Professional Interactions and Hiring Decisions: Evidence from the Federal Judiciary[1]

Marco Battaglini
Cornell University

Jorgen M. Harris
Occidental College

Eleonora Patacchini
Cornell University

Abstract

We study the effect of hearing cases alongside female judicial colleagues on the probability that a federal judge hires a female law clerk. Federal judges are assigned to cases and to judicial panels at random and have few limitations on their choices of law clerks: these two features make the federal court system a unique environment in which to study the effect of professional interactions and beliefs in organizations. For our analysis, we constructed a unique dataset by aggregating federal case records from 2007-2017 to collect information on federal judicial panels, and by merging this data with judicial hiring information from the *Judicial Yellow Book*, a directory of federal judges and clerks. We find that a one standard deviation increase in the fraction of co-panelists who are female increases a judge's likelihood of hiring a female clerk by 4 percentage points. This finding suggests that increases in the diversity of the upper rungs of a profession can shift attitudes in a way that creates opportunities at the entry level of a profession.

JEL Codes: J16, J82, J110

Keywords: Economics of Gender, Discrimination, Labor Force Composition

[1] This work was supported by a grant from the Cornell Institute for Social Sciences small grant competition. This support was essential to the completion of this work. We would also like to thank our fantastic team of undergraduate and post-graduate research assistants. Francesca Guiso, Sam Markiewitz, Becky Cardinali, Cenjia Lu, Ignasi Merediz Sola, Linchen Zhang, and Hongyi Duan: thank you for your hard work, creativity, and patience.

1. Introduction

Prejudiced and discriminatory beliefs are widespread in human relationships, have profoundly negative consequences both for members of out-groups and for society as a whole, and may lead to inefficient economic decisions. There is a long-standing literature in sociology documenting how these beliefs are formed, when they are maintained, and how they can be disrupted. These topics have also recently attracted growing attention among economists. Several theoretical papers have developed models explaining the perpetuation of different beliefs between groups under different hypotheses, including the scenario when one group's beliefs are incorrect.[2] Still the empirical literature has struggled to formally document causal effects of such beliefs on economic outcomes.[3]

In this paper, we exploit a unique institutional feature of the Federal appellate court system to present clear causal evidence on the effect of exposure to out-groups for individual hiring decisions. Appellate court cases are heard by panels of three judges, randomly selected from a pool of appellate court justices and district court justices. Because appellate judges do not choose the cases that come before their courts or the colleagues with whom they hear these cases, their likelihood of working with female colleagues on cases is effectively random. At the same time, appellate judges are broadly unconstrained in their decision of who to hire as a court clerk, a highly prestigious position typically filled by graduates of top law programs. As a result, changes in the likelihood of hiring a female clerk likely reflect changes in a judge's assessment of the likely ability of a female junior colleague. We can thus exploit exposure to female colleagues in the appellate panels as exogenous shocks to a judge's attitude in order to assess its causal impact on the hiring decision of court clerks. An important and unique feature of this data is that the non-

[2] See Fang and Moro [2010] for an overview.
[3] An extensive discussion of the related literature is presented at the end of this section.

voluntary interactions concern high-stakes, sustained, professional interactions between peers at the elite of their occupations.

We find significant positive effects on the likelihood to hire female clerks from professional interactions with female judges. In particular, we find that a one standard deviation increase in the fraction of case interactions a judge carries out with female colleagues increases the likelihood of hiring at least one female clerk in the next year by 4 percentage points. Our finding is robust to a wide range of tests, including placebo regressions in which we match each judge to the most similar judge within the court and regress the judge's exposure to female colleagues on the hiring decision of the match.

Our work relates to three strands of literature. The first is the literature on the effect of attitudes toward gender roles on labor market outcomes. This literature highlights the role played by such attitudes as important determinants of the gender wage gap and labor participation. Guiso, Sapienza and Zingales [2003] and Algan and Cahuc [2004] use international individual value surveys to show that religious beliefs are associated with a less positive attitude toward working women. Using data from the World Value surveys on OECD countries, Fortin [2005] shows that anti-egalitarian views are negatively correlated with female employment rates and positively correlated with gender pay gaps. This pioneering literature has highlighted important relationships between attitudes toward gender roles and labor conditions for women. However, by relying on cross sectional evidence at the country level it has not fully addressed the question of causal identification.

The second literature to which our work is related examines the effect of integration of groups that differ on ethnic, religious or gender dimensions. This body of work has directly addressed the issue of causality relying on quasi-experimental evidence of integration to evaluate its effects on cultural and social attitudes. Boisjoly et et al. [2006] and Corno et al. [2018] use the random allocation of roommates to show that inter-racial interactions make the advantaged group more empathetic to the disadvantaged and improve the academic performance of the disadvantaged group.[4] Carrell et al. [2016] and Dahl, Kotsadam and Rooth [2018] consider the random assignment of female recruits to squads during boot camp and find that exposure to women leads men to adopt more egalitarian attitudes. All these works measure the effect of the quasi-experimental treatment using measures of self-reported attitudes or the Implicit Association Test[5], a tool developed by sociologists to measure attitudes without reliance on self-reporting.

This literature focuses on the effect of integration on beliefs and attitudes; it therefore does not directly establish effects on economically relevant decisions such as hiring decisions or gender and racial wage gaps. Three exceptions are Beaman et al. [2009], Washington [2008] and Glynn and Sen [2015]. Beaman et al. [2009] exploit random assignments of gender quotas for leadership positions in Indian village councils to show that exposure to women in leadership positions affects gender attitudes and that, in the long term, these changes translate in electoral wins for women. Washington [2008] and Glynn and Sen [2015] study the effect of having daughters on the vote of lawmakers and judges in the U.S. court of Appeals. Washington [2008] finds that, conditional on the number of children, the number of daughters is positively correlated to the propensity of lawmakers to vote in a liberal way, especially on reproductive issues. Glyn and Sen [2015] finds that, conditional on number of children, judges with daughters vote in a more liberal or feminist fashion on cases that involve gender issues than do judges who only have sons.

[4] For other work that exploits the random assignment of roommates, see also Van Laar et al. [2005], Boisjoly et al. [2006], Mark and Harris [2012].
5 For discussion of the use of the Implicit Association Test to measure latent stereotypes, see Carlana [2018].

Our paper contributes to the literature discussed above by proposing quasi-experimental evidence on the exposure of judges to female colleagues in professional interactions on an important economic decision: the hiring of court clerks, a highly prestigious position filled by top law students. Our approach is unique because it directly estimates the effect of professional interactions on hiring decisions. Studying the effect of professional interactions on hiring decisions at the upper reaches of the job market is especially important since this is where the gender wage gap is the largest (see Blau and Kahn [2016]).

Finally, our work contributes to the literature studying the decisions of judges. Boyd, Epstein and Martin (2010) have examined the effect of hearing cases alongside female colleagues on a judge's rulings, but have not looked at the effect of interacting with female colleagues on decisions made outside of court, such as hiring. Glyn and Sen [2015], as said, have looked at the gender of judges siblings on their voting decisions. Ash, Chen, and Ornaghi (2020) have examined the role of gender bias, as measured by the use of sexist language in written opinions, on judge's rulings, likelihood of reversing female district judges, and citation of female authors. Our study of the effect of the gender composition of appellate panels on judges' hiring decisions is relevant in this context because provides insight on how the judges' professional interactions affect their beliefs and decisions.

The reminder of this paper is organized as follows. Section 2 presents a description of the institutional setting concerning the random formation of judicial panels in Appellate Courts and the hiring of clerks. Section 3 describes the construction of our key dataset linking appellate judges' exposure to female judges in judicial panels and their hiring of clerks. Section 4 presents the empirical model and discusses the key identification assumption that judges are allocated randomly to appellate panels. Section 5 presents the main results: Section 5.1 presents the main finding; Section 5.2 presents the placebo regressions and the robustness tests; Section 5.3 presents the results on the heterogeneity of the response depending on the judge's characteristics and the characteristics of the female co-panelists. Section 6 concludes.

2. Institutional Setting

2.1. Federal Appellate Courts

The federal court system in the United States is composed of three tiers: district courts, appellate courts, and the Supreme Court. All cases are initially heard in district courts, where evidence is presented, parties appear, and an initial ruling is made. Parties bringing a case in federal court are entitled to appeal decisions to appellate courts, which review the legal reasoning used in district courts. Parties can request an appeal of appellate court decisions in the Supreme Court, but are not entitled to a review—as a result, the appellate court's decision is final in the great majority of cases.

Federal appellate courts are organized into circuits, which are primarily organized geographically. Twelve circuits cover cases heard in specific geographical areas, while the Federal Circuit Court hears cases involving patents, international trade, money claims against the Federal Government, and other particular subject areas originating anywhere in the United States. Each appellate court hears appeals generated by district courts within their jurisdiction. As a consequence of the geographical organization of courts, federal judges are required to maintain expertise in a wide variety of legal areas, and might be expected to learn different information about the ability of their colleagues in each case that they hear.

Most appellate cases are heard in panels of three judges, with a small number of cases heard by larger panels. While the process used to assign judges to cases is intended to be random in each circuit, the specific mechanism by which assignments are made varies by panel, and includes methods such as the use of computer programs and the drawing of lots (Levy 2019). Some circuit courts, such as the Fifth Circuit, choose panels in a manner that avoids having any judge serve too often with any other judge (Levy 2019), limiting the variance in exposure to other judges in our sample.

Appellate panels are composed of regular appellate judges, senior appellate judges (appellate judges who work part time and do not occupy congressional authorized positions), and visiting judges—typically either retired district, appellate, or Supreme Court judges or district judges serving on a district court that is subsidiary to the appellate court. Though the specific uses of visiting judges vary from court to court, circuits typically ensure that there are at least two regular appellate judges on each case, so visiting judges hear cases on panels with other visiting judges only in exceptional circumstances (Levy 2019). In addition, chief justices will often restrict the set of cases for which visiting judges are used, for instance by only allowing them to hear civil cases (Levy 2019). As a result, while the assignment of a specific judge to a specific case should not depend on that judge's characteristics for appellate judges or visiting judges, visiting judges as a group will be exposed to a different set of cases and colleagues than will regular appellate judges as a group. In addition, because visiting judges that hear cases throughout the year may not be available at the same time, the assumption that co-panelists are randomly assigned to visiting judges relies on the assumption that the timing of visits does not reflect judge characteristics. To avoid making this assumption, we restrict our sample only to appellate and senior appellate judges.

Once a panel is formed, judges accept legal briefs from the appellant and appellee, as well as occasional amicus briefs—briefs filed by people or organizations not directly involved in the litigation but with an interest in the precedent set by the litigation. In some cases, panels also convene in person to hear oral arguments given by attorneys for appellants and appellees. After receiving briefs and hearing arguments, the panel confers, and decides to either affirm the decision of the district court, overrule the district court, send the case back to the district court to hear new evidence, or some combination of the above. A majority (typically two members) of a court must agree to this decision, and in 87% of cases, the court reaches a unanimous decision. Once a decision is made, a member of the court's majority is assigned to write the opinion of the court, and dissenting members of the panel write opinions explaining the grounds of their dissent. If the members of the judicial panel believe that the case addresses novel legal reasoning and will thus be useful as precedent, they elect to publish the opinion, making it available in legal registers and online databases. Cases with published opinion are typically more complex cases, and more likely to involve outside parties. As shown in Section 4.1, the publication decision does not depend on the gender of empaneled judges.

2.1.1. Appointment of Judges and Clerks

Judges are appointed to both district and appellate courts by receiving a nomination from the President and confirmation from the Senate[6]. Once appointed, judges have lifetime tenure on the court, barring serious misconduct. Because judges hear cases in particular states, presidents tend to nominate judges recommended by the senators that represent the state in which a judge will hear cases, so long as that senator is from the president's party (Epstein et al, 2007). While there are no explicit legal requirements that federal judges hold specific qualifications, a majority of appellate court judges have prior judicial experience in either district courts or state courts, and 85% have prior experience practicing law (McMillon 2014). Likewise, while the process of

[6] With the exception of magistrate judges—district judges appointed by current members of a district court (McCabe 2014).

nominating and confirming federal judges is often politically contentions, the large majority of nominees have been confirmed under all recent presidents, with George W. Bush holding the lowest confirmation rate at 78%, and Richard Nixon holding the highest Confirmation rate at 99% (Gramlich 2018).

Most Judges in our sample are over 60 years old, and have served in their current positions for more than a decade and a half. As a result, most federal judges were educated and began serving on the court when women were significantly less represented in the legal profession than they are today. While women make up half of new lawyers today, fewer than 5% of law school graduates were women prior to 1968, and only 36% were female in 1981, the year that the median federal judge in our sample finished law school (American Bar Association 2013). In 1992, when the average judge in our sample started their current position, women made up 43% of new law graduates. As a result, federal judges making hiring decisions now do so in a labor market in which women are significantly more numerous and successful than they were when the judges first formed impressions of the legal profession and of the judiciary.

Judges are provided with a budget for a staff, consisting of law clerks and administrative assistants. Typically hired directly out of law school, law clerks typically serve one-year or two-year terms and are responsible for assisting judges with legal research and decision writing (Posner et. al. 2001). Appellate court clerkships are prestigious and highly sought-after positions, and are used as stepping stones to prestigious positions at top law firms, government agencies and the judiciary (Rhinehart 1994). Appellate court clerkships are especially valuable for law students hoping to enter the judiciary—every Supreme Court clerk serving from 2005-2014 had prior experience as an appellate court clerk (Hess 2015).

Despite efforts to push the hiring date for law clerks to the beginning of their third years, a substantial number of law clerks are hired as early as the first semester of their second year of law school or are asked for informal agreements in the first semester of their second year, and begin work after their third year of law school (Avery et. Al. 2007). As a result, there is as much as a two-year gap between the decision to hire a clerk and the clerk's start date. Given this, we examine the effect of interactions with co-panelists in year *t* on the gender of clerks who start work in year *t+3*, on the assumption that clerks will start up to two years after their date of hire.

Due to the structure of the appellate clerk market, Judges have wide latitude to choose the candidate clerk that best matches their preferences (Avery et. al. 2007). Because approximately half of all law third-year law students report applying to clerkship positions, appellate court clerks typically receive thousands of applications (NALP 2019). Offers of clerkships are typically made with very short decision windows (in some cases, as brief as 10 minutes), so most law students take the first clerkship offered.

3. Data description

We pool data from several sources. Our two primary datasets are the *Judicial Yellow Book*—a directory published annually by Leadership Directories Inc. that lists information on judges and clerks working in the US court system (Leadership Directories Inc. 2007-2014), and a directory of case records published by federal appellate courts from 2007 to 2017 (Leagle Inc. 2018). We also incorporate a rating of the conservatism of the president who appointed each judge using the DW Nominate algorithm (Epstein et. al. 2007).

3.1. Primary Data

Judge and clerk information is collected from the *Judicial Yellow Book* published by Leadership Directories Inc. Intended as a resource for attorneys presenting cases in state and federal court, the *Judicial Yellow Book* includes information on the names and backgrounds of judges serving at all levels of the federal court system, as well as the names and limited background information of each judge's clerks. We purchased archived copies of the *Judicial Yellow Book* from Leadership Directories Inc. for the years 2007-2017, in the form of pdf pre-publication masters. We use these data to determine the characteristics of judges and the gender of the clerks hired by each judge in each year.

For each judge, the *Judicial Yellow Book* lists current courthouse, start date, date of birth, appointing president, and provides information on all staff members.[7] It particular, for each staff member the *Judicial Yellow Book* lists name, title, beginning and end of term, and contact information, and education. We use these data to determine which clerks worked for each judge in each year. We do so by taking advantage of the consistent formatting of the yellow book pages in the following way. Each subsection of the yellow books begins with a header "Chambers of Judge [judge's name]," followed by a list of judge characteristics and a list of staff, preceded by a "Staff" sub-header. Law clerks and other staff are then listed with their title, followed by their name, with biographical information indented on following lines. We use this formatting both to determine which staff work for each judge and to exclude staff with titles other than "Law Clerk," such as

[7] The Yellow Books also contain information on the education (degrees earned, year of degree, and alma mater) and prior experience (in government, other judicial offices, law practice, private sector, military, and academia) of judges. In addition, they contain information on the education (degrees earned, year of degree, and alma mater) of judicial staff. This information, however, is largely incomplete, particularly for staff members, and is not used in our analysis. We use judge's college graduation year and law school graduation year to construct the age of the judge in case the date of birth is missing.

"Administrative Assistant" or "Judicial Assistant." We count 7,443 court clerks, with between 1 and 4 clerks working for each judge in each year in 98% of cases.

The gender of judges and clerks are derived using the gender guesser algorithm (Pérez 2016), a tool that determines if a name is male, female or uncertain by comparing it to a database of 40,000 names from 54 countries, primarily in the United States and Europe. We have verified the results of this algorithm both by comparing its results to the relative frequency of men and women with each first name in the US Census and by confirming the genders of judges with ambiguous names through web searches. Using this method, we are able to determine the gender of 95.5% of clerks and 91% of judges. We determined the gender of the remaining 9% of judges by examining biographical information from their court websites and/or news articles. Clerks for whom gender cannot be determined are excluded from the analysis.

Ethnicity of the judges is constructed by comparing judge names to surnames occurring at least 100 times in the 2010 decennial census. If more than 70% of census respondents with a judge's surname are Hispanic, we consider the judge to be Hispanic. If fewer than 30% of census respondents with a judge's surname are Hispanic, we consider the judge to be non-Hispanic. We determined the ethnicity of judges with an indeterminate surname through court websites and news articles.

To determine which judges sat on panels together in each year, we scraped information from the online court records aggregator leagle.com. Leagle stores and categorizes the decisions handed down by the United States courts – trial courts, appellate courts, and the Supreme Court. The library is comprehensive and contains over 5 million cases since 1950. We pool information on the universe of published cases heard between 2004 and 2017, in total 50,813 cases. For each case, Leagle provides the full text of the court's decision, exactly as it appears in published court documents. In addition, these court documents include a set of headers with the case's docket number and name, the date(s) the case was heard and decided, the court in which the case was heard, the names and affiliations of attorneys for the appellant and appellee, and most importantly for our purposes the names of the judges who heard that case. We use these records to identify the judges serving on the appellate panel for a given case.[8] Specifically, we determine which judges heard a case by exploiting the consistent formatting of court document headers to identify the area where presiding judges are typically listed. We then capture each word in this section of the document to determine whether it is a possible judge name or a linking or descriptive word, such as "Before," "Justice," or "Honorable". In following this procedure, we err on the side of including too many potential judge names rather than including too few. Therefore, we further compare the list of potential judge names to the list of surnames held by at least 100 people in the 2010 US Census, and remove all words not in the list of surnames.

We combine these two primary datasets by matching judges appearing in cases originating from each circuit court in each year in the case records data with judges listed in the *Judicial Yellow Book* in that circuit court or a subsidiary district court in that year. Judges from subsidiary district courts are included because judges from district courts are invited to serve on appellate panels (28 U.S. Code § 292)[9]. When merging the case data to the list of judges in the *Judicial Yellow Book*, 66% of potential judge names identified in the case records are also found in the list of judges in the relevant appellate court or subsidiary district courts. An analysis of the remaining 34% of potential names finds that they consist of names of attorneys or parties incorrectly

[8] Once a panel of judges is assigned to a case, the panel remains together until a decision is made. In rare cases, such as in the case of sickness, a judge will be replaced on a panel. In these circumstances, all listed judges are included as interacting colleagues in our measures.
[9] 28 U.S. Code § 44. Appointment, tenure, residence and salary of circuit judges (https://www.law.cornell.edu/uscode/text/28/44)

categorized as judges or, in fewer cases, judges visiting from other circuits and retired judges hearing cases as senior judges. These 34% of names are dropped from the analysis and not used in determining the interactions of each judge. Among appellate court judges listed in the judicial yellow books, 85% of judges appear on at least one case record in the year that they are listed. Among the 15% of judges who do not appear in any published cases, the majority are senior judges, and thus have discretion to hear few or no cases in a year. These judges may have heard cases over the year but heard no published cases, may have taken sick leave, or may be recorded inconsistently in the two data sources. Among district court judges listed in the judicial yellow books, 12% appear on at least one case record in the year that they are listed, consistent with a significant minority of district judges hearing appellate cases in any particular year. Judge names that do not match across these two data sources are eliminated. In total, we identify 298 appellate judges and 589 district judges who served on an appellate court panel that produced a published opinion at least once between 2004-2017.[10] In our final database of 50,484 cases, 70% have three judges that are included in the analysis, 20% have two judges, 6% have one judge, and 4% have more than three judges. Cases can have fewer than three recognized judges if a member of the judicial panel is a visiting judge who is not from a subsidiary district court or if names are recorded improperly in our database. Cases can have more than three judges if they are heard en blanc (before all judges on an appellate court) or if a judge was replaced during the progress of the case due to illness or other circumstances.

[10] Judges are identified in court documents by surname only. For nineteen surnames, multiple judges served simultaneously within a circuit (circuit court judges and district judges in subsidiary districts). For twelve of those surnames, the judges were of different genders. In these cases, interactions with a judge of these surnames was counted based on the "expected" gender of the judge. Because appellate judges hear, on average, 32 published cases per year, and district judges hear, on average, 0.5 published cases per year, we take the average gender of judges with each surname, assigning a weight of 32 to appellate judges and a weight of 0.5 to district judges.

Table A.1 compares the data on judges and cases used in this paper to official court statistics. First, we compare the count of published cases included in our data for each court with the count of published cases reported in annual Judicial Business Tables B.12. Because our data source covers the period January 2007-December 2017 whereas the Judicial Business Tables report cases from November 2006 to November 2017, small deviations in the number of cases across the two datasets are expected. Table A.1 shows that the number of cases in our data are remarkably similar to those reported in official data, with the exception of the Third Circuit, where we recover fewer cases then expected.[11] We also compare the number of judges appearing in the judicial yellow book and hearing cases in each year to the number of judges appearing in the federal judicial center database (Federal Judicial Center 2019) serving in each year. There are an average of 267 appellate judges appearing in each year of our data, compared to 280 appellate judges in each year of the federal judicial center data with start and end dates suggesting service in each year. The discrepancy between our data and the federal Judicial Center data is primarily a consequence of the fact that the *Judicial Yellow Book* does not include all judges serving in appellate courts with senior status. In particular, in the Fourth Circuit, several judges are listed in the Judicial Center database as senior appellate judges who had never served as regular appellate judges—none of these judges appear in the *Judicial Yellow Book* data. Likewise, judges who attained senior status prior to 1995 only occasionally appear in the *Judicial Yellow Book* data. Because these judges also do not appear in our case records data, we believe that these judges have maintained senior status but are not actively hearing cases.

In addition to these core variables, we also analyzed the text of the court's decision reported in the Leagle library to extract further information on the cases heard. First, we extracted

[11] As shown in Section 5.2, results are robust to the exclusion of data from the Third Circuit.

information on how many times each case was cited by the supreme court, appellate courts, and district courts. In addition to the verbatim decisions of each court, Leagle collects and attaches a list of cases that cite each included case. We use this information to count the citations of each case, and to categorize them by court and year. Next, we determine the decision writer for each case using consistent formatting of decisions within appellate circuits. We also collect information on whether a dissent was filed in each case, whether oral arguments were conducted, and whether an amicus brief was filed, but do not use these variables in our analysis[12]. We use these citation rates to construct a measure of "quality" for each judge, defined as the average number of Supreme Court and appellate court citations for cases published by the judge, relative to the average rate of citation for cases published in the same circuit and year. While the citation rate of any particular case likely reflects the importance of that case as much or more as the quality of the decision writer, a high average citation rate is likely to indicate an ability to construct clear, convincing or novel legal arguments.

3.2 Other Data

In addition to our two primary datasets, we incorporate two additional sources of information: the Biographical Directory of Article III Judges, compiled by the Federal Judicial Center, and an indicator of judge ideology, as measured by the DW-Nominate of the judge's appointing president.

[12] We perform this textual analysis by taking advantage of the formatting of decisions and the presence of key words. We determine whether an amicus brief was cited in a case by searching the decision text for the term "amicus", and whether a concurring or dissenting was filed by searching for header text with the term "dissent, dissenting, dissents, concur, concurring, concurs" etc. We determine the decision writer using case formatting—opinions either begin with the decision writer's name or end with the writer's name. Citations are recorded in a standard bibliographic format, allowing us to both count citations and also determine which court and what level of court issued each citing opinion.

The Biographical Directory of Article III Judges contains limited biographical information on each judge confirmed by the senate. We use the directory for three sources of information: judge race and ethnicity, judge gender, and the American Bar Association's rating of their quality.

Prior to confirmation, nominees to the federal courts are rated by the American Bar Association Standing Committee of the Federal Judiciary on the basis of their professional qualifications. According to the rules of the Standing Committee, ratings are made on the basis of a judge's "integrity, professional competence and judicial temperament," and do not reflect the judge's "philosophy, political affiliation or ideology" (American Bar Association 2017). This committee consists of 15 attorneys with standing to represent clients in appellate court circuits. Each member of the standing committee rates a nominee as either well-qualified,[13] qualified, or not qualified, and the committee reports both the opinion of the majority and the opinion of the next largest minority (American Bar Association 2017). We convert these ratings to numeric scores ranging from 0 (for a unanimous rating of not qualified) to 5 (for a unanimous rating of well-qualified). Because 34 judges were confirmed prior to the online dissemination of American Bar Association ratings, our final sample includes 195 judges with qualification ratings. We use the American Bar Association's quality rating for two purposes: (i) to investigate whether a judge qualification affects the responsiveness of a judge's hiring decisions to exposure to female colleagues, and (ii) to examine the effect of exposure to qualified female judges specifically (see Section 7).

We also include a measure of a judge's political ideology as both a control and as a source of heterogeneity in effect, taken from Epstein et. al. (2007). Presidents typically defer to senators from their own party on the nomination of a judge from a senator's state. Epstein et. al. exploit this

[13] In the 1989-1990 term, judges could also be rated Exceptionally Well-Qualified. We collapse this category with the Well-Qualified category.

senatorial courtesy to assign a judge the ideology of same-party home-state senators, when such exist, and the ideology of nominating presidents when home-state senators are of a different party than the president. The ideology of presidents and senators are based on the record of votes (for senators) and stated support (for presidents), calculated using the DW-Nominate algorithm (Lewis et. al. 2019).[14]

3.3. Sample Selection

We construct a panel dataset where an observation consists of an appellate court judge in a particular year.[15] As we mentioned in Section 2.1, We identify 395 distinct appellate judges in the *Judicial Yellow Book* data, of whom 305 hear at least one appellate case in at least one year of the data (the rest consist of inactive senior judges). If we had records for all eight years for each of these 305 judges, we would have a total sample of 3050 observations. In reality, however, 20% of judges start after 2007, and another 15% retire before 2017. Of those who started prior to 2007 and continued in their positions until 2017, 5% heard no cases during at least one year of their service. As a result, only 63% of judges in our sample appear on appellate panels in each year, and we only observe 2827 judge-years of interaction on federal appellate court. Furthermore, 76 judges who hear cases in at least one year do not hire any appellate clerks during the sample period, and all but 8% of the remaining 229 judges have at least one year where they hire no clerks. As a

[14] DW-Nominate scores are computed by examining the likelihood that each member of congress votes with each other member of congress. The DW-Nominate algorithm assumes that the likelihood that each voter votes in favor of a piece of legislation is determined by the distance of that bill's ideological content from the voter's ideal point on a three-dimensional ideology space. The algorithm determines a combination of vote ideological content and voter ideology using maximum likelihood estimation. The algorithm treats stated support or opposition to votes by presidents as votes.

[15] As we explained in Section 2.1, we restrict the sample to appellate court judges only, excluding visiting and district court judges, because the timing of visits and the set of cases on which visiting judges are used is nonrandom.

result, our final sample includes 1339 observations, at the judge by year level, from 229 judges over ten years. As shown in Table A.2, our sample consists of judges in years where the judge was on at least one panel with a published case, hired at least one clerk in the following year, and is not missing any primary covariates. We also include regressions that control for the current gender composition of a judge's law clerks—this covariate is missing when a judge has no law clerks on staff, resulting in missing values for 95 observations, primarily in the first year of a judge's tenure. Characteristics of this sample are available in Table 1. As shown in Table 1, 26% of observations come from female judges. Judges hire an average of 2.9 law clerks per year, and hire at least one female law clerk in 69% of years in which they make a hiring decision. Overall, 42% of clerks are female.

[TABLE 1]

3.4 How do Judges Select Clerks?

We begin our analysis by documenting characteristics of the market for clerks that will guide the specification of our empirical model. There are two pathways through which we might expect judges' exposure to female colleagues to affect their decision of who to hire. First, judges may change their beliefs about the suitability of women to the legal profession after interacting with female colleagues. Second, they may experience social pressure from female colleagues to hire a diverse staff. These two pathways differ both in their implications for policy and in their specific predictions about judges' hiring decisions.

We determine the importance of these two pathways by testing their empirical implications for the market for law clerks. If judges differ from each other in their beliefs about the suitability of women to the practice of law, we would expect some judges to consistently hire more female clerks than others. Meanwhile, if judges preferred to maintain gender-diverse staffs, we would expect judges to hire a larger number of female clerks in years when a larger proportion of their current staff is male. We test these hypotheses using the following regression equation:

$$PctFem_{j,c,t+1} = \beta_1 AFF_{j,c,\mathrm{t}} + \mathrm{B}_2\mathrm{CFF}_{\mathrm{j,c,t}} + \delta X_{j,c,t} + \theta_{c,t} + \varepsilon_{j,c,t} \qquad (1)$$

Where $PctFem_{j,c,t+1}$ represents the fraction of clerks hired by judge *j*, in court *c*, in year *t* who are female. $AFF_{j,c,t}$ represents the average fraction female of judge *j*'s staff in years other than year *t*. It is calculated as the number of female clerks hired by judge *j* in years other than year *t* divided by the total number of clerks hired by judge *j* in years other than year *t*, and is used to reflect a judge's persistent preferences or beliefs influencing their hiring of female clerks. $CFF_{j,c,t}$ represents the current fraction female of a judge's clerks at the time of hiring. It is calculated as the number of female clerks employed by judge *j* in year *t* divided by the total number of clerks employed by judge *j* in year *t*. $X_{j,c,t}$ is a set of judge characteristics and $\theta_{\mathrm{c,t}}$ is a set of court by year fixed effects. Each observation represents a judge's hiring in a particular year. As shown in Table 2, the average gender composition of a judge's staff is positively correlated with their hiring of female clerks in year *t*, but the current gender composition of their staff is unrelated to their hiring of female clerks. This suggests that judges do differ in their beliefs or preferences about female clerks, but do not choose clerks in order to maintain gender-diverse staffs.

[TABLE 2]

We further investigate whether judges care about diversity by examining the hiring decisions of judges who hire multiple clerks in the same year. Were judges indifferent to gender diversity of their staff in any given year, we would expect the gender of the second clerk hired to be unrelated to the gender of the first clerk hired, conditional on the judge's characteristics. As a result, we would expect the distribution of female hires by judge to follow a Bernoulli distribution, where the probability of hiring a female clerk was given by the Average Fraction Female of each judge's hires. For instance, if 50% of a judge's clerks are female over the sample period, we would expect that one of two hires would be female 50% of the time, both hires would be female 25% of the time, and neither hire would be female 25% of the time. In contrast, if judges care about the gender diversity of their staffs, we would expect 1 of 2 hires to be female more than 50% of the time, and 0 or 2 hires to be female less than 25% of the time.

We test this Hypothesis in Table 3. In Table 3, we compare the actual fraction of judges hiring 0, 1, 2, 3, 4, or 5 female clerks in a given year to the fractions predicted by a Bernoulli distribution, where each judge's probability of hiring a female clerk is given by their "Average Fraction Female." Results show that these two distributions are highly similar. In particular, the fraction of judges hiring zero female clerks and hiring only female clerks is similar to the fractions predicted by a Bernoulli distribution.

[TABLE 3]

Together, these two pieces of evidence suggest that preferences or beliefs about female clerks play a substantial role in judges' hiring decisions, but attempts to maintain staff gender diversity do not. As a result, we expect exposure to female colleagues to affect hiring decisions by modifying preferences or beliefs, rather than by increasing judge's desire to maintain a gender-balanced staff.

This in turn suggests that we should expect to see effects of exposure to female colleagues primarily among judges who underestimate their female colleagues. Because we examine the effect of interacting with randomly selected female colleagues, judges whose expectations of female colleagues match their observations of female colleagues would be as likely to be disappointed by a female colleague as they are to be impressed, so interacting with an unusually large number of female colleagues would not change their beliefs in expectation. On the other hand, judges whose expectations of female colleagues are worse than their observations of female colleagues will be impressed more often than they will be disappointed by interactions with a female colleague.

Because we expect the effect of exposure to female colleagues to be greatest for judges who have negative opinions of women in the legal profession, we use a binary indicator taking value of 1 if at least one female clerk is hired and 0 otherwise as our main dependent variable. Because variation in the likelihood of hiring two or more female clerks is disproportionally driven by judges with a high propensity to hire females, we expect to see a larger effect of exposure to female colleagues on a judge's likelihood of hiring one female clerk than on their likelihood of hiring a second or third female clerk. However, we present the effect of exposure to female colleagues on the fraction of hired clerks who are female and the number of hired clerks who are female as a robustness check.

4. Empirical model and identification strategy

Our empirical strategy takes advantage of the random assignment of judges to panels to regress a measure of interaction with female judges in a given year on the likelihood of hiring at least one female clerk in the following year. Because assignment of appellate judges to cases is random conditional on circuit, and year, we control for fixed effects at the court by year level. Variation in the fraction of co-panelists who are female in a year, conditional on court and year, is due entirely to the random assignment of judges to cases, and to the determination of panels that a case is worthy of publication. We estimate the effect of exposure to female colleagues on the likelihood of hiring a female clerk using the following regression equation:

$$Hire_{j,c,t+1} = \beta Inf_{j,c,t} + \delta X_{j,c,t} + \theta_{c,t} + \varepsilon_{j,c,t} \qquad (2)$$

where $Hire_{j,c,t+1}$ is an indicator of whether judge j, in court c, hired at least one female clerk in year $t + 1$; $Inf_{j,c,t}$ is exposure to female judges, $X_{j,c,t}$ is a set of judge characteristics, and $\theta_{c,t}$ is a set of court by year fixed effects. Each observation represents a judge j's hiring decision in a particular year t.

We measure exposure to female judges $Inf_{j,c,t}$ as the fraction of co-panelists on the cases heard by judge j in year t that are female. There are a few noteworthy characteristics of this measure. First, because we calculate the fraction of co-panelists who are female, our measure of exposure to female colleagues does not depend on the volume of cases heard by a judge in a particular year. This decision reflects the assumption that full-time judges with few cases are likely to have more time-consuming cases than those with many cases, as well as the assumption that the salience of individual cases is likely greater when a judge has heard fewer cases in a year. Second, this measure does not distinguish between interactions with a single female colleague on many cases and interactions with multiple female colleagues, each on an individual case. We selected this measure based on a learning model in which each case heard with a co-panelist reveals a small amount of information about that co-panelist's ability, which is then used to inform the likely distribution of legal professionals with the co-panelist's gender. If there is substantial uncertainty about the ability of each judge, repeated interaction with one judge will have similar information content to individual interactions with multiple judges.

In order to demonstrate the feasibility of this approach, we estimate the variation in our main dependent and independent variables that is not accounted for by court by year variation. As shown in A.3, very little variation in either the hiring decisions of judges or in the exposure of judges to female colleagues is explained by differences between courts and years. Likewise, observed judge characteristics do not explain a significant amount of variation in either judge hiring decisions or exposure to female judge.

4.1 Evidence in support of identification strategy

The key identifying assumption in this paper is that variation in the gender composition of co-panelists within a particular circuit and year is unrelated to a judge's preference for female clerks and to a judge's available labor pool. This assumption is justified by the assertion, common to all appellate circuits, that judges are randomly assigned to cases (Stearns and Abramowicz 2005). While violations of pure randomness are inevitable, violations of random assignment are small, unlikely to be sustained over a year, seen only in a few courts, and unlikely to be related to judge's preferences or labor pools. In particular, work by Chilton and Levy (2015) finds that due to scheduling conflicts and similar concerns, the assignment of judges to appellate panels deviates from random assignment in several courts. As a consequence, the distribution of Republican appointees across cases differs slightly from what would be expected by chance in the Second, Sixth and DC circuits, and more substantially in the Ninth Circuit. However, the likelihood that a Republican will serve with another Republican differs from chance by less than a percentage point in all circuits but the Second and Ninth, in which it differs from chance by less than two percentage points. In addition, as shown in Table 4, judges appointed by Republican presidents are as likely to serve on panels with female judges as are judges appointed by Democratic presidents, indicating that any non-randomness in the assignment of judges to panels on the basis of political party does not affect the likelihood of serving on panels with female colleagues. Finally, as shown in Table 5, the inclusion of controls for judge characteristics, including party, does not weaken the measured effect of exposure to female colleagues on hiring decisions. Levy (2017) examines a broader range of potentially non-random scheduling decisions made by the chief justice's office of each appellate circuit, finding, for instance, that one circuit had a tradition of ensuring that judges have the opportunity to be the presiding judge on one case in their first year by constructing a panel with two senior or visiting judges. However, these deviations from strict randomness are small enough that federal judges themselves believe panels to be randomly constructed (Levy 2017).

We test the potential threat of nonrandom case assignment to identification across a number of dimensions by regressing our main independent variable, the fraction of a judge's co-panelists who are female in each year, onto a series of observed judge characteristics—specifically, on a judge's racial and ethnic background, years of experience, age, political party, the ideology of their nominating president, and the gender composition of their current staff, controlling for judge gender and for court by year fixed effects. As shown in Table 4, there is little to no relationship between the exposure of a judge to female colleagues and any observed judge characteristics.

[Table 4]

Table 4 shows the relationship between a variety of judge characteristics and the main variable of interest, for both the full sample (columns 1 and 2) and separately for male and female judges (columns 3 and 4). We separate the sample by judge gender because the expected female share of colleagues is mechanically lower for female than for male judges, due to the fact that judges cannot interact with themselves. While we control for judge gender in column (2), the size of this mechanical effect is larger in small circuits such as the first circuit (with 10 judges) than in the ninth circuit (with 48 judges). Overall, the relationships we observe between judge characteristics and interaction with female colleagues is no greater than would be expected by chance, with the only statistically significant relationship being a lower likelihood of serving with female colleagues for male judges with a larger number of current female staff, significant at the 10% level. Because we perform 24 tests, a single test that is significant at the 10% level would be expected even if there were no true relationship between any of the judge characteristics and interactions with female colleagues. This is confirmed by a test of joint significance for each group of eight regressions, none of which demonstrate deviations from zero estimated effects larger than would be expected by chance.

We conclude this section by documenting the differences between published and unpublished cases. As discussed in section 2.1, we use data from published cases. We do so both because only published cases are systematically available in all courts throughout our sample period and because published cases involve significantly more and higher-stakes interaction with co-panelists than do routine, non-precedential cases. As a result, they provide more information to copanelists about a judge's abilities.

Because the decision to publish a case is made by the judges empaneled to hear the case, it is important to assure that the assignment of a female judges to a case does not affect its likelihood of being published. In particular, one might conjecture that judges who consider female colleagues to be less capable than male colleagues will be less likely to support publication when seated on a panel with female judges. Were this the case, we would underestimate the exposure of sexist judges to female colleagues, resulting in positive bias in our regressions. We explore this hypothesis by examining cases occurring in the years 2016 and 2017, the only years in our sample in which all appellate circuits made a substantial number of unpublished cases available online (roughly half of them).

Using this data, we provide evidence on whether the gender composition of the panel affects the probability that the case is published. We first examine the question of whether female judges publish a smaller fraction of cases than do male colleagues, as would be expected if some judges were reluctant to publish when empaneled with women. As shown in Appendix Table A.4, we find that male and female judges publish at nearly identical rates. We regress the fraction of cases on which a judge is empaneled that are published in each year on a judge's gender, under the regression framework presented in equation 2. We find that female judges publish at slightly higher rates than male judges, with point estimates suggesting that women publish at a rate 2% greater than men, but these differences are neither statistically significant nor meaningful as potential sources of bias. This indicates that on average, judges are no more or less likely to agree to publish a case when empaneled with a female colleague.

Next, we examine the question of whether judges with potential sexist beliefs are less likely to publish cases when empaneled with female colleagues. We do this by calculating the ratio of female colleagues on published case to female colleagues on unpublished cases. Were sexist judges less likely to publish a case when empaneled with female judges, they would have fewer female co-panelists on published cases and more female co-panelists on unpublished cases than would a judge whose publication decisions did not depend on the gender of their co-panelists. As a result, we would expect this ratio to be positively correlated with judge's preferences for female clerks if judges published selectively, and uncorrelated with judge's preferences for male clerks if judges do not publish selectively. We regress this ratio onto judge's "Average Female Fraction of Clerks," which gives the fraction of clerks hired by a judge in all observed years who are female. As shown in Appendix Table A.5, we find a small and statistically insignificant negative relationship between publication ratio and average female fraction of clerks. Because these effects are small and in the opposite direction predicted by an effect of sexist beliefs on publication decisions, this analysis provides no evidence that the publication process reflects gender biases of judges.

In contrast, we find clear evidence that published cases differ substantively from unpublished cases along other observed characteristics. As reported in Appendix Table A.6, 9.4% of published cases had at least one amicus brief filed by an individual or organization not party to a case, compared to 0.3% of unpublished cases, suggesting that published cases are far more likely than unpublished cases to set precedents relevant to outside parties. Likewise, judges requested oral arguments for 37% of published cases, as opposed to 6% of unpublished cases. Published opinions were also nearly four times the length of unpublished cases on average, and included more than three times as many citations. In sum, our evidence indicates the publication probability depends on factors unrelated to the gender of the panel of judges: men and women are equally likely to be on published cases.

5. Findings

5.1. Main Results

Table 5 presents ordinary least squares (OLS) regressions in which the dependent variable is an indicator of whether a judge hired a female clerk in year $t+1$ and the key independent variable is the fraction of the judge's co-panelists who were female in year t. Column (1) includes court by year fixed effects, with no additional covariates. Column (2) adds controls for judge gender, Hispanic ethnicity, and age, and column (3) adds controls for the political party of the judge's nominating president, a quadratic of the DW-Nominate score of the judge's nominating president, and a quadratic of the judge's years of experience on their current court. This table shows that a one standard-deviation increase in a judge's exposure to female colleagues, or an increase of 0.11 in the fraction of judicial interactions with female colleagues, leads to a 5 percentage-point increase in the likelihood that a judge hires a female clerk. The addition of controls slightly increases the precision of the estimate but has no detectable effect on its magnitude.

[TABLE 5]

We also examine the effect of exposure to female colleagues on the fraction of all hires in year *t+1* who are female as well as the number of females hired in year *t+1*. As discussed in section 3.4, the effect of exposure to female colleagues on hiring is expected to be smaller for judges with more positive prior beliefs about the quality of women in the judiciary. As a result, we expect that the effect of exposure to female colleagues on hiring one female clerk to be larger than the effect on hiring more than one female clerk. As shown in Appendix Table A.7, the effect of exposure to female colleagues on the fraction of newly hired clerks who are female is significant and positive, but smaller in magnitude than our main effect, suggesting that a one standard deviation increase in exposure to female colleagues increases the percent of hires who are female by 2.3 percentage points. Likewise, a one standard deviation increase in exposure to female colleagues increases the number of female clerks hired by 0.63. These point estimates are consistent with exposure to female colleagues having a positive effect on the likelihood of hiring at least one female clerk, but little effect on the likelihood of hiring more than one female clerk, conditional on hiring at least one.

5.2: Robustness

We conduct two exercises in order to further increase our confidence that judges' exposure to female clerks is unrelated to unobserved judge characteristics.[16] First, we regress an indicator

[16] Any such relationship would need to result from small deviations from perfect randomization in the assignment of judges to cases. For example, if judges with greater influence have more flexibility in scheduling vacation days, a judge might take vacations that are scheduled while (predominantly non-senior)

of whether judge j hired any female clerks in year *t-1* on judge j's exposure to female colleagues in year *t*. Because staff are hired one to two years in advance of their start-date, exposure to female colleagues in year *t* cannot have a causal effect on hiring in year *t-1*. As shown in columns 1, 2, and 3 of Table 6, exposure to female clerks in the year following a hiring decision is unrelated to that hiring decision, providing evidence in support of our identification strategy.[17]

Second, we match each judge j to the most similar judge within their court (with replacement). We determine matches by regressing the fraction of each judge's staff who are female onto the judge's characteristics and court. We then select the judge with the most similar predicted staff gender composition and regress $Hire_{j,c,t+1}$ on $Inf_{k,c,t+1}$, where *j* is the reference judge and *k* is the match. As shown in columns 4, 5 and 6 of Table 6, the association between the exposure of a judge's most similar colleague and the judge's likelihood of hiring a woman is negative and not statistically significant.[18]

[Table 6]

Apart from questions over the random assignment of judges to copanelists, one might be concerned that judges are induced to hire more female clerks when interacting with female colleagues due to characteristics of their colleagues other than gender. For instance, if judges are

---

female judges are working. Greater seniority might also increase their likelihood of hiring their (predominantly male) favored clerk candidates.

[17] Note that the sample sizes for these placebo regressions are higher than the sample sizes for the primary regressions. Because hiring decisions happen two years prior to staffing starts, staff hired after 2014 cases are heard appear in our data in 2017, requiring us to drop case data from 2015-2017. In contrast, the placebo test only requires us to drop case data from 2007.

[18] This exercise is a further check against the possibility that our results are driven by unobserved shocks affecting particular subsets of judges serving on an appellate circuit. Factors correlated with a judge's ideology and/or experience might affect the labor pool from which they hire, either because judges prefer ideologically similar clerks or because more senior or more ideologically mainstream judges can offer more prestigious positions and are thus able to hire the most sought-after clerks. If a quirk of the assignment of judges to cases led similar judges to be more likely to hear cases with female colleagues in some years, this could produce spurious correlations between exposure to female colleagues and hiring.

more likely to hire liberal clerks after interacting with liberal colleagues, this would create an association between interactions with female colleagues and hiring of female clerks. We address this concern by including additional controls to our main regression for the fraction of co-panelists who are Republican, who have served on the court for fewer than 10 years, who are younger than 60, and who have above-average rates of citation to their published opinions. As shown in Appendix Table A.8, none of these other judge characteristics significantly mediate the relationship between the gender of co-panelists and the gender of newly-hired clerks.

Additionally, as discussed in Section 3.1, case data from the Third Circuit is likely to be incomplete. In order to ensure that our results are not biased by nonrandom selection of cases into the Leagle database, we omit data from the Third Circuit from our main results. As shown in Appendix Table A.9, this omission has no effect on the magnitude of our estimated results.

Finally, we perform a permutation resampling procedure to determine whether our standard errors accurately reflect the distribution of likely effect sizes. To do this, we randomly reassign hiring decisions to judges within each court and year, and estimate the full model (shown in Table 5, column 3) for each random assignment. Figure 1 shows the distribution of 10,000 randomly generated effects against the estimated effect, and shows that our results are unlikely to have occurred by chance.

[FIGURE 1]

5.3. Additional Evidence

We explore whether interactions with female colleagues has a persistent effect on judge's hiring decisions by examining the effect of exposure to female colleagues in year t on hiring decisions made in years $t+2$ and $t+3$. Specifically, we estimate the following regression equation, with values of k ranging from -5 to +3.

$$Hire_{j,c,t+\mathrm{k}} = \beta_k Inf_{j,c,t} + \delta X_{j,c,t} + \theta_{c,t} + \varepsilon_{j,c,t} \tag{3}$$

The coefficients estimated from this equation are presented in Figure 2 below:

[FIGURE 2]

As shown in Figure 2, we find no evidence of a persistent effect of interactions with female colleagues on hiring decisions. This suggests that standard professional interactions are salient in the short-run but do not have substantial effects on judge's long-term beliefs.

We next examine heterogeneity in the effect of interaction with female colleagues by both the characteristics of the influenced judge and the characteristics of the interacting female colleagues. We examine seven sets of characteristics, examining both whether judges with each characteristic are more affected by interactions with female colleagues and whether female colleagues with each characteristic have a greater effect when serving as co-panelists. These characteristics are: judge gender; judge quality, as measured by their rating as qualified by the American Bar Association; the fraction of a judge's current staff that is female; judge age; judge experience, as measured by decades on the appellate court; the political party of the judge's nominating president; and whether the judge is visiting from a district court[19]. The results are shown in Table 7. The effects are imprecisely estimated, and differences in effect are statistically insignificant along all dimensions of heterogeneity examined. However, the point estimates for several dimensions of heterogeneity are very large and broadly in line with a model of judge learning. We thus briefly discuss them below. The results in Panel A suggest that male judges and judges with fewer than 10 years of experience on the court are more strongly affected by exposure to female colleagues than are female judges and judges with more than 10 years of experience on

[19] As discussed in Section 2.1, District Judges are not included in our sample because the rules governing their assignment to cases differ from those governing the assignment of appellate judges. However, each appellate judge's exposure to a visiting district judge is as good as random.

the court. These findings are consistent with a model of judge learning, where judges who are least familiar with their colleagues and most likely to be surprised by competent female colleagues showing the largest changes in hiring.

Panel B shows suggestive evidence on the relationship between the characteristics of female co-panelists and the effect of serving with female co-panelists on hiring. In particular, it reveals that interactions with female district court judges (column 7) and female judges who have served on the court for fewer than 10 years (column 6) have relatively large effects on hiring. This is consistent with the hypothesis that interactions with visiting judges provide more novel information than do interactions with regular colleagues. Likewise, it shows that interactions with female judges who were rated as highly qualified by the American Bar Association have relatively large effects on hiring (column 3). This is consistent with the hypothesis that highly qualified judges are more likely to impress their colleagues than are less qualified judges. Finally, column (2) of panel B shows no evidence that interacting with female judges with a majority female staff has a greater effect on hiring than does interacting with female judges with majority male staff. This makes it somewhat less likely that we are capturing an effect of referrals and information about specific candidates conveyed from a judge's female stuff members, rather than changes in judge's perceptions of women in law stemming from interactions with female judges.

6. Conclusion

This paper presents evidence that contact with female colleagues increases the likelihood that judges hire women at the entry level of their profession. While this work builds on a substantial and growing literature demonstrating the effect of contact with out-groups on attitudes and decisions, this is the first work to demonstrate that interaction with out-groups affects hiring decisions. This finding is particularly significant because it suggests that increases in the diversity of the upper rungs of a profession can shift attitudes in a way that creates opportunities at the entry level of a profession. This in turn suggests that policies aimed at increasing the diversity in the leadership of a profession, such as affirmative action policies or policies requiring that a certain number of board seats be filled by women, may have benefits beyond their immediate beneficiaries.

We find that a one standard deviation increase in the fraction of published cases heard alongside female colleagues increases a judge's likelihood of hiring at least one female clerk by 5 percentage points, increases the fraction of women hired by 0.023, and increases the number of female clerks hired in a year by 0.063. Because judges are broadly unconstrained in who they hire as a clerk, we interpret these changes in hiring practices as a change in judge's assessment of the ability of women in judicial practice and law. These findings are large, and suggests that hiring a female judge to serve on a typical appellate court would be expected to lead the judge's colleagues to hire six additional female clerks over the next decade[20].

Because this work is unique in its ability to estimate the effect of peer interactions among established professionals on their hiring practices, it is reasonable to wonder whether particular characteristics of the legal profession or the judiciary make it more or less susceptible to such effects. One possibility is that the effect of peer exposure may be particularly salient in occupations which have seen dramatic increases in the number of women at the entry level and in higher positions over the career of the current leadership generation. Law is such a profession—while fewer than 5% of American law school graduates were women prior to 1970, 50% of recent law

[20] The average appellate circuit has 13 appointed judges, 25% of whom are women. Thus, hiring a woman as the 13th judge increases the fraction of interactions with a female judge by 1/13 for each of 12 judges. 1/13*0.633*12=5.85

school cohorts have been female.[21] As a result of this generational change in law, the great majority of federal judges began their careers in a cohort composed predominantly of men, and at a time when few or no women occupied prominent positions in law, such as professorships or judicial appointments. It is thus reasonable to suspect that these judges have developed expectations about women in the profession that do not reflect the characteristics of the newest cohort of lawyers. Our results may be large in law precisely because judges have developed attitudes that can be counteracted with new evidence. Due to the dramatic increase in women's education and workforce engagement over the past seventy years (Goldin, Katz, & Kuziemko 2006), this generational change in the representation of women characterizes many high-earning professions (Goldin and Katz 2011).

We may also expect the legal profession to exhibit large effects of peer interaction because of a lack of clear and objective criteria for hiring decisions. As argued in Goldin (2015), negative perceptions of women's abilities within an occupation may be less persistent in occupations where clear tests of abilities are available. While certain credentials are very helpful in securing employment in law, such as graduation from a top law school or membership on a legal review, the legal profession lacks highly objective measures of quality or productivity available in some engineering fields.

21 https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/statistics/enrollment_degrees_awarded.pdf

Bibliography

Algan Y., and Cahuc P. 2006. “Job Protection: The Macho Hypothesis” Oxford Review of Economic Policy, 22(3), 390-410.

Algan Y., Quoc-Anh D, Dalvit N., Le Chapelain A. and Zenou, Y., 2015. "How Social Networks Shape Our Beliefs: A Natural Experiment among Future French Politicians," Sciences Po publicationsinfo:hdl:2441/78vacv4udu9, Sciences Po.

American Bar Association Required Disclosures 2013. “Enrollment and Degrees Awarded 1963 – 2012 Academic Years.”

American Bar Association Standing 2017. “American Bar Association Standing Committee on the Federal Judiciary: What it is and How it Works.” ISBN: 978-1-62722-880-0

Ash, E, Chen, D, and Ornaghi, A 2020. “Stereotypes in High-Stakes Decisions: Evidence from U.S. Circuit Courts.” The Warwick Economics Research Paper Series (TWERPS) 1256, University of Warwick, Department of Economics.

Avery C., Jolls C. , Posner R.A. and Roth A. E. 2007. “The New Market for Federal Judicial Law Clerks.” University of Chicago Law Review 74, no. 2: 447–486.

Beaman L., Chattopadhyay R., Duflo E., Pande R. and Topalova P.. 2009. “Powerful Women: Does Exposure Reduce Bias,” Quarterly Journal of Economics, 124(4): 1497-1540.

Blau F. and Kahn L. 2016. “The Gender Wage Gap: Extent, Trends, and Explanations.” NBER Working Paper No. 21913.

Boisjoly J., Duncan G. J., Kremer M., Levy D. M. and Eccles J. 2006. “Empathy or Antipathy? The Impact of Diversity.” American Economic Review, 96(5): 1890-1905.

Boyd C. L., Epstein, L. and Martin, A. D. 2010. “Untangling the Causal Effects of Sex on Judging,” American Journal of Political Science (54:2), pp. 389--411.

Carlana, M. 2019. “Implicit Stereotypes: Evidence from Teachers' Gender Bias.” Quarterly Journal of Economics.

Chilton A. S. and Levy, M. K.. 2015. “Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals”. Duke Law School Public Law and Legal Theory Series No. 2015-1; U of Chicago, Public Law Working Paper No. 529. Available at SSRN: https://ssrn.com/abstract=2520980 or http://dx.doi.org/10.2139/ssrn.2520980

Dahl G., Kotsadam A. and Rooth D.-O. 2018. “Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams” (24351), Technical report, National Bureau of Economic Research

District Judges 28 U.S. Code § 292, 1986

Epstein L., Martin A. D., Segal, J. A. and Westerland C. L. 2007. “The judicial common space.” Journal of Law, Economics, and Organization, 23(2), 303-325. https://doi.org/10.1093/jleo/ewm024

Fang, H. and Moro A.. 2010. “Theories of Statistical Discrimination and Affirmative Action: A Survey, in Handbook of Social Economics.” edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson. North-Holland.

Federal Judicial Center. 2019. “Biographical Directory of Article III Federal Judges” https://www.fjc.gov/history/judges/biographical-directory-article-iii-federal-judges-export

Fortin N. M. 2005. “Gender Role Attitudes and the Labour-market Outcomes of Women across OECD Countries,” Oxford Review of Economic Policy, Volume 21, Issue 3, Autumn , Pages 416–438, https://doi.org/10.1093/oxrep/gri024

Glynn A. N. and Sen M. 2015. “Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?” American Journal of Political Science, 59: 37–54. doi:10.1111/ajps.12118

Goldin C. 2015. “A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings.” In Human Capital in History: The American Record (pp. 313–348). Chicago, IL: University of Chicago Press.

Goldin C., & Katz L. F. (2011). “The Cost of Workplace Flexibility for High-Powered Professionals.” The Annals of the American Academy of Political and Social Science, 2011, 45–67.

Goldin C., Katz L. F., and Kuziemko, I. (2006). “The Homecoming of American College Women: The Reversal of the College Gender Gap.” The Journal of Economic Perspectives, 20, 133–56. Retrieved from http://www.jstor.org/stable/30033687.

Gramlich J. 2018. “Federal judicial picks have become more contentious, and Trump’s are no exception.” Pew Research Center.

Guiso, L. , Sapienza, P. and Zingales, L. 2003. “People's opium? Religion and economic attitudes,” Journal of Monetary Economics, 50, issue 1, p. 225-282, https://EconPapers.repec.org/RePEc:eee:moneco:v:50:y:2003:i:1:p:225-282.

Hess, A. G. 2015. “The Collapse of the House that Ruth Built: The Impact of the Feeder System on Female Judges and the Federal Judiciary, 1970-2014,” Journal of Gender, Social Policy and the Law (24:1), pp. 1-42.

History of the Federal Judiciary. 2019. http://www.fjc.gov. Web site of the Federal Judicial Center, Washington, DC.

Judicial Yellow Book: Who's Who in Federal and State Courts. New York, N.Y: Leadership Directories, Inc, 2007-2017. Publication Masters.

US Courts. (2016). Judicial Business: Table B-12. Available At: https://www.uscourts.gov/data-table-numbers/b-12

Leagle, 30 Nov. 2019, http://www.leagle.com/.

Levy, M. K. 2017. “Panel Assignment in the Federal Courts of Appeals,” 103 Cornell Law Review 65-116

Levy M. K. 2019. "Visiting Judges," California Law Review, Vol. 107, No. 67, 2019; Duke Law School Public Law and Legal Theory Series No. 2019-11. Available at SSRN: https://ssrn.com/abstract=3328571

Lewis J. B., Poole K., Rosenthal H., Boche A., Rudkin A., and Sonnet L. 2019. “Voteview: Congressional Roll-Call Votes Database.” https://voteview.com/

Corno L., La Ferrara E. and Burns J. 2019. “Interaction, stereotypes and performance. Evidence from South Africa,” IFS Working Papers W19/03, Institute for Fiscal Studies.

Mark, N. and Daniel R. H.. 2012. “Roommate's race and the racial composition of white college students' ego networks.” Social science research. 41. 331-42. 10.1016/j.ssresearch.2011.11.012.

McCabe P. G. 2014. “A Guide to the Federal Magistrate Judges System,” Technical report, Federal Bar Association.

McMillon B. J. 2014. “US Circuit Court Judges: Profile of Professional Experiences Prior to Appointment,” Technical report, Congressional Research Service.

NALP. 2019. “Clerkship Study Law Student Findings,” https://www.nalp.org/clrkfind3

Pérez I. S. 2016 “gender-guesser 0.4.0” https://pypi.org/project/gender-guesser/

Posner R. and Roth A. E. and Avery C. and Jolls C. 2001. “The Market for Federal Judicial Law Clerks.” University of Chicago Law Review 68, no. 3: 793–902.

Rhinehart L. K. 1994. “Is There Gender Bias in the Judicial Law Clerk Selection Process,” 83 Geo L.J. 575

Stearns M. L. and Abramowicz M.. 2005. “Defining Dicta,” Faculty Scholarship. 119.https://digitalcommons.law.umaryland.edu/fac_pubs/119

Van Laar C., Levin S., Sinclair S. and Sidanius J. 2005. “The effect of university roommate contact on ethnic attitudes and behavior.” Journal of Experimental Social Psychology, Volume 41, Issue 4, Pages 329-345, ISSN 0022-1031, https://doi.org/10.1016/j.jesp.2004.08.002.

Washington E. L. 2008. “Female Socialization: How Daughters Affect Their Legislator Fathers.” American Economic Review, 98(1): 311-32.

Tables and Figures

Tables

Table 1: Characteristics of Judges and Clerks in Sample

	Full Sample Mean (SD)	Male Mean (SD)	Female Mean (SD)	Full Sample Min / Max	Sample Size Male / Female	M/F Diff
	(1)	(2)	(3)	(4)	(5)	(6)
Judge characteristics						
Female	0.2562 (0.4367)	0.0000 (0.0000)	1.0000 (0.0000)	0 / 1	996 / 343	***
Asian	0.0127 (0.1120)	0.0151 (0.1219)	0.0058 (0.0762)	0 / 1	996 / 343	
Black	0.0941 (0.2921)	0.0894 (0.2854)	0.1079 (0.3107)	0 / 1	996 / 343	
Hispanic	0.0635 (0.2439)	0.0653 (0.2471)	0.0583 (0.2347)	0 / 1	996 / 343	
Age (decades)	6.4277 (1.0089)	6.5443 (1.0186)	6.0892 (0.8996)	3.6 / 9.2	996 / 343	***
Decades on current court	1.5999 (0.9882)	1.7105 (0.9849)	1.2787 (0.9269)	0 / 4.4	996 / 343	***
Ideology score	0.0681 (0.3614)	0.0888 (0.3550)	0.0081 (0.3736)	-0.521 / 0.693	996 / 343	***
Republican	0.5400 (0.4986)	0.5763 (0.4944)	0.4344 (0.4964)	0 / 1	996 / 343	***
Number of clerks hired in year	2.8656 (1.2262)	2.8072 (1.2167)	3.0350 (1.2397)	1 / 7	996 / 343	***
Number of cases heard in year	57.7371 (40.4221)	58.1908 (41.1069)	56.4198 (38.3920)	1 / 208	996 / 343	
% of years with at least one female clerk hired[a]	0.6923 (0.4617)	0.6807 (0.4664)	0.7259 (0.4467)	0 / 1	996 / 343	
% of clerks that are female	0.4205 (0.4937)	0.4210 (0.4938)	0.4192 (0.4936)	0.0000 / 1.0000	3637 / 1274	

Notes: This table presents the average values of judge-level covariates in the analysis sample. For all variables other than % of clerks that are female, the sample is at the judge by year level. For % of clerks that are female, the sample is at the clerk by year level. *Source*: Judicial yellow books, case dataset collected by authors (see data section)

Table 2: Persistent Preferences and Staff Gender Composition on Female Hiring

Dep Var: Fraction of hires in next year who are female	(1)	(2)	(3)
Persistent hiring rate	0.3389*** (0.0805)	0.3259*** (0.0808)	0.2697*** (0.0849)
% of current staff female	0.0209 (0.0408)	0.0228 (0.0407)	0.0265 (0.0406)
Female	0.0142 (0.0217)	0.0072 (0.0222)	-0.0015 (0.0229)
Asian		0.0091 (0.0771)	-0.0224 (0.0786)
Black		0.0465 (0.0351)	0.0246 (0.0367)
Hispanic		0.0378 (0.0389)	0.0218 (0.0398)
Age		0.1162 (0.1013)	0.1494 (0.1184)
Age^2		-0.0001 (0.0001)	-0.0001 (0.0001)
Years on current court			0.0092 (0.0892)
Years on current court^2			-0.0423 (0.0430)
Ideology score			-0.0371 (0.0563)
Ideology score^2			-0.1941 (0.1390)
Republican			0.0072 (0.0114)
Court by year fixed effects	Yes	Yes	Yes
Observations	1262	1262	1262
Dependent variable mean	0.4052	0.4052	0.4052

Notes: This table reports OLS estimation results from a regression of the fraction of hires in year *t* who are female on the fraction of all hires in years other than year *t* who are female and on the fraction of staff in year *t* who are female, as described in Equation 1 in the text. Standard errors are robust and clustered at the judge level. The dependent variable is the fraction of clerks hired in the following year who are female, conditional on hiring any clerk. Significance levels are: * 10%, ** 5%, *** 1%.

Source: Judicial yellow books, case dataset collected by authors (see data section).

Table 3: Actual and Predicted Gender Composition of Hires

		Number of Female Clerks					
Number of Clerks Hired		0	1	2	3	4	5
1	Actual	61.4%	38.6%				
(Observations:223)	Predicted	60.0%	40.0%				
2	Actual	39.5%	44.3%	16.2%			
(Observations:314)	Predicted	41.1%	41.5%	17.4%			
3	Actual	21.5%	43.2%	29.4%	5.8%		
(Observations:377)	Predicted	25.7%	38.6%	27.0%	8.7%		
4	Actual	14.4%	29.9%	32.0%	18.9%	4.8%	
(Observations:334)	Predicted	16.1%	31.2%	30.9%	17.3%	4.5%	
5	Actual	21.0%	28.4%	22.2%	22.2%	6.2%	0.0%
(Observations:81)	Predicted	13.6%	24.4%	28.0%	21.3%	10.2%	2.5%

Notes: This table reports the number of judges hiring 0, 1, 2, 3, 4, and 5 female clerks in a single hiring cycle, among judges hiring 0, 1, 2, 3, 4, or 5 clerks. It also reports the number of judges hiring 0, 1, 2, 3, 4, or 5 female clerks was the gender of each clerk in the hiring cycle an independent event with a probability given by fraction of clerks hired in all other cycles by that judge that are female. *Source:* Judicial Yellow Books, author's calcuations.

Table 4: Balance Tests for Random Assignment to Panels

Dependent variable: % co-panelists who are female	(1)	(2)	(3)	(4)
Age	0.0019 (0.0045)	0.0015 (0.0029)	0.0002 (0.0035)	0.0054 (0.0056)
Asian	-0.0136 (0.0315)	-0.034 (0.0313)	-0.037 (0.0302)	-0.0002 (0.0198)
Black	0.0044 (0.0193)	-0.0028 (0.0074)	0.0052 (0.0088)	0.01 (0.0145)
Hispanic	0.0148 (0.0155)	0.0008 (0.0114)	-0.0161 (0.0103)	0.0163 (0.0317)
Years on current court	0.0106** (0.0046)	-0.0004 (0.0027)	-0.0004 (0.0031)	0.0005 (0.0049)
Ideology score	0.0244* (0.014)	-0.0017 (0.0088)	-0.0039 (0.0106)	0.0117 (0.0198)
Republican	0.0112 (0.0096)	0.0018 (0.0057)	0.0003 (0.0066)	0.008 (0.0132)
% of current staff female	-0.0017 (0.0139)	0.0157 (0.0099)	0.0204* (0.012)	0.0161 (0.022)
Court by year fixed effects	No	Yes	Yes	Yes
Sample	All	All	Male	Female
F-Stat (P-val)	1.4215 (0.1894)	0.7416 (0.6547)	1.0727 (0.385)	0.9399 (0.4928)
Observations	1339	1339	996	343

Notes: The table reports OLS estimation results from regressions of the fraction of co-panelists who were female in a year on a series of judge characteristics. Standard errors are robust and clustered at the judge level. Column (2) controls for whether the judge is female. Columns (2)-(4) include court by year fixed effects. Column (3) shows regression results for male judges, column (4) shows regression results for female judges. Significance levels are: * 10%, ** 5%, *** 1%.

Source: Judicial yellow books, case dataset collected by authors (see data section for details).

Table 5: Effect of Serving with Female Judges on Hiring Decisions

Dep Var: Probability of hiring any female clerk in next year	(1)	(2)	(3)	(4)
Fraction of co-panelists who are female	0.4832*** (0.1681)	0.5222*** (0.1654)	0.4950*** (0.1617)	0.5021*** (0.1665)
Female	0.0775*** (0.0294)	0.0576* (0.0299)	0.0308 (0.0305)	0.0463 (0.0316)
Asian		0.1071 (0.0940)	0.0121 (0.0954)	0.0825 (0.0877)
Black		0.0314 (0.0463)	-0.0204 (0.0474)	-0.0167 (0.0483)
Hispanic		0.0949* (0.0487)	0.0571 (0.0492)	0.0611 (0.0531)
Age		0.1086 (0.1288)	0.1322 (0.1491)	0.2208 (0.1704)
Age^2		-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0002 (0.0001)
Years on current court			-0.0453 (0.1162)	-0.0334 (0.1217)
Years on current court^2			-0.1058** (0.0527)	-0.1049* (0.0562)
Ideology score			-0.0522 (0.0751)	-0.0507 (0.0781)
Ideology score^2			-0.2382 (0.1840)	-0.2617 (0.1899)
Republican			0.0095 (0.0138)	0.0106 (0.0147)
% of current staff female				0.1345*** (0.0463)
Court by year fixed effects	Yes	Yes	Yes	Yes
Observations	1339	1339	1339	1244
Dependent variable mean	0.69231	0.69231	0.69231	0.6881

Notes: This table reports OLS estimation results from the regressions described in equation (2) in the text. Standard errors are robust and clustered at the judge level. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the results of regressions of the dependent variable on the fraction of co-panelists who were female in each year.Significance levels are: * 10%, ** 5%, *** 1%.

Source: Judicial yellow books, case dataset collected by authors (see data section).

Table 6: Placebo Tests

	Prob. **Actual Judge** Hired any female clerks in **past** year			Prob. **Matched judge** hired any female clerks in **next year**		
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction of co-panelists who are female	-0.1184 (0.0988)	-0.1310 (0.0973)	0.0561 (0.0763)	-0.1983 (0.1218)	-0.1748 (0.1173)	-0.0095 (0.1123)
Female	0.0443 (0.0354)	0.0218 (0.0364)	0.0207 (0.0274)	0.0479 (0.0380)	0.0287 (0.0398)	0.0146 (0.0349)
Asian		0.0460 (0.1049)	0.0164 (0.0575)		0.1812*** (0.0465)	0.0848* (0.0496)
Black		0.0381 (0.0423)	-0.0365 (0.0430)		0.0122 (0.0447)	-0.0343 (0.0448)
Hispanic		0.1051** (0.0531)	0.1050*** (0.0363)		0.1298** (0.0587)	0.0782 (0.0616)
Age		0.0650 (0.1366)	-0.0599 (0.1213)		-0.0409 (0.1581)	0.0462 (0.2168)
Age^2		-0.0086 (0.0109)	0.0048 (0.0096)		-0.0000 (0.0001)	-0.0000 (0.0002)
Years on current court			-0.0854 (0.0850)			-0.0952 (0.1272)
Ideology score			-0.0568 (0.0418)			-0.0795 (0.0611)
Ideology score^2			0.0231 (0.0537)			-0.0359 (0.0833)
Republican			-0.1303 (0.1618)			-0.1570 (0.1865)
Years on current court^2			-0.0005 (0.0096)			0.0034 (0.0160)
% of current staff female			0.9683*** (0.0319)			0.1130** (0.0469)
Court by year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1408	1404	1393	1312	1312	1243
Dependent variable mean	0.70	0.70	0.70	0.69	0.69	0.69

Notes: This table reports OLS estimation results from placebo regressions. Standard errors are robust and clustered at the judge level. In columns 1-3, the dependent variable is an indicator of whether the judge hired at least one female clerk in year *t-1*. In columns 4-6, the dependent variable is an indicator of whether the most similar judge within a court, based on characteristics predicting the employment of female clerks, hired at least one female clerk in year *t+1*. The table reports the results of regressions of the dependent variable on the fraction of co-panelists who are female in each year. Judge characteristics include quadratics of judge age, experience in current position, ideology, judge gender, judge ethnicity, and party of nominating president. Significance levels are: * 10%, ** 5%, *** 1%. *Source:* Judicial yellow books, case dataset collected by authors (see data section).

Table 7: Heterogeneity in Main Effect

Dep Var: Probability of hiring any female clerk in next year

Panel A: Characteristics of judge

Var. Z:	Female	>50% Fem Staff	Qualified (ABA)	Republican	Age < 60	< 10 yrs on court	
	(1)	(2)	(3)	(4)	(5)	(6)	
Frac co-panelists female	0.5665***	0.4287*	0.4910***	0.3941*	0.4379**	0.3801**	
	(0.1755)	(0.2206)	(0.1888)	(0.2193)	(0.1716)	(0.1787)	
Frac co-panelists female X var. Z	-0.4149	0.0532	-0.0797	0.1023	0.0517	0.2771	
	(0.2662)	(0.2648)	(0.2573)	(0.2581)	(0.2580)	(0.2639)	
Var. Z	0.1182*	0.0440	-0.0022	-0.0740	0.0416	-0.0606	
	(0.0654)	(0.0689)	(0.0727)	(0.1160)	(0.0770)	(0.0821)	
Court by year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	1339	1339	1339	1339	1339	1339	
Dependent Variable Mean	0.69	0.69	0.69	0.69	0.69	0.69	

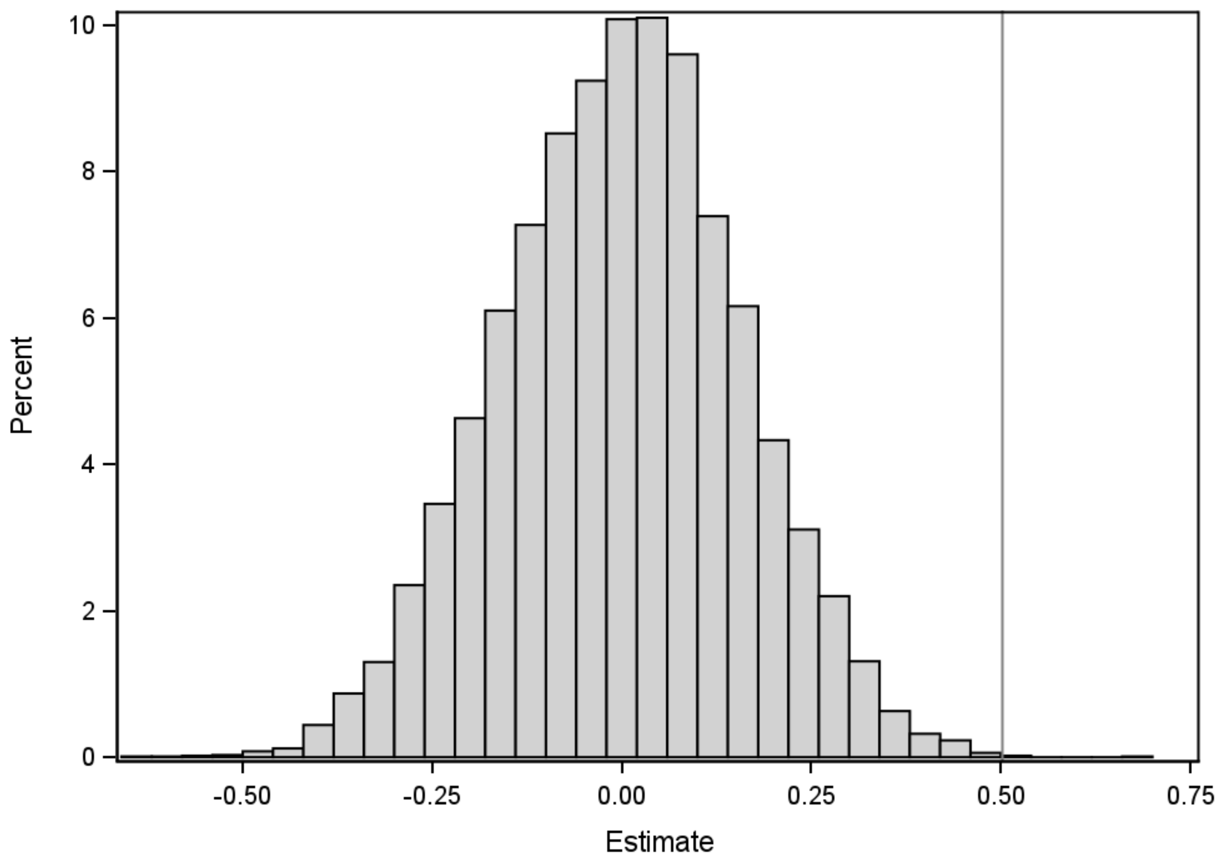
Panel B: Characteristics of co-panelists

Var. Z:		>50% Fem Staff	Qualified (ABA)	Republican	Age < 60	< 10 yrs on court	District Judge
		(2)	(3)	(4)	(5)	(6)	(7)
Frac co-panelists female		0.5841***	0.2432	0.5942***	0.3126	0.3338*	0.4523***
		(0.1879)	(0.2349)	(0.1852)	(0.2616)	(0.1984)	(0.1605)
Frac co-panelists female AND var. Z = yes		-0.2488	0.3456	-0.3593	0.2355	0.3145	1.2380
		(0.2150)	(0.2756)	(0.2969)	(0.3039)	(0.3040)	(2.7433)
Court by year fixed effects		Yes	Yes	Yes	Yes	Yes	No
Observations		1339	1339	1339	1339	1339	1339
Dependent Variable Mean		0.69	0.69	0.69	0.69	0.69	0.69

Notes: Panel A reports estimated coefficients on the interaction of the main independent variable with 6 interaction variables describing judge characteristics. Column (1) is an indicator of whether the judge is female. Column (2) is the fraction of the judge's staff that is female in year t, prior to the new hire. Column (3) is an indicator for whether the was rated as highly qualified by the American Bar Association. Column (4) is an indicator of whether the judge was appointed by a Republican president. Column (5) is the age of the judge (decades). Column (6) is the decades of experience of the judge. Column (7) is whether the judge is a district judge. Panel B reports the estimated coefficients on the interaction of the main independent variable with 6 characteristics of interacting judges. The main independent variable in each regression is the fraction of a judge's co-panelists who are female in each year. All results control for quadratics of judge age, experience in current position and ideology, judge gender, judge race/ethnicity, and party of nominating president. Standard errors are robust and clustered at the judge level. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

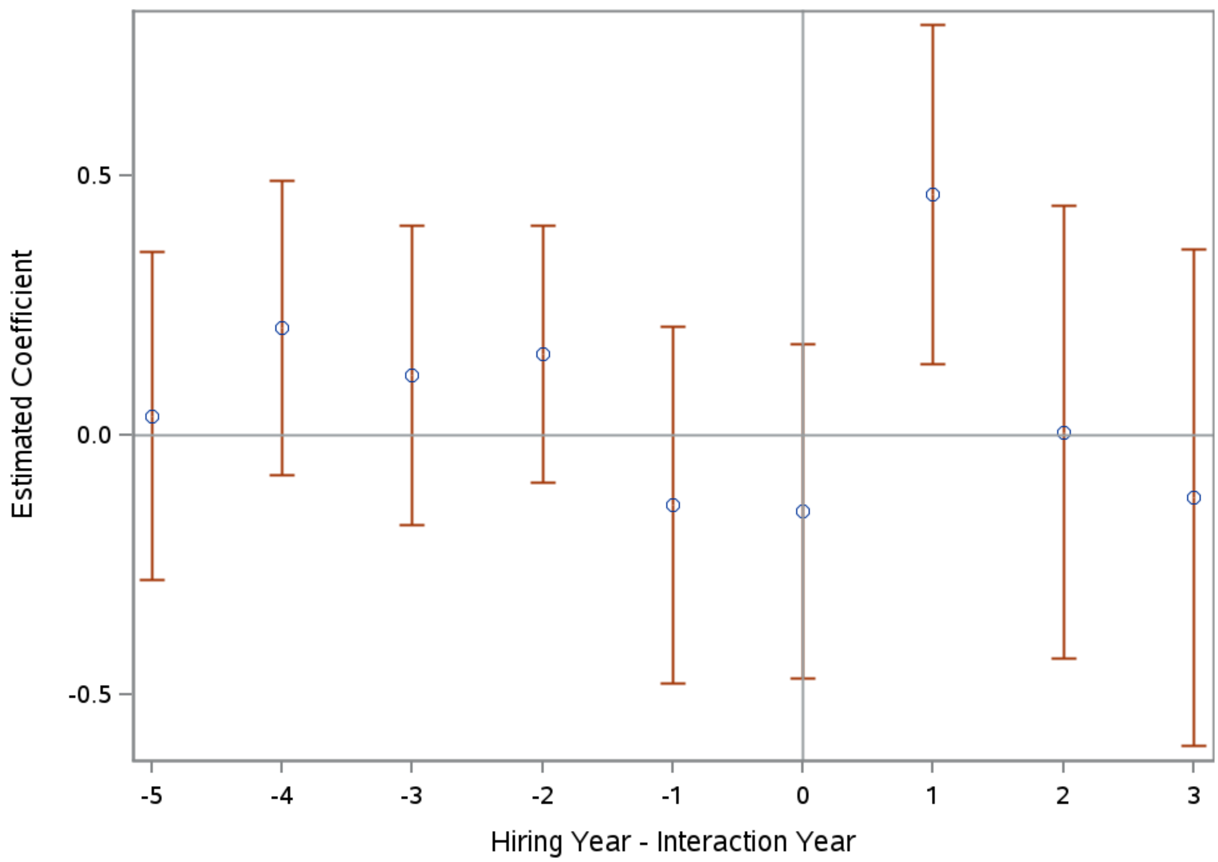
Figures

Figure 1: Randomization-Based Inference for Fraction of Co-Panelists who are Female



Notes: The figure shows distribution of coefficients obtained from the final OLS specification in Table 2 while replacing the fraction of co-panelists who are female for a judge with the fraction of co-panelists who are female from a randomly selected judge from the same court and year. Vertical line represents the actual estimate obtained in the final specification in Table 2.

Figure 2: Event Study: Effect of Fraction Female Copanelists in Year T on Probability of Hiring at Least One Female Clerk in Year T+k



Notes: The figure the effect of interactions with female colleagues in year *t* on the likelihood of hiring at least one female colleage in year t+k, for k ranging from -5 to +3, as described in Equation 2.

Source: Judicial yellow books, case dataset collected by authors (see data section).

Online Appendix Figures and Tables

Table A.1: Validation of Case and Judge Records Data

Court	Published cases in administrative records	Case records in database	Number of judges: Judicial Yellow Books	Number of judges: Federal Judicial Center
	(1)	(2)	(4)	(5)
First Circuit	350	324	9.6	10.4
Second Circuit	310	313	22.4	24.3
Third Circuit	262	157	21.5	22.6
Fourth Circuit	205	200	13.9	17.4
Fifth Circuit	430	391	22.1	22.6
Sixth Circuit	364	351	28.1	29.1
Seventh Circuit	623	611	14.6	15.5
Eighth Circuit	608	591	19.4	20.1
Ninth Circuit	638	613	48.3	47.3
Tenth Circuit	309	271	19.8	21.4
Eleventh Circuit	280	269	16.0	17.1
District of Columbia Circuit	220	216	15.5	14.9
Federal Circuit	-	-	15.9	17.0
Total	4600	4307	267.1	279.6

Notes: Column (1) reports the number of published cases completed in each appellate circuit from 2007 to 2017, according to US Judicial Business Statistics. Column (2) reports the number of cases taken from the leagle database from January 1, 2007 to December 31, 2017 collected by the authors. Column (3) gives the percent of cases reported in administrative data that are included in the leagle database collected by the authors. Column 4 gives the average number of appellate judges included in the judicial yellow book data used in this paper in each year. Column 5 gives the average number of appellate judges reported by the Federal Judicial Center in each year.

Table A.2: Sample Selection

Sample	Observations	Appellate Judges
Appellate judges in Judicial Yellow Books, 2007-2017	3513	395
Appearing on at least one published appellate panel, 2007-2017	2827	305
Hired at least one clerk in year after appearing on panel	1339	229
With known gender	1339	229
With known race and age	1339	229
Appointed by President (non-magistrate judge)	1339	229
With known staff % female	1244	210

Notes: This table reports the number of observations (judge X year) and the number of distinct appellate judges included in the analysis under each set of restrictions imposed on the data. The primary sample consists of 1339 observations from 215 appellate judges. *Source*: Judicial yellow books, case dataset collected by authors (see data section).

Table A.3: Raw and residual variation in female hires, female co-panelists

Panel A: fraction of co-panelists who are female	Mean	Std.dev.	Min	Max	Obs
Raw variable	0.2257	0.1007	0.0000	1.0000	1339
Residuals: net of court by year fixed effects	0.0000	0.0755	-0.2110	0.7685	1339
Residuals: net of court by year fixed effects, judge characteristics	0.0000	0.0724	-0.2260	0.7451	1339
Residuals: net of court by year fixed effects, judge fixed effects	0.0000	0.0579	-0.2352	0.2957	1339

Panel B: Hired a female clerk	Mean	Std.dev.	Min	Max	Obs
Raw variable	0.6923	0.4617	0.0000	1.0000	1339
Residuals: net of court by year fixed effects	0.0000	0.4378	-0.9091	0.8333	1339
Residuals: net of court by year fixed effects, judge characteristics	0.0000	0.4276	-0.9273	0.8777	1339
Residuals: net of court by year fixed effects, judge fixed effects	0.0000	0.3739	-0.9830	0.9201	1339

Notes: This table reports descriptive statistics for the key dependent and independent variables in this analysis. The key dependent variable is an indicator for whether a judge hired a female clerk in each year, conditional on hiring. The key independent variable is the fraction of a judge's co-panelists who were rated as highly qualified by a majority of American Bar Association raters who are female in each year. Judge characteristics include quadratics of judge age, experience in current position and ideology, judge gender, Hispanic ethnicity and party of nominating president. *Source*: Judicial yellow books, case dataset collected by authors (see data section).

Table A.4: Fraction of Cases Published on Judge Characteristics

Dep Var: Fraction of cases published	(1)	(2)	(3)
Female	-0.0107	-0.0107	-0.0006
	(0.0167)	(0.0161)	(0.0215)
Asian	0.0404	0.0354	0.0571
	(0.0604)	(0.0626)	(0.0744)
Black	-0.0183	-0.0298	-0.0365
	(0.0259)	(0.0254)	(0.0347)
Hispanic	-0.0371	-0.0368	-0.0197
	(0.0287)	(0.0303)	(0.0346)
Age	0.0555	0.0663	0.1023
	(0.0527)	(0.0623)	(0.0856)
Age^2	-0.0049	-0.0080	-0.0111
	(0.0046)	(0.0056)	(0.0075)
Years on current court		0.0348	0.0586
		(0.0298)	(0.0375)
Years on current court^2		0.0011	-0.0023
		(0.0070)	(0.0087)
Ideology score		0.0524	0.0627
		(0.0675)	(0.0822)
Ideology score^2		-0.2899***	-0.3086**
		(0.1077)	(0.1278)
Republican		-0.0614	-0.0658
		(0.0455)	(0.0556)
% of current staff female			0.0199
			(0.0305)
Court by year fixed effects	Yes	Yes	Yes
Observations	453	453	332
Dependent variable mean	0.54319	0.54319	0.55424

Notes: This table reports OLS estimation results from regressions of the fraction of each judge's cases in 2016 and 2017 that were published on judge characteristics. Standard errors are robust and clustered at the judge level. The dependent variable is calculated as the number of published cases in 2016 and 2017 on which the judge was empaneled divided by the total number of cases in which an opinion was made available online and on which the judge was empaneled in 2016 and 2017. The table reports the results of regressions of the dependent variable on a binary indicator of whether the judge is female, binary indicators of whether the judge is Asian, Black, and Hispanic, and quadratics of judge age, experience, and ideology. Significance levels are: * 10%, ** 5%, *** 1%.

Source: Judicial yellow books, case dataset collected by authors (see data section).

Table A.5: Interactions with Female Copanelists on Published vs Unpublished Cases

Dep Var: Fraction of current staff female	(1)	(2)	(3)
Female Colleagues on Published cases / Female Colleagues on Unpublished Cases	-0.1130 (0.0874)	-0.1245 (0.0884)	-0.0887 (0.0897)
Female	-0.0338 (0.0376)	-0.0431 (0.0385)	-0.0597 (0.0388)
Asian		0.1342 (0.1560)	0.1023 (0.1513)
Black		0.0467 (0.0618)	0.0355 (0.0609)
Hispanic		-0.0702 (0.0607)	-0.0834 (0.0615)
Age		0.1430 (0.1906)	0.1601 (0.2183)
Age^2		-0.0132 (0.0156)	-0.0112 (0.0179)
Years on current court			-0.0636 (0.0901)
Years on current court^2			0.0023 (0.0209)
Ideology score			-0.0816 (0.1860)
Ideology score^2			0.1416 (0.2773)
Republican			0.0220 (0.1212)
Court by year fixed effects	Yes	Yes	Yes
Observations	294	294	294
Dependent variable mean	0.4305	0.4305	0.4305

Notes: This table reports OLS estimation results from regressions of staff gender composition on the fraction of female co-panelists on published cases. Standard errors are robust and clustered at the judge level. The primary independent variable represents the number of female co-panelists on published cases in 2016 and 2017 divided by the number of co-panelists on unpublished cases in 2016 and 2017.The table reports the results of regressions of the dependent variable on a binary indicator of whether the judge is female, binary indicators of whether the judge is Asian, Black, and Hispanic, and quadratics of judge age, experience, and ideology. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

Table A.6: Characteristics of Published and Unpublished Cases

	Published Cases	Unpublished Cases
At least one amicus brief filed	9.4%	0.3%
Oral Argument Requested	37.7%	6.2%
Number of Cases Cited	23.0	6.9
Word Count	4145.8	1016.8

Notes: This table reports the average characteristics of cited and uncited cases available on the Leagle database in 2016 and 2017.

Source: Leagle database, author's calcuations.

Table A.7: Effect of Serving with Female Judges on Hiring Decisions: Alternative Dependent Variables

	% of Clerks Female		Number of Clerks Female	
Dep Var: Percentage/Number of clerks hired in next year who are female	(1)	(2)	(3)	(4)
Fraction of co-panelists who are female	0.2161 (0.1350)	0.2343* (0.1340)	0.5829* (0.3444)	0.6333** (0.3216)
Female	0.0266 (0.0221)	0.0103 (0.0229)	0.1241* (0.0641)	0.0067 (0.0661)
Asian		-0.0384 (0.0742)		0.0695 (0.2790)
Black		0.0339 (0.0369)		0.1291 (0.1112)
Hispanic		0.0147 (0.0376)		0.2633** (0.1273)
Age		0.1365 (0.1082)		0.0025 (0.3160)
Age^2		-0.0001 (0.0001)		-0.0001 (0.0002)
Years on current court		-0.0186 (0.0926)		-0.4214 (0.2679)
Years on current court^2		-0.0255 (0.0399)		0.0162 (0.1101)
Ideology score		-0.0357 (0.0592)		-0.0207 (0.1680)
Ideology score^2		-0.1912 (0.1376)		-0.0394 (0.3993)
Republican		0.0014 (0.0107)		-0.0273 (0.0278)
Court by year fixed effects	Yes	Yes	Yes	Yes
Observations	1339	1339	1339	1339
Dependent variable mean	0.4010	0.4010	1.1188	1.1188

Notes: This table reports OLS estimation results from the regressions described in equation (2) in the text. Standard errors are robust and clustered at the judge level. The dependent variable in columns (1) and (2) is the fraction of hires in year *t+1* who are female. The dependent variable in columns (3) and (4) is the number of hires in year *t+1* who are female. The table reports the results of regressions of the dependent variable on the fraction of co-panelists who were female in each year. Significance levels are: * 10%, ** 5%, *** 1%.

Source: Judicial yellow books, case dataset collected by authors (see data section).

Appendix Table A.8: Escalating interaction controls

Dep Var: Probability of hiring any female clerk in next year	(1)	(2)	(3)	(4)	(5)	(6)
Fraction of co-panelists who are female	0.4950*** (0.1636)	0.4738*** (0.1630)	0.4963*** (0.1667)	0.4534*** (0.1643)	0.4954*** (0.1635)	0.4463*** (0.1664)
Fraction of co-panelists who are Republican		-0.1230 (0.1460)				-0.1613 (0.1505)
Fraction of co-panelists <10 years in current position			0.3234** (0.1512)			0.2483 (0.1834)
Fraction of co-panelists <60 years old				0.3093** (0.1492)		0.1661 (0.1816)
Fraction of co-panelists above average citations					-0.2234 (0.1482)	-0.2159 (0.1492)
Judge Controls	No	Yes	Yes	Yes	Yes	Yes
Court by year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1339	1339	1339	1339	1339	1339
Dependent Variable Mean	0.69	0.69	0.69	0.69	0.69	0.69

Notes: This table reports OLS estimation results from the regressions described in equation (2) in the text. Standard errors are robust and clustered at the judge level. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the results of regression of the dependent variable on the fraction of co-panelists who were female in each year. Significance levels are: * 10%, ** 5%, *** 1%. *Source:* Judicial yellow books, case dataset collected by authors (see data section).

Appendix Table A.9: Effect of Serving with Female Judges on Hiring Decisions, Omitting Third Circuit

Dep Var: Probability of hiring any female clerk in next year	(1)	(2)	(3)	(4)
Fraction of co-panelists who are female	0.4287** (0.1741)	0.4541*** (0.1725)	0.4454*** (0.1689)	0.4528*** (0.1752)
Female	0.0715** (0.0312)	0.0489 (0.0321)	0.0242 (0.0328)	0.0419 (0.0341)
Asian		0.0991 (0.0951)	0.0012 (0.0964)	0.0716 (0.0878)
Black		0.0288 (0.0478)	-0.0181 (0.0486)	-0.0161 (0.0495)
Hispanic		0.0785 (0.0536)	0.0451 (0.0538)	0.0476 (0.0591)
Age		-0.0006 (0.1402)	0.0442 (0.1570)	0.1049 (0.1849)
Age^2		-0.0000 (0.0001)	-0.0000 (0.0001)	-0.0001 (0.0001)
Years on current court			-0.0917 (0.1208)	-0.0888 (0.1273)
Years on current court^2			-0.1144** (0.0547)	-0.1121* (0.0584)
Ideology score			-0.0196 (0.0772)	-0.0148 (0.0807)
Ideology score^2			-0.2884 (0.1939)	-0.2986 (0.1998)
Republican			0.0103 (0.0145)	0.0111 (0.0154)
% of current staff female				0.1331*** (0.0490)
Court by year fixed effects	Yes	Yes	Yes	Yes
Observations	1229	1229	1229	1138
Dependent variable mean	0.68836	0.68836	0.68836	0.68366

Notes: This table reports OLS estimation results from the regressions described in equation (2) in the text, omitting data from the Third Circuit. Standard errors are robust and clustered at the judge level. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the results of regressions of the dependent variable on the fraction of co-panelists who were female in each year.Significance levels are: * 10%, ** 5%, *** 1%.

Source: Judicial yellow books, case dataset collected by authors (see data section).