Stereotypes in High Stake Decisions:

Evidence from U.S. Circuit Courts

Arianna Ornaghi, University of Warwick

Elliott Ash, ETH Zurich

Daniel L. Chen, Toulouse School of Economics*

October 1, 2019

Abstract

Attitudes towards social groups such as women and racial minorities have been shown to be important determinants of individual's decisions but are hard to measure for those in policy-making roles. We propose a way to address the challenge in the case of U.S. appellate court judges, for whom we have large corpora of written text (their published opinions). Using the universe of published opinions in U.S. Circuit Courts 1890-2013, we construct a judge-specific measure of gender-stereotyped language (*gender slant*) by looking at the relative co-occurrence of words identifying gender (male versus female) and words identifying gender stereotypes (career versus family). We find that female and younger judges tend to use less stereotyped language in their opinions. In addition, the attitudes measured by gender slant matter for judicial decisions: judges with higher slant vote more conservatively on women rights' issues. These more slanted judges also influence workplace outcomes for female colleagues: they are less likely to assign opinions to female judges, they cite fewer female-authored opinions, and they are more likely to reverse lower-court decisions if the lower-court judge is a woman. Our results expose a possible use of text to detect decision-makers' stereotypes that predict behavior and disparate outcomes.

*Arianna Ornaghi, University of Warwick, a.ornaghi@warwick.ac.uk (corresponding author); Elliott Ash, ETH Zurich, ashe@ethz.ch; Daniel Chen, Toulouse School of Economics, daniel.chen@iast.fr. We thank Jacopo Bregolin, David Cai, Christoph Goessmann, and Ornelie Manzambi for helpful research assistance. We thank David Mimno and Mirko Draca for helpful comments.

1 Introduction

Stereotypes can be a powerful determinant of individual decisions. Attitudes toward social groups – most notably women and racial minorities – have been shown to be highly predictive of judgments and choices (Bertrand et al., 2005), including 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, 2018). Do attitudes toward social groups shape the decisions of those in high-stakes policy-making roles as well? We address this question by studying the role played by gender attitudes of federal appellate judges in U.S. courts.

U.S. Appellate Courts (in particular, the Circuit Courts of Appeal) play a major common-law policy-making role given their ability to set precedent. This has motivated an expansive literature on the determinants of judicial decisions in federal and state courts (see Posner, 2008), which has shown that ideological and biographical characteristics of judges (Boyd and Spriggs II, 2009; Glynn and Sen, 2015; Kastellec, 2013; Sunstein et al., 2006) and other non-legal factors (Danziger et al., 2011; Chen, 2014; Chen et al., 2016) matter for their decisions. In addition, some judges have been found to be systematically more or less lenient for equivalent cases (Dunn et al., 2017), suggesting a role for pre-determined judgements and the use of heuristics or stereotypes about salient attributes in judicial decisions (Bordalo et al., 2015; Gabaix, 2014).

Research concerning stereotypes in social psychology, economics, and other social sciences generally measures attitudes using Implicit Association Tests (Greenwald et al., 1998), where subjects are asked to assign words to categories and reaction times are compared across trials when pairings are consistent with stereotypes vs. when they are not. Unfortunately, IATs are rarely available for actors in policy-making roles such as judges, which makes measurement a major challenge to answering the research question.

In this paper, we propose a way to address this challenge 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) (Pennington et al., 2014; Caliskan et al., 2017; Kozlowski et al., 2018; Antoniak and Mimno, 2018) and proxy judges' attitudes toward gender by measuring their use of gender-stereotyped language. 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 are word embeddings (Mikolov et al., 2013a; Pennington et al., 2014), which are algorithms that distribute words in a vector space based on their co-occurrence in a corpus. Word embeddings preserve semantic relationship. First, words with similar meaning have similar representations. Second, dimensions induced by word differences can be used to identify cultural concepts in the space (Kozlowski et al., 2018). For example, the vector difference $\overrightarrow{man} - \overrightarrow{woman}$ isolates a gender dimension in the space..

The dimensions are useful because they produce quantitative measures of similarity between their associated concepts and specific words, in the corpus that is being represented. For example, we can understand the gender connotation of a given word by taking the cosine of the angle between the vector representation of the word and the (differenced) vector representing the gender dimension (Kozlowski et al., 2018). Words with male connotations – e.g. male first names – are going to be positively correlated with $\overrightarrow{man} - \overrightarrow{woman}$. Female first names, in turn, will be negatively correlated with the dimension.

This framework provides an intuitive approach to measuring stereotypical associations in a given corpus. In particular, we measure the intensity of gender stereotyping by looking at how similar the gender dimension ($\overrightarrow{man} - \overrightarrow{woman}$) is to the dimension representing the stereotypical gender roles of interest ($\overrightarrow{career} - \overrightarrow{family}$). 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, stereotypes are not present in the text.

To construct a judge specific gender slant measure, we consider the majority opinions authored by a given judge as a separate corpus. A separate embeddings model is trained for each judge corpus, which then allows us to calculate a judge-specific gender slant measure. Descriptively, we find that female and younger judges display lower gender slant. In addition, having a daughter reduces gender slant.

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 not just in terms of gender stereotypes but along 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 global 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, which allows us to study how slant impacts real-world behavior.

The central research question of the paper is how gender slant impacts the behavior of judges. We focus on two behaviors. First, we study how gender slant impacts judicial decisions, with a particular focus on gender-related cases. Second, we study how gender slant affects the treatment of female judges. The empirical strategy relies on the 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' demographic characteristics, which ensures that gender slant is not acting as a proxy for some other judge characteristics.

We begin by studying how gender slant impacts decisions in gender-related cases. 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 result is robust to including additional interacted controls, and ensuring that we are only comparing judges that were exposed to similar types of cases in the past. Gender slant has limited effects on non-gender related cases. By affecting how judges vote in appellate cases, and therefore impacting how precedent is set, the gender attitudes we measure in written language have the potential to impact real-word outcomes even outside the judiciary (Chen and Sethi, 2011; Chen and Yeh, 2014b,a). Gender slant *is* policy relevant.

If gender slant does indeed proxy for attitudes towards gender (as suggested by the decision results), we might also expect it to influence how judges interact with female colleagues in the court. We explore this question by studying three sets of interactions with female judges: opinion assignment, citations, and reversals of district court decisions. As before, identification relies on the random assignment of judges to cases and conditioning on detailed demographic characteristics.

We find that gender slant affects the treatment of female judges. First, we consider the decision of a senior judges tasked with assigning the writing of the majority opinion to a panel member. Assigning judges with higher gender slant are less likely to assign the opinion to a female judge. Second, more slanted judges are also less likely to cite opinions of female judges. Finally, we show that more slanted judges are more likely to vote to reverse lower-court decisions authored by female district judges. The magnitude of the effects are sizable, corresponding to around 5%-10% of the baseline mean. Our estimates are robust across a number of specifications, and as before, we find a limited role for gender slant in interactions with judges with given ideological leaning, minority status, and age. 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 district judges relative to male district judges.

These results build on the literature on the determinants of judicial decisions (Boyd and Spriggs II, 2009; Kastellec, 2013; Sunstein et al., 2006; Cohen and Yang, 2019), which have shown that demographic and ideological characteristics of judges predict their decisions. 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. We take a different approach by taking as our starting point a proxy of stereotypical gender attitudes and asking how this impacts decisions, instead of inferring bias from the decisions themselves.[1] This allows us to directly study the role played by stereotypical gender attitudes, at least as far as these can be measured from judges' writing. The existing evidence on stereotypes 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; Esteves et al., 2018). Instead, our measure importantly allows us to link stereotypical attitudes to real world outcomes.

If we think of courts as a workplace, the paper speaks to the literature on how gender shapes the labor market outcomes of women (see among others Bohren et al., 2018; Card et al., 2018; 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: we are not aware of existing work that takes this approach to the courts, in particular to study the potential for gender discrimination toward female judges. In addition, we contribute to this literature by providing evidence that stereotypical 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 stereotypes in decision making. For example, Glover et al. (2017) show that stereotypes regarding minorities influence manager-employee interactions is such a way to impact performance, and Carlana (2018) 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 stereotypical attitudes allows us to study the role played by stereotypes in determining the behavior of high-skilled professionals, who might otherwise be hard to reach for researchers interested in administering surveys or tests traditionally used in the literature. Stereotypes

[1] 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.

matter, even for high-stakes decisions that have the power to influence public policy for years to come.

The remainder of the paper is organized as follows. Section 2 describes our measure of gender slant. Section 3 provides descriptive statistics. Section 4 shows the effect of gender slant on judicial decisions, while Section 5 discusses how gender slant impacts female judges. Finally, Section 6 concludes.

2 Gender Slant: Measuring Gender Stereotypes 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 judges' 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.

2.1 Word Embeddings

In order to construct our measure, we use word embeddings, a language modeling technique from natural language processing that relies on word co-occurrence to create a representation in a (relatively) low dimensional Euclidean space that preserves semantic meaning (Mikolov et al., 2013a; 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 dimensionals – dense vectors which can accommodate large vocabularies 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 encodes 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, the direction of the vectors, and the differences between vectors that identify directions in the space, also convey meaning. As Figure 1 illustrates, going from 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 $w_i \in w$ to minimize the following objective function:

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

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), so we follow that literature and train 300-dimensional vectors. In turn, we choose a standard window size of 10, which is 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.[2]

A practical feature of GloVe is that the algorithm goes through the full corpus only once, 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., 2013b), 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.

2.3 Word Vectors and Gender Slant

We use word embeddings to identify cultural dimensions in language (Kozlowski et al., 2018). 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. The difference between male and female can be used to identify a gender dimension in the space. 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.

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

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

[2]Appendix Figure 1 shows that embeddings trained using 5, 10, or 15 word windows produce highly correlated measures of gender slant.

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

We put these dimensions in service of constructing a gender slant measure. The goal is to capture the strength of the association between gender and stereotypical attitudes – identifying men more closely with career, and women with family. Word embeddings provide an intuitive metric. 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 (Panel (A)). If there is no association between the two, then $\theta \approx 90^\circ$, and the slant measure is 0 (Panel (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 (Panel (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.[3] 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 by dropping one word at a time in the career versus family dimension.

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

2.4 Measuring Judges' Gender Slant

Our goal is to produce measures of gender stereotypes in the writing of Circuit Court judges. Our starting corpus is the universe of published opinions in U.S. Circuit Courts for the years 1890-2013, which consists of 380K opinions from in thirteen courts.

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.[4]

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.

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

[4]To ensure convergence, we train vectors for 20 iterations with a learning rate of 0.05.

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

Figure 3 shows the distribution of the coefficient and t-statistic 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-statistic is rarely above conventional significance level and is at times even negative. As the number of tokens increases, the t-statistics start to become significant and are never lower than zero, showing that the gender dimension identified in the embeddings does indeed contain meaningful gender information. Based on these experiments, all of our main results focus on the 139 judges whose corpus includes at least 1,500,000 tokens.[5]

2.6 Discussion of Alternative Measures

There are (perhaps many) other approaches that we could have used to measure gender stereotypes 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 14,056,877 sentences).

Second, we could have looked at stereotypical language more directly,for example by asking whether career or family related words tend to appear more frequently with gender identifiers. The advantage of word embeddings (relative to the count approach) is that the global co-occurrences do not require words to appear directly next to each other to register an association. Therefore they are able to take into account more information contained in judges' writing. A count-based approach would likely miss implicit and nuanced gender stereotyping and therefore provide a less precise measure.

[5]Around 1000 judges served in Circuit Courts from 1890-2013. Of these, 475 (272) have at least 500,000 (1,000,000) tokens in their corpus.

3 Gender Slant and Demographic Characteristics

This section explores descriptively how the gender slant measure varied based on with 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: if anything, contrary to expectations, judges that were nominated by a Democratic President – and therefore we would expect to hold more liberal values – have higher slant, although the coefficient is not statistically different than zero. Looking at the entire distribution of gender slant separately by party (Figure 4), shows that indeed, Democratic appointed judges appear to display stronger stereotypical associations on gender in their language, but again the difference is not statistically significant according to a Kolmogorov-Smirnov test for equality of distribution functions ($p = 0.164$).

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. Examining the full distribution (Figure 4) shows that all judges tend to associate men more strongly to career than family with respect to women the gender slant measure is overwhelmingly positive for judges of both genders. However, 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. We also see some geographic variation in slant, based on the region in which judges were born (column (5)): judges from the Midwest appear to have lower gender slant than judges born in the Northeast.

Column (6) 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 the Circuit Courts, and circuit fixed effects. The previously discussed correlations remain, although the difference across regions of birth is no longer statistically significant. 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.[6] 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:

$$slant_j \quad = \quad \beta daughter_j + X_j^{'}\gamma + \delta_c + \delta_n + \varepsilon_j \tag{1}$$

where X_j are demographic controls (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), δ_c are circuit fixed effects and δ_n are number of children fixed effects. Standard errors are robust.

Table 3 reports the estimates. We find that, conditional on the number of children, having a daughter lowers gender slant by 0.47 standard deviations. In comparison, female judges tend to have about 0.67 lower gender slant than male judges in this sample. The effect is only significant at the 10% level, and is not robust to additionally controlling for two-way interactions for gender, party, and race. These estimates therefore are potentially consisten 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 towards specific groups (Lowe, 2018; Alesina et al., 2018; Corno et al., 2019).

To put the gender slant measure in prospective, we might be interested in understanding how our gender slant correlates with measures of implicit and explicit bias traditionally used in the literature. The existing

[6]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).

evidence on this is limited, to the best of our knowledge, to recent work showing that languages that display higher gender slant have speakers that display higher implicit and explicit gender bias as measured by IAT scores (Lewis and Lupyan, 2019). In a similar fashion, we ask whether the demographic correlations we are seeing in gender slant are reflected in the implicit and explicit biases of a sample of individuals working in the legal profession using data from a large-scale administration of the Implicit Bias Test (Project Implicit).[7] As shown in 1, implicit and explicit bias correlate differently to demographic characterstics than does gender slant. While these correlations clearly have to be taken with a grain of salt, they suggest that our slant measure might be picking up attitudes towards stereotypes differently than traditional measures of gender bias.

4 Effect of Gender Slant on Judicial Decisions

This section asks whether judges' slant is related to how they vote in gender-related cases. This question is first-order to establish whether slant is policy-relevant. If slant does not proxy for gender preferences – or judges are able to correct for it when voting on cases – then it should have no impact of decisions, and on the real world outcomes that are directly impacted by them(Chen and Sethi, 2011; Chen and Yeh, 2014b,a).

To answer this question, we use existing legal datasets with hand-coded vote direction and topic, produced 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.[8] In gender-related topics (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 to maximize power.[9]

[7] More details on the exercise are reported in Appendix A.2.

[8] 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.

[9] 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 depends on the issues, 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 pulished 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.

At the Circuit Court level, cases are randomly assigned to panels formed by three judges.[10] This ensures no self-selection of judges into cases: the estimates are not driven by judges with varying slant being assigned different types of cases. In other words, the unobservable characteristics of gender-related cases that judges with higher and lower slant face should be comparable.

An additional threat to identification is that gender slant could be confounded with other judge characteristics that affect decisions in gender-related cases. As documented in the previous section, for example, male judges might be more conservative towards gender norms and also have higher slant. Our empirical strategy to address this issue is selection on observables, as we condition on detailed demographic characteristics of the judge. In short, identification requires that conditional on the covariates, slant is not systematically correlated with other omitted factors related to liberal voting.

We estimate the following specification:

$$liberal\,vote_{ictj} \quad = \quad \beta slant_j + X_j^{'}\gamma + \delta_i + \delta_{ct} + \varepsilon_{ictj} \tag{2}$$

where $liberal\,vote_{ictj}$ is a dummy equal to 1 if judge j of circuit c voted liberally (in favor of expanding women's rights) in case i during year t, $slant_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of judge j, X_j are demographic controls (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), δ_i are issue fixed effects (reproductive rights, gender discrimination, and sexual harassment), and δ_{ct} are circuit-year fixed effects. The dataset is at the vote level and standard errors are clustered at the judge level.

Table 4 reports the results for how slant affects decisions on gender-related cases. Columns (1) estimates the baseline specification with circuit-year fixed effects, topic fixed effects, and demographic controls. 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 4.6 percentage points. The coefficient is significant at the 1% level. The results are virtually unchanged if we additionally control for two-way interactions of gender, party and race (column (2)). Moreover, the result are also robust to controlling for 'career fixed effects' (column (3)) – indicator

[10] Ash et al. (2017) discuss in detail how the random assigment works. Judges are chosen from a pool of 8-40 judges, depending on the specific circuit. Before oral arguments, available judges (including visiting judges) are assigned to cases by a computer program. The program places some limits on judges serving on the same panel and on how many cases assigned to senior judges. It is possible that judges recuse themselves at times, and rarely randomization is not used for some specialized cases (e.g. cases involving the death penalty).

variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit, which ensure that we are comparing judges who served in the same years and circuits, and therefore have been exposed to similar types of cases.[11,12]

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 4.6 percentage points. This corresponds to a 12% decrease over the outcome mean. To put this in perspective, the effect of increasing the gender slant measure by one standard deviation has around one-half of the effect of the judge being being female and one-third of the effect of the judge being nominated by a Democratic President. Given that gender slant in measured with error, meanwhile, the estimates are likely to be attenuated toward zero. In short, gender slant *is* policy-relevant.[13]

These results do not depend on the specific choice of words used for constructing the male-female and career-family dimension. We experiment with expanding or restricting the word sets, or dropping single words at a time and present the results in 2. In particular, the graph to the left shows the coefficient on slant from the baseline specification (equation (1)), 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 graph to the right show coefficients and 95% confidence intervals from separate regressions where slant is measured dropping one attribute word at the time. Overall, the results appear to be robust to the choice of the word set. Smaller word sets give larger confidence intervals and weaker explanatory power of decisions in gender related cases. 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.

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 on those cases, and judges with higher slant simply happen to have been more exposed to such cases in

[11] Career fixed effects are different than circuit-year fixed effect: they are a biographical characteristic that refers to the career of the judge, and not a characteristic of the specific case.

[12] Appendix Table 2 displays the result of the test proposed in Oster (2019) for bounding selection of unobservable based on selection on observables.

[13] Appendix Table 6 shows the results are the same if we separetely estimate the regression for the Epstein et al. (2013) and Glynn and Sen (2015) datasets.

their careers.[14] We believe this to be unlikely for the following reasons. First, the specification controls for demographic characteristics such as judge cohort fixed effects and career fixed effects, which work to condition on the types of cases judges have been exposed to. Second, there are less than 4500 gender-related cases in these datasets, making up a small proportion of the texts in the full corpus of 114,702 circuit court cases on which we train word embeddings. Third, the bootstrap procedure we employ to train the word embeddings helps ensure that the measure is not driven by any specific opinions.

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 study non-gender related cases in the Epstein et al. (2013) data. Appendix Table 7 shows that gender slant has a negative effect on voting liberally also in non-gender related cases, although the magnitude of the effect is smaller than that on gender-related cases. Importantly, the effect is cut by half and not significant if we control for 'career fixed effects', which allow us to compare judges that have been exposed to similar cases in the past. In addition, the effect of gender slant is robust to controlling for the share of liberal votes that judges cast in non-gender related cases (Appendix Table 8). Second, we can test whether gender slant has an effect on decisions in a 5% random sample of Circuit cases that were hand-coded for vote valence (liberal, conservative or neutral/hard to code). Appendix Table 9 shows that gender slant has no effect on the probability of casting a liberal vote across all specifications. Taken together, these results reinforce the idea that while gender slant captures attitudes that are specific to gender, and is not a simple proxy of being more liberal or conservative across the board.

5 Effect of Gender Slant on Female Judges

Female judges are the minority in U.S. Courts. In our data, around 10% of Circuit judges are women, and only 20% of panels include at least one female judge. As shown in Figure 5, there were very few female judges before the 1970s, with Judge Florence Einwood Allen being the notable early exception. There are likely many factors contributing to the gender disparity in judgeships. This section explores the role of differential treatment by judges toward their female colleagues. Existing evidence on this front includes Jacobi and Schweers (2017), who show that female Supreme Court judges tend to be interrupted more

[14] Importantly, random assignment of judges to cases implies that cases are comparable within a given circuit and year, but judges serving across circuits or time might be expossed to different types of cases.

frequently than their male colleagues.

Our specific research question is whether gender slant manifests itself in how judges respond to female colleagues. We focus on dimensions that are relevant to a judge's career. In particular, we explore the following outcomes: whether female judges are assigned the writing of majority opinions, forward citations by future judges, and whether lower-court decisions by female judges are more likely to be reversed.

To run this analysis, we use detailed records from all U.S. Circuit Court cases for the years 1890-2013. The records include metadata on year, circuit, and topic, as well as the panel of judges assigned to a case. For each assigned judge, we have information on whether he/she voted to affirm or reverse the district court decision, authored the majority opinion, and dissented/concurred. For a subset of cases, we have the identity of the district judge who decided the lower-court case which is currently on appeal. Finally, we parse the table of referenced cases in each opinion to construct the citation network, and then count the number of forward citations to each opinion by future circuit court judges.

As before, each circuit judge is linked to a measure of gender slant constructed from his/her full portfolio of authored opinions. The identification strategy once again relies on the random assignment of cases to judges, and on conditioning on an extensive set of biographic controls. We estimate slightly different specifications for each of the outcomes, which we discuss in detail below.

5.1 Effect of Gender Slant on Disparities in Opinion Authorship

Written opinions are at the core of the courts' ability to shape policy. These opinions seek to articulate the the principles behind decisions, which are binding law on for lower courts (Rohde and Spaeth, 1976). 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 (either two or three) becomes the majority position, and the judges in the majority then have to decide who is going to author the association 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). Given the policy stakes of opinion assignment, a relevant question is whether the preferences of the senior judges affect this procedure. In particular, in this section we investigate whether slanted senior judges are differentially likely to assign opinion writing to female judges.

We estimate the following specification:

$$female\,author_{ictj} = \beta slant_j^{SENIOR} + X_j^{SENIOR'}\gamma + \delta_{ct} + \varepsilon_{ictj} \tag{3}$$

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 with senior judge j is a female judge, $slant_j^{SENIOR}$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of the most senior judge on the panel, X_j^{SENIOR} are demographic controls for the senior judge, and δ_{ct} are circuit-year fixed effects. The dataset is at the case level. Standard errors are clustered at the senior judge level.[15]

Identification comes from random assignment of cases to judges, which ensures that conditional on circuit-year fixed effects, variation across cases in the gender slant of the panel's senior judge is randomly assigned. As above, a second threat to identification is that our estimates pick up other judge characteristics that are correlated with gender slant, which we address by conditioning on a number of demographic characteristics.

For authorship 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.[16] Since the decision to dissent or concur is possibly endogeneous, we also exclude cases that contain either a dissent or a concurrence.[17]

Table 5 reports the estimated effect of the gender slant of the assigning judge on the 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 2 percentage points, corresponding to a 5% decrease over the baseline probability.[18]Perhaps surprisingly, if the senior judge was appointed by a Democratic President, there is a 6.5 percentage point (17%) decrease in the probability of the author being female. If the senior judge is a woman instead, there is 13.7 percentage points (35%) higher probability that the author of the opinion is a woman, possibly because

[15]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.

[16] In Appendix Table 10 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.

[17]In Appendix Table 10, 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)).

[18]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.

senior judges are more likely to assign themselves opinions. It appears that 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 other judge characteristics.

Column (2) shows that the result is unchanged when we include interacted controls. Interestingly, from the size of the interaction coefficient we can see that the positive effect of a female senior judge is driven largely by Republican female senior judges (not Democrat female senior judges). In Column (3), similarly, the main effect is unchanged (although slightly smaller in magnitude) if we condition on 'career fixed effects', which ensure we are comparing senior judges that have been exposed to similar cases. Column (4) shows that these results are not driven by confounded ideology of judges: if we control for a measure of how liberal a judge's voting record is, the coefficient on slant is unchanged.[19]

In column (5) we show that the main result is not sensitive to the inclusion of cases that had dissents or concurrences. In column (6) we re-estimate the main specification excluding cases where the most senior judge on the panel is a woman. The effect is unchanged, showing that the result is not driven solely by self-assignment by female senior judges.

Finally, in column (7) we take an even more conservative specification and condition on all time-invariant characteristics of the female judge who could be potentially assigned an opinion. This specification restricts the sample to cases with a single female judge and includes fixed effects for the identity of that judge. This specification effectively compares panels with the same female judge but with different levels of assigning-judge gender slant. We find that the same female judge is less likely to be assigned authorship when the assigning judge has higher slant (altough the coefficient is about half the magnitude of the baseline specification and significant at the 10% level). Finally, Appendix Table 3 shows that the results are robust to using different word sets to identify the gender and career-family dimensions. In short, judges with stronger stereotypical association towards gender are less likely to assign an important career-relevant task to female judges.[20]

[19]The liberal score is defined as the average share of cases in which the judge cast a liberal vote according to the non-gender cases in the Epstein et al. (2013) dataset.

[20]This result raises the question of whether slanted judges also assign different types of cases to female judges. In Appendix A.7, we show that this does not appear to be the case: the cases that are assigned to female judges by higher slant judges do not have differential predicted citations based on baseline case characteristics, and are not concentrated in specific areas of the law.

5.2 Effect of Gender Slant on Disparities in Citations

The second 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. In the words of Posner (2000), "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." 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.

The specification we estimate is:

$$cites\,female\,judge_{ictj} = \beta slant_j + X_j^{'}\gamma + \delta_{ct} + \varepsilon_{ictj} \tag{4}$$

where $cites\,female\,judge_{ictj}$ is a dummy equal to 1 if opinion i authored by judge j in circuit c during year t cites at least one opinion authored by a female judge, $slant_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of judge j, X_j are demographic controls, and δ_{ct} are circuit-year fixed effects. Standard errors are clustered at the judge level. The sample is restricted to cases in which the opinion is authored by a specific judge.

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. Choice of authorship is endogenous, which we address by adjusting for a number of judge characteristics to improve comparability across authoring judges. These controls also, as before, are designed to address the issue of other relevant judge characteristics being confounded with gender slant. Still, 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; the results in this sections have to be interpreted carefully.

Table 6 shows the estimates from Equation (4) 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 0.9 percentage points less likely to cite any opinions authored by female judges. The effect is significant at the 10% level, but relatively small in magnitude (2.3% of the outcome mean) compared to the effect of the author being a woman (12.3 percentage points or 30%). Including interacted controls does not affect the results (column (2)), and neither does including career fixed effects (column (3)). Similarly, the results are unchanged by controlling for a judge's vote record (column (4)): the coefficient is very similar in magnitudes, although it is no longer statistically significant (p-value = 0.154). Given that part of the effect could be explained by self-citations, in column (5) we estimate the baseline specification excluding self-cites. The coefficient on gender slant is negative, but slightly lower in magnitude and again not statistically significant (p-value = 0.154). Interestingly, when self-cites are excluded female judges appear to be less likely to cite other female judges. The effect is robust to using different word sets (see Appendix Figure 4). Overall, judges with higher gender slant appear to be less likely to cite opinions authored by female judges, although the result has to be interpreted with caution.

5.3 Effect of Gender Slant on Disparities in Reversals

This section explores gender disparities in the probability of voting to reverse a lower-court decision. 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). Reversals matter for career outcomes: as shown in Appendix Figure 7, 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. In particular, we estimate a differences-in-differences specification in which we compare appealed cases decided by female and male district judges that are assigned to circuit judges with different levels of gender slant. The quasi-random assignment of cases to panels at the circuit level ensures that cases assigned to higher or lower slant are comparable. More precisely, district court cases decided by a female judge are not systematically assigned to higher-slant circuit judges if they are more likely to be reversed in the first place. Note that the identification strategy allows for cases decided by female and male judges 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.

For this part of the analysis, we complement the dataset of Circuit Court cases with information of the corresponding district court case. In particular, we obtain for each appealed case the name of the associated district court judge, with which we match to gender data from the Federal Judicial Center. The judge's name was obtained either directly from the district case's metadata (when linking was feasible) or by parsing the name from from the circuit opinion's case history.[21] We are able to assign a unique district judge (and associated gender) to 107699 circuit cases (28% of all cases).

We estimate the following baseline specification:

$$\begin{aligned} voted\,to\,reverse_{jictk} =& \pi female\,district\,\,judge_k * slant_j \\ &+ female\,district\,\,judge_k * X_j^{'}\gamma + \delta_k + \delta_j + \delta_{ct} + \varepsilon_{jictk} \end{aligned} \tag{5}$$

where $voted\,to\,reverse_{id\tau ct\,j}$ is a dummy 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, $slant_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of judge j, $female\,district\,\,judge_k$ is a dummy equal to 1 if the district judge is female, X_j are demographic controls for the circuit court judge, δ_k are district judge fixed effects, δ_j are circuit judge fixed effects, and δ_{ct} are circuit-year fixed effects. The dataset is at the case level. Standard errors are clustered at the circuit judge level. Note that $slant_j$ and the X_j are not independently identified upon inclusion of circuit judge fixed effects δ_j, but they are identified when interacted with the female-district-judge indicator.

Table 7 reports the estimates from this last set of results. Column (1) shows the estimates from the baseline specification. Circuit judges with a one standard deviation higher slant are 0.01 percentage points (5.6%) 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, column (2) shows that this result hides substantial heterogeneity. When we additionally control for two(three)-way interactions between the district judge being female and the circuit judge being female

[21] 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 they score is above 70 (out of a maximum 100).

and Democratic, we see that male Democratic circuit judges and female Republican judges are less likely to reverse female district judges, but female Democratic judges are more likely to do so. The main effect is unchanged.

Similarly, the results do not change if we additionally control for the judge's vote record (column (3)). The results thus far do not exploit the fact that district cases are also quasi-randomly assigned in each district and year, as it is not necessary for identification. In column (4), we show that when we nonetheless include district-year fixed effects the main result does not change. The results are insensitive to particular word sets used to identify language dimensions

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. On a back-of-the-envelope calculation, a female judge that 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.[22]

5.4 Effect of Gender Slant on Other Judge Characteristics

To strengthen the argument that lexical gender slant is indeed proxying for gender preferences, we explore whether gender slant also affects treatment of judges with different demographic characteristics – namely, political leaning, minority status, and age.

First, Appendix Table 13 explores the effect of gender slant of the most senior judge of the panel on opinion assigment to whether the authoring judge is Democrat, whether the authoring judge is minority and age of the authoring judge. We find that judges with higher slant are 2.7 percentage points (7%) less likely to assign the opinion of the court to a Democratic judge. However, the magnitude of the coefficient decreases by half when we additionally control for a proxy of being liberal (column (2); p-value = 0.10), which is reassuring that the measure is indeed a meaninful proxy for attitudes toward gender specifically. Consistent with this idea, column (3) and (4) show that slant does not impact the probability of the judge assigning the opinion to a minority judge, or the age of the judge.

[22] In fact, Appendix Table 12 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.

Second, Appendix Table 14 shows tha judge with higher gender slant are less likely to cite opinions authored by Democrat-appointed judges (column (1)). Column (2) shows that the effect is, if anything, larger when we control for a proxy of a liberal a judge is. Instead, gender slant does not affect the probability of citing minority judges (column (3)) or the average age of the judges cited (column (4)).

Finally, we check whether judges with higher gender slant are also more likely to reverse decisions of district judges with different characteristics. Appendix Table 15 shows that when circuit judges with higher slant decide on cases of district judges that were appointed by a Democratic President, they are more likely to reverse them than cases of district judges that were appointed by a Republican President (column (1)), although the effect is killed by controlling fro how liberal a judge is (column (2). Circuit judges with higher slant also appear slightly more likely to reverse cases in which the district judge is a minority (column (3)).

Overall, this results suggest that gender is the salient characteristic to which judges with higher gender slant respond. Although gender slant does appear on the surface to influence how judges treat judges appointed by a Democrat President, the effect does not survive controlling for proxy of how liberal a judge is – whereas treatment of female judges is generally not affected by the inclusion of this control. The results can therefore be taken as supporting the idea of gender slant proxying for attitudes towards women.

6 Conclusions

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

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 doubt, 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 stereotypes but large corpora of text available.

References

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti, "Revealing Stereotypes: Evidence from immigrants in schools," Technical Report, National Bureau of Economic Research 2018. 3

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). 5.2

_ , Daniel L. Chen, and Suresh Naidu, "Ideas have Consequences: The Impact of Law and Economics on American Justice," Technical Report, NBER 2017. 10

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," 2018. 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, Nicola Gennaioli, and Andrei Shleifer, "Salience Theory of Judicial Decisions," *The Journal of Legal Studies*, 2015, *44* (S1), S7–S33. 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 US Courts of Appeals*, University of Virginia Press, 2014. 5.1

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

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri, "Are Referees and Editors in Economics Gender Neutral?," *October https://eml. berkeley. edu/˜ sdellavi/wp/EditorGender_Oct24. pdf*, 2018. 1

Carlana, Michela, "Implicit Stereotypes: Evidence from Teachers' Gender Bias," 2018. 1

Chen, Daniel L., "This Morning's Breakfast, Last Night's Game: Detecting Extraneous Factors in Judging," Working Paper, ETH Zurich 2014. 1

_ and Jasmin Sethi, "Insiders and Outsiders: Does Forbidding Sexual Harassment Exacerbate Gender Inequality?," Working Paper, University of Chicago October 2011. 1, 4

_ and Susan Yeh, "The Construction of Morals," *Journal of Economic Behavior and Organization*, August 2014, *104*, 84–105. 1, 4

_ and _ , "How Do Rights Revolutions Occur? Free Speech and the First Amendment," Working Paper, ETH Zurich December 2014. 1, 4

_ , Tobias J. Moskowitz, and Kelly Shue, "Decision-Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," *Quarterly Journal of Economics*, 2016. 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," Technical Report, IFS Working Papers 2019. 3

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, 2011, *108* (17), 6889–6892. 1

DiMaggio, Paul, "Culture and cognition," *Annual review of sociology*, 1997, *23* (1), 263–287. 2.6

Dunn, Matt, Levent Sagun, Hale Şirin, and Daniel Chen, "Early predictability of asylum court decisions," in "Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law" ACM 2017, pp. 233–236. 1

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. 5.3

Epstein, Lee, William M. Landes, and Richard A. Posner, *The Behavior of Federal Judges*, Harvard University Press, 2013. 4, 9, 13, 4, 19

Esteves, Ana Beatriz Dillon, Luiza Lopes Franco Costa, Roxana Kreimer, Noel Struchiner, and Ivar R Hannikainen, "Gender Stereotypes Underlie Child Custody Decisions," 2018. 1

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

Gabaix, Xavier, "A sparsity-based Model of Bounded Rationality," *The Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710. 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

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. 1, 3, 6, 4, 9, 13

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

Greenwald, Anthony G, Debbie E McGhee, and Jordan LK Schwartz, "Measuring individual differences in implicit cognition: the implicit association test.," *Journal of personality and social psychology*, 1998, *74* (6), 1464. 1

Hengel, Erin, "Publishing while Female. Are women held to higher standards? Evidence from peer review.," 2019. 1

Jacobi, Tonja and Dylan Schweers, "Justice, interrupted: The effect of gender, ideology, and seniority at

Supreme Court oral arguments," *Virginia Law Review*, 2017, pp. 1379–1485. 5

Jakiela, Pamela and Owen Ozier, "Gendered language," 2019. 1

Kastellec, Jonathan P., "Racial Diversity and Judicial Influence on Appellate Courts," *American Journal of Political Science*, January 2013, *57* (1), 167–183. 1

Kozlowski, Austin C, Matt Taddy, and James A Evans, "The Geometry of Culture: Analyzing Meaning through Word Embeddings," *arXiv preprint arXiv:1803.09288*, 2018. 1, 2.3

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

Lewis, Molly and Gary Lupyan, "What are we learning from language? Associations between gender biases and distributional semantics in 25 languages," 2019. 3

Lowe, Matt, "Types of contact: A field experiment on collaborative and adversarial caste integration," *Unpublished Manuscript*, 2018. 3

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, "Distributed representations of words and phrases and their compositionality," in "Advances in neural information processing systems" 2013, pp. 3111–3119. 1, 2.1

_ , _ , _ , _ , and _ , "Distributed Representations of Words and Phrases and their Compositionality," in C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds., *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc., 2013, pp. 3111–3119. 2.2

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 (EMNLP)" 2014, pp. 1532–1543. 1, 2.1, 2.2

Posner, Richard, *How Judges Think*, Harvard University Press, 2008. 1, 5.2

Posner, Richard A, "An economic analysis of the use of citations in the law," *American Law and Economics Review*, 2000, *2* (2), 381–406. 5.2

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

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. 5.1

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 [Job Market Paper]," 2019. 1

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. 3

Figure 1: Gender Dimension

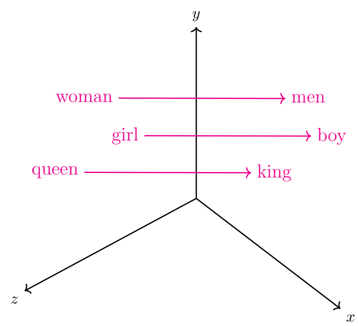


Figure 2: Measuring Gender Stereotypes using Cosine Similarity

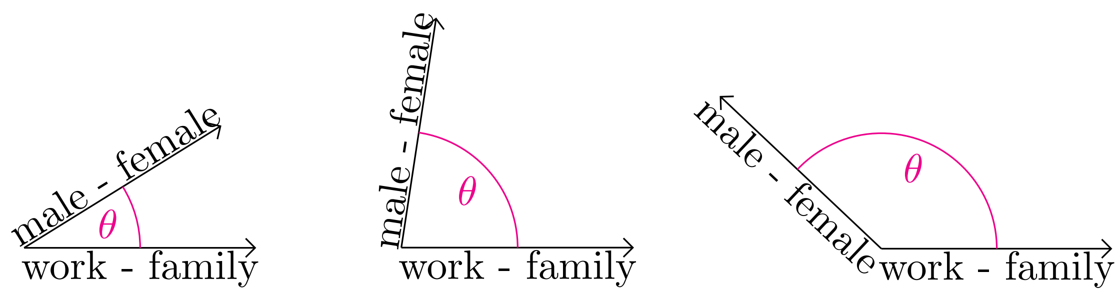
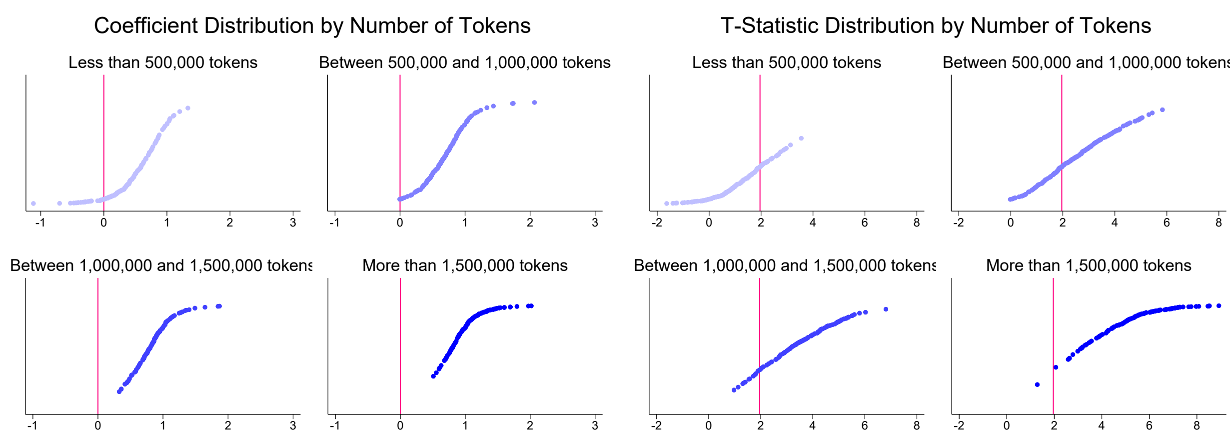
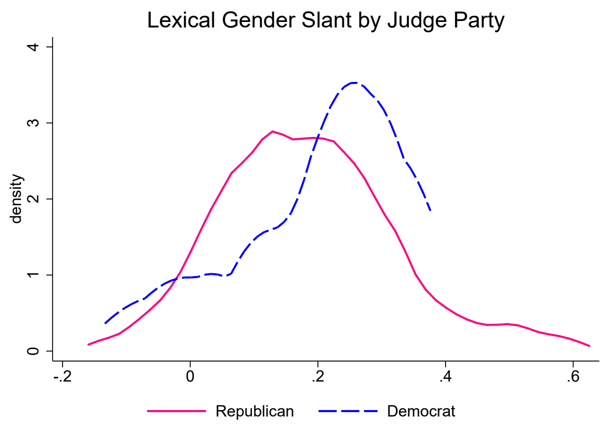


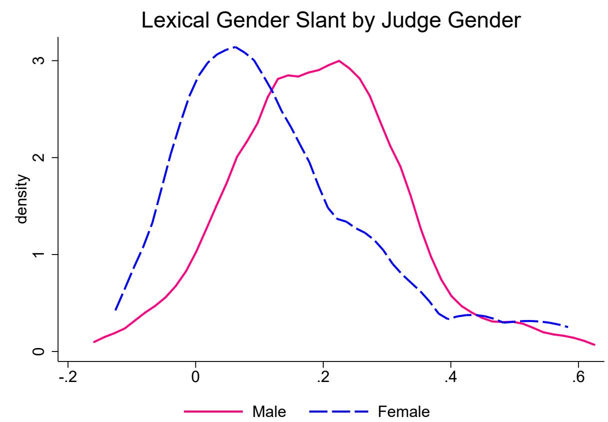
Figure 3: Judge Specific Word Embeddings Capture Gender Information



Notes: The graphs show the distribution of the coefficient and the t-statistic resulting from a regressions of 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, for sets of judges with different number of tokens. Each observation corresponds to a different judge.

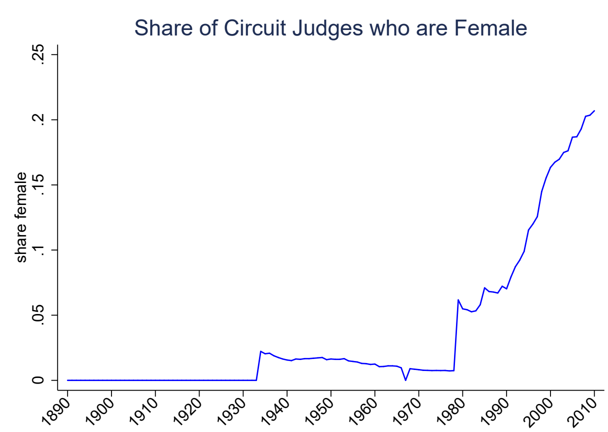
Figure 4: Gender Slant, by Demographic Characteristics





Notes: The graphs show the distribution of the slant measure, by party of nominating President (graph to the left) and by judge gender (graph to the right).

Figure 5: Share of Circuit Judges who are Female



Notes: The graph shows the share of Circuit Judges who are female by year, for the original 10 circuits.

Table 1: Word Sets

Female	his, he, him, mr, himself, man, men, king, male, fellow
Male	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	0.109					0.308
	(0.261)					(0.310)
Female		-0.502*				-0.621***
		(0.288)				(0.181)
Minority			-0.098			-0.128
			(0.329)			(0.181)
Born in 1920s				-0.069		0.122
				(0.191)		(0.208)
Born in 1930s				-0.765***		-0.682***
				(0.203)		(0.226)
Born after 1940				-0.537**		-0.518**
				(0.229)		(0.243)
Born in the Midwest					-0.412*	-0.319
					(0.223)	(0.224)
Born in the South					-0.056	-0.402
					(0.221)	(0.286)
Born in the West					-0.248	-0.672
					(0.262)	(0.423)
Observations	139	139	139	139	139	139
Outcome Mean	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R2	-0.006	0.020	-0.007	0.087	0.013	0.447
Circuit FE						X
Additional Demographic Controls						X

Notes: The table shows the correlation between judges' demographic characteristics and gender slant. We regress slant on demographic controls in separate regressions (columns (1) to (5)) and in a multivariate regression that includes additional controls (column (6)). Additional demographic controls are dummies for religion, law school attended, and whether the judge had federal experience prior to being appointed to an Appellate Court. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Judges with Daughters have Lower Gender Slant

Dependent Variable	Gender Slant
	(1)
Daughter	-0.477*
	(0.274)
Democrat	-0.016
	(0.535)
Female	-0.659***
	(0.232)
Democrat * Female	
Observations	98
Outcome Mean	-0.085
Adjusted R2	0.528
Circuit FE	X
Number of Children FE	X
Additional Demographic Controls	X

Notes: The table shows the effect of having a daughter on gender slant. We regress slant on an indicator variable for having at least one daughter, demographic controls, number of children fixed effects 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Standard errors are clustered at the judge level. Data on judges' family composition is from Glynn and Sen (2015). Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Effect of Gender Slant on Decisions in Gender Related Cases

Dependent Variable	Liberal Vote		
	(1)	(2)	(3)
Gender Slant	-0.046***	-0.047***	-0.083***
	(0.013)	(0.013)	(0.014)
Democrat	0.174***	0.167***	0.234***
	(0.030)	(0.030)	(0.032)
Female	0.095***	0.111***	0.055**
	(0.027)	(0.033)	(0.022)
Democrat * Female		0.036	
		(0.049)	
Observations	3086	3086	3086
Clusters	113	113	113
Outcome Mean	0.395	0.395	0.395
Circuit-Year FE	X	X	X
Topic FE	X	X	X
Additional Demographic Controls	X	X	X
Interacted Demographic Controls		X	
Career FE			X

Notes: The table shows the effect of gender slant on decisions in gender related cases. We regress an indicator variable equal to 1 if a judge voted liberally gender-related cases on the gender slant of the judge, demographic controls, circuit-year fixed effects, topic fixed effects, and dataset 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Career fixed effects are indicator variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit. Standard errors are clustered at the judge level. Data on votes on gender-related cases are from Epstein et al. (2013)'s update of the Sustein's (2006) data, and Glynn and Sen (2015). *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effect of Assigning Judge's Slant on Author's Gender

Dependent Variable	Author is Female						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender Slant	-0.020**	-0.020**	-0.015*	-0.023***	-0.023***	-0.014*	-0.010*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.006)	(0.005)
Democrat	-0.065**	-0.033	-0.080**	-0.095***	-0.059**	-0.053*	-0.054**
	(0.029)	(0.034)	(0.033)	(0.029)	(0.026)	(0.031)	(0.019)
Female	0.137***	0.146***	0.160***	0.134***	0.135***		0.042***
	(0.015)	(0.018)	(0.016)	(0.015)	(0.016)		(0.010)
Democrat * Female		-0.120***					
		(0.039)					
Liberal Score Epstein et al. (2013)				0.006			
				(0.041)			
Observations	32052	32052	32052	30614	36939	22828	27695
Clusters	125	125	125	111	125	108	124
Outcome Mean	0.383	0.383	0.383	0.387	0.383	0.347	0.338
Circuit-Year FE	X	X	X	X	X	X	X
Additional Demographic Controls	X	X	X	X	X	X	X
Interacted Demographic Controls		X					
Career FE			X				
Includes Dissents/Concurrences					X		
Excludes Female Senior Judges						X	
Female Judge on Panel FE							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 authoring judge. We regress an indicator variable equal to 1 if the authoring judge is female on the gender slant of the most senior judge on the panel, demographic controls for the most senior judge, 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Career fixed effects are indicator variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit. The liberal score is defined as the average share of liberal votes of the judge in non gender-related cases from the Epstein et al. (2013) data. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that are decided unanimously. Column (5) includes cases with dissents or concurrences. Column (7) further restricts the sample to cases with exactly one female judge on the panel. 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 Probability of Citing a Female Judge

Dependent Variable	Cites at Least One Female Judge				
	(1)	(2)	(3)	(4)	(5)
Gender Slant	-0.009*	-0.008*	-0.010*	-0.009	-0.007
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Democrat	-0.021	-0.030*	-0.046***	-0.049**	-0.045***
	(0.015)	(0.015)	(0.015)	(0.021)	(0.015)
Female	0.123***	0.107***	0.134***	0.117***	-0.096**
	(0.015)	(0.017)	(0.013)	(0.015)	(0.017)
Democrat * Female		0.049*			
		(0.027)			
Liberal Score Epstein et al. (2013)				0.039	
				(0.033)	
Observations	107923	107923	107923	86910	107923
Clusters	139	139	139	112	139
Outcome Mean	0.383	0.383	0.383	0.452	0.383
Circuit-Year FE	X	X	X	X	X
Additional Demographic Controls	X	X	X	X	X
Interacted Demographic Controls		X			
Career FE			X		
Excludes Self-Citations					X

Notes: The table shows the effect of gender slant 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 judge, demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Career fixed effects are indicator variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit. The liberal score is defined as the average share of liberal votes of the judge in non gender-related cases from the Epstein et al. (2013) data. 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.

Table 7: Effect of Gender Slant on Reversals of District Court Cases if Lower Court Judge is Female

Dependent Variable	Votes to Reverse District Decision			
	(1)	(2)	(3)	(4)
Gender Slant * Female District Judge	0.010***	0.010***	0.011***	0.012***
	(0.004)	(0.004)	(0.004)	(0.004)
Democrat * Female District Judge	-0.009	-0.024**	-0.001	-0.007
	(0.014)	(0.009)	(0.016)	(0.013)
Female * Female District Judge	-0.009	-0.022***	-0.0046	-0.011
	(0.009)	(0.008)	(0.010)	(0.010)
Democrat * Female * Female District Judge		0.152***		
		(0.015)		
Liberal Score Epstein et al. (2013) * Female District Judge			-0.017	
			(0.027)	
Observations	145862	145862	129677	145563
Clusters	133	133	106	133
Outcome Mean for Male Judges	0.180	0.180	0.167	0.180
Outcome Mean for Feale Judges	0.157	0.157	0.157	0.157
Circuit-Year FE	X	X	X	X
Judge FE	X	X	X	X
District Judge FE	X	X	X	X
Additional Demographic Controls	X	X	X	X
Interacted Demographic Controls		X		
District-Year FE				X

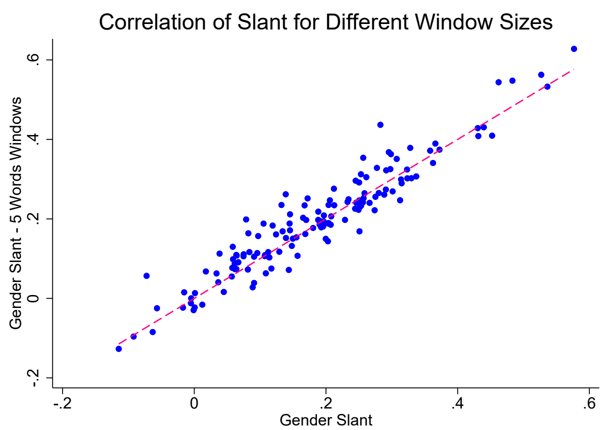
Notes: The table shows the effect of gender slant on probability of reversal of district court decisions. We regress an indicator variable equal to 1 if the judge votes to reverse the district decision on the gender slant of the judge interacted with an indicator variable for whether the district judge is female, demographic controls interacted with an indicator variable for whether the district judge is female, judge fixed effects, district 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. The liberal score is defined as the average share of liberal votes of the judge in non gender-related cases from the Epstein et al. (2013) data. 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.

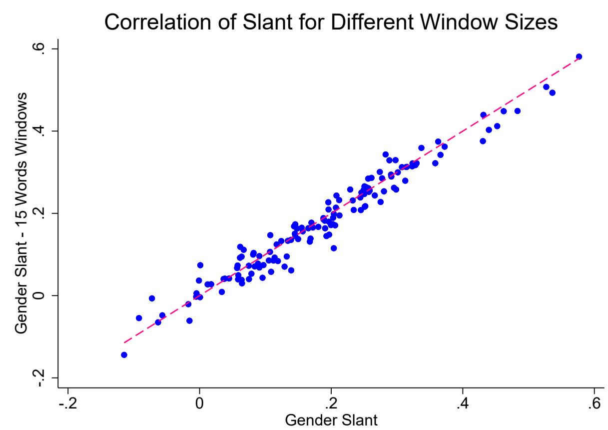
A Appendix

A.1 Correlation of Gender Slant for Embeddings Based on Different Window Sizes

A key hyperparameters 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.

Appendix Figure 1: Correlation of Gender Slant for Embeddings Based on Different Window Sizes





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.

A.2 Demographic Correlates of Implicit Bias in Legal Sector

In this section, we explore how two measures of gender bias correlate with demographic characteristics in a sample of individuals working in the legal profession. In particular, we exploit the fact that Project Implicit, a large-scale administration of an Implicit Bias Test, makes their data available to researchers. We focus on a sample of approximately 10,000 individuals from the United States who self-report having a Juris Doctor degree and working in the legal sector ("Lawyers, judges, and related workers"), and took the test between 2005 and 2018.

Appendix Table 1 shows the coefficient from a regression on implicit gender bias (Columns (1) to (3)) and explicit gender bias (Columns (4) to (6)) on the gender, political leaning and age of the individual. We use the two measures of implicit and explicit bias commonly used in the literature. Implicit bias is defined as the standardized IAT score for the Career-Family IAT (difference in the mean reaction times for the stereotypical and non-stereotypical association sequences divided by the pooled standard deviations). The explicit bias score is defined as the standardized difference between the self-reported association between men and career and women and family.

Contrary to what we see for gender slant, members of the legal profession who are more liberal tend to display lower implicit and explicit gender bias. Female respondents present instead higher implicit bias. Finally, younger respondents display higher implicit bias but lower explicit bias.

Appendix Table 1: Correlates of Implicit and Explicit Bias

Dependent Variable	Implicit Bias			Explicit Bias		
	(1)	(2)	(3)	(4)	(5)	(6)
Liberal	-0.070***			-0.170***		
	(0.024)			(0.026)		
Female		0.118***			0.022	
		(0.021)			(0.021)	
Age			0.004***			-0.005***
			(0.001)			(0.001)
Observations	9954	9954	9954	9954	9954	9954
Outcome Mean	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R2	0.001	0.003	0.002	0.005	0.000	0.004

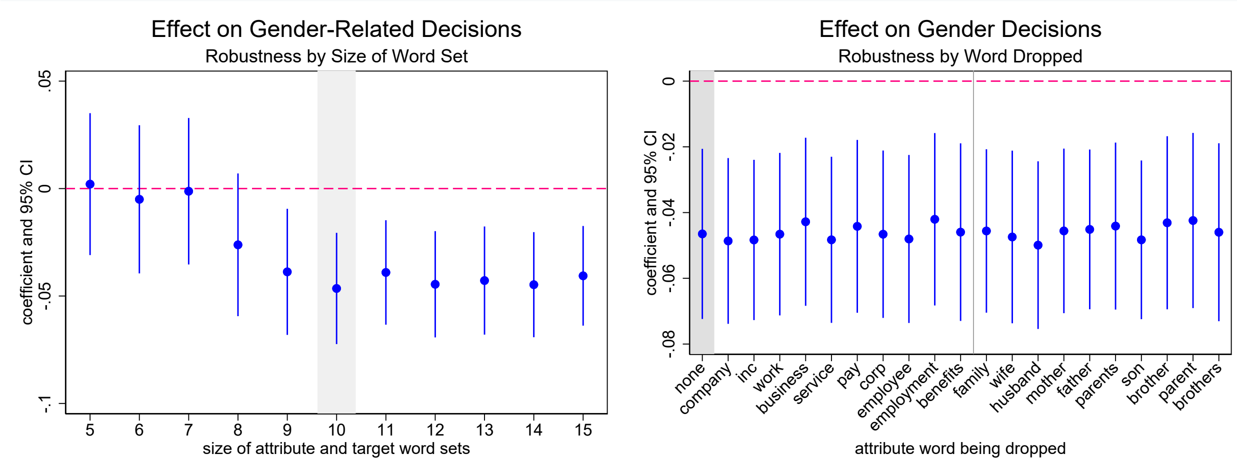
Notes: The table shows the correlation between demographic characteristics and implicit and explicit bias for a sample of self-identified professionals working in the legal sector. The data is from the Career-Family IAT dataset provided by Project Implicit. It provides IAT scores, questions on attitudes towards gender, and self-reported biographic characteristics of individuals who took the Career-Family IAT on the Project Implicit website 2005-2018. The sample is restricted to individuals who self-reported working in the legal sector ("Lawyers, judges, and related workers"), having completed a J.D. and living in the United States. Implicit bias is defined as the standardized IAT score for the Career-Family IAT (difference in the mean reaction times for the stereotypical and non-stereotypical association sequences divided by the pooled standard deviations). The explicit bias score is defined as the standardized difference between the self-reported association between men and career and women and family. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

A.3 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.

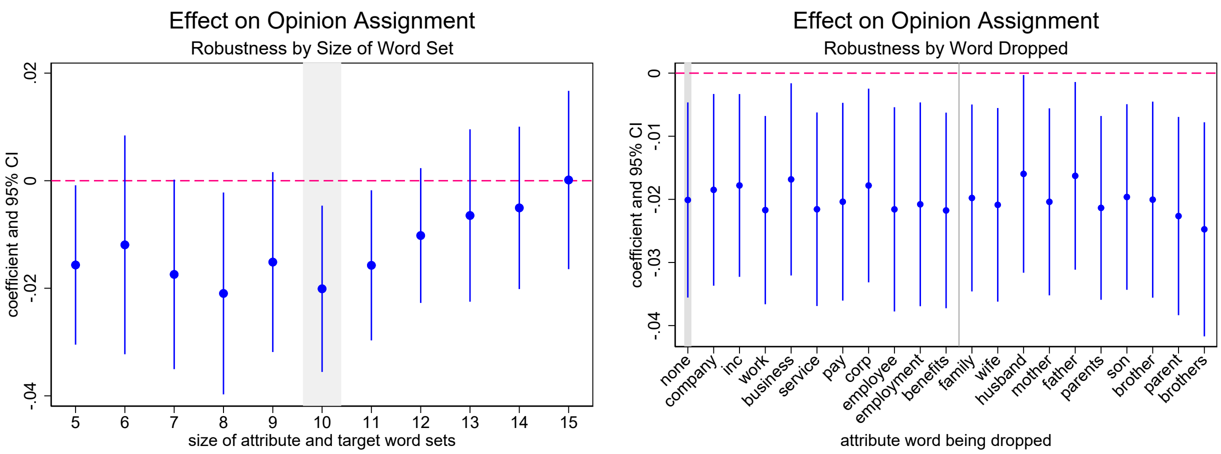
In this section, we explore how the main results of the paper are affected by perurbation of the word sets used. First, we show how the results are affected if we choose the top 5 to top 15 most frequent words appearing in the judicial corpus. Second, we show how the results change if we drop one word 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 2: Robustness of the Effect of Gender Slant on Gender Related Decisions



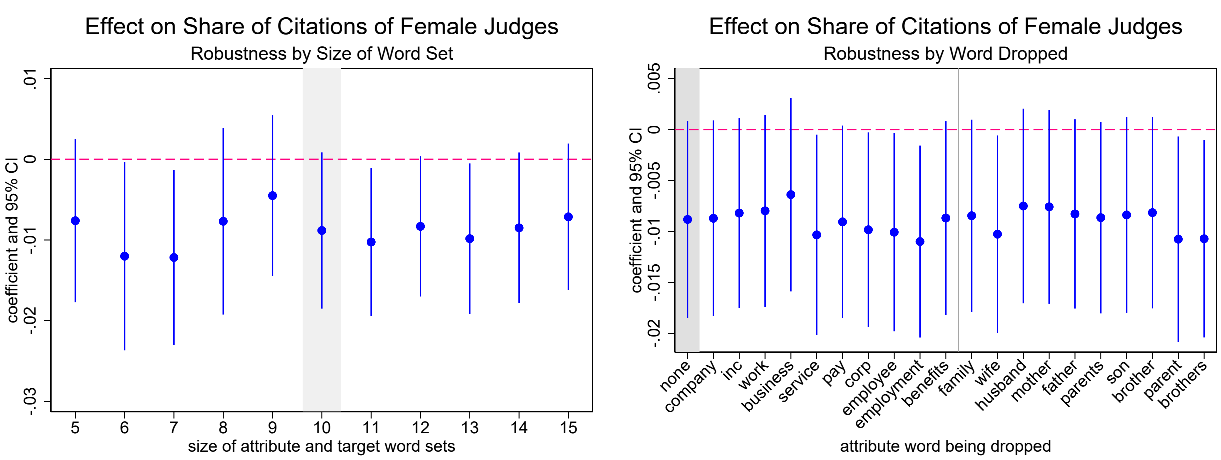
Notes: The graphs show how the effect of gender slant on decisions in gender related cases varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the rigth shows robustness to dropping 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 liberally in gender related cases on the gender slant of the judge, demographic controls, circuit-year fixed effects, topic fixed effects, and dataset 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 an Appellate Court. Standard errors are clustered at the judge level. Data on votes on gender-related cases are from Epstein et al. (2013)'s update of the Sustein's (2006) data.

Appendix Figure 3: Robustness of the Effect of Gender Slant on Authorship Assignment



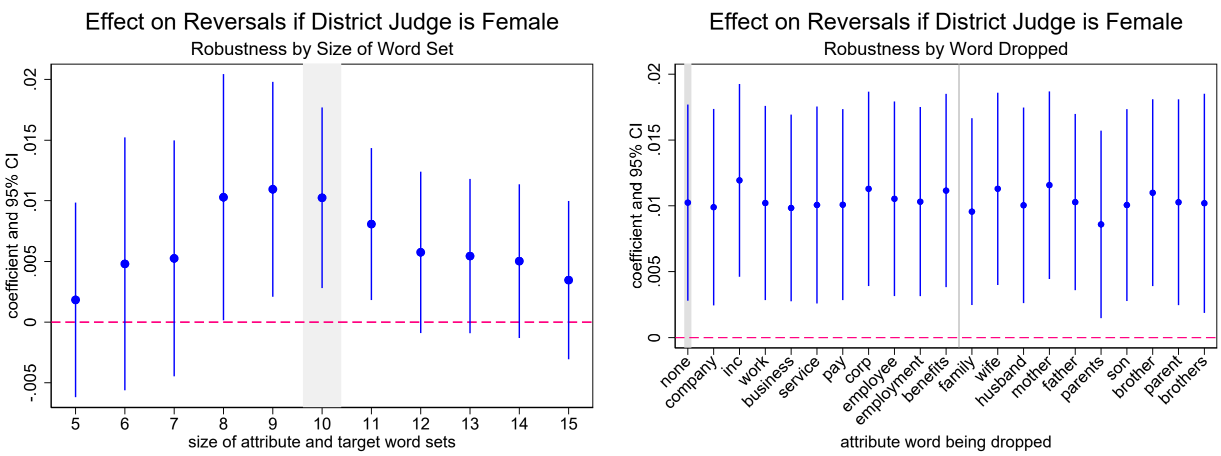
Notes: The graphs show how the effect of the gender slant of the most senior judge on the panel, who is in charge of assigning opinion authorship, on the gender of the authoring judge varies based on the word sets used to identify the gender and attribute dimension. The graph on the left shows robustness to using word sets of different sizes; the graph on the rigth 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 opinion is assigned to a female author on the gender slant of the most senior judge on the panel, demographic controls, circuit-year fixed effects and topic 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 an Appellate Court. The sample is restricted to cases with a specific author and with one female judge on the panel. Standard errors are clustered at the judge level.

Appendix Figure 4: Robustness of the Effect of Gender Slant on Citations



Notes: The graphs show how the effect of the gender slant of the author of the 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 rigth 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 opinion cites at least one female judge on the gender slant of the judge, demographic controls, circuit-year fixed effects and topic 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 an Appellate Court. Standard errors are clustered at the judge level.

Appendix Figure 5: Robustness of the Effect of Gender Slant on Reversals



Notes: The graphs show how the effect of gender slant on probability of reversal of district court judges of different genders 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 rigth 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 votes 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, judge fixed effects, district 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 an Appellate Court. The sample is restricted to cases for which we were able to determine the gender of the district judge. Standard errors are clustered at the judge level.

A.4 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 characteristic 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 observabls 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 of unobservable were only half as large as the amount of selection on observables.

Appendix Table 2: Effect of Gender Slant on Decisions in Gender Related Cases

β Uncontrolled	β Controlled	$\overline{R_{max}}$	δ for $\beta = 0$
(1)	(2)	(3)	(4)
-0.040	-0.046	0.084	4.161
		0.112	2.195
		0.280	0.572

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 characteristcs of the judge. Column (2) shows the estimate of the coefficient from the baseline specification (equation (2)). Column (3) shows the value of R_max for which δ (Column (4)) is computed. δ is the degree of selection on unobservables as a proportion of the selection on observabls that would be needed to obtain a bias-adjusted coefficient equal to 0 is computed.

Appendix Table 3: Effect of Gender Slant on Opinion Assignment

β Uncontrolled	β Controlled	$\overline{R_{max}}$	δ for $\beta = 0$
(1)	(2)	(3)	(4)
-0.028	-0.020	0.025	1.205
		0.034	0.616
		0.084	0.157

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 characteristcs of the judge. Column (2) shows the estimate of the coefficient from the baseline specification (equation (3)). Column (3) shows the value of $R_m ax$ for which δ (Column (4)) is computed. $R_m ax$ is set to be 1.5, 2, or 5 times the adjusted R^2 from the baseline specification. δ is the degree of selection on unobservables as a proportion of the selection on observabls that would be needed to obtain a bias-adjusted coefficient equal to 0 is computed.

Appendix Table 4: Effect of Gender Slant on Citations

β Uncontrolled	β Controlled	$\overline{R_{max}}$	δ for $\beta = 0$
(1)	(2)	(3)	(4)
-0.024	-0.009	0.022	0.605
		0.030	0.305
		0.074	0.076

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 characteristcs of the judge. Column (2) shows the estimate of the coefficient from the baseline specification (equation (3)). Column (3) shows the value of $R_m ax$ for which δ (Column (4)) is computed. $R_m ax$ is set to be 1.5, 2, or 5 times the adjusted R^2 from the baseline specification. δ is the degree of selection on unobservables as a proportion of the selection on observabls that would be needed to obtain a bias-adjusted coefficient equal to 0 is computed.

Appendix Table 5: Effect of Gender Slant on Reversals of District Court Cases

β Uncontrolled	β Controlled	$\overline{R_{max}}$	δ for $\beta = 0$
(1)	(2)	(3)	(4)
0.006	0.010	0.0003	53.762
		0.0004	20.511
		0.0010	5.618

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 characteristcs of the judge. Column (2) shows the estimate of the coefficient from the baseline specification (equation (3)). Column (3) shows the value of $R_m ax$ for which δ (Column (4)) is computed. $R_m ax$ is set to be 1.5, 2, or 5 times the adjusted R^2 from the baseline specification. δ is the degree of selection on unobservables as a proportion of the selection on observabls that would be needed to obtain a bias-adjusted coefficient equal to 0 is computed.

A.5 Additional Results on Judicial Decisions

Appendix Table 6: Effect of Gender Slant on Decisions in Gender Related Cases, by Dataset

Dependent Variable	Liberal Vote					
Dataset	Epstein et al. (2013) Data			Glynn and Sen (2015) Data		
	(1)	(2)	(3)	(4)	(5)	(6)
Gender Slant	-0.041***	-0.041***	-0.066***	-0.050**	-0.0541***	-0.059**
	(0.013)	(0.013)	(0.018)	(0.019)	(0.019)	(0.023)
Democrat	0.150***	0.142***	0.185***	0.247***	0.253***	0.266***
	(0.031)	(0.031)	(0.035)	(0.043)	(0.044)	(0.054)
Female	0.122***	0.143***	0.089***	0.084**	0.114***	0.092**
	(0.026)	(0.036)	(0.022)	(0.036)	(0.039)	(0.042)
Democrat * Female		0.038			-0.006	
		(0.057)			(0.069)	
Observations	2335	2335	2335	1620	1620	1620
Clusters	112	112	112	109	109	109
Outcome Mean	0.417	0.417	0.417	0.374	0.374	0.374
Circuit-Year FE	X	X	X	X	X	X
Topic FE	X	X	X	X	X	X
Additional Demographic Controls	X	X	X	X	X	X
Interacted Demographic Controls		X			X	
Career FE			X			X

Notes: The table shows the effect of gender slant on decisions in gender related cases. We regress an indicator variable equal to 1 if a judge voted liberally gender-related cases on the gender slant of the judge, demographic controls, circuit-year fixed effects and topic 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Career fixed effects are indicator variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit. Standard errors are clustered at the judge level. Data on votes on gender-related cases are from Epstein et al. (2013)'s update of the Sustein's (2006) data, and Glynn and Sen (2015). *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 7: Effect of Gender Slant on Decisions in Non-Gender Related Cases

Dependent Variable	Liberal Vote		
Dataset	Epstein et al. (2013) Data		
	(1)	(2)	(3)
Gender Slant	-0.030***	-0.0321***	-0.017
	(0.010)	(0.010)	(0.011)
Democrat	0.078**	0.106***	0.073
	(0.038)	(0.032)	(0.045)
Female	0.094***	0.112***	0.097***
	(0.019)	(0.024)	(0.016)
Democrat * Female		-0.139***	
		(0.045)	
Observations	5477	5477	5477
Clusters	112	112	112
Outcome Mean	0.431	0.431	0.431
Circuit-Year FE	X	X	X
Topic FE	X	X	X
Additional Demographic Controls	X	X	X
Interacted Demographic Controls		X	
Career FE			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 a judge voted liberally gender-related cases on the gender slant of the judge, demographic controls, circuit-year fixed effects and topic 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Career fixed effects are indicator variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit. Standard errors are clustered at the judge level. Data on votes on non-gender related cases are from Epstein et al. (2013)'s update of the Sustein's (2006) data. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 8: Effect of Gender Slant on Decisions in Gender Related Cases, Additionally Controls for Liberal Score

Dependent Variable	Liberal Vote
	(1)
Gender Slant	-0.043***
	(0.013)
Democrat	0.158***
	(0.035)
Female	0.084***
	(0.028)
Liberal Score	0.140
	(0.117)
Observations	3078
Clusters	111
Outcome Mean	0.395
Circuit-Year FE	X
Topic FE	X
Additional Demographic Controls	X

Notes: The table shows the effect of gender slant on decisions in gender related cases. We regress an indicator variable equal to 1 if a judge voted liberally gender-related cases on the gender slant of the judge, demographic controls, circuit-year fixed effects, topic fixed effects, and dataset 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 an Appellate Court. The liberal score is defined as the average share of liberal votes of the judge in non gender-related cases from the Epstein et al. (2013) data. Standard errors are clustered at the judge level. Data on votes on gender-related cases are from Epstein et al. (2013)'s update of the Sustein's (2006) data, and Glynn and Sen (2015). *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 9: Effect of Gender Slant on Decisions in All Cases, Songer Auburn Data

Dependent Variable	Liberal Vote		
Dataset	Songer-Auburn 5% Dataset		
	(1)	(2)	(3)
Gender Slant	-0.003	-0.003	-0.012
	(0.007)	(0.007)	(0.011)
Democrat	-0.005	-0.005	-0.071***
	(0.018)	(0.018)	(0.026)
Female	-0.008	-0.003	0.003
	(0.020)	(0.023)	(0.018)
Democrat * Female		0.056	
		(0.04)	
Observations	13280	13280	13280
Clusters	136	136	136
Outcome Mean	0.380	0.380	0.380
Circuit-Year FE	X	X	X
Topic FE	X	X	X
Additional Demographic Controls	X	X	X
Interacted Demographic Controls		X	
Career FE			X

Notes: The table shows the effect of gender slant on decisions. We regress an indicator variable equal to 1 if a judge voted liberally on the gender slant of the judge, demographic controls, circuit-year fixed effects and topic 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. Career fixed effects are indicator variables equal to 1 if the judge sat on at least one case in a given period of 25 years and circuit. Standard errors are clustered at the judge level. The sample is restricted to cases where votes could be coded to be liberal or conservative. Data on votes are from the Songern-Auburn 5% dataset. *** p<0.01, ** p<0.05, * p<0.1.

A.6 Additional Results for Opinion Assignment

Appendix Table 10: Effect of Language Slant of Senior Judge on Panel on whether 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	0.001	0.003	-0.000	-0.001	0.002	0.000
	(0.005)	(0.004)	(0.003)	(0.003)	(0.006)	(0.005)
Democrat	-0.000	-0.020	-0.020*	0.009	-0.018	-0.021
	(0.015)	(0.016)	(0.010)	(0.013)	(0.021)	(0.019)
Female	0.000	0.009	0.003	-0.003	0.012	0.009
	(0.011)	(0.008)	(0.004)	(0.004)	(0.009)	(0.008)
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
Additional Demographic Controls	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)), if the panel decided unanimously (columns (5) and (6)) on the gender slant of the most senior judge on the panel, demographic controls for the most senior judge, 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 an Appellate Court. Standard errors are clustered at the senior judge level. *** p<0.01, ** p<0.05, * p<0.1.

A.7 Do Higher Slant Senior Judges Assign Different Cases to Female Judges?

Given that senior judges with higher lexical slant assign fewer opinions to female judges, we might want to know whether they also assign different opinions to them. In this section, we ask whether female judges are assigned cases that different along topic or a proxy of case importance when the assigning judge has higher or lower gender slant. We explore case topic by looking at whether the case is about the following topics: economic activity, first amendment, due process, labor law, civil rights, privacy, and criminal appeals. The topic characterization comes from Bloomberg. We proxy for importance by looking at predicted citations based on party names and case topics.

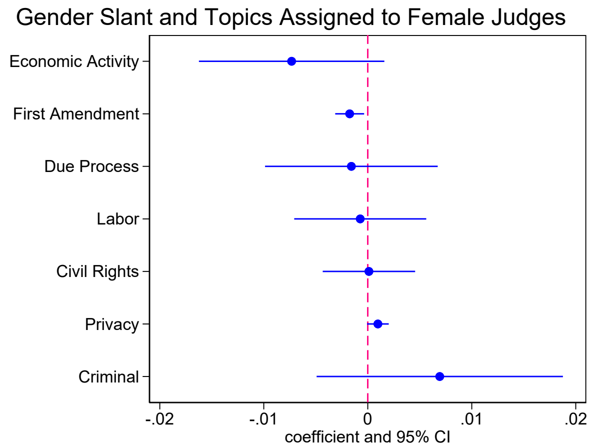
In particular, we estimate the following difference-in-difference specification:

$$\begin{aligned} y_{ict} = &\pi \, female\, author_c * slant_j \\ &+ female\, author_c * X_j' \gamma + \delta_j + \delta_{ct} + \varepsilon_{jictk} \end{aligned} \tag{6}$$

where y_{ict} is a characteristic of case i in circuit c in year t (either as a dummy equal to 1 if the case belongs to a give topic, or predicted forward citations based on party names and topic). On the right-hand side, $slant_j$ is the slant (standardized cosine similarity between the gender and career-family dimensions) of judge j, who is the most senior judge on the panel, $female\, author_k$ is a dummy equal to 1 if the district judge is female, X_j are demographic controls for the senior judge, δ_j are senior judge fixed effects, and δ_{ct} are circuit-year fixed effects. The dataset is at the case level. Standard errors are clustered at the circuit judge level.

Figure 6 shows that overall, slanted senior judges do not appear to systematically assign cases on different topics to female judges, although they might be less likely to assign cases related to regulations of economic activity to female judges. This take-away is confirmed by 11, which shows that slanted senior judges do not appear to be systematically assigning less important cases to female judges, an interesting result in light of the literature on discrimination in task assignments.

Appendix Figure 6: Effect of Gender Slant on Liberal Voting, by Case Topic



Notes: The graph explores whether slanted senior judges assign different types of cases to female judges. We regress a dummy for the case to be regarding a given topic on 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 an Appellate Court. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that are decided unanimously. Standard errors are clustered at the judge level.

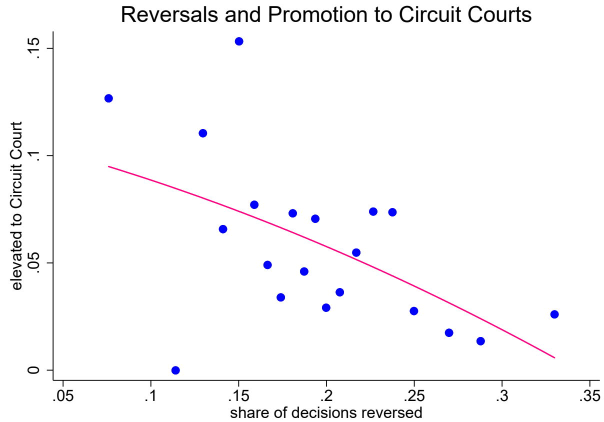
Appendix Table 11: Effect of Gender Slant on Reversals of District Court Cases if Lower Court Judge is Female

Dependent Variable	Forward Predicted Citations
	(1)
Gender Slant * Female Author	0.001
	(0.002)
Democrat * Female Author	-0.016
	(0.011)
Female * Female Author	-0.006
	(0.007)
Observations	31616
Clusters	123
Outcome Mean	1.739
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 citations on 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 an Appellate Court. Interacted demographic controls additionally includes two-way interactions for gender, party and race. The sample is restricted to cases with a specific author, with at least one female judge on the panel, and that are decided unanimously. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

A.8 Additional Results for Reversals

Appendix Figure 7: 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, and law school attended. 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: Reversals and Promotion from District to Circuit Courts

Dependent Variable	Promoted to Circuit Court	
	(1)	(2)
Share of Decisions Reversed on Appeal	-0.351***	
	(0.136)	
Share of Votes to Reverse on Appeal		-0.372***
		(0.116)
Female	0.036	0.037
	(0.028)	(0.029)
Democrat	-0.022	-0.018
	(0.0191)	(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, and law school attended. 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 judge level. *** p<0.01, ** p<0.05, * p<0.1.

A.9 Effect of Gender Slant on Other Judge Characteristics

Appendix Table 13: Effect of Language Slant of Senior Judge on Panel on Opinion Assignment, other Judge Characteristics

Dependent Variable	Author is Democrat	Author is Democrat	Author is Minority	Author Age
	(1)	(2)	(4)	(5)
Gender Slant	-0.027***	-0.014*	0.006	0.069
	(0.010)	(0.007)	(0.008)	(0.168)
Democrat	0.156***	0.0923***	0.019	1.176**
	(0.021)	(0.019)	(0.024)	(0.566)
Female	-0.045**	-0.027**	0.025	-0.009
	(0.019)	(0.011)	(0.015)	(0.499)
Liberal Score		0.033		
		(0.034)		
Observations	46735	27350	23436	120365
Clusters	137	110	126	139
Outcome Mean	0.366	0.333	0.340	63.030
Circuit-Year FE	X	X	X	X
Additional Demographic Controls	X	X	X	X
Panel Includes Democrat Judge	X	X		
Panel Includes Democrat and Female Judge				
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 characteristics of the authoring judge. We regress an indicator variable equal to 1 if the authoring judge was appointed by a Democratic President (columns (1) and (2)), if the authoring judge is minority (column (3)) and age of the authoring judge (column (4)) on the gender slant of the most senior judge on the panel, demographic controls for the most senior judge, 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 an Appellate Court. The liberal score is defined as the average share of liberal votes of the judge in non gender-related cases from the Epstein et al. (2013) data. The sample is restricted to cases with a specific author that are decide unanimously. Columns (1) and (2) additionally restricts the sample to cases with one democratic judge on the panel and column (4) 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 14: Effect of Language Slant on Citations

Dependent Variable	Cites Democrat	Cites Democrat	Cites Minority	Average Age	Average Bias
	(1)	(2)	(3)	(4)	(5)
Gender Slant	-0.011**	-0.015***	-0.005	-0.069	0.112***
	(0.005)	(0.006)	(0.005)	(0.083)	(0.012)
Democrat	0.014	-0.018	-0.032*	0.010	0.003
	(0.018)	(0.022)	(0.019)	(0.153)	(0.034)
Female	0.027**	0.023**	0.049***	-0.017	-0.025
	(0.011)	(0.012)	(0.010)	(0.156)	(0.020)
Liberal Score		0.0105			
		(0.033)			
Observations	107923	86910	107923	107923	98435
Clusters	139	112	139	139	139
Outcome Mean	0.607	0.612	0.336	61.407	0.052
Circuit-Year FE	X	X	X	X	X
Additional Demographic Controls	X	X	X	X	X

Notes: The table shows the effect of gender slant 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 (columns (1) and (2)), at least one case authored by a minority judge (column (3)), the average age of the authors of cited opinions (column (4)), and the average slant of the authors of the cited opinions (column (5)) on the gender slant of the judge, demographic controls and circuit-year fixed effects. Demographic controls are gender, party of appointing President, region of birth, cohort of birth, religion, law school attended, and whether the judge had federal experience prior to being appointed to an Appellate Court. The liberal scores are the average share of liberal votes of the judge in the 5% hand-coded Songer Auburn data and in the non gender-related cases in the Epstein et al. (2013) data respectively. The sample is restricted to cases in which the opinion is authored by a specific judge. Standard errors are clustered at the judge level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 15: Effect of Gender Slant on Reversals of District Court Cases, by Characteristics of Lower Court Judge

Dependent Variable	Votes to Reverse District Decision		
	(1)	(2)	(3)
Gender Slant * Democrat District Judge	0.006*	0.004	
	(0.004)	(0.004)	
Democrat * Democrat District Judge	-0.022	-0.011	
	(0.014)	(0.015)	
Female * Democrat District Judge	-0.007	-0.001	
	(0.008)	(0.008)	
Gender Slant * Minority District Judge			0.011**
			(0.005)
Democrat * Minority District Judge			-0.009
			(0.010)
Female * Minority District Judge			0.018*
			(0.010)
Liberal Score * Democrat District Judge		-0.054**	
		(0.022)	
Observations	145862	129677	145862
Clusters	133	106	133
Outcome Mean	0.177	0.166	0.177
Circuit-Year FE	X	X	X
Judge FE	X	X	X
District Judge FE	X	X	X
Additional Demographic Controls	X	X	X

Notes: The table shows the effect of gender slant on probability of reversal of district court decisions. We regress an indicator variable equal to 1 if the judge votes 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)), judge fixed effects, district 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 an Appellate Court. The liberal scores are the average share of liberal votes of the judge in the 5% hand-coded Songer Auburn data and in the non gender-related cases in the Epstein et al. (2013) data respectively. 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.