the baseline specification. $\delta$ is 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.

Appendix Table 5: Oster Test: Effect of Assigning-Judge Gender Slant on Author Gender

| $\beta$ Uncontrolled | $\beta$ Controlled | $\overline{R_{max}}$ | $\delta$ for $\beta = 0$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| -0.028 | -0.017 | 0.025 | 0.877 |
| | | 0.033 | 0.445 |
| | | 0.082 | 0.113 |

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 (4)). Column (3) shows the value of $R_max$ for which $\delta$ (column (4)) is computed. $R_max$ is set to be 1.5, 2, or 5 times the adjusted $R^2$ from the baseline specification. $\delta$ is 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.

Appendix Table 6: Oster Test: Effect of Gender Slant on Probability of Citing a Female Judge

| $\beta$ Uncontrolled | $\beta$ Controlled | $\overline{R_{max}}$ | $\delta$ for $\beta = 0$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| -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 the demographic controls. Column (2) shows the estimate of the coefficient from the baseline specification (equation (5)). Column (3) shows the value of $R_max$ for which $\delta$ (column (4)) is computed. $R_max$ is set to be 1.5, 2, or 5 times the adjusted $\tilde{R}^2$ from the baseline specification. $\delta$ is 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.

## A.9 Additional Results on Judicial Decisions

Appendix Table 7: Effect of Gender Slant on Decisions in Gender-Related Cases, by Dataset

| Dependent Variable | Conservative Vote | |
|---|---|---|
| Dataset | Epstein et al. (2013) Data | Glynn-Sen (2015) Data |
| | (1) | (2) |
| Gender Slant | 0.037*** | 0.040 |
| | (0.014) | (0.024) |
| Democrat | -0.145*** | -0.082 |
| | (0.026) | (0.054) |
| Female | -0.054 | 0.030 |
| | (0.033) | (0.056) |
| Observations | 2335 | 738 |
| Clusters | 112 | 104 |
| Outcome Mean | 0.583 | 0.675 |
| Circuit-Year FE | X | X |
| Additional Demographic Controls | X | X |

Notes: The table shows the effect of gender slant on decisions in gender-related cases separately by dataset. We regress an indicator variable equal to 1 if the judge voted conservatively in a genderrelated 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: Effect of Gender Slant on Decisions in Gender-Related Cases, Additional Robustness Checks

| Dependent Variable | Conservative Vote | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Gender Slant | 0.039*** | 0.041*** | 0.040*** | 0.039*** |
| | (0.013) | (0.013) | (0.013) | (0.012) |
| Democrat | -0.157*** | -0.143*** | -0.137*** | -0.157*** |
| | (0.031) | (0.025) | (0.027) | (0.026) |
| Female | -0.010 | -0.031 | -0.030 | -0.024 |
| | (0.035) | (0.032) | (0.032) | (0.032) |
| Share Female Clerks | 0.006 | | | |
| | (0.088) | | | |
| Association Between Gender and +/- Attributes | | 0.010 | | |
| | | (0.014) | | |
| Conservative Score (Epstein et al. 2013) | | | 0.059 | |
| | | | (0.098) | |
| Observations | 2348 | 3086 | 3078 | 3086 |
| Clusters | 72 | 113 | 111 | 113 |
| Outcome Mean | 0.612 | 0.606 | 0.606 | 0.606 |
| Circuit-Year FE | X | X | X | X |
| Additional Demographic Controls | X | X | X | X |
| Weights by Inverse of Slant Variance | | | | X |

Notes: The table shows the effect of gender slant on decisions in gender-related cases with additiona controls. 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) for the standardized cosine similarity between the gender and career-family dimensions, and column (3) for the judge's share of conservative votes in non gender-related cases from the Epstein et al. (2013) data. Column (4) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample. 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: Effect of Gender Slant on Decisions in Non-Gender-Related Cases

| Dependent Variable | Conservative Vote | | | |
|---|---|---|---|---|
| Dataset | Epstein et al. (2013) Data | | | |
| | (1) | (2) | (3) | (4) |
| Gender Slant | 0.027** | 0.027*** | 0.004 | 0.018* |
| | (0.012) | (0.012) | (0.012) | (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 Demographic Controls | X | X | X | X |
| Year of Appointment | | X | | |
| Exposure FE | | | X | |
| Slant Excludes Gender-Related Cases | | | | X |

Notes: The table shows the effect of gender slant on decisions in non-gender-related cases. We regress an indicator variable equal to 1 if the judge voted conservatively in a genderrelated 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.

We test whether gender slant has larger effects on gender-related cases as opposed to non-gender-related cases by estimating the following differences-in-differences specification:

$$\begin{aligned} \textit{Conservative Vote}_{jcit} = &\alpha \textit{Gender Related Case}_i + \pi \textit{Gender Related Case}_i * \textit{Gender Slant}_j \\ &+ \textit{Gender Related Case}_i * X_j^{'} \gamma + \delta_{ct} + \delta_j + \varepsilon_{jcit} \end{aligned}$$

where *Gender Related Case*$_i$ is an indicator variable equal to one if the case is on a gender-related issue and zero otherwise, and all other variables are defined as before.

Appendix Table 10: Differential Effect of Gender Slant on Decisions by Whether a Case is Gender-Related, Differences-in-Differences Specification

| Dependent Variable | Conservative Vote |
|---|---|
| Dataset | Epstein et al. (2013) Data |
| | (1) |
| Gender Slant * Gender-Related Case | 0.027** |
| | (0.013) |
| Democrat * Gender-Related Case | -0.082*** |
| | (0.029) |
| Female * Gender-Related Case | 0.030 |
| | (0.039) |
| Observations | 8565 |
| Clusters | 113 |
| Outcome Mean | 0.582 |
| Circuit-Year FE | X |
| Judge FE | X |
| Exposure FE | X |

Notes: The table tests whether slanted senior 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.

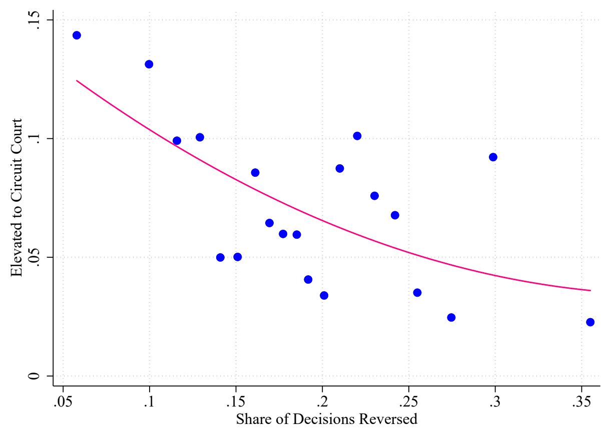
Appendix Table 11: Effect of Gender Slant on Decisions in All Cases, U.S. Courts of Appeals Data

| Dependent Variable | Conservative Vote | | | |
|---|---|---|---|---|
| Dataset | Songer 5% Dataset | | | |
| | (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 Demographic Controls | | X | X | X |
| Controls for Year of Appointment | | X | | |
| Includes Exposure FEs | | | X | |
| Slant Excluding Gender-Related Cases | | | | X |

Notes: The table shows the effect of gender slant on decisions in non-gender-related 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.

## A.10 Additional Results for Reversals

Appendix Figure 14: 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 (this requires that the case was appealed, and that we were able to match the circuit court case to the respective district judge).

Appendix Table 12: Differential Effect of Gender Slant on Reversals of District Court Cases by Gender of District Judge, Additional Robustness Checks

| Dependent Variable | Votes to Reverse District Decision | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Gender Slant * Female District Judge | 0.009** | 0.010*** | 0.012*** | 0.016*** | 0.010** |
| | (0.005) | (0.004) | (0.004) | (0.004) | (0.004) |
| Democrat * Female District Judge | -0.023** | -0.010 | -0.010 | -0.012* | -0.010 |
| | (0.009) | (0.006) | (0.007) | (0.007) | (0.006) |
| Female * Female District Judge | 0.018 | -0.003 | -0.000 | -0.003 | -0.003 |
| | (0.012) | (0.010) | (0.010) | (0.010) | (0.010) |
| Share Female Clerks * Female District Judge | 0.035 | | | | |
| | (0.039) | | | | |
| Association Between Gender and +/- Attributes | | -0.001 | | | |
| * Female District Judge | | (0.005) | | | |
| Conservative Score (Epstein et al. 2013) * Female District Judge | | | 0.007 | | |
| | | | (0.026) | | |
| Observations | 83751 | 145862 | 129677 | 145862 | 130381 |
| Clusters | 68 | 133 | 106 | 133 | 119 |
| Outcome Mean for Male Judges | 0.163 | 0.180 | 0.167 | 0.180 | 0.168 |
| Outcome Mean for Female Judges | 0.151 | 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 Demographic Controls | X | X | X | X | X |
| Weights by Inverse of Slant Variance | | | | X | |
| After 1980 | | | | | X |

Notes: The table shows the differential effect of gender slant 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) additionally controls for the share of clerks that are female, column (2) for the standardized cosine similarity between the gender and career-family dimensions, and column (3) for the share of conservative votes of the judge in non gender-related cases from the Epstein et al. (2013) data. Column (4) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample. Column (5) additionally restricts the sample to cases decided after 1980. 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 Demographic 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.

## A.11 Additional Results on Opinion Assignment

Appendix Table 14: Effect of Assigning-Judge Gender Slant on 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 the gender slant of the most senior judge on the panel, who assigns opinion authorship, on whether the opinion has a specific author, on whether the opinion 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. 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 15: Effect of Assigning-Judge Slant on Author Gender, Additional Robustness Checks

| Dependent Variable | Author is Female | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Slant | -0.016 | -0.017** | -0.018** | -0.029*** | -0.012* | -0.020** | -0.017** |
| | (0.011) | (0.008) | (0.009) | (0.009) | (0.007) | (0.008) | (0.008) |
| Democrat | -0.016 | -0.002 | -0.002 | 0.005 | 0.001 | -0.001 | -0.001 |
| | (0.025) | (0.014) | (0.017) | (0.015) | (0.014) | (0.013) | (0.014) |
| Female | 0.172*** | 0.134*** | 0.134*** | 0.126*** | | 0.133*** | 0.134*** |
| | (0.018) | (0.017) | (0.017) | (0.017) | | (0.017) | (0.016) |
| Share Female Clerks | -0.021 | | | | | | |
| | (0.055) | | | | | | |
| Association Between Gender and | | 0.004 | | | | | |
| +/- Attributes | | (0.008) | | | | | |
| Conservative Score (Epstein et al. 2013) | | | 0.023 | | | | |
| | | | (0.040) | | | | |
| Observations | 20543 | 32052 | 30614 | 32052 | 22828 | 36939 | 31998 |
| Clusters | 72 | 125 | 111 | 125 | 108 | 125 | 124 |
| Outcome Mean | 0.396 | 0.383 | 0.387 | 0.383 | 0.347 | 0.383 | 0.383 |
| Circuit-Year FE | X | X | X | X | X | X | X |
| Controls for Demographics | X | X | X | X | X | X | X |
| Weights by Inverse of Slant Variance | | | | X | | | |
| Excludes Female Senior Judges | | | | | X | | |
| Includes Dissents/Concurrences | | | | | | X | |
| After 1980 | | | | | | | X |

Notes: The table shows the effect of the gender slant of the most senior judge on the panel, who is in charge of assigning opinion authorship, on the gender of the judge authoring the majority decision. 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)). 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, column (2) for the standardized cosine similarity between the gender and career-family dimensions, and column (3) for the share of conservative votes of the judge in non gender-related cases from the Epstein et al. (2013) data. Column (4) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample. Column (5) excludes panels in which the most senior judge is female. Column (6) does not restrict the sample to cases decided unanimously, but includes cases with dissents or concurrences. Column (7) restrict the sample to post-1980 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 (with the exception of column (6). Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

## A.12 Do Higher Slant Senior Judges Assign Different Cases to Female Judges?

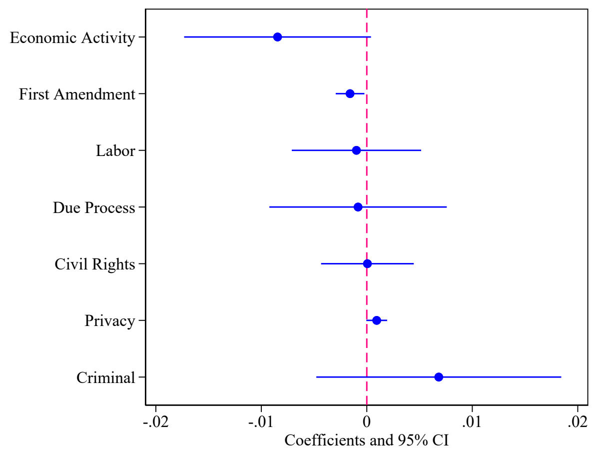
It is possible that more slanted senior judges not only assign fewer opinions to female judges, but also assign them opinions for cases that are different in topic or expected relevance, as proxied by predicted forward citations. We estimate the following difference-in-difference specification:

$$\begin{aligned} Case\,Characteristic_{ict} =& \pi Female\ Author_c * Gender\ Slant_j^{SENIOR} \\ & + Female\ Author_c * X_j^{SENIOR\prime}\gamma + \delta_{ct} + \delta_j + \varepsilon_{jictk} \end{aligned} \tag{6}$$

where $Case\,Characteristic_{ict}$ is a characteristic of case $i$ in circuit $c$ in year $t$ with senior judge $j$, $Female\ Author_k$ is a dummy equal to 1 if the district judge is female, $Gender\ Slant_j$ is the gender slant (standardized cosine similarity between the gender and career-family dimensions) of the most senior judge on the panel $j$, $X_j^{SENIOR}$ are demographic controls for the senior judge,$\delta_{ct}$ are circuit-year fixed effects, and $\delta_j$ are senior judge fixed effects. The dataset is at the case level. Standard errors are clustered at the senior judge level.

First, we examine whether slanted senior judges systematically assign cases on different topics to female judges. Cases are classified as belonging to one of seven broad topics: economic activity, First Amendment, due process, labor law, civil rights, privacy, and criminal appeals. Appendix Figure 15 shows that overall, slanted senior judges do not appear to assign cases on different topics to female judges, although they might be less likely to assign cases related to regulations of economic activity and First Amendment rights to female judges. Second, we study whether slanted senior judges assign cases with systematically different expected relevance to female judges. Given that citations might be endogeneous to the gender of the judge, we proxy for relevance using forward citations predicted based on circuit-year of the case, topic, and party names. Again, as shown in Appendix Table 16, we find little evidence that slanted senior judges are assigning different cases to female judges.

Appendix Figure 15: Differential Effect of Assigning-Judge Gender Slant on Topic of Cases Assigned to by Author's Gender




Notes: The graph explores whether slanted senior judges assign different types of 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. 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.

Appendix Table 16: Differential Effect of Assigning-Judge Gender Slant 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 Demographic Controls | X |

Notes: The table tests whether slanted senior 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. 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.

## A.13 Additional Results for Citations

Appendix Table 17: Effect of Gender Slant on Probability of Citing a Female Judge, Additional Robustness Checks

| Dependent Variable | Cites at Least One Female Judge | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gender Slant | -0.006 | -0.009* | -0.009 | -0.008 | -0.010 | -0.009* |
| | (0.009) | (0.005) | (0.006) | (0.006) | (0.006) | (0.005) |
| Democrat | -0.038** | -0.011 | -0.020 | -0.012 | -0.013 | -0.027*** |
| | (0.018) | (0.010) | (0.013) | (0.011) | (0.012) | (0.010) |
| Female | 0.157*** | 0.128*** | 0.125*** | 0.138*** | 0.128*** | -0.084** |
| | (0.022) | (0.016) | (0.017) | (0.017) | (0.017) | (0.018) |
| Share Female Clerks | -0.018 | | | | | |
| | (0.042) | | | | | |
| Association Between Gender and | | -0.008 | | | | |
| +/- Attributes Attributes | | (0.005) | | | | |
| Conservative Score (Epstein et al. 2013) | | | -0.039 | | | |
| | | | (0.035) | | | |
| Observations | 54301 | 107923 | 86910 | 107923 | 83680 | 107923 |
| Clusters | 73 | 139 | 112 | 139 | 125 | 139 |
| Outcome Mean | 0.536 | 0.383 | 0.452 | 0.383 | 0.487 | 0.383 |
| Circuit-Year FE | X | X | X | X | X | X |
| Additional Demographic Controls | X | X | X | X | X | X |
| Weights by Inverse of Slant Variance | | | | X | | |
| After 1980 | | | | | X | |
| Excludes Self-Citations | | | | | | X |

Notes: The table shows the effect of the gender slant of the author of the majority opinion 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, column (2) for the standardized cosine similarity between the gender and career-family dimensions, and column (3) for the judge's share of conservative votes in non gender-related cases from the Epstein et al. (2013) data. Column (4) weights the regression by the inverse of the variance of the gender slant measure across bootstrap sample. Column (5) restricts the sample to cases decided after 1980. Column (6) defines the outcome excluding self-citations. 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.

## A.14 Effect of Gender Slant on Other Judge Characteristics

Appendix Table 18: Differential Effect of Gender Slant on Reversals of District Court Cases by District Judge 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 Demographic Controls | X | X |

Notes: The table shows the differential effect of gender slant 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: Effect of Assigning-Judge Gender Slant on Author's 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 Demographic Controls | X | X | X |
| Panel Includes Democrat Judge | X | | |
| Panel Includes Minority Judge | | X | |

Notes: The table shows the effect of the gender slant of the most senior judge on the panel, who is in charge of assigning opinion authorship, on the the characteristics of the authoring judge. 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)). 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: Effect of Gender Slant on Probability of Citing a Judge with 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 | 0.113*** |
| | (0.004) | (0.005) | (0.082) | (0.012) |
| Democrat | 0.008 | -0.021* | -0.071 | 0.013 |
| | (0.007) | (0.011) | (0.105) | (0.018) |
| Female | 0.023** | 0.058*** | 0.026 | -0.032 |
| | (0.009) | (0.011) | (0.173) | (0.022) |
| 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 Demographic Controls | X | X | X | X |

Notes: The table shows the effect of the gender slant of the author of the majority opinion 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.