Journal of Empirical Legal Studies Volume 8, Issue 3, 504–532, September 2011

Judging Women

Stephen J. Choi, Mitu Gulati, Mirya Holman, and Eric A. Posner*

Justice Sonia Sotomayor's assertion that female judges might be better than male judges has generated accusations of sexism and potential bias. An equally controversial claim is that male judges are better than female judges because the latter have benefited from affirmative action. These claims are susceptible to empirical analysis. Using a data set of all the state high court judges in 1998–2000, we estimate three measures of judicial output: opinion production, outside state citations, and co-partisan disagreements. For many of our tests, we fail to find significant gender effects on judicial performance. Where we do find significant gender effects for our state high court judges, female judges perform better than male judges. An analysis of data from the U.S. Court of Appeals and the federal district courts produces roughly similar findings.

I. INTRODUCTION

Justice Sonia Sotomayor's suggestion, prior to her elevation, that women might be better judges than men ignited an inferno of criticism in the months leading up to her confirmation hearings, and she backed away from it (Lithwick 2009; Dickerson 2009). The claim contradicts a more familiar notion that presidents and other elected officials must engage in affirmative action favoring women in order to ensure that the judiciary has a sufficient mix of women and men. Justice Sotomayor's claim that, because of their backgrounds, women are better judges than similarly qualified men, implies that presidents do not appoint less competent women but merely engage in a kind of statistical reverse discrimination by treating femaleness as a proxy for judicial quality.

The idea that women might be better—or no worse—judges than are men breaks from taken-for-granted assumptions of the recent past. Female judges were rare before the 1970s (Schafran 2005). In 1977, Rose Bird was the first woman appointed to the California Supreme Court and by 1980, 14 women sat on state high courts among several hundred men (Curriden 1995). Sometime after 1980, the political establishment decided that women should have greater representation on the courts. By 1995, more than 50 female

^{*}Address correspondence to Eric Posner, University of Chicago Law School, 1111 E. 60th St., Chicago, IL 60637; email: eric_posner@law.uchicago.edu. Choi is at New York University Law School; Gulati is at Duke University Law School; Holman is at Florida Atlantic University—Political Science; Posner is at University of Chicago Law School.

Thanks to Rick Abel, Christina Boyd, Ros Dixon, Maxine Eichner, Laura Gomez, Sung Hui Kim, Jack Knight, Ann McGinley, David Levi, Carrie Menkel-Meadow, Jeff Rachlinski, Un Kyung Park, Gowri Ramachandran, Jon Tomlin, and participants at workshops at Duke, UNM, and Southwestern law schools for comments. Special thanks to Charles Clotfelter and Lee Epstein for comments on multiple occasions. Choi wishes to thank the Filomen D'Agostino and Max E. Greenberg Research Fund for providing summer research support for this article.

judges had joined the state high courts (Curriden 1995; Songer & Crews-Meyer 2000). We find that, in the period from 1998 to 2000, more than 100 women sat on the state high courts, roughly a quarter of the total. The federal courts similarly witnessed a dramatic increase in the fraction of female judges during the past two decades (Hurwitz & Lanier 2008).

Much of this change no doubt resulted simply from the increasing numbers of women who have entered the legal profession since the 1970s. Since this time, politicians likely engaged in affirmative action, giving preference to female candidates who are less qualified than men on the basis of standard measures, such as length of time in the profession. Women serving on state high courts starting in the late 1990s generally went to law school in the mid 1970s, where they were the distinct minority in law schools and in the legal profession. If there is a smaller pool of women from which to select judges (compared with the pool of potential male judges), then forcing the selection of a substantial number of women may result in more qualified men being passed over, thereby reducing overall court performance (Buchanan 2009; Shapiro 2009).

The bulk of the literature on gender and judging examines what we call the "differential viewpoints" question.¹ This literature focuses on the subject areas where female judges are likely to bring a distinctive perspective to bear. The most prominent finding is that female judges are more likely to favor plaintiffs in sex discrimination cases (Peresie 2005; Boyd et al. forthcoming). This result does not, however, cast light on whether female judges are better or worse than men. The empirical research has not established that the female judges are legally correct in these cases; it is possible that those plaintiffs should have lost.

Our focus is on the relationship between the gender of judges and judicial quality. We cannot directly test Justice Sotomayor's claim that female (or Latina) judges exercise superior judgment because we do not have a direct measure of the quality of opinions, but proxies for the overall quality of judicial performance are available. Drawing on prior work on judicial output, we focus on opinion publication, citations, and disagreements with co-partisans as metrics of judicial performance. Using three data sets—justices sitting on the highest courts of the 50 states from 1998 to 2000, federal appellate judges from 1998 to 2000, and federal district judges from 2001 to 2002—we test whether gender has a significant effect on judicial performance. With qualifications that will be discussed below, for many of our tests on data from the state high courts we are unable to reject the null hypothesis of no gender effects and instead find only insignificant gender-related differences. Where we do reject the hypothesis that gender has no effect for our sample of state high court judges, we find that female judges in fact perform better than male judges. Our analysis of U.S. federal circuit and district court judges produces roughly similar results.

¹See Beiner (1999), Davis (1993), and Sherry (1986). Research in this area has asked whether there are variations in the outcomes of cases in certain areas due to the different perspectives women bring to the bench (Davis 1993; Allen & Wall 1993). Scholars have examined whether female judges rule differently in subject areas perceived to involve women's issues or areas where women's supposed liberal leanings will make a difference, such as criminal law matters (Songer et al. 1994; Jackson 1997; Martin & Pyle 2000). Although the overall picture is unclear (Palmer 2001), the general story appears to be that female judges support the rights of women more strongly than do their male colleagues (Martin & Pyle 2005; McCall & McCall 2007; McCall 2008).

II. PREDICTING GENDER DIFFERENCES

One distinctive characteristic of U.S. courts, as compared to their counterparts in most other countries, is that judges come to the bench later in life, roughly around age 50, after significant experience outside the judiciary. The aggregation of these prior experiences constitutes a judge's human capital—in effect, his or her training to become a judge. A lawyer with more legal experience should be a better judge than a lawyer with less legal experience. In addition, attending a better law school should, theoretically, provide better training for the tasks associated with judging. Further, because judicial candidates coming to the bench have a major portion of their professional career behind them, they have likely passed through numerous selection screens already.

These factors suggest two opposite sets of predictions. Under what we call the *preference story*, women who are less qualified than men are selected to be judges, with the result that female judges perform less competently than do male judges. Our empirical tests focus on the preference story, which has support in the literatures on women in the legal profession. Alternatively, under the *screening story*, prejudicial barriers to entry—including sex discrimination and employment conditions that are hostile to the needs and interests of women—screen out less competent women. Even though the resulting pool of women candidates for the judiciary might be smaller than the pool of men, the women who remain in that pool after the informal screening might be of higher quality than the men. Assuming that there are but a small number of judicial positions, a small pool of women might still provide more than adequate numbers of candidates to select from as the large pool of men. Ultimately then, the screening story implies that female judges should be as competent as—or more competent than—male judges.

A. The Preference Story

1. Women Law Students and Lawyers

Research on gender and legal education suggests that women have a lower quality experience in law school than do their male colleagues. They participate less in classroom discussion, feel more alienated, and underperform in terms of the traditional indicators of success in law school such as grades, law review membership, and publications (Banks 1990; Guinier et al. 1994; Mertz et al. 1998; Mertz 2007; Leong 2009). Female students also are disproportionately excluded from social networks among students, faculty, and alumni and ultimately receive less value from their educations (Guinier et al. 1994; Leong 2009).

This pattern of limited access continues at the next stage, early legal employment. The initial job for lawyers out of law school provides basic apprenticeship for fresh law graduates; after a few years they can take that training to other jobs (Garth & Sterling 2009). In this apprenticeship, women may receive fewer advantages than men do. Within private law firms, research suggests that women (and other outsiders) receive less in terms of training, mentoring, and networking opportunities (Garth & Sterling 2009; Wilkins & Gulati 1996, 1998). In addition, multiple studies find that women are more likely to enter into the public sector than their male counterparts, thus forgoing the apprenticeship experience altogether (Hull & Nelson 2000; Wood et al. 1993).

2. Female Judges

Our statistical analysis focuses on individuals who were judges in the late 1990s and early 2000s, and (for the most part) went to law school in the 1970s or before. As of the early 1970s, the fraction of women in law schools was in the 10–20 percent range (Epstein 1993; Savage 2009). Because women in this cohort likely dropped out of the legal profession at a greater rate than men to care for family members or pursue other opportunities, the effective pool of women qualified for judgeships was even smaller by the 1990s. Despite the relatively small pool of potential female judges, the fraction of female judges in our data set of state high court judges from 1998 to 2000 was 24.1 percent. Under the preference story, the disproportionate selection of women judges—given the lower training among women attorneys both at law school and in their early employment—leads to less qualified judges.

There is also the matter of discrimination women might face after they become judges. A series of reports produced by gender bias taskforces around the country starting approximately two decades ago suggested the presence of bias against women participating in the judicial system, including female judges (for overviews, see Resnik 1996; Kearney & Sellers 1996). Facing such discrimination requires female judges to expend greater effort than their male colleagues to have their views heard and requests fulfilled (Barteau 1997; Gandy et al. 2003; Haddon 2008). Justice Ginsburg recently observed:

It was a routine thing [in the past] that I would say something and it would just pass, and then somebody else [who was male] would say almost the same thing and people noticed. I think the idea in the 1950s and '60s was that if it was a woman's voice, you could tune out, because she wasn't going to say anything significant. There's much less of that. But it still exists, and it's not a special experience that I've had. I've talked to other women in high places, and they've had the same experience. (Bazelon 2009)

Research on other professional settings suggests the possibility that women and other outsiders can sometimes become stuck with disproportionate shares of administrative burdens (Carbado & Gulati 2000); this might occur on the courts as well. The prospect of bad working conditions might deter more qualified women (with a resulting higher opportunity cost) from pursuing or accepting judgeships—further diminishing the quality of women judges.

3. Women, Risk Aversion, and Conflict Avoidance

The third body of literature relevant to our predictions concerns women generally, as opposed to women lawyers or judges. Multiple studies find that women display a greater degree of risk aversion than do men (Levin et al. 1988; Powell & Ansic 1997; Corrigan 2009). Women are also found to be less competitive, more averse to conflict, and less prone to aggression than are men (Stuhlmacher & Walters 1999; Gneezy et al. 2003; Croson & Gneezy 2008). The implications of these studies generally are ambiguous for judicial performance. For example, risk-averse judges might be better because they take greater care with their opinions, or worse because they fear offending colleagues or powerful people. There are implications for certain specific aspects of judicial performance, such as the willingness to openly disagree with a co-partisan. The literature suggests that judges

prefer not to dissent because dissenting invites outside scrutiny of the court and creates more work for colleagues (Posner 2008). As a result, judicial colleagues can sometimes take umbrage at dissenting behavior (Posner 2008) and a number of courts have norms against dissenting (Brace & Hall 2005). Risk-averse and conflict-averse judges, therefore, are likely to dissent less (Dumas 2010). In contrast to the preference story, one might not expect the women in the screening story to be risk averse or conflict averse. Given the hurdles they have had to clear, those women who remain probably have a greater inclination toward taking risks and tackling conflict. Further, having had to succeed in male environments might mean that these women are not primarily interested in certain "women's" topics such as family law; instead, they are probably interested in, and adept at, tackling a wide range of issues.

B. Data and Measures

Our data set has information on several objective metrics of judicial performance for all state high court judges in the United States for the years 1998–2000. There are 409 judges, of whom 103 (or 25.18 percent) are female. For each judge, we collected data on three separate measures: the number of published opinions, the number of citations from outside the state (that is, nonprecedent driven citations) to majority opinions, and open disagreements (dissents) with those from the same political party background (our measure of judicial independence).² Others have questioned the value of the objective measures and some have suggested alternate measures (Cross & Lindquist 2009; Baker et al. 2009). For purposes of this article, we tie our predictions of gender differences to the objective measures as opposed to general notions of quality. Although the measures are rough proxies, we have found in other work that they are correlated with other factors in a theoretically sound way (Choi et al. 2009a, 2009b, 2010, forthcoming)³ and so provide at least a starting point in assessing gender differences in judicial production. We also assume that the inadequacies of our objective measures are not a function of gender, allowing us to assess how men and women perform differentially on our measures. In analyzing the results, we control for variations among the states.

C. Predictions

The predictions below are simplified hypotheses based on the preference story and the screening story.

1. Opinion Publication Rates

Publishing an opinion, as opposed to issuing an unpublished disposition, we assume, takes greater effort (Choi et al. forthcoming). Further, the designation of the opinion as

²The information on numbers of published opinions, dissents, and citations was collected from the LEXIS database for the years in question. For specific coding schema, see Choi et al. (forthcoming).

³For example, elected judges and appointed judges differ in a systematic way. In addition, judges close to retirement are less productive and judges with more court experience are more productive (Choi et al. forthcoming).

published brings greater external scrutiny and, therefore, greater risk of criticism. We predict under the preference story that female judges will publish fewer opinions than their male colleagues because they are likely to have received lower amounts of legal training and are more likely to be risk averse. The screening story makes the opposite prediction— women judges should publish either more (or at least no fewer) opinions compared with male judges.

The publication of an opinion gives it greater precedential weight. If women are more interested in advancing the law in certain areas, they will focus their publication efforts in those areas. Given the scholarship cited above that suggests that female judges are more likely to hold in favor of female plaintiffs in civil rights cases than are male judges, women might devote more effort to cases involving civil rights and family law, resulting in more opinions or more frequently cited opinions in those areas. The subject matter prediction is not tied to either the preference story or the screening story, but is supported by the scholarship, is related to a number of debates, and worth testing.

2. Citations

Citations by outside authorities are a commonly used measure of influence (Landes et al. 1998). We collect data on the number of outside state citations to majority opinions by a variety of outside actors, including other state courts and the federal courts outside the relevant circuit.⁴ Citations to judicial opinions have been described as measuring multiple characteristics of the underlying opinions, including quality of analysis (Choi et al. 2009a, 2009b, forthcoming), nimbleness in writing (Vladeck 2005), and creativity (Posner 2005).

If, as predicted under the preference story, women lawyers ascend to the bench with fewer legal skills and are more risk averse than their male colleague, then female judges should write less frequently cited opinions. Women may be less likely than their male colleagues to have built up networks among lawyers and other judges, resulting in fewer citations. If the techniques of reasoning and the perspectives of female judges are markedly different from those of male judges, then the majority of judges (who are men) will likely prefer to cite opinions by male judges (Lithwick 2009). In contrast, the screening story suggests that the opinions of female judges will receive the same if not greater number of citations compared with male judges.

Beyond the preference and screening stories, other predictions are possible. Women judges may receive a differential number of citations in certain subject matter areas, also driven by stereotypes. If there is a perception that women understand better and pay more attention to issues in certain areas—what are considered "women's" domains, such as family law or sex discrimination cases—we would expect to see more citations to female judges in these areas. Conversely, we would expect fewer citations to female judges in areas considered "male," such as business law.

⁴We define the number of outside-state citations to a majority opinion as the sum of outside federal court + other state court + U.S. Supreme Court citations to that majority opinion. All citations are from the LEXIS Shepard's database and are tracked up until January 1, 2007.

3. Disagreement

Our third measure, an assessment of judicial independence, looks at the willingness of a judge to disagree with co-partisans, either by dissenting against their opinions or writing majority opinions that induce dissents. We look first at the number of disagreements by a judge against co-partisans divided by the total number of disagreements by the judge against a judge of either party (Same Party). We used the following methodology to determine the political parties for each of our state high court judges (used in determining whether a disagreement is against a co-partisan). We looked to three sources of information on party membership. First, we searched NEXIS and the Internet (using Google) for any news reports on the political affiliation of the each judge. Second, we searched for information on political contributions at the opensecrets.org website. We used the political party of the donee candidate as a proxy for the political party of judges who contributed. Where a judge contributed to candidates from more than one political party, we did not use the opensecrets data to assign a political affiliation to the judge. Third, we used the party of the governor (if any) who appointed the judge as a proxy for the judge's political party. In most of the cases where we had multiple sources of information on political party, the party was consistent across these sources. Where we found no data on the judge's political affiliation or the judge's affiliation was neither a Democrat nor a Republican (but was instead an Independent), we ignored the judge for purposes of calculating the independence measure. When our three sources reported different parties, we gave first priority to the party identified through our NEXIS and Internet searches, second to the party identified in the opensecrets.org database, and third to the party of the appointing governor.

This measure gives us a raw sense of how often a particular judge is in open disagreement with co-partisans. A highly partisan judge, for example, may never come in disagreement with a co-partisan (preferring to save his or her dissents primarily for judges from the opposite political party). How often a judge opposes a co-partisan, of course, will depend on the number of co-partisans on the same bench. If a judge is the lone Democrat on a specific court, the judge will necessarily oppose opposite-party judges (due to the lack of any co-partisans).

To control for court composition, we look second at the total number of majority opinions by co-partisans (opportunities to dissent) over the total number of majority opinions by all judges on the court (Same_Pool).⁵ We then define independence as the difference between (1) the number of disagreements by a judge against co-partisans divided by the total number of disagreements by the judge (Same_Party) and (2) the total number of majority opinions by co-partisans (opportunities to dissent) over the total number of majority opinions by co-partisans (opportunities to dissent) over the total number of majority opinions by all judges on the court (Same_Pool). A more negative score corresponds to a judge who writes opposing opinions against opposite-party judges more

⁵There are problems with this measure that we document in Choi et al. (2009a, 2009b, 2010). Among them is that our measure does not work for the handful of states where all the judges are of the same party. Accordingly, we drop those states from our independence calculations (including Georgia, Maryland, New Mexico, South Carolina, and South Dakota). Further, as a function of the number of judges of each party on a court, the potential scores for a judge are bounded. To adjust for this, we calculated a simpler, alternate measure of independence.

frequently than the background pool of majority opinions authored by opposite-party judges. Conversely, a more positive score corresponds to an authoring judge who writes opposing opinions less frequently against opposite-party judges compared with the background pool of opinions (and thus more frequently against co-partisans).

We treat a more positive score as indicative of a more independent judge. Others might view disagreement among judges as a negative—a sign of disagreeability or cantankerousness. Regardless of perspective, the prediction under the preference story is that women will disagree less. If the preference story is correct, female judges, because they are less likely to be willing to engage in open conflict, particularly with co-partisans, should receive lower scores on our independence measure. Further, their relatively lower levels of legal training (from discrimination in school and in the workplace) should also make them less willing to engage in conflict, since their (mostly male) opponents will have greater skill and experience. In contrast, the screening story predicts that women judges will receive a higher independence score.

To summarize, we have five predictions regarding gender differences to show up in our measures if the preference story is correct. Female judges will publish fewer opinions overall (Hypothesis 1), but more opinions on topics of specific interest to women such as family law (Hypothesis 2). Female judges will be cited less overall by outside courts (Hypothesis 3), but more on topics of specific interest to women such as family law (Hypothesis 4). Women will score lower on their willingness to disagree with co-partisans (Hypothesis 5). Three of these predictions (Hypotheses 1, 3, and 5) address the question of whether female judges underperform their male counterparts. The other two (Hypotheses 2 and 4) test whether (any) differential performance on the part of female judges is explainable due to a specific subject matter focus on the part of female judges.

III. DIFFERENT PATHWAYS

A. Education and Training

The preference story assumes that female judges have less experience and lower-quality training than male judges. We test whether this assumption is true. In our data set, female judges have worse educational credentials than do male judges. Panel A of Table 1 reports summary statistics for our sample of state high court judges. The average *U.S. News* rank⁶ of the law school attended by a male judge is approximately 52 and that for a woman judge is 63 (difference not significant). The rankings difference is larger for undergraduate education, where the average college ranking for a woman judge is 154 and that for male judge is 125 (difference significant at the 1 percent level).

⁶To have consistent and reliable information about the rankings of the schools that these judges attended, we used data from 2002. *U.S. News and World Report* data on college rankings is available only back to 1983. In other words, we do not have information on the rankings at the time these judges attended college and law school but, as these rankings tend to be fairly stable over long periods of time, we are confident in using the 2002 version of the rankings (see the Appendix for details).

Table 1: State High Court Judges

Panel A: Background Characteristics for State High Court Judges

	Men		Women			
	Mean	SD	Mean	SD	p Value	
Chief judge	0.1809	0.0221	0.1667	0.0371	0.735	
Court experience	7.8268	0.4231	4.6408	0.5551	0.0001	
Post-law-school experience	32.9186	0.4815	25.6842	0.8504	0.0000	
Close to retirement	0.3750	0.0278	0.1863	0.0387	0.000	
Age	58.5809	0.4851	52.9314	0.7933	0.0000	
Private practice	0.8355	0.0213	0.7647	0.0422	0.112	
PAJID	36.9277	1.2898	38.8382	2.2411	0.4463	
US News BA ranking	124.6352	4.9459	154.2937	10.3061	0.00289	
US News JD ranking	52.4013	2.3747	62.8700	4.5186	0.1524	
In-state school	0.6213	0.0280	0.6000	0.0492	0.723	
Married	0.6494	0.0164	0.5778	0.0301	0.132	
Number of children	1.9141	0.0659	1.0556	0.0822	0.0000	
Divorced	0.0459	0.0072	0.0556	0.0140	0.490	
Selection method						
Appointed	0.1993	0.4001	0.2524	0.4365	0.255	
Merit selection	0.3300	0.4710	0.2233	0.4185	0.041	
Nonpartisan elections	0.2614	0.4401	0.3689	0.4849	0.037	
Partisan elections	0.2092	0.4073	0.1553	0.3639	0.234	

Panel B: Gender and Production, Citations, and Independence for State High Court Judges

	Men		Women			
	Mean	SD	Mean	SD	p Value	
Total number of published opinions per year	26.145	0.598	24.086	0.938	0.0100	
Number of majority opinions per year	18.846	11.909	16.783	10.209	0.0415	
Number of outside-state citations per majority opinion Independence score	$0.7084 \\ -0.0516$	$0.0148 \\ 0.0118$	$0.8138 \\ 0.0093$	$0.0295 \\ 0.0190$	$\begin{array}{c} 0.0891 \\ 0.0087 \end{array}$	

NOTES: For Panel A, p value is calculated from chi-square (CHIEF, CLOSE TO RETIREMENT, PRIVATE PRACTICE, IN-STATE SCHOOL, MARRIED, as well as selection methods), t test (COURT and POST-LAW-SCHOOL EXPERIENCE, AGE, PAJID, US News BA and JD RANKING, CHILDREN), and Fisher's exact (DIVORCED) tests. For Panel B, p value is from a two-sided t test of the difference in means between male and female judges; data for the total number of published opinions, majority opinions, and citations are logged in the calculations of p values but left unlogged for mean and standard deviation comparison in the table.

The gender differences in the types of schools attended appear to be within those law schools ranked 50 or under, which is approximately half the law schools. Between men and women who attended the top five law schools, there are few differences; similar results are seen among those judges who attended top-10 schools. However, among those judges who attended a top-50 school, there are significant gender differences, as 47 percent of men attended these schools, compared to 38 percent of women (difference significant at the 5 percent level). In effect, it is when one goes below the "elite" law schools that a difference emerges in the quality of institution that male and female judges attended.

One other variable that we also examined was whether the judge was employed in a judicial clerkship after graduating from law school. Unreported, we find that men are more likely to have done judicial clerkships, but the data are available for only a small group of judges. A clerkship is not only a sign of high performance in law school, but also is a source of legal training.

B. Prior Professions

Panel A of Table 1 reports the primary prior professions of the state high court judges in our sample. One might expect that women judges would come more often from publicsector jobs, consistent with the patterns for women lawyers more generally (Dau-Schmidt et al. 2007; NALP Foundation 2004). There are several possible explanations for why women are more likely to work in the public sector than men: first, women have more difficulty in tackling the work-life conflict presented by modern law firm jobs (Garth & Sterling 2009). Second, women-because of discrimination or less mentorship-are less likely to receive either training or promotion in the law firm context (Garth & Sterling 2009). There is also research on entering women law students suggesting that they are initially more interested in public interest work than their male colleagues are (Dau-Schmidt et al. 2007) but, by the end of law school, the expectations of men and women students appear to converge in favor of private-sector jobs (Dau-Schmidt et al. 2007; Ku 2008). To obtain information on the primary prior professions of a judge, we obtained information on their prior jobs reported in Who's Who (2007). Lacking consistent information on the length of experience in the private versus public sector for each judge, we instead track the job that a judge held immediately prior to becoming a judge. As this job is a measure of a judge's pre-judicial experience, we use it as a rough measure of prior employment. Panel A of Table 1 reports that 83.6 percent of male judges were in private practice, compared to 76.5 percent of female judges—a difference that is not statistically significant.

C. Marriage, Children, and Age

Background variables such as marriage and number of children, although not necessarily part of the preference story, are potentially relevant control factors as gender differences in these variables could have an impact on performance. Age is also a potentially revealing variable: younger judges are likely to have less experience and training.

The women in our state high court judge data set are less likely to be married than the men and more likely to be divorced (see Table 1). This is consistent with reports on professionals (including lawyers), where women have both lower marriage rates and higher divorce rates than their male counterparts (Wilson 2008). The differences in marriage and divorce rates among men and women judges in our sample, however, are not significant. We find also that male judges have more children than female judges. Reported in Panel A of Table 1, the average is one child for the women versus just under two children for the men (difference significant at the 1 percent level). Women also are less likely to have children than men (43 percent of the women have children vs. 57 percent of the men). These numbers are perhaps more indicative of the screening story than the preference story: women who succeed at becoming judges at a high level are those who have chosen their careers over a family.

In terms of age, the women in our state high court judge sample attended and graduated from law school later than their male colleagues. The average [D date for women is 1972 versus 1965 for the men. Given the years of graduation, it is safe to assume that many of these women likely faced significant barriers when they were law students; in 1972, women made up only 10 percent of the JDs (Catalyst Report 2009). Law school environments were hostile to women during the early 1970s, when their numbers were small (Epstein 1993). Measured in 2000, the mean age of male judges in our sample is 58.6 and the mean age of women judges is 52.9 (difference significant at the 1 percent level). Comparing the judge's age at graduation from law school to his or her age at becoming a judge, we see that women rise more quickly to judgeships; unreported, it takes female judges, on average, 21 years from JD to judgeship, while it takes male judges over 26 years (difference significant at the 1 percent level). As a result, women are younger (48 years old) than their male counterparts (51.5 years old) when they become state high court judges (difference significant at the 1 percent level). We also find, unreported, that women are older when they graduate from law school, regardless of the year of graduation. The foregoing is consistent both with the preference story and with the screening story. Looking at the preference story, the smaller pool of available women lawyers to choose from probably meant that those selecting judges had to go deeper into the pool-hence, selecting female judges who were younger and less experienced than their male counterparts. On the other hand, women who are overachievers might take less time to accomplish professional goals, which fits the screening story.

D. Type of Judicial Selection System

Finally, we examine the type of judicial selection systems for state high court judges most likely to yield female judges. The bottom portion of Panel A of Table 1 reports that female judges are most numerous in states with nonpartisan election systems (and, to a lesser extent, appointment systems) and less present in merit selection states. It is hard to make much out of this, except perhaps that officials are more likely to engage in affirmative action than is the public.

IV. TESTING THE HYPOTHESES

A. Predictions of Gender Underperformance

Panel B of Table 1 reports the raw differences in publication rates, outside citations, and independence for our sample of state high court judges.⁷ Generally, men publish more, writing and publishing an average of 26.15 opinions per year, while women write and

⁷We use slightly different levels of analysis for each of these measures: citations are measured at the individual-opinion level; production is measured for each judge for each year; and independence is for each judge with all years combined.

publish 24.09 opinions per year (difference significant at the 1 percent level).⁸ Looking just at majority opinions, we find that male judges published 18.85 majority opinions per year; female judges published only 16.78 majority opinions per year (difference significant at the 5 percent level). However, women are cited⁹ more than their male counterparts (0.81 outside state citations per opinion for women and 0.71 for men, a difference that is significant at the 10 percent level) and are more independent (significant at the 1 percent level). Unreported, we also examined the page numbers of opinions published (both majority and total) as an alternate measure of productivity and found no significant gender differences. At the first cut, then, women outperform men on two of three measures. However, the various states differ in terms of the characteristics of their legal systems and the types of disputes they receive. To say anything meaningful about gender differences, therefore, one has to correct for state differences.

States vary along a number of dimensions, including differences in population, crime rates, court structures, and judicial salaries. As there is no reason to expect big variations in these state-specific variables in the three years in our sample (1998–2000), we estimate using a state-fixed-effects model, which controls for state-level differences in independent variables. We estimate the following equations using regressions on pooled judge-level data (Independence), judge-year-level data (Production), and opinion-level data (Citations).¹⁰

Independence (ordinary least squares model with robust standard errors):

Independence_i = $\alpha + \beta_{1i}$ Female + State Fixed Effects + ε_i

Production (negative binomial regression with errors clustered by judge):

Number of Majority Opinions per Year_i = $\alpha + \beta_{li}$ Female + State Fixed Effects + Year Fixed Effects + ϵ_i

Citations (negative binomial regression with errors clustered by judge):

Number of Outside State Citations per Majority Opinion_i = $\alpha + \beta_{li}$ Female + State Fixed Effects + Year Fixed Effects + Subject Matter Controls + ϵ_i

We include year fixed effects for both the Production and Citation models but not for the Independence model, which is estimated on data pooled over the 1998 to 2000

 $^{^{8}}P$ values are calculated using the log of citation and publication variables to control for the nonnormal distribution of the data.

⁹We use citations from courts outside the state throughout the article. We also test a variety of citation types, including law review citations and citations from different types of citing courts. Unreported, women are cited at the same level or more than are their male counterparts regardless of whether looking law review citations or a specific type of court citation. Law review citations are for law reviews as tracked by the LEXIS Shepard's database (until January 1, 2007).

¹⁰As the Production and Citations data presented here are count data, we use negative binomial regression with standard errors clustered at the judge level to estimate the Production and Citations models.

	Independence	Production (Number of Majority Opinions per Year)	Citations (Number of Outside-State Citations per Majority Opinion)
Female	0.0641**	-0.056	0.0145
	(3.29)	(-1.37)	(0.28)
Constant	-0.0252	4.117**	-0.806**
	(-0.62)	(43.99)	(-3.58)
Subject matter controls	No	No	Yes
State fixed effects	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes
Ν	350	1080	25025
R^2	0.299	N/A	N/A

 Table 2:
 Gender and Production, Outside Citations, and Independence for State High

 Court Judges

NOTES: *T* statistics in parentheses; +p < 0.10; *p < 0.05; **p < 0.01. The Independence model is estimated on judgelevel data with ordinary least squares with robust standard errors. The Production (judge-year-level data) and Citations (opinion-level data) models are estimated using a negative binomial regression. R^2 is unavailable with negative binomial regressions. Errors in the Production and Citations models are clustered by state judge. Subject matter controls include indicator variables for the following case subject matter areas: administrative, attorney and client, capital punishment, church and state, commercial, criminal, family, First Amendment, labor, property, rights, and torts (with "other" as the base category). The subject matter areas are defined in the Appendix.

sample time period. For the Citations model, estimated on opinion-level data, we include controls for the subject matter of the case. A criminal law case will generate a different number of outside state citations, all else equal, compared with a commercial law case. Subject matter controls (defined in the Appendix) include the following case subject matter areas: administrative, attorney and client, capital punishment, church and state, commercial, criminal, family, First Amendment, labor, property, rights, and torts (with "other" as the base category).

As Table 2 shows, once we correct for state fixed effects, the gender differences for both publications and outside citations disappear, suggesting that men and women are performing at roughly the same levels. Differences remain in the independence regressions after inserting state controls, with female judges scoring higher on independence (significant at the 1 percent level). Thus far, we find little support in the data for the preference story's prediction (Hypotheses 1, 3, and 5) that female judges will underperform—if anything, female judges outperform male judges, at least on our independence measure. To examine the question of the importance of gender effects and judicial performance, we estimate separate models for each of our measures with a variety of control variables.

B. Controlling for Judge Background Characteristics

The state high court judges in our sample vary on a number of individual characteristics, all of which might affect judicial outcomes. Some of these variables are proxies for human capital, including education, years of experience, or prior profession. An important element of the preference story is that those female lawyers who become judges have

a lower accumulated amount of human capital from their careers (including law school and private practice) and, therefore, will not perform as well as the male judges. We find (see Table 1) that the women judges graduated from lower-ranked law schools and undergraduate institutions, have less post-law-school experience or experience on the court, and are generally younger. This suggests that the assumptions underlying the preference story have support. However, our state-fixed-effects findings lead us to ask alternate questions about why we see either insignificant or positive effects for women on our measures of judicial output. The first question is whether the traditional measures of human capital have purchase in the gender and judging narrative. It may be, for example, that the effect of gender is indirect. Judges who graduated from lower-ranked law schools may perform worse due to their relatively lower human capital; women, in turn, are more likely to have graduated from lower-ranked law schools (and thus have lower human capital). If the answer is yes, that the preference story is correct, we should expect to find significance for our judge background variables in our production, quality, and independence models. If the answer is no, and focusing on traditional measures of human capital is the wrong approach, we should see no significant effects of any background variable in the models.

Note that the results already reported in Table 2 suggest that the preference story, with its emphasis on traditional human capital measures, does not hold up. If it had, women would have had scores significantly lower than those for men in our state-fixed-effects models in Table 2. Instead, we found that while women did have lower levels of human capital (on the traditional measures), they still scored just as well as the men on our measures, even without controlling for background differences.

To test the importance of traditional human capital, we add independent variables for a variety of judge-level background factors, collectively referred to as "judge controls," to the Independence, Production, and Citations models of Table 2. Our judge controls include whether the judge was the chief judge of the high court. A judge who is chief judge may have less time to author opinions, but also may command greater respect and receive greater numbers of citations as a result, for her opinions. The chief may also be able to assign herself the more important opinions and garner more citations that way (Langer 2003). For experience, we include the number of years between 1998 and the year in which the judge received her law degree (POST-LAW-SCHOOL EXPERIENCE) and the number of years the judge has been on the high court (COURT EXPERIENCE). More experienced judges may decide opinions with greater skill, leading to more citations. We include a variable valued at 1 for whether a judge retired from the bench in 2001 or earlier and 0 otherwise (RETIREMENT CLOSE).

We also include a number of variables specific to the background of the individual judge measured as of 2000. These include the age of the judge (AGE), whether the judge was married (MARRIED), the judge's number of children (NUMBER OF CHILDREN), whether the judge was divorced (DIVORCED), and whether the judge's primary experience before becoming a judge was in private practice (PRIVATE PRACTICE). We include the PAJID score for each judge as developed by Brace et al. (2000). These scores locate judges on a political continuum from highly conservative (0) to highly liberal (100). We lastly

	Independence	Production (Number of Majority Opinions per Year)	Citations (Number of Outside State Citations per Majority Opinion)
Female	0.0809**	-0.0384	0.0518
	(3.62)	(-1.41)	(1.08)
Chief judge	-0.0071	-0.0536*	-0.0463
	(-0.28)	(-1.99)	(-0.92)
Court experience	0.0021	0.0144**	0.00237
×.	(1.04)	(5.78)	(0.63)
Post-law-school experience	0.0001	-0.0000929	-0.00240
	(0.08)	(-0.03)	(-0.51)
Retirement close	0.0271	-0.254**	-0.0590
	(1.11)	(-8.29)	(-1.22)
Age	0.0001	-0.00161	-0.00210
0	(0.05)	(-0.54)	(-0.48)
Married	0.0286	0.0451	-0.0686
	(1.05)	(1.25)	(-1.25)
Number of children	-0.00338	0.0174*	0.0319*
	(-0.37)	(2.04)	(2.08)
Divorced	0.0638	0.0369	-0.0701
	(1.58)	(1.12)	(-0.83)
Private practice	-0.0344	0.0696+	0.0335
*	(-1.04)	(1.96)	(0.59)
PAJID	0.00004	0.00100	0.00118
5	(0.07)	(1.53)	(1.15)
US News JD ranking	-0.0006	0.000325	0.0000995
0 0	(-1.64)	(0.82)	(0.15)
In-state law school	0.0286	-0.0704*	0.00360
	(1.18)	(-2.31)	(0.07)
Constant	-0.00460	4.031**	-0.397
	(-0.04)	(28.11)	(-1.60)
Subject matter controls	No	No	Yes
State fixed effects	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes
Ν	327	943	23629
R^2	0.339	N/A	N/A

 Table 3: Gender and Production, Citations, and Independence for State High Court

 Judges (with Judge Controls)

NOTES: *T* statistics in parentheses; +p < 0.10; *p < 0.05; **p < 0.01. The Independence model is estimated on judgelevel data with ordinary least squares with robust standard errors. The Production (judge-year-level data) and Citations (opinion-level data) models are estimated using a negative binomial regression. R^2 is unavailable with negative binomial regressions. Errors in the Production and Citations models are clustered by state judge. Subject matter controls include indicator variables for the following case subject matter areas: administrative, attorney and client, capital punishment, church and state, commercial, criminal, family, First Amendment, labor, property, rights, and torts (with "other" as the base category).

include variables relating to the judge's education, including the U.S. News ranking of the judge's law school measured in 2002 (US NEWS JD RANKING), and whether the judge went to an in-state law school (IN-STATE LAW SCHOOL). Table 3 reports the results of the models. We include in the Appendix a description of the sources for all our variables.

1. Publications

We estimate a negative binomial regression model for production, with the number of majority opinions per year as the dependent variable, adding judge controls to the model in Table 2 (errors clustered by judge). In the model, as reported in Table 3, FEMALE is insignificant. For all judges, whether the judge was the chief judge and whether the judge was close to retirement turn out to be relevant; both have a negative effect on publication rates. This is not surprising, as chief judges have additional responsibilities, while a judge who is close to retirement may be slowing down. None of the traditional human capital measures, such as prior employment or law or undergraduate school rankings, are significant. The years of court experience and the number of children a judge has, as well as whether the judge attended an in-state law school or has private practice experience, are also significant.

2. Citations

We estimate a negative binomial regression model for production, with the number of outside-state citations to majority opinions as the dependent variable, adding judge controls to the model in Table 2 (errors clustered by judge). In the model, as reported in Table 3, FEMALE is insignificant. Moreover, except for the number of children the judge has, none of the judge "control" variables are significant. The level of analysis here is the individual opinion, so the number of observations is much higher than in the Production model of Table 3; state, subject matter, and year controls are included.

3. Independence

In the Independence model of Table 3 with the addition of judge controls, we find that the coefficient on FEMALE remains positive and significant at the 1 percent level. The Independence model is estimated with ordinary least squares with robust standard errors. As with the Independence model in Table 2, the Independence model in Table 3 provides evidence in support of the view that female judges perform as well if not better than male judges. To summarize, the above three sets of findings are inconsistent with Hypotheses 1, 3, and 5. We find little support for the preference story, as almost none of the background variables are significant.

Overall, these findings suggest that women serving on state supreme courts are either able to overcome their lack of training, or that the job of being a state high court judge simply does not require skills learned in higher-ranked law schools and private practice. These results call into question the focus on traditional measures of human capital in predicting the performance of female (and male) judges.

C. Predictions of Differential Interests

Our next two hypotheses (2 and 4) draw on the idea that women might have different subject area interests than men and, therefore, might invest effort in law making in different areas than men. One possible criticism of our results is that women are on par with

	Number of Majority Opinions per Year—Men	Number of Majority Opinions per Year—Women	z Value	Female Significant in Full Model? ^a
Administrative	1.354	1.139	0.0729	No
Attorney	0.578	0.566	0.6102	No
Capital	0.738	0.629	0.1538	No
Church	0.006	0.000	0.7639	Yes, negative
Commercial	2.809	2.386	0.2071	No
Criminal	6.162	5.562	0.7115	Yes, negative
Family	1.417	1.457	0.9577	No
First Amendment	0.062	0.037	0.1495	Yes, negative
Labor	1.565	1.270	0.3524	Yes, negative
Property	1.156	1.015	0.7088	No
Rights	0.298	0.330	0.8983	No
Torts	2.296	2.097	0.7527	No
Other	0.405	0.296	0.1853	No
Total	18.846	16.783	0.3381	No

Table 4: Gender and Subject Matter Differences in Production for State High Court Judges

^aEach model used the log of the number of published majority opinions within a subject matter category for each judge as the dependent variable, with FEMALE, judge controls, and state and year fixed effects as independent variables. This column indicates whether the FEMALE gender variable is a significant predictor of the log of the number of published majority opinions within the specific subject matter category, and whether the variable has a positive or negative effect.

NOTES: Z values calculated using a Mann-Whitney test. The dependent variable in the Full model is the number of majority opinions in a particular subject matter area summed for the 1998 to 2000 period for each judge. Female, judge controls, and state and year fixed effects as independent variables (estimated with negative binomial regressions on judge-level data). We estimate the Full model using a negative binomial regression model with robust standard errors on judge-level data.

men only because they excel in certain traditionally female-focused areas of law (such as family law). Outside these areas, the preference story may still prevail. To examine this question, we examined publication and citation numbers as a function of specific subject areas.¹¹

Table 4 reports summary statistics on the number of majority opinions published per year categorized by gender and by subject matter. We borrow the subject areas from Epstein and Segal (2000) (see Appendix for definition of subject matter categories). We find a wide variety of significant differences with simple difference of means tests, although there are no clear patterns of gender-specific subject specialization. Generally, female judges publish fewer majority opinions in the administrative category of cases. This difference—or the lack of other differences—may be driven by underlying differences in caseloads across the different states and other factors. To control for this, we estimate a series of negative binomial regression models with robust standard errors on judge-level data using the number of published majority opinions within each specific subject matter category

¹¹For independence, because the measure is a function of cases where the judges openly disagreed in writing, the number of data points is relatively small as compared to the data on citations and publications. The result of having fewer data points on the independence variable is that it is not meaningful to break those data down by subject area.

summed for the 1998 to 2000 period for each judge as the dependent variable and include FEMALE, judge controls, and state and year fixed effects as independent variables (referred to in Table 4 as the "Full" model). Table 4 reports that in the series of Full models, female judges publish fewer majority opinions in the church, criminal, First Amendment, and labor categories. Based on these models, women do seem to publish less than men in several areas, but none of these are in "traditional" female-focused subject matter areas as predicted in Hypothesis 2. Moreover, there is no indication that women are publishing more cases in the family law area.

Turning to Hypothesis 4, we examine whether women are cited less or more in specific subject areas. As women may be seen as experts in areas relating to family law or gender-based rights, we expect that women will be cited more in these areas, but less in areas such as business law that are outside of women's stereotypical domain. Looking first at the number of outside state citations to majority cases published in each subject area, we find no differences in citation rates between men and women in any particular subject area (although women are cited more in the residual "other" category; significant at the 10 percent level).

We estimate a series of negative binomial regressions with robust standard errors on judge-level data with the sum of the number of outside-state citations to all majority opinions in a particular subject matter area written by a judge during the 1998 to 2000 period as the dependent variable. We include gender, judge controls, and state and year fixed effects as independent variables (referred to in Table 5 as the "Full" model). We find that FEMALE is not significant in any of these models in explaining the number of outsidestate citations for specific subject matter areas. Female judges are not cited significantly less than are their male counterparts in any subject area, suggesting that other judges view female judges' opinions as holding the same weight as their male counterparts' opinions.

V. Gender in the Federal Courts

To evaluate whether our results are unique to the state high courts, where there is tremendous variation in terms of court systems and state effects, we report data on the federal courts of appeals and district courts for roughly the same time periods (1998–2000 for the courts of appeal and 2001–2002 for the district courts).¹² Owing to constraints in the data sets, we are able to estimate gender comparisons for only a subset of the hypotheses. Further, because of the relatively small size of the appeals court data set, we were unable to use as many controls as we did with the state court data.

A. Appeals Courts

The data for the courts of appeals, collected for a prior project (Choi & Gulati 2008) have information for all the active federal circuit court judges during the period 1998 to 2000

¹²The time periods for the different data sets do not perfectly overlap because the federal court data were collected for different projects; see Choi et al. (2010) and Choi and Gulati (2008).

	Number of Outside-State Citations per Majority Opinion—Men	Number of Outside-State Citations per Majority Opinion—Women	z Value	Female Significant in Full Model?"
Administrative	0.452	0.488	0.5402	No
Attorney	0.707	0.736	0.7922	No
Capital	0.786	1.170	0.6452	No
Church	_	_		_
Commercial	0.983	1.133	0.5588	No
Criminal	0.662	0.716	0.9002	No
Family	0.625	0.939	0.7548	No
First Amendment	1.191	1.182	0.7967	No
Labor	0.436	0.478	0.7964	No
Property	0.455	0.536	0.4596	No
Rights	1.203	0.976	0.4870	No
Torts	0.954	1.056	0.1591	No
Other	0.471	0.662	0.0928	No
Total	0.708	0.814	0.0009	No

Table 5: Gender and Subject Matter Differences in Citation Rates for State High Court Judges

^aEach model used the number of outside-state citations for majority cases in each subject area as the dependent variable, with FEMALE, judge controls, and state and year fixed effects as independent variables. As with the publication table, this column indicates whether the FEMALE gender variable is a significant predictor of the level of citations from outside the state a case receives, and whether the variable has a positive or negative effect.

NOTES: Z values are calculated using a Mann-Whitney test. There were no majority opinions authored by a female judge in the church category. The dependent variable in the Full model is the sum of the number of outside-state citations to all majority opinions in a particular subject matter area written by a judge during the 1998 to 2000 period (estimated with negative binomial regressions on judge-level data). Female, judge controls, and state and year fixed effects as independent variables. We estimate the Full model using a negative binomial regression model with robust standard errors on judge-level data.

who had been on the bench at least two years and were under the age of 65 at the time. Data were collected for three measures similar to our measures of state judge quality: number of majority opinions published per judge (for the 1998 to 2000 time period), the average outside federal circuit citations to majority opinions per judge, and co-partisan disagreements, controlling for the political makeup of a particular circuit court (as a measure of independence).¹³ We estimate ordinary least square regression models with INDEPENDENCE and the log of 1 + the average outside federal circuit citations per judge-level data. We estimate a negative binomial regression with the number of majority opinions published per judge in the 1998 to 2000 time period as the dependent variable on judge-level data.¹⁴ In all our models we include FEMALE as an independent variable and controls for circuit effects since the circuits likely differ in both behavioral norms and caseloads.

¹³We did not have data on subject areas so as to be able to test whether there were gender differences in the types of cases the judges wrote opinions on or for which they received citations.

¹⁴The number of majority opinions published per circuit court judge in the 1998 to 2000 time period is count data, making the negative binomial regression model appropriate.

	Independence	Production (Number of Majority Opinions per Judge)	Citations (Log of 1 + Average Outside Federal Circuit Citations per Opinion)
Female	-0.0285	-0.048	0.00207
	(-0.76)	(-0.86)	(-1.76)
Constant	-0.0615	4.875**	1.800**
	(-0.54)	(89.09)	(26.62)
Circuit effects	Yes	Yes	Yes
Ν	98	98	98
R^2	0.144	N/A	0.403

Table 6: Federal Court of Appeals Judges

NOTES: T statistics in parentheses; +p < 0.10; *p < 0.05; **p < 0.01. The Independence and Citations models are estimated on judge-level data using ordinary least squares with robust standard errors. The Production (number of majority opinions) model is estimated on judge-level data using negative binomial regression with robust standard errors.

The coefficients for FEMALE in all the models of Table 6 are not significantly different from zero. We find that female circuit court judges do not have significant differences in independence, production (measured by the number of majority opinions), and citations from courts outside their federal circuit compared with male judges.

B. District Court

For the district courts, we used data for the 629 federal district judges who were active in the 2001–2002 period.¹⁵ Because these judges sit individually, we are unable to calculate independence scores in a fashion similar to the state high courts; instead, we focus on two dependent variables. We use the average publications per filing as our measure of production, assessed using district-judge-level data. We define the average publications per filing for a federal district court judge as the total number of published opinions for the judge divided by the average number of filings per judge in that judge's district (total filings for the district divided by number of judgeships in that district).¹⁶ We use as our measure of citations the number of outside positive citations per majority opinion assessed on opinion-level data. We track the number of positive outside-circuit federal court citations to

¹⁵We lack data on all our control variables for each judge, leading to fewer observations in our regression models in Table 7. For example, we have full data for only 533 district court judges in the publications per filings model in Table 7.

¹⁶By published opinions, we mean opinions that are available in the published reports issued by Westlaw. Although Westlaw can publish whatever opinions it wants to publish, anecdotal reports suggest that Westlaw simply publishes whatever opinions judges choose to designate as published opinions. In recent years, because of the widespread availability of judicial decisions in electronic databases, and particularly due to the passage of the E-Government Act, the distinction between published and unpublished opinions may have become less important. However, we suspect that the choice to send an opinion for inclusion in the print version is still an important one that reveals information about the case in question and the judge. That said, we constrain our database of opinions to roughly the period immediately prior to the passage of the E-Government Act in late 2002. See E-Government Act of 2002 (Pub. L. 107-347, 116 Stat. 2899, 44 U.S.C. § 101, H.R. 2458/S. 803) (enacted Dec. 17, 2002, with an effective date for most provisions of Apr. 17, 2003).

	Production: (Log of 1 + Average Publications per Filing)	Citations (Number of Outside Positive Citations per Majority Opinion)
Female	-0.000641	0.01794*
	(-0.18)	(3.13)
Constant	0.0204**	0.4563**
	(3.37)	(14.56)
District judge controls	Yes	Yes
N	533	12,173
R^2	0.064	N/A

Table 7: Federal District Court Judges

NOTES: *T* statistics in parentheses; +p < 0.10; *p < 0.05; **p < 0.01. District judge controls are defined in the text in Section V.B. The Production model is estimated on judge-level data using ordinary least squares with robust standard errors. The Citations model is estimated on opinion-level data using negative binomial regression. Errors in the Citations model are clustered by federal district judge.

a federal district judge's published majority opinions using Westlaw. As is common in the citation literature, we use outside-circuit citations rather than total citations (including in-circuit citations) because in-circuit citations might reflect intracircuit norms.¹⁷

We estimate an ordinary least square regression model with the log of 1 + average publications per filings as the dependent variable on district-judge-level data. We estimate a negative binomial regression with the number of outside positive citations per majority opinion as the dependent variable on opinion-level data.¹⁸ Our key independent variable in our regression models for district judges is FEMALE. We use the following for district judge controls: we include indicator variables for black judges (BLACK), and judges of other racial minority groups (OTHER RACE). Our experience variables include indicator variables for the judge's prior profession immediately before becoming a federal district court judge as follows: whether the judge worked as a judge, such as a magistrate judge, prior to becoming a federal district court judge (PRIOR JUDGE), the judge worked as a prosecutor (PRIOR PROSECUTOR), and the judge worked in private practice (PRIOR PRIVATE PRACTICE). To capture the salience of a judge's mix of cases, we develop a variable (SALIENT) by dividing the judge's number of salient published cases by the judge's total number of published cases. Salient cases are those involving church and state, campaign finance, federalism, First Amendment, and other constitutional rights (Choi & Gulati 2008, which relies on the methodology of Epstein & Segal 2000). For our political controls, we use an indicator variable for whether the judge was appointed by a Democratic president (JUDGE DEMO-CRAT) and a variable for the judge's experience in years defined as the difference between 2002 and the appointment year of the judge (JUDGE EXPERIENCE). We also include in our district judge controls an indicator variable for chief judge status during either 2001 or 2002 or both (CHIEF JUDGE) and an indicator variable for whether the judge attended one of the

¹⁷This number includes citations by state courts that are outside the circuit in question.

¹⁸The number of outside positive citations to a majority opinion is count data, making the negative binomial regression model appropriate.

	Production (Number of Majority Opinions)	Citations (Number of Outside Federal Circuit Citations to Majority Opinions)	
	2004–2006		
Calabresi	72	784	
Clement	81	240	
Garland	65	264	
Garza	112	255	
Jones	77	335	
Lynch	215	998	
McConnell	119	630	
McKeown	67	404	
Raggi	53	438	
Schroeder	60	120	
Sotomayor	90	706	
Wardlaw	51	207	
Wilkinson	88	537	
Williams	123	397	
Wood	156	831	

Table 8: Sotomayor Data

NOTES: Both the number of majority opinions and the number of outside federal circuit citations are measured for the 2004 to 2006 time period.

three top law schools as measured by U.S. News in 1992—Harvard, Yale, and Stanford which also were the three law schools most frequently represented among the circuit court judges in our sample (TOP SCHOOL).

We do not find significant gender differences in the average publications per filing, but we do find significant gender differences in the outside positive citations for published majority opinions, with women outperforming men (significant at the 5 percent level).¹⁹

C. Justice Sotomayor Versus the Others

As Justice Sotomayor's statements and the reactions they generated were the starting point for our project, we examined data on her as well. To estimate a meaningful comparison, we calculated the number of majority opinions (published) and the number of outside federal circuit citations to all majority opinions for the 2004 to 2006 period for then-Judge Sotomayor. As a control, we estimated similar measures for six court of appeals judges who were rumored either to have been on President Obama's short list or President Bush's short list. In addition, we also included two other Second Circuit judges who were active during the same period, Judges Calabresi and Raggi.

¹⁹It is possible that gender may work indirectly through judge characteristics (e.g., if a female judge is more or less likely to become a chief judge). Including district judge controls may therefore understate the impact of gender. As a robustness test, we removed the district judge controls and substituted district court effects. Unreported, we obtained the same qualitative results for both models of Table 7.

The comparisons here are necessarily rough because there are not enough judges to control for factors such as circuit effects. That said, then-Judge Sotomayor's outside-circuit citation scores are among the highest of any of the judges on either president's short list (Posner 2009; cf. Anderson 2009); so are the scores of Diane Wood (who was among the leading candidates for selection to the Court to replace Justice Stevens (Bazelon & Lithwick 2010)).

VI. CONCLUSION

We find little to no support for the preference story's predictions that female judges would underperform male judges (Hypotheses 1, 3, and 5). Indeed, the prediction that women will underperform men in terms of independence scores was false. Women were more independent than men (directly contradicting Hypothesis 5), supporting the screening story. We also find that the equivalent performance of women and men judges is not driven by any specific subject matter area effects (refuting Hypotheses 2 and 4). Women judges do not perform well because of outsized performance in traditionally women-focused subjects.

Perhaps our most striking finding is that the premise of the preference story is true (female judges have weaker credentials and less experience) but its conclusion is false (female judges and male judges perform about the same). What might account for this outcome? First, the measures of credential and experience might be inaccurate. We have been told by some female judges that they went to lower-rank law schools in order to accommodate their husbands but did very well at those schools. Our measures do not capture this phenomenon. It might also be the case that the rank of the law school, a few extra years of practice, and so forth make little difference for the quality of judging. Second, the measures of judicial performance might be inaccurate. As we noted before, our objective measures of performance might not capture high-quality judicial performance. If so, we have a "garbage-in, garbage-out" problem. Third, it is possible, as Justice Sotomayor suggested before backtracking, that women are naturally more gifted judges. The various psychological differences between men and women might favor women, so that even if women have less training and experience, they end up being superior judges. It might also be the case that women's experiences in a gender-biased world give female judges a distinctive perspective that enhances their judicial talents.

A couple of points regarding gaps in our analysis are in order. Although we frame the threshold question in terms of the value of gender diversity, we only arrive at that question indirectly. Judges on the state high courts always sit in teams; an estimation of the value of gender diversity should compare the performance of gender-diverse teams versus those of homogenous teams. There are also likely intergenerational differences embedded within the reported gender differences. The performance predictions for the female judges who attended law school in the late 1960s and early 1970s may be different compared with those who attended law school one decade later, in the early 1980s, and yet different again for those who were in school in the early 1990s. Our data set was not large enough to make these comparisons.

References

- Allen, David, & Diane Wall (1993) "Role Orientations and Women State Supreme Court Justices," 77 Judicature 156.
- Anderson, Robert (2009) Distinguishing Judges: An Empirical Ranking of Judicial Quality in the U.S. Court of Appeals. Available at http://srn.com/abstract=1433442>.
- Baker, Scott, Adam Feibelman, & William P. Marshall (2009) "The Continuing Search for a Meaningful Model of Judicial Rankings and Why it (Unfortunately) Matters," 58 *Duke Law J.* 1645.
- Banks, Taunya Lovell (1990) "Gender Bias in the Classroom," 14 Southern Illinois Law Rev. 527.
- Barteau, Betty (1997) "Thirty Years of the Journey of Indiana's Female Judges 1964–1994," 30 Indiana Law Rev. 43.
- Bazelon, Emily (2009) "The Place of Women on the Court," July 17 New York Times Magazine.
- Bazelon, Emily, & Dahlia Lithwick (2010) "Who Should Replace Justice Stevens?" April 10 Slate. Available at http://www.slate.com/id/2250251>.
- Beiner, Theresa (1999) "What Diversity on the Bench Could Mean for Justice," 6 Michigan J. of Gender & Law 113.
- Boyd, Christina L., Lee Epstein, & Andrew D. Martin (forthcoming) "Untangling the Causal Effects of Sex on Judging," *American J. of Political Science*.
- Brace, Paul, & Melinda Gann Hall (2005) "Is Judicial Federalism Essential to Democracy? State Courts in the Federal System," in K. L. Hall & K. T. McGuire, eds., *The Judicial Branch*. New York: Oxford Univ. Press.
- Brace, Paul, Laura Langer, & Melinda Gann Hall (2000) "Measuring the Preferences of State Supreme Court Judges," 62 J. of Politics 387.
- Buchanan, Pat (2009) "Miss Affirmative Action, 2009," June 6 HumanEvents.com. Available at http://www.humanevents.com/article.php?id=32264>.
- Carbado, Devon, & Mitu Gulati (2000) "Conversations at Work," 79 Oregon Law Rev. 103.
- Catalyst Report (2009) Women in Law in the United States. Available at http://www.catalyst.org/ublication/246/women-in-law-in-the-us.
- Choi, Stephen J., & Mitu Gulati (2008) "Bias in Judicial Citations: A New Window into the Behavior of Judges," 37 J. of Legal Studies 87.
- Choi, Stephen J., Mitu Gulati, and Eric A. Posner (2009a) "Are Judges Overpaid?" 1 J. of Legal Analysis 47.
- (2009b) "Judicial Evaluation and Information Forcing: Ranking the State High Courts and Their Judges," 58 Duke Law J. 1313.
- ----- (2010) "Judicial Ability and Output in Securities Class Actions," Univ. of Chicago Working Paper.
- (forthcoming) "Professionals or Politicians?" J. of Law, Economics & Organization. Available at http://jleo.oxfordjournals.org/cgi/content/abstract/ewn023v1>.
- Corrigan, Tracy (2009) "Women Should be Wary of Financial Flattery," May 16 Daily Telegraph. Available at http://www.telegraph.co.uk/finance/comment/tracycorrigan/5002408/Women-should-be-wary-of-financial-flattery.html>.
- Croson, Rachel, & Uri Gneezy (2008) "Gender Differences in Competition: The Role of Socialization," 47 J. of Economic Literature 448.
- Cross, Frank, & Stefanie Lindquist (2009) "Judging the Judges," 58 Duke Law J. 1383.
- Curriden, Mark (1995) "Judicial Barriers Quickly Breaking Down," 81 ABA J. 24.
- Dau-Schmidt, Kenneth Glenn, Marc S. Galanter, Kaushik Mukhopadhaya, & Kathleen E. Hull (2007) "Gender and the Legal Profession: The Michigan Alumni Data Set 1967–2000," *Indiana Legal Studies Research Paper* 104.
- Davis, Sue (1993) "The Voice of Sandra Day O'Connor," 77 Judicature 134.
- Dickerson, John (2009) "More Better Judging," June 3 *Slate.com*. Available at ">http://www.slate.com/id/2219699/>.
- Dumas, Tao (2010) "Gender and State Supreme Courts: Explaining Male and Female Judges' Concurring and Dissenting Behavior," paper presented at the Western Political Science Association 2010 Annual Meeting. Available at http://ssrn.com/abstract=1580487>.

Epstein, Cynthia (1993) Women in the Law. Urbana-Champaign, IL: Univ. of Illinois Press.

Epstein, Lee, & Jeffrey A. Segal (2000) "Measuring Issue Salience," 44 American J. of Political Science 66.

- Gandy, Evelyn, Deanne Mosley, & Amy Whitten (2003) "Supreme Court of Mississippi Task Force on Gender Fairness," *Mississippi Lawyer* 23.
- Garth, Bryant, & Joyce Sterling (2009) "Exploring Inequality in the Corporate Law Firm Apprenticeship: Doing the Time, Finding the Love," 22 Georgetown J. of Legal Ethics 1361.
- Gneezy, Uri, Muriel Niedele, & Aldo Rustichini (2003) "Performance in Competitive Environments: Gender Differences," 118 Q. J. of Economics 1049.
- Guinier, Lani, M. Fine, J. Balin, A. Bartow, & D. Stachel (1994) "Becoming Gentlemen: Women's Experience at One Ivy League Law School," 143 Univ. of Pennsylvania Law Rev. 1.
- Haddon, Phoebe (2008) "Final Report of the Pennsylvania Supreme Court Committee on Racial and Gender Bias in the Justice System. Report on Race and Gender Bias in the Justice System, 2004," SNO31 ALI-ABA 489.
- Hull, Kathleen, & Robert Nelson (2000) "Assimilation, Choice, or Constraint—Testing Theories of Gender Differences in the Careers of Lawyers," 79 Social Forces 229.
- Hurwitz, Mark, & Drew Noble Lanier (2008) "Diversity in State and Federal Appellate Courts: Change and Continuity Across 20 Years," 29 Justice System J. 47.
- Jackson, Vicki C. (1997) "What Judges Can Learn from Gender Bias Task Force Studies," 81 *Judicature* 21.
- Kearney, Richard, & Holly Sellers (1996) "Sex on the Docket: Reports of State Task Forces on Gender Bias," 56 Public Administration Rev. 587.
- Ku, Manwai C. (2008) Similar or Different? Women's and Men's Aspirations in Law Schools. Available at http://www.stanford.edu/group/scspi/pdfs/rc28/conference_2008/p431.pdf>.
- Landes, William M., Lawrence Lessig, & Michael E. Solimine (1998) "Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges," 27 J. of Legal Studies 271.
- Leong, Nancy (2009) A Noteworthy Absence. Available at http://ssrn.com/abstract=1345165>.
- Levin, I. P., M. A. Snyder, & D. P. Chapman (1988) "The Interaction of Experiential and Situational Factors and Gender in a Simulated Risky Decisionmaking Task," 122 J. of Psychology 173.
- Lithwick, Dahlia (2009) "White Men Can't Judge?" June 11 Slate.com. Available at http://www.slate.com/id/2220225/>.
- Martin, Elaine, & Barry Pyle (2000) "Gender, Race, and Partisanship on the Michigan Supreme Court," 63 Albany Law Rev. 1205.
- (2005) "State High Courts and Divorce: The Impact of Judicial Gender," 36 Univ. of Toledo Law Rev. 923.
- McCall, Madhavi (2008) "Structuring Gender's Impact," 36 American Politics Research 264.
- McCall, Madhavi, & Michael McCall (2007) "How Far Does the Gender Gap Extend? Decision Making on State Supreme Courts in Fourth Amendment Cases. 1980–2000," 44 Social Science J. 67.
- Mertz, Elizabeth (2007) The Language of Law School: Learning to "Think" Like a Lawyer. London: Oxford Univ. Press.
- Mertz, Elizabeth, Wamucii Njogu, & Susan Gooding (1998) "What Difference Does it Make? The Challenge for Legal Education," 48 J. of Legal Education 1.
- NALP Foundation (2004) After the JD: First Study of a National Study of Legal Careers. Available at http://www.nalpfoundation.org/webmodules/articles/articlefiles/87-After_ID_2004_web.pdf>.
- Palmer, B. (2001) "'To Do Justly?' The Integration of Women into the American Judiciary," 34 PS: Political Science & Politics 235.
- Peresie, Jennifer L. (2005) "Female Judges Matter: Gender and Collegial Decision Making in the Federal Appellate Courts," 114 Yale Law J. 1759.
- Posner, Eric (2009) More Data and a New Conclusion. Available at http://www.volokh.com/posts/1243482653.shtml>.
- Posner, Richard A. (2005) "Judicial Behavior and Performance: An Economic Approach," 32 Florida State Univ. Law Rev. 1259.

(2008) How Judges Think. Cambridge, MA: Harvard Univ. Press.

- Powell, M., & D. Ansic (1997) "Gender Differences in Risk Behavior in Financial Decisionmaking: An Experimental Analysis," 18 J. of Economic Psychology 605.
- Resnik, Judith (1996) "Asking About Gender in Courts," 21 Signs 952.
- Savage, Charlie (2009) "Wider World of Choices to Fill Souter's Vacancy," May 1 N.Y. Times A1.
- Schafran, Lynn Hecht (2005) "Not from Central Casting: The Amazing Rise of Women in the American Judiciary," 36 Univ. of Toledo Law Rev. 953.
- Shapiro, Illya (2009) "Sotomayor Pick Not Based on Merit," May 27 CNN.com. Available at http://www.cnn.com/2009/POLITICS/05/27/shapiro.scotus.identity/index.html>.
- Sherry, Suzanna (1986) "Civic Virtue and the Feminine Voice in Constitutional Adjudication," 72 Virginia Law Rev. 543.
- Songer, Donald R., & Kelley A. Crews-Meyer (2000) "Does Judge Gender Matter? Decision Making in the State Supreme Courts," 81 Social Science Q. 750.
- Songer, Donald R., Sue Davis, & Susan Haire (1994) "A Reappraisal of Diversification in the Federal Courts: Gender Effects in the Courts of Appeals," 56 *J. of Politics* 425.
- Stuhlmacher, A. F., & A. E. Walters (1999) "Gender Differences in Negotiation Outcome: A Meta-Analysis," 52 Personnel Psychology 653.
- Vladeck, David (2005) "Keeping Score: The Utility of Empirical Measurements in Judicial Selection," 32 Florida State Univ. Law Rev. 1415.
- Who's Who (2007) Biography Resource Center Online. Available at <www.marquiswhoswho.com/>.
- Wilkins, David, & Mitu Gulati (1996) "Why Are There So Few Black Lawyers in Corporate Law Firms," 84 California Law Rev. 493.
- (1998) "Reconceiving the Tournament of Lawyers," 84 Virginia Law Rev. 1582.
- Wilson, Robin Fretwell (2008) "Keeping Women in Business (and Family)," Washington & Lee Legal Studies Paper 2008–34. Available at http://ssrn.com/abstract=1115468>.
- Wood, Robert G., Mary E. Corcoran, & Paul N. Courant (1993) "Pay Differences Among the Highly Paid: The Male-Female Earnings Gap in Lawyers' Salaries," 11 J. of Labor Economics 417.

Appendix A: State High Court Judge Variable Definitions

Variable	Definition	Origin of Data
Number of majority opinions per year	Total number of majority opinions authored by a particular judge in one year (ranging from 1998 to 2000).	Westlaw
Number of outside-state citations per majority opinion	Total number of citations to a majority opinion from (1) federal courts outside the circuit that includes the state in question and (2) courts in other states. Citations are measured for opinions authored up until January 1, 2007 (as tracked in the LEXIS Shepard's database).	LEXIS Shepard's database
Same_Party	The total number of opposing opinions written against a same-party judge divided by the total number of opposing opinions written against either a judge of the opposite or same party as the state high court judge in question for the 1998 to 2000 time period. Opposing opinions include dissents written against a majority opinion and majority opinions where a dissenting opinion exists.	Westlaw; NEXIS; Internet (including Google Searches); Opensecrets.org
Same_Pool	Total number of majority opinions written by the state high court judges of the same political party (from the perspective of the judge in question) divided by the total number of majority opinions written by judges of both the same and opposite parties from 1998 to 2000.	Westlaw; NEXIS; Internet (including Google Searches); Opensecrets.org
Independence	Defined as Same_Party minus Same_Pool. A more negative independence score indicates an increased tendency to write an opposing opinion against an opposite-party judge. Conversely, a more positive independence score indicates a decreased tendency to write an opposing opinion against an opposite-party judge.	
Female	Indicator variable equal to 1 if the judge in question is female and 0 if the judge is male.	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)
Chief judge	For year-level data, indicator variable equal to 1 if the judge in question is the chief judge of the court in the year in question and 0 otherwise. For pooled data, indicator variable equal to 1 if the judge in question is the chief judge of the court for any year from 1998 to 2000 and 0 otherwise.	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)
Court experience	For year-level data, the difference between the year in question and the year the judge first joined the high court. For pooled data, the difference between 1998 and the year the judge first joined the high court (if the judge started on the court in 1998 or later, the court experience is set to 0).	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)

Variable	Definition	Origin of Data
Post-law-school experience	The difference between 1998 and the year the judge graduated law school.	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)
Retirement close	Indicator variable equal to 1 if the judge in question retired from the bench in 2001 or earlier and 0 otherwise.	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)
Age	For year-level data, the difference between the year in question and the judge's birth year. For pooled data, the difference between 1998 and the judge's birth year.	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)
Married	Indicator variable equal to 1 if the judge is married as of the year 2000 and 0 otherwise.	Who's Who (2007)
Number of children	The number of children a judge had as of the year 2000.	Who's Who (2007)
Divorced	Indicator variable equal to 1 if the judge is divorced as of the year 2000 and 0 otherwise.	Who's Who (2007)
Private practice	Indicator variable equal to 1 if the judge had private practice experience before becoming a judge and 0 otherwise.	Who's Who (2007)
PAJID score	PAJID score for each judge as developed by Brace et al. (2000). These scores locate judges on a political continuum from highly conservative (0) to highly liberal (100).	Brace et al. (2000)
US News JD and BA rankings	 (JD) The U.S. News rankings of the judge's law school measured as of 2002. (BA) The U.S. News rankings of the judge's undergraduate institution measured as of 2002.²⁰ 	U.S. News and World Report 2002 Edition
In-state law school	Indicator variable equal to 1 if the judge is went to an in-state law school and 0 otherwise.	Website for the state highest court; NEXIS; Internet searches; Who's Who (2007)

APPENDIX A Continued

²⁰To compare undergraduate programs across USNWR's categories (national, liberal arts, masters, and baccalaureate colleges), we assigned weights to each category and tier. National universities were given their actual weight (i.e., if someone attended Stanford, he or she received a ranking of 4), liberal arts colleges were given their rank plus 50 points, and baccalaureate colleges and masters were given their rank plus 100 points. Third- and fourth-tier schools were given an even ranking within each category. Third-tier national schools were given a ranking of 200. Third-tier liberal arts schools were ranked at 200, while the fourth tier received a ranking of 250. Third-tier masters and baccalaureate schools received a ranking of 300, and the fourth tier received a ranking of 350. Unranked schools were given a ranking of 400.

Appendix B: Subject Matter Categories for State Judge Opinions

Variable	Definition
Administrative	Review of agency/government decision making (not in another subject matter category). Also includes government actions (e.g., state suit to comply with state statute that does not fit in other categories); private actions suing state actors for negligence, etc.
Attorney and client	Attorney misconduct; attorney fees (unless fits in one of above categories); disbarment; contempt of court order against attorney.
Capital punishment	Capital-punishment-related actions.
Church and state	Pledge of Allegiance; funding for private religious schools; prayer in school; Ten Commandments.
Commercial	Contracts; insurance; private arbitration; creditor versus debtor; lessor-lessee; usury laws; franchise versus franchisor; employment contractual disputes; corporate law; piercing the corporate veil; tax; bankruptcy; enforcement of mechanic's lien; implied warrant of merchantability.
Criminal	Sentencing guidelines; prisoners' rights; murder; rape; drugs/controlled substances; attorney-client privilege in criminal context; grand jury related; juvenile criminals. Excludes capital punishment cases.
Family	Divorce; adoption; child support; probate/inheritance.
First Amendment	Employment issues (excluding employment contractual disputes); ERISA; National Labor Relations Board (NLRB); Occupational Safety and Health Act (OSHA); Fair Labor Standards Act (FLSA); wrongful discharge; Labor Management Relations Act (LMRA); Family and Medical Leave Act (FMLA); employee benefits; workers' compensation claims; retaliatory discharge claims.
Labor	Employment issues (excluding (1) employment contractual disputes that are not workers' comp or state administrative wage rate related (commercial) and (2) excluding discrimination-type claims (civil rights)); ERISA; NLRB; Occupational Safety and Health Act (OSHA); Fair Labor Standards Act (FLSA); wrongful discharge; Labor Management Relations Act (LMRA); Family and Medical Leave Act (FMLA); employee benefits; workers' compensation claims; retaliatory discharge claims; state wage rate claims.
Property	Takings claims; zoning issues; property rights; property licensing related or permit related; landlord-tenant related.
Rights	Race discrimination; sex discrimination; affirmative action; civil rights; age discrimination; privacy; handicap discrimination; abortion (includes discrimination in employment context cases); voting rights-voting related.
Torts	Federal Tort Related Act; medical malpractice; products liability; wrongful death; libel; etc.
Other	All other cases.