Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Nancy Bays Arrington	Date

Inference, Accountability, and Gender Diversity on the Bench

By

Nancy Bays Arrington
Doctor of Philosophy

Political Science

Dr. Jeffrey Staton, Ph.D
Advisor

Dr. Adam Glynn, Ph.D.
Committee Member

Dr. Beth Reingold, Ph.D
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D
Dean of the James T. Laney School of Graduate Studies

Date

Inference, Accountability, and Gender Diversity on the Bench

By

Nancy Bays Arrington
MIP, UGA, 2012
BA, UT Austin, 2009

Advisor: Dr. Jeffrey Staton, Ph.D

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Political Science
2018

Abstract

Inference, Accountability, and Gender Diversity on the Bench

By Nancy Bays Arrington

This dissertation project presents a framework for explaining judicial diversity that links selection institutions to diversity outcomes through a process of citizen learning and accountability. For citizens to hold elites accountable for homogeneous selections to the bench, citizens must be able to (1) make accurate inferences about whether or not bias is actually occurring, (2) accurately attribute blame for perceived bias, and (3) hold those responsible for bias accountable. Importantly, judicial selection institutions shape each of these three steps: prior beliefs in the fairness of institutions will shape how observers interpret information; institutions that affect the size of the court and turnover shape how much data citizens have with which they can update their beliefs; the presence of multiple actors/steps in selection obscures blame attribution; finally, institutions that shelter selectors from sanctions undermine accountability for diversity. I test implications of this framework using evidence from survey experiments, observational data on state supreme courts in the United States, and observational data on peak courts cross-nationally.

Inference, Accountability, and Gender Diversity on the Bench

By

Nancy Bays Arrington
MIP, UGA, 2012
BA, UT Austin, 2009

Advisor: Dr. Jeffrey Staton, Ph.D

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Political Science
2018

Acknowledgement

Many thanks to the Jeff Staton, Beth Reingold, and Adam Glynn for their patient and generous mentorship. Jeff, you've set a standard as an advisor that I hope I can one day replicate. Thank you to Kirsten Widner, Anna Gunderson, Bethany Morrison, Lauren Mattioli, Jonathan Nash and the other members of the Law and Courts working group for your helpful and constructive criticisms of this project. Thank you to the members of the FCJ working group for your outpouring of both practical and emotional support; I'm proud I can call so many of my colleagues "wonderful friends" as well. Thank you to Jeff Staton, Leeann Bass, the many research assistants, and the Varieties of Democracy project for the Judicial Selection and Gender Diversity data collection project. Thank you to Dani Villa, Ben Arrington, Helen Arrington, and Robert Arrington for your proof-reading prowess. And, of course, shout out to the whole Arrington lot. I'm happy and lucky to be related to you all, and I'm endlessly grateful that I was able to live with and around so many of you during the last six years. I love you!

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

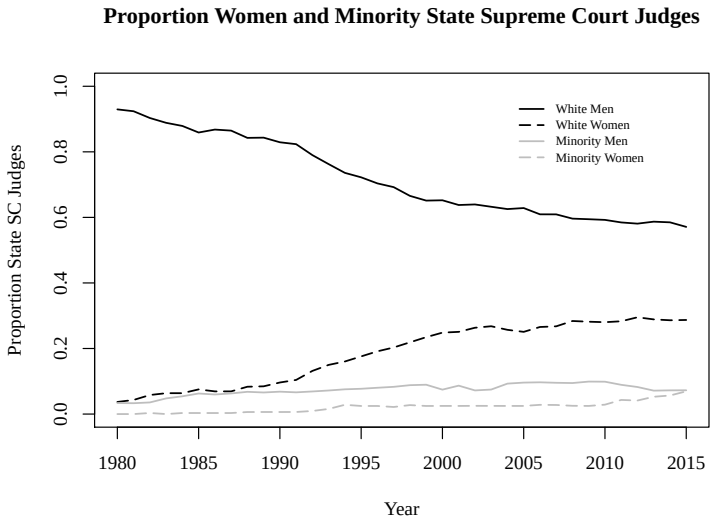
The exclusion of certain classes of people and their lived experiences from public office limits the scope of political discourse and sends a signal that the political process is only available to certain segments of the population. Yet for many years, women and non-white people have been excluded or overlooked for political office. In recent decades, grassroots organizations and high political officials have called for the diversification of office, including the diversification of judicial benches. Progress has been made.[1] Figure 1, for example, shows the increase in the presence of women and minority judges on U.S. state supreme courts over time. Despite improvements, however, most courts still do not descriptively reflect the populations they judge. Persistent disparity in prestigious offices suggests that selection processes for such posts disproportionately favors certain types or groups of people over others (Milyo and Schosberg, 2000; Lawless, 2004; Fox and Oxley, 2003; Richard L. Fox, 2004).

Normative political theorists emphasize the relationship between descriptive representation, substantive representation, and symbolic representation (Pitkin, 1967; Phillips, 1995) and argue that full representation requires more than just the representation of ideas or policy preferences. "Full" representation requires the inclusion of people in office with different experiences and backgrounds in a way that reflects the experiences and backgrounds of the population.[2] Only by having diverse seg-

[1] For example, in an interview for the New Yorker, President Obama said, "I think there are some particular groups that historically have been underrepresented – like Latinos and Asian-Americans – that represent a large and larger portion of the population. And so for them to be able to see folks in robes that look like them is going to be important." (Toobin, 2014). In the same quotation President Obama references the selection of openly gay judges as well, a descriptive category not addressed here: "When I came into office, I think there was one openly gay judge who had been appointed. We've appointed ten."

[2] In this paper I focus on the representation of gender and, to a lesser extent, race. There are many other characteristics along which individuals vary including economic, geographic, and professional background, for example. See Cahill-O'Callaghan (2015) for a discussion of diversity of values on the bench.

Figure 1.1: Diversity on U.S. State Supreme Courts



The proportion of judges selected to state supreme courts over time who are minority women, minority men, white women, and white men. Most judges are white men, and much of the diversity on the bench gained in recent decades is due to the increased presence of white women.

ments of society present in office, proponents argue, can policy discourse reflect the varied preferences of the constituency. As such, women officials may be better able to represent women due to their shared experiences as women, and African American officials may be more aware of and responsive to issues of particular importance to African American constituents. Moreover, the presence of women and minorities in the political process can increase perceptions of legitimacy (Kenney, 2012) and increase the likelihood that historically under-represented constituents will be engaged in and participate in the political process (Griffin and Keane, 2006; Reingold and Harrell, 2010; Atkeson, 2003)

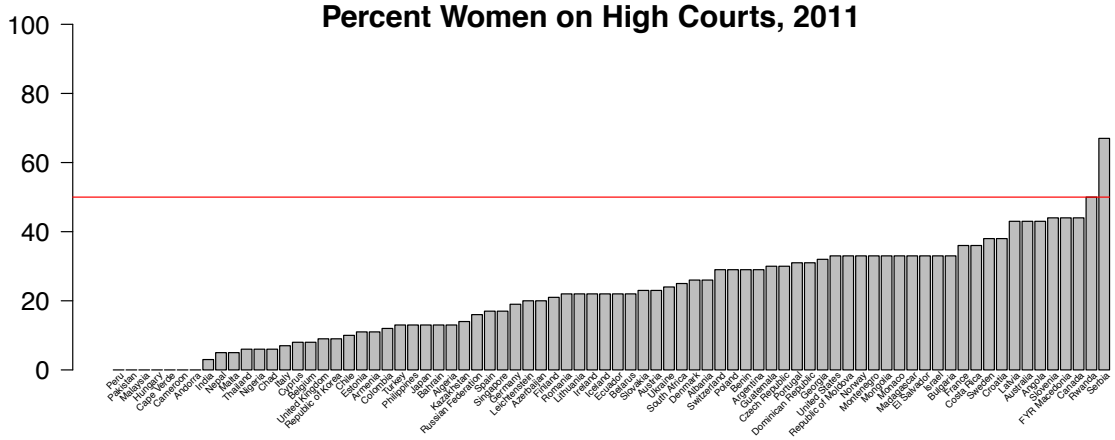
In the judicial context, empirical evidence suggests that a diverse bench can alter the attitudes of litigants, civilians, and judges themselves. Scherer and Curry (2010), for example, find that descriptive representation of Black judges on U.S. federal courts increases perceptions of legitimacy among African Americans. Farhang and Wawro (2004) find that the presence of a woman judge on a three judge U.S appellate panel

"influences the behavior of male judges" (p 321); male judges voted more liberally in anti-discrimination cases when there was a woman serving with them on the panel. Similarly, Boyd, Epstein and Martin (2010) find that male appellate judges are more likely to decide in favor of plaintiffs in sex-discrimination cases when a woman is serving on a panel alongside them. Finally, there is some evidence that for certain legal issues women and minority judges make different judgments than their white-male counterparts (Collins, Manning and Carp, 2010).

These racial and gender differences are particularly important in the context of courts because citizens are more likely to come into contact with and have their lives directly affected by representatives of the judiciary than other representatives. Ifill (1998) writes, "judges have a more direct and irrevocable impact in the lives of many Americans than local or even national legislators. This is particularly true for African Americans, who are disproportionately involved with the judicial system" (p 407-408). Moreover, the idea that the life and group experiences of justices shape how they perform their duties is not limited to academic debate. Justice Sotomayor, prior to her appointment to the Supreme Court, said, "Whether born from experience of inherent physiological or cultural differences... our gender and national origins may and will make a difference in our judging" (Sotomayor, 2002, p.. 92).

Granted, some scholars claim that the gender differences in judging have been overstated (Dixon, 2009; Kenney, 2008), and others emphasize the importance of gender diversity on the bench *absent* any gender differences or changes in outcomes. Kenney (2013) argues that the inclusion and continued presence of women on judicial benches is important in its own right because it "normalizes women's authority and power" and demonstrates judicial legitimacy (p. 9, 175). Regardless of the extent to which diversity on the bench alters outcomes, persistent disparity on the bench requires theoretical and empirical inquiry.

Figure 1.2: Gender Composition of Constitutional Courts Cross-Nationally



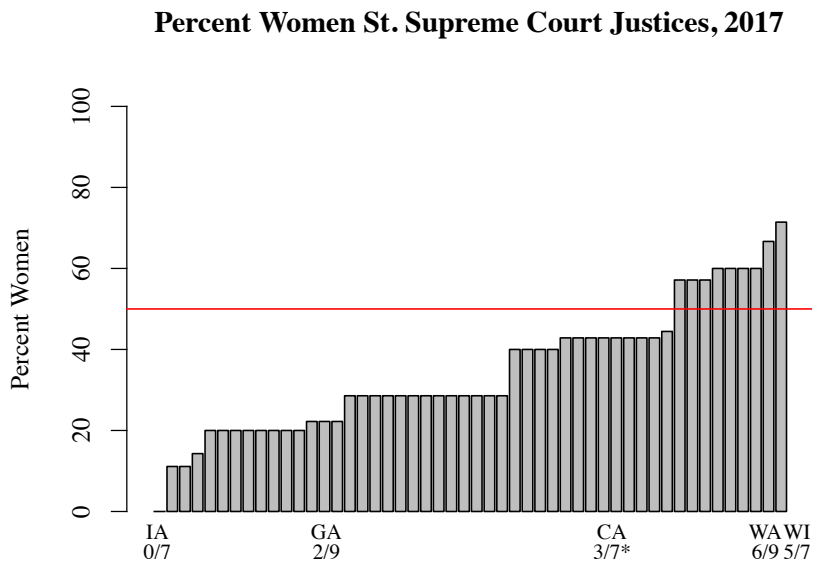
The percent of high court justices who are women cross-nationally. Data from Turguet et al 2012.

1.1 Variation in Diversity

Despite the benefits of diverse courts, there is substantial variation in diversification over time and across place. Figure 1.2 shows the gender composition of peak courts cross-nationally and Figure 1.3 shows gender diversity on state supreme courts in the U.S. There is substantial variation across space in both the U.S. and cross nationally. Several courts have *no* women justices (Peru, Pakistan, Malaysia, Hungary, Cape Verde, Cameroon, and Andorra. In the U.S., Iowa.), while some courts have reached gender parity (Serbia and Rwanda. In the U.S., Washington and Wisconsin, among others). Why does disparity persist in some contexts but not others?

There are several potential explanations for the variation in diversity depicted in Figure 1.2 and 1.3. For one, disparity could depend entirely on the absence of qualified women for particular positions. If disparity stems from generational gaps and the time required for women to move into fields that feed into high courts, then the "solution" is patience and the encouragement of women to pursue qualifications. Justice Sumption of the Supreme Court of the United Kingdom has taken this stance. Speaking about the diversification of the UK Supreme Court, he has said, "These things simply cant be transformed overnight, not without appalling consequence in

Figure 1.3: Gender Composition of U.S. State Supreme Courts, 2017



The percent of state supreme court justices who are women by state. Note, as of Jan 2017, California has a current vacancy, so there are 3 women out of 6 serving judges

other directions.... we have got to be very careful not to do things at a speed which will make male candidates feel that the cards are stacked against them. If we do that, we will find that male candidates don't apply in the right numbers" (Walker, 2015).

If disparity simply reflects the absence of women or minorities in candidate pools, the selection processes that converts candidates to justices may be perfectly "fair" in the sense that the selection process does not favor certain types of qualified judicial candidates over other types of qualified candidates.[3] In this case, cross-national variation would stem from variation in the number of women qualified for high judicial office.

However, evidence from U.S. legislative scholarship suggests that women may need to be *more qualified* than their male counterparts in order to perceive themselves as qualified and to be recruited for office (Fox and Lawless, 2010; Lawless and Fox, 2010; Richard L. Fox, 2004). Similarly, women incumbents in the U.S. House of Represen-

[3]In this case bias could still occur in the process of one *becoming* a qualified candidate.

tatives are more likely to face challengers than their male peers (Milyo and Schosberg, 2000). Taken together, this evidence suggests that qualified women candidates and incumbents are viewed and treated differently than men.

If differential treatment in the recruitment and retention process exists in judicial office as it does in legislative office, then variation in the level of diversity and in the persistence of disparity on high courts may be more complicated than a simple reflection of variation in candidate pools for high judicial office. Indeed, expectations for successful selection to the bench will shape the extent to which women and minority individuals pursue the necessary qualifications for office (see chapter 5 for a discussion of candidate pools).

This goal of this project is to understand and explain variation in diversity and in the persistence of bias on high courts as a function of institutional selection procedures. As I will demonstrate, three conditions must be met for courts to diversify. First, observers must have enough information about the composition of courts to infer if bias is occurring. Second, observers must be able to accurately identify who is responsible for bias, and third, they must be able to exert pressure on those responsible. Institutional features of the selection process mediate each of these steps. First, institutional features of the court determine the rate at which citizens accumulate information and are able to make inferences about bias on the court. Second, the complexity of the judicial selection process shapes the ability of citizens to accurately identify who or what is responsible for bias. Finally, the institutional configuration of judicial selectors determines how, when, and to what extent citizens can hold selectors accountable for bias. As I will show, several institutional features of judicial selection complicate the process in ways that make diversification *less* likely. Taken together, satisfying the three conditions presents a significant hurdle for diversification, and the failure to satisfy any condition allows for the persistence of bias.

1.1.1 Institutions and the Selection of Women to High Courts

Myriad social and political phenomena have been associated with diversity in office. Some explanations are cultural, and focus on cross national variation in attitudes towards gender roles (Norris, 1987; Siaroff, 2000; Inglehart and Norris, 2003). Others emphasize how political socialization affects the willingness of women to be engaged in, knowledgeable about, and willing to participate in political office (Burns, Schlozman and Verba, 2001; Verba, Burns and Schlozman, 1997; Chhibber, 2002; Richard L. Fox, 2004; Lawless and Fox, 2010).

Some scholars of diversity in the judiciary focus on diffusion of norms across space and institutions. Hoekstra, Kittilson and Bond (2014), for example, find that countries with higher percentages of women in legislatures have greater gender diversity on high courts, and Williams and Thames (2008) find that countries with legislative quotas tend to have a greater presence of women on the bench. The presence of women in the legislative branch, these scholars argue, helps shift norms about women in office and serves as a signal to judicial selectors and appointers that women's presence in office is important.

Other explanations focus on the gendered nature of courts: Williams and Thames (2008) note that less prestigious courts are more likely to be diverse than highly prestigious courts. Using the cases of the United States and France, Remiche (2015) argues that courts in which judges are expected to wield power are more likely to exclude women than courts in which judges are expected to mechanically apply law. Similarly, Schultz and Shaw (2013) argue that civil law systems have more women judges than common law systems in part because judicial qualifications are more transparent in civil law systems (i.e., examination results versus professional achievement and visibility).[4]

[4]The authors also note that there are more judges in civil law systems and that judges begin their careers earlier in civil law systems as well. These features might also explain some of the difference in gender diversity.

This project contributes to a growing literature that focuses on the effects of *selection institutions* on political diversity. Institutional explanations are a useful point of investigation for several reasons. First, extant literature suggests that appointment institutions are important factors shaping diversification. In the legislative context, there is a persistent relationship between proportional representation electoral systems and higher levels of gender parity in legislative office (Paxton, 1997; Kenworthy and Malami, 1999; Salmond, 2006; Matland and Brown, 1992; Rule, 1987) both because of how voters respond to party lists and how parties respond to each other in multi-party systems (Kittilson, 2006; Lovenduski and Norris, 1993; Matland and Studlar, 1996; Murray, 2010). In the judicial context, scholars hypothesize that variation in how judges are selected by presidents or governors, elections, council, or merit may shape the gender and racial diversity of courts due to variation in how different selectors view qualifications and seek to gain electoral advantage through diversifying (Williams and Thames, 2008; Carbon, Houlden and Berkson, 1982; Bratton and Spill, 2002; Gill, 2012; Alozie, 1988, 1990; Slotnick, 1984).

Second, empirical evidence of the effects of selection institutions in the judicial context thus far has been inconsistent. Williams and Thames (2008) find that among OECD countries, systems in which presidents appoint judges are associated with a greater presence of women judges. Similarly, Carbon, Houlden and Berkson (1982) find that more female judges came to their positions on state courts via gubernatorial selection than election (see also Bratton & Spill 2002). However, the relationship between executive appointment and judicial diversity is not universal. Hoekstra, Kittilson and Bond (2014) find only weak support that selection method influences the presence of women on courts cross-nationally, and Gill (2012) finds the likelihood a woman will be appointed increases when more people are involved in the nominating and appointing process. She argues that a more diverse nominating or appointing body is more likely to consider a more diverse group of candidates. Others, however,

find no effect of formal selection procedures on women or minority judges (Alozie, 1988, 1990; Slotnick, 1984). In all, scholars have been unable to establish clear patterns between selection methods and the appointment of women and minority judges.

Third, institutional selection procedures – unlike underlying political culture or or norm diffusion – can be manipulated. Judicial selection procedures can and *do* change. U.S. states frequently alter the method of judicial selection. Arkansas, for example, switched from partisan to nonpartisan elections in 2000. Colorado switched to merit selection[5] in 1966, and Florida switched to merit selection in 1976. State Supreme Court justices in Georgia were elected by the general assembly in 1865, were elected by the people in 1868, were elected by the general assembly in 1877, were elected again by the people in 1896, and then in 1983 judicial elections were made non partisan.

There is substantial change over time in the selection process of some national peak courts as well. Afghanistan, for example, has changed the selection process of its peak court five times since 1976. In 1976 the process changed from appointment by the king to appointment by the president. In 1980 it changed to appointment by the Presidium of the Revolutionary Council. In 1987 the process reverted to having the president appoint judges before switching back to the King appointing in 2001. In 2004 the process changed to a two-step process: the president appoints judges and then House of People must endorse those selections (VDem Judiciary).

In Benin, the selection of peak court justices changed four times in 26 years. In 1964 candidates for the Supreme Court were proposed by the Chief Justice of the Supreme Court and then appointed by the Republic. In 1970, justices were selected by a presidential council. In 1984, the peak court changed from the Supreme Court to the Popular Central Court and professional judicial candidates are now nominated by

[5]A Judicial nominating committee provides a list of candidates determined by "merit" to the governor. The Governor chooses a judge who then faces a retention election at the next general election.

the executive council and elected by the national assembly. In 1990, the court changed back to the Supreme Court and the process added a third step: the superior council of magistrates consults on the selection, then the Guard of Seals of the Minister of Justice nominates candidates, and then the President approves them (VDem Judiciary).

Selection institutions can and do change. Understanding the effects – both intended and unintended – of those changes will allow for better informed institutional manipulation and can have important consequences for the politics of judicial reform.

Fourth, existing institutional explanations for diversity (or the absence of diversity) conflate distinct mechanisms through which institutions shape outcomes. Theoretical ambiguity has lead to inconsistent empirical evidence, which obscures our ability to determine how judicial selection mechanisms shape outcomes. The goal of this project is to combine a clearer theoretical approach to the role of selection institutions and diversification with research strategies that allow for causal inference.

Specifically, I argue that selection institutions influence four specific processes that shape diversity in office: First, selection institutions determine what people can learn about the fairness of the selection process. Second, institutions affect how and whether observers interpret disparity as bias. Third, selection institutions shape the ability of citizens to determine who is responsible for perceived bias in the selection process, and fourth, selection institutions shape what people can do to rectify perceived bias in the selection process. In other words, selection institutions condition citizens' abilities to gather information and make inferences about the fairness of their political institutions, and selection institutions determine the ability of citizens to hold judicial selectors accountable for providing descriptive representation on the bench. As such, a key focus of this theoretical story is the ability of citizens to hold selectors accountable for insufficient levels of diversity in office.

In chapter 2, I detail a framework that links selection institutions to judicial diversity through a process of citizen inference and accountability. Chapter 3 reports

evidence from a survey experiment that shows how institutions affect prior beliefs in the fairness of judicial selection. Specifically, observers view merit selection procedures in which a commission generates a short list from which the executive selects a judge to fill a vacancy as *more fair* than a selection procedure in which the executive has sole authority to choose judges. Importantly, preferences for "merit" selection are robust to a different name for the procedure, which means the strategic name of the selection procedure is not the only feature swaying respondents. Qualitative explanations for respondents' beliefs suggest that it is the presence of multiple actors that serve as checks on each other that lead observers to favor merit selection over executive selection.

In chapter 4, I test several of the main features of the framework outlined in chapter 2. Using evidence from a survey experiment, I find that institutions do affect whether observers perceive disparity as bias. When observers – a priori – trust the institution, in this case merit selection, they must observe more disparity before concluding that the process is biased. In addition, under merit selection, blame for perceived bias is diffuse: it is shared between the governor and the commission. Under gubernatorial selection, the majority of blame is concentrated on the governor. Taken together, these two findings suggest that procedures with multiple actors involved in selection can undermine the process of diversification via (1) the perception of bias and (2) blame attribution/accountability. Using observational data collected through the Varieties of Democracy judiciary project, I find preliminary evidence that an institutional change *away* from unitary selection to a process with multiple actors slows the process of gender diversification.

Chapter 5 tests an implication of accountability for judicial diversity. I find a pattern in which judicial turnover for U.S. state supreme courts is gendered – women disproportionately replace women and men disproportionately replace men. This pattern is consistent with a narrative in which those tasked with selecting judges feel

pressure to maintain minimum levels of diversity. I compare the rate of selection of women to a simple proxy for the estimate of the candidate pool and find no evidence that the gendered pattern of turn over has suppressed the overall level of women justices on state supreme courts.

The 6th chapter addresses an institutional feature that might increase opportunities for judicial diversity: slate selection. Drawing on the logic of party list PR elections in the legislative context, I argue that the selection of judges as a group (that is, a slate) rather than on a rolling, one-by-one basis might encourage observers to more easily interpret disparity as bias. Evidence from a survey experiment suggests that the selection of judges as a group rather than one-by-one *does* affect how observers make inferences about bias; observers are more critical of homogeneity when judges are selected simultaneously rather than one-by-one. Specifically, observers are more likely to even *notice* gender when judges are selected as a group. Cross-national observational evidence of institutional changes to slate selection suggest a positive but statistically insignificant effect.

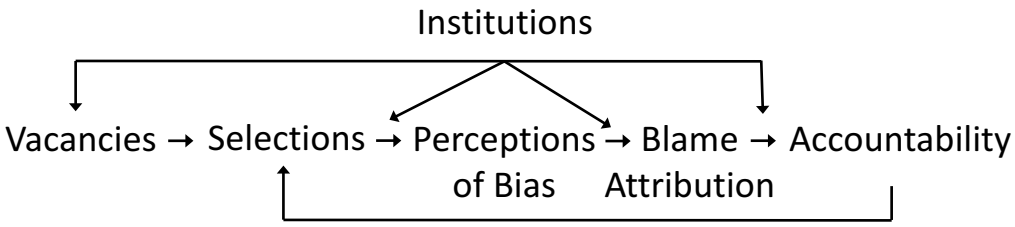
Chapter 7 addresses (one aspect) of intersectionality on the bench: minority women state supreme court justices. In this chapter, I compare several characteristics of minority women judges to white women judges and minority men to asses whether standards for selection vary systematically across gender and racial groups. Across the six categories where minority women can be directly compared to minority men and white women (ivy league education, prior judicial experience, age, selection method, party, and length of tenure), minority women are indistinguishable from minority men in all six categories. In contrast, minority women differ from white women across four categories (ivy league education, selection method, ideology, and tenure). Finally, chapter 8 offers final conclusions of the project and outlines avenues for future research.

Chapter 2 Theoretical Framework: Institutions, Accountability, and Diversity

In the theoretical story I tell here, the process starts with vacancies on a court. Judges are selected to fill vacancies. Observers see selections to the bench and use this information to update their beliefs about whether or not the process is biased. If they infer that the process is indeed biased, they must accurately attribute blame and then hold those they blame responsible for bias. In turn, future selections should conform to popular preferences for diversity. Importantly, judicial selection institutions shape how this process works at several stages; institutions like the size of the court and term limits shape outcomes such as the number of vacancies and the timing of vacancies. Institutions, such as merit selection procedures, that affect prior beliefs about fairness (addressed in more detail in chapter 3) shape how observers interpret information. The number of actors and steps involved in the selection process shape if and how observers can accurately attribute blame for perceived bias. The number of actors and their exposure (Valdini and Shortell, 2016) to potential sanctions shape whether elites are induced to change their behavior (i.e., accountability). Figure 2.1 shows the steps of the framework and where institutions intervene.

The main focus of this chapter is to link judicial selections to perceptions of bias. In other words, this chapter describes how observers use information gleaned from past and current judicial selections to make inferences about bias in the judicial selection process. There are many ways in which observers can learn and make inferences about bias. To ground the analyses of this dissertation, I model this process with a Bayesian learning model, described below.

Figure 2.1: Theoretical Framework: Steps



This figure outlines the steps of the theoretical framework of this project. The process starts with vacancies. Judges are selected to fill vacancies. Observers see selections to the bench and use this information to update their beliefs about whether or not the process is biased. If they infer that the process is indeed biased, they must accurately attribute blame and then hold those they blame responsible for bias. In turn, future selections should conform to popular preferences for diversity. Most of the arrows in the figure point to other arrows. This represents that institutions shape the process of moving from one step to another. For example, institutions shape how the selections *of judges translates to* perceptions of bias*. In contrast, the arrow points directly to* vacancies *because institutions determine – through dictating court size and turnover – the quantity and timing of vacancies directly.*

2.1 Components of the Model

The Bayesian learning model that describes how selections affect perceptions of bias works in the following way: observers have prior beliefs about the fairness of the selection institution and individual thresholds for disparity. They then observe selections to the courts. Using this observed information, they update their beliefs about the fairness of selection. If their posterior beliefs about the fairness of the process are sufficiently different from their prior beliefs – specifically, beyond their threshold for bias – they infer that the observed disparity is unacceptable and is the outcome of a biased process. This section details the components of the model.

2.1.1 The Candidate Pool

The candidate pool is the set of people who have the intellectual and professional criteria for judicial office. Critically, qualifications and recruitment criteria are them-

selves an outcome of political bargaining that benefits some over others.[1] As such, qualifications may be selected precisely to *exclude* women and minorities from office. Prior to emancipation in the United States, for example, there were no Black judicial candidates.[2] Moreover, the boundaries between formal and informal qualifications is blurry: one could argue, for example, that being a personal relative to the executive, may count as a "qualification."[3]

In this project, the qualified candidate pool refers to the intellectual and professional characteristics that shape the ability of an individual to serve as an effective judge. The composition of the candidate pool refers to the gender composition of candidates. What is most important for the purposes of the model, though, is that observers' beliefs about the candidate pool are *fixed* in the short term. Of course the composition of the candidate pool changes over time, particularly over the years during which candidate pools transitioned from total homogeneity to moderate diversity after formal restrictions were removed. In addition, the composition of the candidate pool is likely endogenous to the prevalence of unfairness in the selection process. If women and minorities do not expect to be selected for particular positions, they should be less willing to incur the costs of obtaining specific qualifications for those posts. However, absent any unusual shocks, the endogeneity of the candidate pool is inconsequential in the short term.[4]

[1] Remiche (2015) writes,"the choice of a given [selection] procedure is dictated by the cultural values attached to the judiciary in a legal system, by the way judges are represented in it, and by the foundation of their legitimacy within it" (p.96).

[2] Jonathan Jasper Wright was the first African American to serve on a state supreme Court; he was selected to the South Carolina Supreme Court in 1870. Robert Heberton Terrell was the first African American judge to serve on a federal bench. He was appointed by President Taft to the District of Colombia Municipal Court in 1910.

[3] See Kalantry (2012) for a discussion of how systems in which personal connections are required for selection favor male candidates over female candidates. Similarly, see Feenan (2008) for a discussion of how " old-boys' " networks excluded women judges from information and socialization required for success.

[4] Practically, though, this assumption is necessary to isolate how observers learn about the selection process. If beliefs about the candidate pool and beliefs about the selection process varied at the same time, citizens could use the same data to make inferences about both the selection process and the candidate pool. Without fixing one of these features, it would be very difficult to know how citizens *learn*, at least without making many other assumptions in the process. Moreover, if

2.1.2 Bias

An *unbiased* selection process in this story occurs when selectors choose from the pool of qualified candidates without preference to a specific gender or race. In other words, a non-biased selection process would mirror taking repeated random samples from the qualified candidate pool. If the composition of the pool is 50% women and 50% men, the probability of selecting a woman under a non-biased/fair process is .5. Not every string of selections will be 50% male and female, but for a fair process, over repeated samples, the mean of the sampling distribution would converge to the population proportion of women in the candidate pool.

While random variation in the representativeness of a judicial bench is expected, *persistent* deviation of the proportion of women or minorities selected from the proportion of women or minorities in the candidate pool would suggest, however, non-random draws, defined here as *bias*. Specifically, bias is the difference between the expected probability of selecting a woman under the existing selection mechanism and the expected probability of randomly drawing a woman or minority candidate from the qualified candidate pool.

There are, of course, other ways of conceptualizing bias. For example, we could talk about bias in the recruitment institutions that stems from biased standards for qualification – if a qualification for judicial office is a law degree from a prestigious school that favors men over women in the admissions process, then the pool of qualified candidates will be predominately male. Or, men and women may expe-

citizens do not know the composition of the candidate pool *or* the fairness of the selection process perfectly, I expect citizens to have *better* information and knowledge about the composition of the candidate pool than the fairness of the selection process. Qualifications such as a prestigious law degree or professional experience that qualify one for the candidate pool can be verified in ways that bias cannot be. Furthermore, by interacting with their community and observing media coverage, citizens can develop expectations about the diversity of the candidate pool. If citizens have female or minority doctors, teachers, and lawyers, for example, or observe women and minority judges on trial benches, observers can develop beliefs and make inferences about the gender composition of high court candidate pools as well. I simplify the model, therefore, by assuming that beliefs about the candidate pool are fixed in the short term.

rience "bias" on the bench that takes the from of differential treatment of men and women judges. Understanding the various ways in which women and other historically excluded populations experience bias in the pursuit of and service in offices is an important task. In this project, however, bias is limited to the final selection stage and does not take into account bias in the accumulation of qualifications or bias once in office.

Given a fixed set of qualifications and selection institutions, the observer's task lies in determining when deviance from the expected probability of selecting female or minority judges reflects innocuous, random variance and when deviance reflects bias. How citizens disentangle random variance and bias depends on (1) citizens' prior beliefs and (2) the amount of information citizens observe.

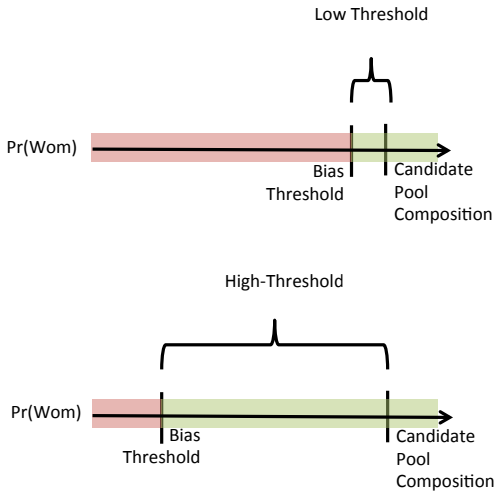
2.1.3 Thresholds for Bias

Citizens want qualified, competent judges to be selected for office. This implies, then, that citizens generally prefer a fair, unbiased selection process. If a selector favors a certain type of judge over another, then that selector is overlooking judicial qualifications in favor of descriptive or personal qualifications, which may result in the selection of less qualified candidates.

While all citizens want a fair selection process, citizens vary according to how much disparity they must observe before concluding that the process is indeed biased. That is, citizens have different thresholds for bias. A threshold for bias describes the individual acceptance for disparity between the probability of selecting a woman under a fair system (that is, the composition of the candidate pool) and the expected probability of selecting a woman judge under the current selection procedures.[5] Take, for example, a process in which the probability a woman is selected is .475 but the probability that a woman would be selected from random draws from the candidate

[5]For now, I address bias that occurs when the probability of selecting a woman or minority is too low. Of course, bias can also occur when the probability of selecting a woman or minority is too *high*.

Figure 2.2: Bias Threshold



Citizens may vary according to how much bias in the selection process they are willing to accept. The top panel shows a citizen with a low tolerance for bias. This citizen will only accept very little bias. The high-threshold citizen (bottom panel) will accept much more bias.

pool (i.e., a fair process) is .5. Under the definition of bias in this project, this selection process is biased: the probability a woman is selected is less than it would be under a fair process. However, some citizens may find the magnitude of this bias too small to be concerning. Figure 2.2 shows two examples of bias thresholds, one small and one large.

Thresholds for bias reflect how willing individuals are to accept disparity. Someone with a high threshold or tolerance for bias accept high levels of homogeneity on the court. These individuals do not prioritize diversity in office and will only find fault with a court that is extremely homogeneous (such as one that is consistently all-male). In contrast, someone with low-bias tolerance cares about diversification and will conclude that a process is biased at much higher levels of diversity than others. These individuals will be relatively quick to interpret disparity as bias.[6] The amount

[6] At the extreme may be those citizens who believe that women or minorities should be selected at a rate greater than their composition of the candidate pool in order to alter norms around political institutions or to rectify past injustices. Particularly for individuals who are skeptical of diversity, they may experience "backlash" to a belief that women are being selected too frequently.

of bias citizens are willing to accept is important because bias acceptance determines how much information citizens require and how much disparity they must observe before inferring a process is unfair/biased.[7]

2.2 Learning about Bias from Appointments

In this model, citizens know the composition of the candidate pool but can be uncertain about the probability that the selection process will select women or minority candidates. As stated previously, in reality citizens are uncertain about *both* the composition of the candidate pool and the selection process, but people should have *better* information about the candidate pool than the selection process: qualifications are easier to verify than the fairness of the selection process.

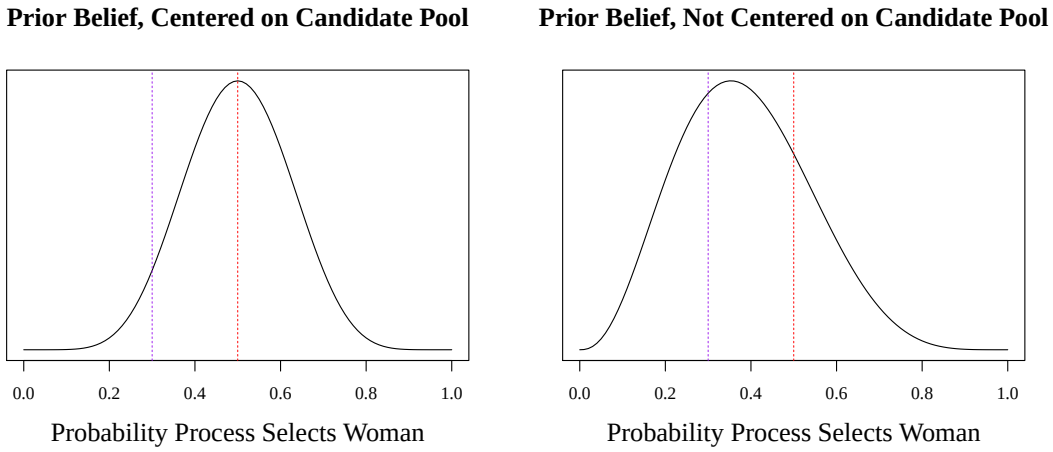
Prior Beliefs

Citizens' *prior* beliefs reflect beliefs about the selection process – that is, beliefs about the probability the process will select women or minority judges – before they observe selections to the court. These beliefs vary in two important ways. First, the central tendency of the prior beliefs may vary. For example, citizens can have a prior belief that the probability a woman is selected is centered on the composition of the candidate pool. Or, they may believe that the probability a woman is selected is most likely lower than the composition of women in the candidate pool, in which case the

[7] In addition to a threshold for bias, citizens can vary by how "sure" they must be before concluding that the process is biased. That is, they vary according to how much density of their posterior belief (addressed in section 4) must be beyond their threshold of bias before concluding that the process is unfair. This feature of citizen decision-making captures variation in risk aversion or how willing citizens are to conclude that a process is biased when it might not be. In order to simplify the number of moving parts, I set the density requirement to the expected value/mean. When the expected probability of selecting a woman is below the threshold for bias, observers infer that the process is biased. While observers surely vary in how they make decisions in this way, in the aggregate, a "more likely than not" decision rule may be accurate. For each citizen that is quick to decide the probability of bias is too high, there is another citizen who may require more certainty. Furthermore, changes in the density required for rejection have the same general implication as an increased bias threshold – the posterior belief must shift farther from the composition of the candidate pool and a fair process for rejection. Fixing the density required for rejection simplifies the moving pieces of this theoretical story but does not substantively change how the process of learning about bias works.

prior would not be centered on the composition of the candidate pool. Figure 2.3 shows graphical depictions of two prior beliefs, one centered on the composition of the candidate pool and one not centered on the composition of the candidate pool.

Figure 2.3: Prior Beliefs and Central Tendency

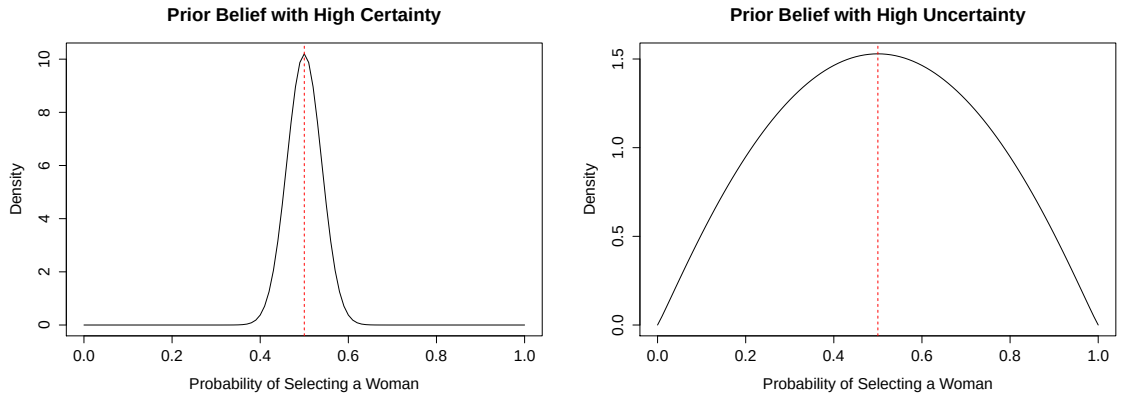


These two figures show how the central tendency of prior beliefs may vary. The red dashed lines show the composition of the candidate pool (.5). The purple dashed line shows the threshold for bias (.3). The left panel shows a prior belief centered on the composition of the candidate pool. The observer's best guess is that the probability a woman is selected is equal to the composition of women in the candidate pool. In the right panel, the observer's best guess is that the process selects women with a probability less than the composition of women in the candidate pool.

Second, the variance or precision of the prior belief can vary depending on the degree of uncertainty citizens have about the probability that the selection process will select a woman or minority judge. For example, a citizen who is confident that the process is fair will have a prior belief centered on the composition of the candidate pool with little variance (see the left panel of Figure 2.4) . In other words, a citizen with low variance (or high precision) believes there is some – but little – chance that women will be selected at a rate very different from their composition in the candidate pool.

A citizen who believes the process to be fair but is *uncertain* will have a prior belief centered at the composition of the candidate pool with high variance (see the right panel of Figure 2.4). This citizen assigns greater probability to the possibility that

Figure 2.4: Prior Beliefs and Certainty/Uncertainty



These two figures shows prior beliefs about the probability that the process will select a woman. Both beliefs are centered on a composition of the candidate pool (here, .5). In the left panel, the citizen is highly certain about her prior belief – 80% of the density is between .45 and .55. This citizen does not expect the probability that a woman is selected to be very far from .5. In contrast, the citizen in the right panel is relatively uncertain that the probability a woman will be selected is close to .5. Here, 80% of the density is between .2 and .8. So, while the expected probability that a woman will be selected is .5, this citizen expects the process to select a woman with a probability somewhere between about .2 and .8.

women will be selected with a probability that is greater or less than the composition of the candidate pool. Figure 2.4 depicts two prior beliefs about the selection process. While both of these prior beliefs are centered on the composition of the candidate pool, one shows a citizen who is confident in their belief about the probability with which women are selected to the bench and one shows a citizen who is highly uncertain about the probability with which women are selected.

2.2.1 Updating Beliefs: Posterior Beliefs

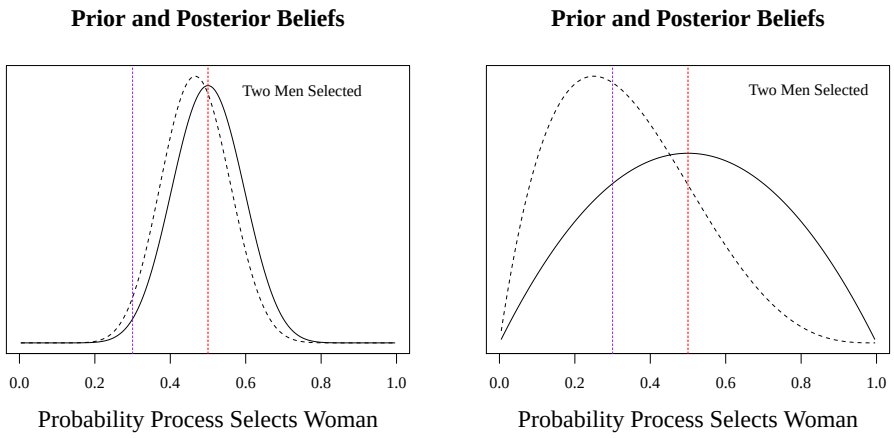
The shape of the prior belief is important because it affects how citizens reconcile their beliefs with observations about the selection of women judges to courts. Specifically, the mean and variance of the prior belief determines how much information a citizen needs for her posterior belief[8] to diverge from her prior belief.

Take, for example, a court with ten judges where all ten judges are selected at

[8]Posterior beliefs are updated via Bayes' rule in which the distribution of the posterior depends on both observed information and the prior belief: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

once as a slate. Citizens have a prior belief about the selection process, and then they observe the selection of three women and seven men to the court. Then, each citizen updates her beliefs about the probability with which women judges are selected. The extent to which the posterior belief reflects observed information depends on the shape of the prior belief.

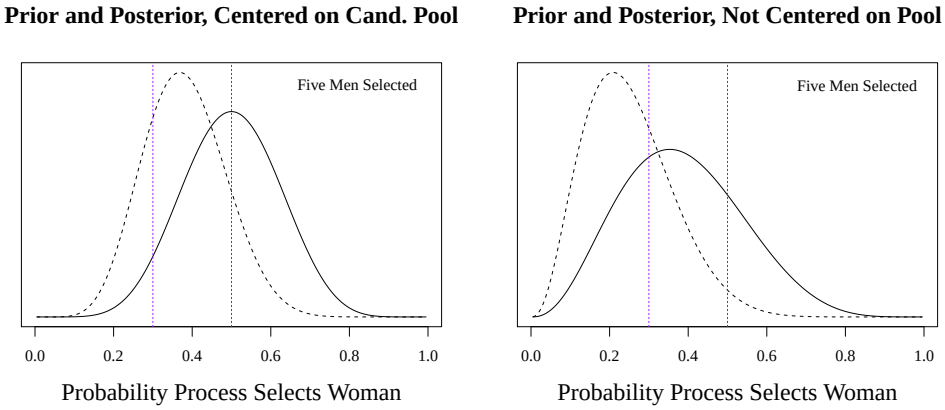
Figure 2.5: Prior and Posterior Beliefs, (Un)Certainty



These figures show the prior and posterior beliefs for a citizen who has observed the selection of two men to the court. The left panel shows a citizen who had a highly certain (low variance) prior belief. Her posterior belief only shifts slight towards the observed data. The citizen's confidence that the process is fair leads to relatively little updating of beliefs. The right panel shows a citizen with uncertain prior beliefs. The posterior belief here shifts quite drastically towards the observed data. because this citizen had uncertain beliefs about the the probability that the selection process selects a woman, she updates her beliefs about the process quite substantially.

Figure 2.5 shows the prior and posterior beliefs for two citizens having observed the selection of three women and seven men to a ten-judge court. The vertical, red-dashed lines show the belief about the composition of the candidate pool. The black-dashed lines show the actual proportion of women selected. The solid black curves show the prior beliefs, and the black dashed curves show the posterior beliefs. As you can see, the posterior belief in the left panel is not much changed from the prior belief. This citizen was and is so confident that the process is fair that observing the selection of ten judges was not enough information for her to change her belief, even though the selection of those ten judges does not reflect the composition of the

Figure 2.6: Prior and Posterior Beliefs, Centered and Not Centered



These figures show the prior and posterior beliefs for a citizen who has observed the selection of five men to the court. For the left panel, the prior belief was centered on the composition of the candidate pool, and the right panel shows a prior belief not centered on the candidate pool. Both observers see the selection of five men and update their beliefs via Bayes' rule. The observer in the left panel is still content with the disparity; the observer in the right panel concludes that the process is biased because sufficient density is the the left of her threshold for bias.

candidate pool. In contrast, the citizen represented in the right panel was less certain about the fairness of the process. Observing two men selected to the bench shifts her posterior belief to the left quite substantially. Her posterior belief is weighted more heavily towards the observed information than the posterior belief of the more certain observer in the left panel.

2.2.2 Institutional Effects on Learning

There are two ways in which judicial selection institution shape how and whether citizens can update their beliefs and make inferences about bias. First, institutions determine how much information citizens observe by dictating the rate of turnover for judges. For the U.S. Supreme Court, for example, judges are appointed for life and judicial turnover is very slow. As of this writing, only three judges have been selected to the U.S. Supreme Court since 2009 (three judges in nine years)[9] Four of the nine justices on the bench have been serving for at least 23 years, and the longest serving

[9] Sotomayor, 2009; Kagan, 2010; Gorsuch, 2017

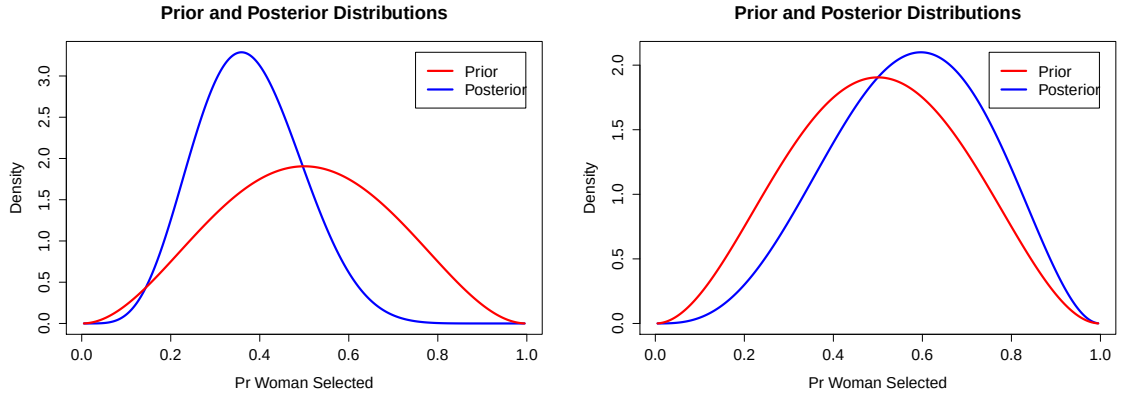
judge, Justice Kennedy, has been on the bench since 1988.[10] For voters younger than 29 (born 1989 or later), at least one member of the U.S. Supreme Court was selected *before* they were born. In Burkina Faso, contrast, one third of the pusine judges of the Constitutional Chamber of the Supreme Court are replaced every three years (VDem, Judiciary Project). Observers in Burkina Faso have far more information with which to update their beliefs about the selection process.

Institutional features such as life tenure, term limits, court size, prestige, salary, or other features that induce judges to either leave or remain on the bench all affect the rate of turnover and, therefore, the amount of information citizens have about past selections. Courts with high turnover allow for more accurate information, which in turn allows for more accurate inferences. Figure 2.7 shows the same prior distribution and two different posterior distributions based on different amounts of information. The observer in the left panel observed the selection of 10 judges and, thus, had more information with which to update her beliefs relative to the observer in the right panel who only observed the selection of one judge.

Second, features of judicial selection – such as the involvement of multiple actors or strategic names such as "merit" selection – that lead observers to trust an institution can affect the shape of observers' prior beliefs. Chapter 3 describes how respondents' perceive merit selection procedures as more fair than gubernatorial selection procedures a priori (that is, before observing the outcome of any selections). As Figures 2.5 and 2.6 show, the shape of the prior belief, in turn, determines how much disparity an observer must see before inferring the process is biased (for a fixed bias threshold).

[10] Kennedy, 1988; Thomas, 1991; Ginsburg 1993; Breyer, 1994

Figure 2.7: Prior and Posterior Beliefs, Different Types of Selections



These figures show the same prior belief updated after observing two different selections. The left panel shows a posterior distribution after the citizen observed seven men and three women selected to the bench. The right panel shows the same prior belief but a posterior belief after observing the selection of one woman. The posterior distribution shifts to the right, but very slightly. Citizens have much less information with which they can update their beliefs when judges are selected one at a time (right panel) versus as a slate (left panel).

2.3 Institutional Effects on Accountability

Once observers update their beliefs, make inferences about bias, and decide that the selection process is biased, the selection institutions shape whether or not citizens are able to address their concerns. Specifically, two things must happen for citizens to be able to pressure selectors to rectify bias. First, citizens must be able to accurately *identify* who is responsible for bias. Second, citizens must be able to exert pressure on whomever is responsible, and they must be able to exert *enough* pressure to change selector behavior. Extant literature on voter evaluation and attribution suggests that voter evaluation of poor performance and attribution of blame does shape voter choice across diverse issue areas (Marsh and Tilley, 2010; Anderson, 2007) although this is the first time (to the author's knowledge) that evaluation and attribution has been applied to judicial selection.

2.3.1 Identification of Responsibility

For citizens to hold selectors accountable for bias, they must be able to identify who is responsible for observed bias. In some settings this can be quite clear. For example, in Bahrain, constitutional court justices are appointed by the king (VDem Judiciary, Bahrain 2002) and in Bangladesh, supreme court justices are appointed by the president after he consults with the chief justice (Vdem Judiciary, Bangladesh 2011).

In both of these cases one individual is responsible for selecting judges. If an observer determines that the process is biased, it is clear who is responsible – either the king or the president. Of course, whether or not observers can hold those responsible for bias accountable varies across these two examples, a feature addressed in section 2.3.2.

In contrast to a unitary appointer, a more complex selection process may make it more difficult for observers to determine responsibility for bias. For example, the seven members of the Constitutional Court of Niger are all chosen in different ways or according to different qualifications. From the 2010 Constitution, judges are appointed by presidential decree in the following ways and with the following qualifications:

1. One "notable [person] with great professional experience in juridical or administrative matters ...is proposed by the President of the Republic"
2. One "notable [person] with great professional experience in juridical or administrative matters ... is proposed by the Bureau of the National Assembly"
3. One "magistrate [is] elected by [his] peers ... of the first grade"
4. One "magistrate [is] elected by [his] peers ... of the second grade"
5. "One lawyer with at least ten years of exercise of the profession [is] elected by his peers"

6. "One professor-researcher holder of a doctorate in public law [is] elected by his peers"

7. "One representative of the associations of defense of human rights and promotion of democracy, holder of at least a diploma of the third cycle in public law, elected by the singular or plural collectives of these associations"

In Niger, even if citizens observe a completely homogenous court, each judge came to his or her position in a different way. Observers must disentangle responsibility, a task more complicated under Niger's institutions than under the selection process in Bangladesh or Bahrain.

Granted, Niger is an extreme example of a process where responsibility is difficult to assign. However, difficulty in assigning blame should make sanctions less likely. Cross-national evidence of government and executive policy responsibility shows that voters responded to economic performance more strongly in countries where responsibility for performance is relatively more clear (G. Bingham Powell, 1993; Anderson, 1995; Kevin M. Leyden, 1995; Robert C. Lowry, 1998). If voters care about judicial selections, then they should be better able to exert sanctions when responsibility is easier to identify.

While institutional variation in judicial selection is vast, there are at least two relatively simple institutional features that complicate the assignment of blame: selector turnover and the presence of multiple actors.

Identification of Responsibility: Selector Turnover The first common institutional feature that complicates the identification or responsibility is when the person responsible for selection changes over time. Using the U.S. supreme Court as an example, the president is only completely responsible for the one or two judges he may select during his tenure. Even so, there are expectations that the one judge should help keep the overall make-up of the court balanced according to certain de-

scriptive features. Abraham (1992) describes, for example, an expectation that the U.S. Supreme Court should include certain "types" of judges. He writes, "There is no doubt that there now exists a black seat on the bench that is, in effect, far more secure than a Catholic seat or a Jewish seat. That is unquestionably true of a woman's seat" (64-65).

Reconciling the expectation of diversity with the fact that a president is only solely responsible for the selection of one or two judges complicates how observers are able to hold a selector accountable; a president is responsible for the makeup of the court, but his agency is restricted by the choices of earlier presidents.

Identification of Responsibility: Multiple Stages or Multiple Actors A second complication is the presence of a multi-stage process, in which one person or group is responsible for nominating or selecting a candidate and then another individual or group is responsible for appointing or approving of the candidate. In this scenario, observers know that the choices of both actors are restrained by the other. If the preferences of the selector and nominator diverge, both actors may have to moderate their selection. A liberal president, for example, might need to choose a supreme court candidate that a conservative Senate will confirm, in which case he may be induced to appoint a moderate judge. The presence of these two steps complicates the observer's ability to determine who is responsible for observed bias.

Closely related are those processes in which multiple actors are involved in determining the composition of the court, such as the Niger example above. In this case each individual actor may be able to avoid blame by spreading the blame among many, [11] which will obscure responsibility.

Complicated or diffuse selection procedures, including those where selectors change over time or where there are multiple stages, make it difficult for observers to deter-

[11] Weaver (1986) classifes the tactic of speading blame about among multiple actors as "circling the wagon." Two other blame-avoiding tactics that apply to a multiple selector situation are referred to as "passing the buck" and "finding a scapegoat" (page 385).

mine who is responsible for bias when they observe it. If observers cannot determine who is responsible for bias, they cannot pressure those selectors to change their behavior and select judges proportionate to the candidate pool. In other words, even if observers believe a process is biased, that bias can persist if observers are unable to determine who or what is responsible for the bias.

If observers are able to (1) determine that bias is occurring in the selection process and (2) identify who is responsible for bias, there is one more set of features that shape whether or not that bias can be rectified: whether or not observers can hold those responsible for bias accountable.

2.3.2 Accountability

For citizens to be able remedy bias they must (1) recognize bias (2) be able to assign responsibility for bias, *and* (3) pressure those responsible to change their behavior. Several institutional features shape the extent to which observers are able to hold selectors accountable.

For example, in the section above, the examples of the king selecting judges in Bahrain and the president selecting judges in Bangladesh were used to illustrate processes in which identification of responsibility was very clear: in both cases, one identifiable person is responsible for the composition of the court. These two cases vary, however, in the extent to which citizens can hold those selectors accountable for bias. In Bahrain, citizens may be able to communicate disapproval of a homogenous court, but they can not exert electoral pressures on the king. In contrast, the president in Bangladesh is elected by Parliament, so should be sensitive to pressure (via members of parliament) to select judges fairly. In this section I will outline four institutional features that shape whether observers are able to hold selectors accountable for bias.

Accountability: Group Selectors, Multiple Stages, and Multiple Actors The first institutional feature that complicates the ability of observers to hold selectors accountable is when the selector is a collective institution, such as a legislature or commission, or when there are multiple actors involved in the process.

In either case, responsibility for the composition of the court is shared between each member of the collective institution or multiple actors. Any punishment, therefore, will be distributed across all members, insulating each member from acute, concentrated pressure. This shared accountability may make each selector less responsive to the preferences of constituents. If each actor involved in the selection process is less responsive to constituent pressure then, all else equal, citizens will have to exert *more* pressure on selectors to induce the same level of responsiveness.

Accountability: Group/Party Alignment The relationship between selectors and their constituencies will shape the ability of constituents to pressure selectors. Selectors often have multiple constituencies and may respond to one or a few of those constituencies rather than others. In the United States, for example, some state supreme court justices are selected by the state governor. Depending on the party composition of the electorate, a governor may only depend on the support of citizens from one political party. As such, she should only respond to citizen pressure from her party's support base. This means that even if many citizens conclude that the process is biased, have correctly identified the governor as the selector accountable for bias, and have pressured the governor to rectify bias in the selection process, the governor may simply ignore that pressure if she does not need their support to remain in office.

The political and institutional environment of a selector shapes the extent to which different groups are able to hold that selector accountable. Selectors should only respond to constituent pressure when those constituents are important to the

selector.

Accountability: Breadth of Citizen Knowledge and Specificity of Selector Role A third factor that conditions the ability of citizens to hold selectors accountable is the specificity of the selector role and the breadth of citizen knowledge about the selector's role and performance.

For some selectors, choosing judges is just one of many tasks they complete. Governors, for example, have many other appointments to make in addition to judicial appointments. In contrast, a selection committee convened specifically to select a judge has one very specific role. When citizens observe the outcome of judicial selections, they may weigh the decisions according to how important that selection is relative to other selections or actions by the same actor.

If a citizen strongly dislikes a governor's judicial selection but is fond of the the rest of her gubernatorial work, then the citizen might discount the unfavorable judicial selection. However, if a committee only has one task, and a citizen disapproves of the outcome of that task, then the citizen will not discount that decision.

Of course, the extent to which citizens discount unfavorable judicial selections depends on citizens *knowing* who selectors are, what other decisions selectors make, and which judges those selectors choose. Moreover, judicial selections need to be an issue that constituencies care about. Someone could have a strong preference for diversification in office but may care little for the judicial branch or judicial office.

Accountability: How Far Removed Selectors are from the Population Some selectors are directly accountable to the population; selectors who come to office through popular election and aim to run for re-election have a clear mechanism through which citizens can apply sanctions. Voters can simply refuse to vote for the selector in the next election. Other officials are further removed from popular sanctions.

For example, some members of judicial commissions are themselves selected by elected officials. These judicial selection commissions are accountable to citizens, but their accountability may be mediated through the official that chooses members for the commission. That is, for citizens to punish members of a judicial selection committee who were selected by an elected official, citizens would have to first punish or pressure the elected official. In turn, the elected official could then punish the commission by dissolving, reprimanding, removing members, or withholding resources (including political favor) from the commission.

An extreme case may be when nominators are not accountable to the public at all. For example, on some courts members of certain professional associations elect a peer to the bench. In those cases, citizens may not be able to exert any pressure. Throughout the 1980s and 1990s in Ecuador, for example, members of the Constitutional Tribunal were nominated by labor unions, indigenous organizations, and the Chambers of Commerce (VDem, Ecuador 1995). While nominated justices were selected by the National Congress, the nominating bodies were potentially far removed from some segments of society, which would make accountability difficult.

Accountability and the Persistence of Bias All of the above institutional features shape the extent to which selectors will respond to pressure to diversify from their constituencies. If accountability is sufficiently diffuse such that observers cannot pressure selectors to alter their selections, bias can persist even if observers recognize/believe that the process is unfair.

2.3.3 Selector Choices

Thus far, the explanations for the persistence of bias have centered on the behavior of citizens and whether or not citizens are able to identify bias, accurately assign blame for the bias, and pressure those responsible to change their behavior. However, selectors have a role to play in this narrative too. Selectors are people and have pref-

erences over diversity in office. Some prefer the status quo of homogeneity while other may go out of their way to select female or minority judges. Moreover, selectors are strategic actors: they expect to be punished for perceptions of poor performance, and they know that the institutional arrangements within which they work will condition the severity of that punishment.

In this section, I address the selection of judges from the perspective of the selectors. If a population generally has a low-bias threshold and the selectors are also pro-diversity, then diversity in office should reflect the composition of the candidate pool. To the extent that there are qualified women and minorities in the candidate pool, then we should observe diversification in office. Likewise, if selectors are skeptical of diversity and citizens have high bias-tolerance thresholds, we should not expect selectors to diversify judicial benches. The interesting cases are when preferences of the constituency and the selectors diverge.

Diversity Skeptical Selectors, Low-Bias Tolerance Population The case implicitly addressed in this project so far is when a population prefers diversification while the selectors prefer to maintain the status quo. This situation can happen, for example, if the process of choosing selectors is itself biased in favor of a subset of the population that it systematically less likely to value diversification than the rest of the population. In the U.S. states, governors are often key selectors for state supreme courts, and governors are overwhelmingly white men who may be less likely to favor diversification than female or minority governors. It is in these cases where the population prefers diversification but selectors prefer the homogeneous status quo that the ability of observers to learn about bias, assign responsibility for bias, and exert pressure on selectors is most important. In these cases, selectors may be induced to select diverse candidates even when they would prefer not to select female or minority judges. Importantly, this responsiveness depends on the ability of citizens to identify

who is responsible for selections and to the ability of citizens to pressure selectors. Specifically, we expect selectors to respond to pressure for diversification (1) when there is high turnover on courts, (2) when one or a few easily identifiable actors select judges, and (3) when those selectors are accountable directly to voters rather than indirectly.

Pro-Diversity Selectors, High-Bias Tolerance Population Although the focus thus far has been on the ability of citizens to hold selectors accountable for selecting too few women or minorities, the relationship could exist in reverse. That is, it is possible that political elites who select judges could be pro-diversity while the population is skeptical of diversity. In this case, citizens can hold selectors accountable for perceive bias *against* majority-male candidates rather than bias against women or minority candidates.[12]

2.3.4 A Framework for Diversification: Conclusions

In the preceding sections, I have laid out a framework that links selection institutions to diversity through a process of citizen learning and accountability. For citizens to hold selectors accountable for (the absence of) diversity, the must be able to (1) make accurate inferences about whether or not bias is occurring in selection, (2) accurately attribute blame for perceived bias, and (3) hold those responsible for bias accountable to induce diverse selections. Each of the three steps required for citizens to induce diversification in political offices is mediated by the specific institutional arrangements of the selection process. Prior beliefs in the fairness of institutions will shape how observers interpret information. Institutions that affect the size of the court and turnover shape how much data citizens have with which they can update their beliefs. The presence of multiple actors/steps in selection obscure blame attribu-

[12]Or, pro-diversity selectors may institutionalize their preferences for diversification through institutional change, such as the imposition of gender or racial/ethnic quotas.

tion, and institutions that shelter selectors from sanctions undermine accountability for diversity.

In the next chapters, I test implications of this framework. In chapter 3, I test whether institutions affect observers' prior beliefs. In chapter 4, I test how institutions affect the interpretation of information and blame attribution. Chapter 5 addresses gendered patterns of replacement, which is consistent with selectors responding to popular pressure to maintain minimum levels of diversity on the bench. Chapters 6 and 7 offer extensions to an alternative institution and intersectionally disadvantaged judges, respectively.

Chapter 3 Empirical Evidence: Prior Beliefs in the Fairness of Institutions

In the previous chapter, I outlined a framework for understanding gender diversity in the judiciary as a process of citizen learning and political accountability shaped by selection institutions. Briefly: observers have prior beliefs about the fairness of institutions – that is, the probability with which the selection procedure will result in the selection of a woman. Then, observers see the selection of judges to vacancies. They update their beliefs about the process. If their posterior beliefs indicate that the probability with which the process selects a woman is sufficiently low (that is, enough density is below the observer's threshold for bias), the observer concludes that the process is gender biased. Then, the observer is tasked with identifying who or what is responsible for the bias and, finally, holding those responsible for bias accountable to induce more diverse appointments.

In this chapter, I present survey experimental evidence that tests (1) the assumption that people have prior beliefs in the fairness of institutions, and that these beliefs vary across institution. Chapter 4 tests (2) the hypothesis that observers update their beliefs when confronted with new information about the world; (3) the hypothesis that the institutions affect how much information observers must see before concluding that the process is biased, and (4) the hypothesis that selection institutions will shape the ability of observers to attribute blame for perceived bias.

3.1 Prior Beliefs in the Fairness of Institutions

Implicit in the framework detailed in the previous chapter is the idea that people's beliefs in the fairness of institutions vary across institutions. When it comes to gender diversity in office, anecdotal evidence suggests a belief that some ways of selecting

judges are indeed more fair than other ways of selecting judges. In the U.S., there has been a debate about "merit selection" procedures as an alternative to gubernatorial selection and popular election. Typically under merit selection procedures, a merit commission of several individuals accepts applications and seeks candidates for a vacancy. Then the commission generates a short list of candidates. The governor then chooses a judge from the list to fill the vacancy. Proponents of merit selection argue that members of the –presumably bi-partisan – commission will focus on qualifications of judges, which will de-emphasize the role of politics in judicial selection and will ensure only the most qualified candidates will be selected.

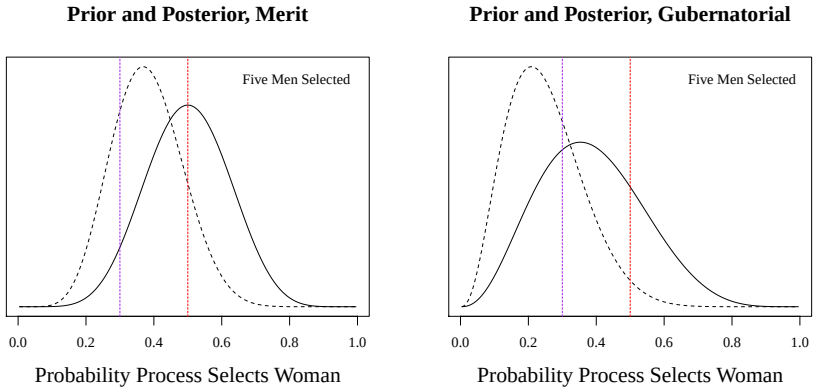
When courts began transitioning to merit selection procedures, many believed the change would benefit gender diversity in the judiciary. For example, Mary Mullarky, the second woman selected to Colorado's state supreme court writes about the history of women in the Colorado Judiciary:

> "[A]nother important factor in Colorado was a new method of selecting judges adopted by Colorado voters in 1966. Partisan political elections of judges were replaced by a merit selection system. Under this system, which remains in effect today, a citizen commission screens applicants for each judicial vacancy, and the Governor appoints the new judge from a list of nominees. *Hopes were high that women would do better under the merit selection system than they had under partisan political elections*" (Mullarky, 2012, emphasis mine).

The idea that "hopes where high" indicates a belief that the new system would be more likely to select a woman judge, presumably, because the new selection system would limit opportunities for bias or discrimination.[1] Similarly, Berkson, Carbon and Neff (1979) write, "When President Jimmy carter established the U.S. Circuit

[1]The alternative could be that the existing system was "fair" and the new merit system is biased in the favor of women. The absence of women on Colorado's supreme court prior to the change, however, makes this alternative unlikely.

Figure 3.1: Prior and Posterior Beliefs, Merit and Gubernatorial Selection



These figures show examples of theorized prior and posterior beliefs for merit and gubernatorial selection procedures. The left panel shows a prior belief centered on the composition of the candidate pool. The dashed line shows the posterior belief after observing the selection of five men. The right panel shows a prior belief centered to the left of the composition of the candidate pool but still to the right of the threshold for bias. The dashed line shows the posterior belief after observing the same selection of five men. For the left panel (merit selection) most of the density of the posterior belief is to the right of the threshold for bias, which means the observer is still content with the process. For gubernatorial selection – the right panel – most of the density of the posterior distribution is to the left of the threshold for bias, which means this observer would infer that the process is unfair.

Judge Nominating Commission on February 14, 1977, he gave it a two-fold mission: to select circuit judges on the basis of professional merit [...] and to correct for past discrimination by affirmatively seeking women and members of minority groups for the bench" (p. 105).

Drawing from the theoretical framework of chapter 2, if observers do believe that merit selection procedures are more fair – that is, more likely to result in the selection of women at a rate commensurate with the composition of the candidate pool – than alternative selection procedures, observers would need to observe more evidence of disparity before they infer that the process is *unfair*.

Figure 3.1 shows hypothetical prior and posterior beliefs for this example. The left panel shows a prior belief for a Merit selection process centered on the composition of the candidate pool. This reflects a prior belief in which the observer's best guess about the probability the process selects a woman is equal to the composition of the candidate pool (i.e. "fair"). With this prior belief, the observer sees five men selected

to the bench and updates her posterior belief. It shifts to the left, but most of the density of the posterior belief is to the right of the threshold, indicating that the observer is still content with this selection process. In contrast, the right panel shows a prior belief for a gubernatorial selection process. Here, the observer's best guess about the probability with which the process selects a woman is slightly *below* the composition of the candidate pool. The observer see the same five selections of male judges and updates her posterior belief, which also shifts to the left. Unlike in the right panel, however, the posterior belief shifts to the left to such a degree that the majority of the density of the posterior belief is to the left of the threshold for bias. This observer, therefore, is likely to conclude that the process is biased.

The important lesson here is that the shape of the prior belief determines how many disparate selections the observer must observe before concluding the process is biased. When someone believes, a priori, that a process is fair, they will tolerate more disparity in selection before inferring that the process is biased. In contrast, if a person is skeptical to begin with, he or she will more quickly conclude that the process is biased after observing the selection of a few male judges.

Implicit here, however, is the idea that observers have different prior beliefs about the fairness of institutions. In the next section, I present a survey experiment in which I test whether variation in prior beliefs about the fairness of selection institutions exists. In addition, I test whether differences in prior beliefs about the fairness of institutions stem from strategic naming – that is, the term "merit" in merit selection – or the procedure. Respondents do perceive merit selection systems as more fair than gubernatorial selection systems. While the term "merit" does have a small effect on perceptions of fairness, most perceptions of fairness stem from the procedure of merit selection.

3.2 Survey Experiment

To test the assumption that people's prior beliefs in the fairness of institutions depend on the institutions themselves, I conducted a survey experiment fielded in December 2017. The survey was designed in Qualtrics and fielded through Amazon's Mechanical Turk (MTurk).

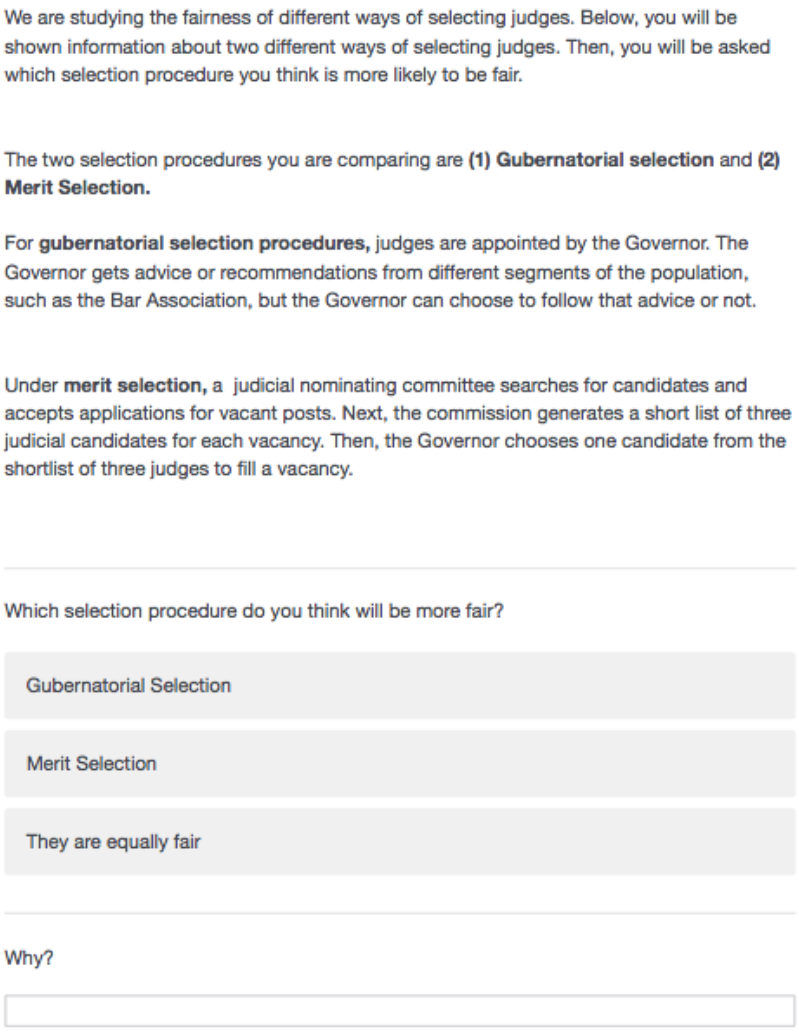
Respondents were shown information about two selection procedures to a hypothetical court. In one selection procedure, the governor is tasked with selecting judges to vacancies (gubernatorial selection). In the other procedure, a commission generates a short list of candidates from which the governor selects a judge to fill the vacancy (merit selection).

After reading brief summaries of the selection procedures, respondents were asked "Which selection procedure do you think will be more fair?" If selection institutions do not affect respondents' prior beliefs in the fairness of institutions, respondents should select each type equally or indicate that they two systems are equally fair (a third option). In contrast, if prior beliefs about fairness systematically vary across institution, respondents should favor one institution over the other. Given the debate surrounding merit selection and the argument that the commission will decrease the role of politics in selection, I expect respondents to perceive merit selection as more fair than gubernatorial selection. Figure 3.2 shows the survey instrument.

Hypothesis 1: If respondents believe that merit selection procedures are more fair than gubernatorial selection procedures, more respondents will indicate that merit selection is more fair than gubernatorial selection.

This experiment tests whether respondents view one selection procedure as more fair than another, but this survey experiment does not test *why* observers prefer one institution over another. One reason respondents may believe merit selection is more fair than gubernatorial selection is due the the name: "merit." It is possible

Figure 3.2: Survey Text, Merit Group



respondents might use the normative term "merit" as a heuristic for the quality of the institution. While the framework presented in chapter 2 is agnostic to the source of observers' prior beliefs, understanding observers' motivations for their beliefs will be useful for better understanding a process of learning and inference and for understanding the implications of institutional design.

To test whether respondents' beliefs in the fairness of institutions depend on the name of the institutions or the process, I add a control group in which respondents are shown the exact same information, but the "merit" selection process is referred to as

"commission assisted" selection. If the term "merit" is responsible for differences in perceptions of fairness, then respondents in the "commission assisted" group will be less likely to select "commission assisted" as more fair than gubernatorial. Likewise, comparing across groups, fewer respondents should select the commission process in the "commission assisted" group relative to the "merit selection" group. In contrast, if the process of the selection procedure is more important than the label, the pattern of responses between merit and gubernatorial selection should be the same or similar to the pattern of responses between commission assisted and gubernatorial selection.

Hypothesis 2: If respondents' beliefs in the fairness of a two-stage selection process stems from the process itself rather than the name, respondents will indicate that the two-stage process is more fair than a gubernatorial selection process at similar rates regardless of whether the process is labeled "merit" selection or "commission assisted" selection.

Table 3.1 reports summary statistics of the two survey groups. Respondents were randomly assigned into the two groups. I removed respondents who were unable to pass a simple attention check in which respondents were asked to identify via multiple choice the selection institutions they read about.

Table 3.1: Sample Characteristics

Gender	Male	Female			
Merit	.46 (144)	.53 (165)			
Commission	.42 (124)	.58(174)			
Age	<25	25-34	35-49	50+	
Merit	.09 (28)	.44 (136)	.31 (97)	.16 (49)	
Commission	.06(19)	.42(124)	.36(108)	.16 (48)	
Ideol.	Very Cons.	Smwht Cons.	Moderate	Smwht Lib.	Very Lib.
Merit	.06 (17)	.29 (90)	.21 (66)	.33 (103)	.10 (32)
Commission	.13 (37)	.29 (86)	.29 (85)	.25(74)	.04 (14)
Edu.	High School	Some Cllge	BA/BS	Masters	Doctorate
Merit	.06 (19)	.34 (104)	.45 (140)	.12 (38)	.02 (7)
Commission	.10 (31)	.31 (94)	.38 (115)	.18(55)	.01(4)

Descriptive characteristics of the survey sample. There were 310 respondents in the group in which the two step selection procedure was called "merit" selection. There were 299 respondents in the group in which the two-step selection procedure was called "commission assisted" selection. These statistics describe the respondents who were able to pass a simple attention check.

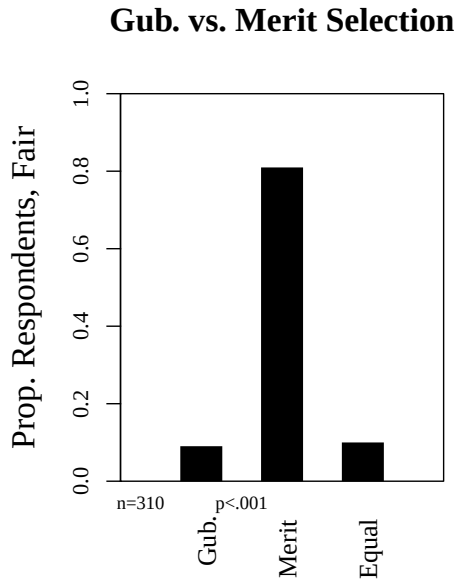
3.2.1 Results, Hypothesis 1: Merit v. Gubernatorial Selection

Quantitative Results

To test whether respondents view merit selection as more fair than gubernatorial selection prior to information about outcomes, I compare the proportion of respondents in the merit group who selected merit selection as more fair to those who selected gubernatorial selection as more fair. Figure 3.3 shows the proportion of respondents in the merit group who indicated each answer.

Just over 80% of respondents indicated that Merit Selection is more fair; 10% of respondents indicated that the two processes were equally fair, and only 9% of respondents indicated that gubernatorial selection was more fair. Moreover, it is highly unlikely that this difference in responses is due to chance. A difference in proportions test comparing gubernatorial selection to merit selection produces a p-value less than .001.

Figure 3.3: Merit vs. Gubernatorial Selection



Proportion of respondents in the "merit" group who indicated each process as more fair. Respondents overwhelmingly indicated that the merit selection process was more fair than the gubernatorial selection process (p¡.001).

Respondents, on a whole, did demonstrate variation in their prior beliefs in the fairness of different selection systems. Overwhelmingly, respondents indicated that merit selection procedures were more fair than gubernatorial selection procedures, absent any information about outcomes.

Qualitative Results

After respondents indicated which institution they perceived as more fair, they had an opportunity to describe the reason they made their choice. Assessing these qualitative responses can shed light on why respondents perceived merit selection as more fair than gubernatorial selection. Qualitative explanations for respondents' choices were classified into seven categories representing the most common themes in the responses:

1. merit selection provides a check on the power of the governor,
2. merit selection allows for more input in selection (having more people involved),

3. merit selection will limit corruption, the role of politics, nepotism, etc,
4. in merit selection, *experts* choose judges,
5. merit selection will result in better qualified judges,
6. gubernatorial selection has more accountability, and
7. gubernatorial selection allows the governor to get input from the broad population.

Of course, there is some overlap in these responses. For example, for many answers classified as "providing a check on the governor," the mechanism through which the governor is limited is the presence of multiple actors, which overlaps with the idea of having "more input" in selection. Table 3.2 gives examples of typical qualitative responses for each category for those who chose merit selection and those who chose gubernatorial selection.

Table 3.2: Qualitative Response Examples

Institution Selected	Category	Response
Merit	Check on Gov	"There is less of a chance of a Governor simply appointing one of his friends to the position."
Merit	Check on Gov	"With merit selection, the governor gets the final say but since he can only select from a short list, he doesn't have too much control over the pick."
Merit	More Input	"Because more people are involved in the decision so more opinions can be heard about the candidates"
Merit	More Input	"More eyes on the matter; less reliance on a single individual and the biases they possess."
Merit	Less Corrupt	"Depending on the committee, it could limit cronyism and political appointments."
Merit	Experts Select	"The committee is comprised of experts and will have a good list of qualified candidates."
Merit	Better Qualified	"Because it is more fair to be judged based on actual merits rather than just someones opinion."
Gov	Accountability	"People elect the governor to make choices like that. I have no clue who the committee is and probably didn't vote for them. I prefer the choice be made by the guy who was elected over some random people with no accountability to the voters."
Gov	Accountability	"[B]ecause the Governor gets to appoint judges who align with his or her political views, with the committee they could nominate people who don't align with the public's choice of politics."
Gov	Popular Input	"I like that they get recommendations from various segments of the population."
Gov	Popular Input	"Because he get advice from different people"

Qualititave explanations for why respondents chose the institution they selected as more fair.

Table 3.3 shows a summary of the distribution of types of qualitative responses. The most common explanation among those who indicated merit selection was more fair was that merit selection allowed for more input in the selection process. Respondents with responses in this category identified the presence of multiple actors as a justification for preferring merit selection over gubernatorial selection. Closely related were those who indicated that their preference stemmed from the committee's check on the power of the governor. These respondents generally mistrusted the (seemingly) unconstrained power of the governor in selecting judges. These two categories – more input in the selection process and a check on the power of the governor – together comprise 48% of explanations for those respondents who chose merit selection as more fair than gubernatorial selection.

Among those who indicated gubernatorial selection was more fair, responses were split rather evenly between emphasizing the accountability of the governor, the ability of the governor to get broad input from the population, and other or unclear reasons.

Table 3.3: Qualitative Responses: Merit Selection

Response Category	**Number Responses**	**Proportion Responses**
Merit Selection		
Check on Governor	53	19%
More Input in Selection	79	29%
Less Corruption/Politics/Nepotism/etc	36	13%
Experts are Selecting	9	3%
Better Qualified Judges	35	13%
Other, Unclear	65	23%
Gubernatorial Selection		
Electoral Accountability	9	35%
Broad Input	8	30%
Other, Unclear	9	35%

A summary of the qualitative explanations for why respondents indicated that merit or gubernatorial selection was more fair than the other.

In summary, respondents overwhelmingly indicated that they perceived merit selection procedures as more fair than gubernatorial selection procedures. Those who indicated that merit selection was more fair generally identified the presence of additional actors and constraints on the governor as their justification. For the few who indicated that gubernatorial selection was more fair, justifications included accountability of the governor and the governor's ability to get advice from a broad swath of the population.

3.2.2 Results, Hypothesis 2: the Importance of "Merit" in Merit Selection

Although most respondents provided substantive explanations for their perception of merit selection as more fair than gubernatorial selection, it is possible that respondents were swayed by the the normatively laden term "merit" in the name. Jones (2012) describes the modern trend of naming short titles for federal statutes with "evocative language" to attract support for the bills (p. 455). He notes how "many short titles imply that measures will be successful (for example, that they will "prevent" certain actions or "protect" certain populations) or contain various subjective characteristics (such as "responsibility" or "accountability")" (p. 458).[2] These names help garner popular support for bills and make voting against the bills potentially more costly for legislators. The term "merit" in "merit selection" could serve a similar function by implying that the process *is necessarialy* meritorious, regardless of the actual procedures or outcomes of the process. If this is the case, the respondents

[2]Examples include, "Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act; the No Child Left Behind Act of 2001; the Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today (PROTECT) Act; the Unborn Victims of Violence Act 2004; the Prison Rape Elimination Act of 2003; the Partial-Birth Abortion Ban Act of 2003; the Adam Walsh Child Protection and Safety Act of 2006; and the Emergency Economic Stabilization Act of 2008... the Lilly Ledbetter Fair Pay Act of 2009; the American Recovery and Reinvestment Act of 2009; the Serve America Act; the Helping Families Save Their Homes Act of 2009; the Credit Card Accountability Responsibility and Disclosure (Credit CARD) Act of 2009; the Patient Protection and Affordable Care Act; and the Dodd-Frank Wall Street Reform and Consumer Protection Act"(Jones, 2012, p. 457-458)

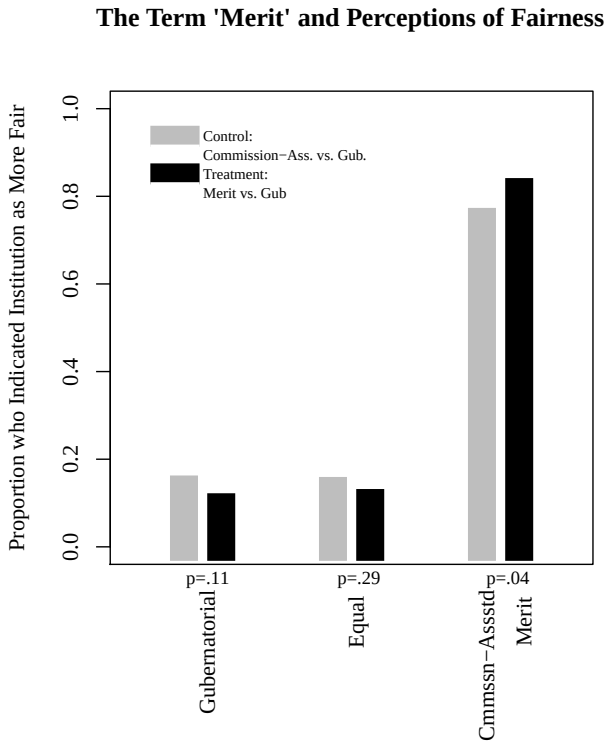
who indicated a preference for "merit" selection procedures may have been persuaded by the name of the procedure rather than the process of the procedure.

It is worth noting, however, that prior beliefs in the fairness of selection institutions should affect how respondents interpret observed information, regardless of how those prior beliefs were formed. Even so, having a better understanding of why observers believe certain institutions are more fair than others can inform the policy discourse and recommendations for selection procedures that might promote inclusion.

To determine if the term "merit" explains why respondents favored merit selection, I compare the proportion of respondents who indicated that "merit" selection is more fair than gubernatorial selection to a treatment group in which respondents were given the exact same information except "merit" selection was labeled "commission assisted" selection. Figure 3.4 plots responses for both groups of respondents. The black bars show responses for those who saw "merit" and gubernatorial selection, and the grey bars show responses for those who saw "commission assisted" and gubernatorial selection. The p-values along the x-axis are for the difference in proportion across treatment groups. Respondents were more likely to choose the two-stage selection method as more fair when the process was labeled as "merit" selection (p=.04), but even with this difference, respondents still overwhelmingly chose "commission assisted" selection as fair rather than gubernatorial selection: 74.2% of respondents chose commission assisted (compared to 81% under "merit"); 13% chose gubernatorial selection, and 12.7% indicated that the two systems were equally fair.

These findings indicate that the term "merit" may affect some respondents' beliefs about the fairness of that selection system. However, most respondents in the "commission assisted" group still indicated that the selection procedure was more fair than gubernatorial selection, absent the normative/subjective language. This suggests that the procedure – that is, the presence of additional actors who serve as a check on the governor – is what leads (most) people to perceive merit selection as

Figure 3.4: Perceptions of Fairness: "Merit" vs "Commission Assisted" Selection



The proportion of respondents who indicated each answer in response to the question "Which selection procedure do you think will be more fair?" The black bars show respondents who were given information about "merit" and gubernatorial selection. The grey bars shows respondents who were given the same information, but were told the processes were called "commission assisted" and gubernatorial selection.

more fair than gubernatorial selection. In the next chapter, I test whether gubernatorial or merit selection procedures affect how respondents interpret information about the outcome of judicial selection.

Chapter 4 Empirical Evidence: Updating Beliefs, Blame Attribution, and Diversification

In chapter 2, I outlined a framework for understanding diversity as a process of citizen learning and accountability: citizens observe selections to the bench, update their beliefs about the fairness of the system, make inferences about bias, and – if they perceive bias – attribute blame and hold those responsible for bias accountable. Importantly, judicial selection institutions mediate each step. In the last chapter, I showed that institutions affect the observers *prior* beliefs about the fairness of selection institutions. Specifically, observers perceived merit selection procedures in which a commission generates a short list of candidates from which the governor selects a judge to fill a vacancy as more fair (or more likely to be fair) than gubernatorial selection in which the governor is solely responsible for selecting judges. Respondent's explanations for their preferences for merit selection over gubernatorial selection emphasized the presence of multiple actors involved in the process.

In this chapter, I asses how selection institutions affect (1) how observers make inferences about bias; (2) how observers attribute blame for perceived bias; (3) and whether an institutional change away from unitary selection slows the process of diversification.

How Observers make Inferences about Bias First, I test whether variation in prior beliefs, in turn, affect how observers interpret observed information. As Figure 3.1 showed, if observers' prior beliefs about merit selection are centered on the composition of the candidate pool (a "fair" process), then observers will have to observe more disparity before concluding the process is biased relative to a process

in which their prior beliefs about the probability the process selects a woman are below the composition of the candidate pool (as some may believe is the case for gubernatorial selection, for example.) If prior beliefs affect how observers interpret information as theorized, then more respondents should perceive the outcomes of judicial selection as unfair at lower levels of diversity under gubernatorial selection than under merit selection.

As expected, survey respondents are more critical of the selection process when there are no or few women on the bench. In addition, at low levels of diversity, respondents are more critical of disparity under gubernatorial selection relative to merit selection. This suggests that –as theorized – prior beliefs in the fairness of a process require observers to see more disparity before concluding the process is biased.

How Observers Attribute Blame for Bias Second, I test how the presence of more actors involved in selection affects the how respondents attribute blame for perceived bias. If observers do determine that a selection process is biased, to rectify that bias citizens must be able to hold those responsible for bias accountable. Valdini and Shortell (2016) demonstrate that systems with "exposed" selectors have greater gender diversity. However, there are two distinct components to an "exposed" system: blame attribution and accountability. For citizens to hold selectors accountable for perceived bias, selectors must be susceptible to electoral or reputational rewards or sanctions. In the U.S. case, some judges are directly elected while others are selected by governors and/or merit commissions. While governors are electorally sensitive, members of the merit commission may be isolated from push-back because members come to hold their posts in various – often unelected or unaccountable – ways.

Even when selectors are electorally or politically accountable, for citizens to hold selectors accountable for bias, they must be able to *identify who is responsible for observed bias*. In some settings this can be quite clear. For example, in Bahrain,

constitutional court justices are appointed by the King (VDem Judiciary, Bahrain 2002) and in Bangladesh, Supreme Court justices are appointed by the President after he consults with the Chief Justice (VDem Judiciary, Bangladesh 2011). In both of these cases one individual is responsible for selecting judges. If an observer determines that the process is biased, it is clear who is responsible – either the king or the president. Of course, whether observers can hold those responsible for bias accountable varies across these two examples.

In contrast to a unitary appointer, a more complex selection process may make it more difficult for observers to determine responsibility for bias. For example, the seven members of the Constitutional Court of Niger are all chosen in different ways or according to different qualifications. In Niger, even if citizens observe a completely homogenous court, each judge came to his or her position in a different way.[1] Observers must disentangle responsibility, a task more complicated under Niger's institutions than under the selection process in Bangladesh or Bahrain.

Niger is, granted, a extreme example of a process where responsibility is difficult to assign, but existing literature suggests that visibility is related to blame attribution; cross-national evidence of government and executive policy responsibility shows that voters responded to economic performance more strongly in countries where responsibility for performance is relatively more clear (G. Bingham Powell, 1993; Anderson, 1995; Kevin M. Leyden, 1995; Robert C. Lowry, 1998). In selection procedures with multiple selectors, each individual actor may be able to avoid blame by spreading the blame among many, which will further obscure responsibility.[2]

The presence of multiple actors in judicial selection, then, may obscure the citizenry's ability to accurately attribute blame. If this is true, bias on the bench could persist even after observers infer that bias is occurring. In the context of the U.S.

[1] The 2017 constitutional court of Niger has 6 men and 1 woman serving. The female judge, Abdoulaye Diori Kadidiatou, is currently the president of the court.

[2] Weaver (1986) classifies the tactic of spreading blame among multiple actors as "circling the wagon". Two other blame avoiding tactics would apply to a multiple selector situation: "passing the buck" and "finding a scapegoat" (page 385).

courts, the debate between gubernatorial selection and merit selection is relevant, too, for blame attribution. Under gubernatorial selection, observers can easily identify who is responsible for disparity: the governor. Under merit selection, however, the governor's choices are limited by the commission. Not only might observers, then, spread blame between the governor and the commission, but among the commission any potential blame will be diffused across members.

Evidence from the survey experiment described in this chapter supports this expectation: under merit selection, blame for perceived bias is more diffuse than under gubernatorial selection.

Institutional Change and Slowing the Process of Diversification Third, I use observational data on selection procedures for peak courts cross-nationally to test whether institutional changes away from unitary appointment procedures slow the process of diversification, as theorized. I find preliminary evidence that moving away from unitary appointment slows the process of diversification.

4.1 Survey Experiment

To test how selection institutions affect the interpretation of information and blame attribution, I use evidence from a survey experiment. The survey was fielded in the summer of 2017, was developed and hosted on Qualtrics, and was completed by respondents recruited from Amazon's MTurk. Table 4.1 presents summary characteristics of the sample. Respondents for the analysis presented were required to reside in the U.S.A at the time they completed the survey and must have accurately answered a simple attention check question.

Respondents were shown information about a hypothetical court with seven members. They were told that judges were drawn from large and deep candidate pool that included many women. The names of seven judges were listed along with an arbi-

Table 4.1: Survey Respondent Characteristics

	Male	Female			
Gender	.43 (422)	.57 (556)			
	<25	25-34	35-49	50+	
Age	.08 (84)	.35 (346)	.34 (340)	.21 (208)	
	Very Cons.	Smwht Cons.	Moderate	Smwht Lib.	Very Lib.
Ideol.	.06(60)	.21 (213)	.23 (223)	.31 (301)	.18 (175)
	High School	Some Cllge	BA/BS	Masters	Doctorate
Edu.	.07(70)	.32 (313)	.40(393)	.15 (155)	.04 (40)

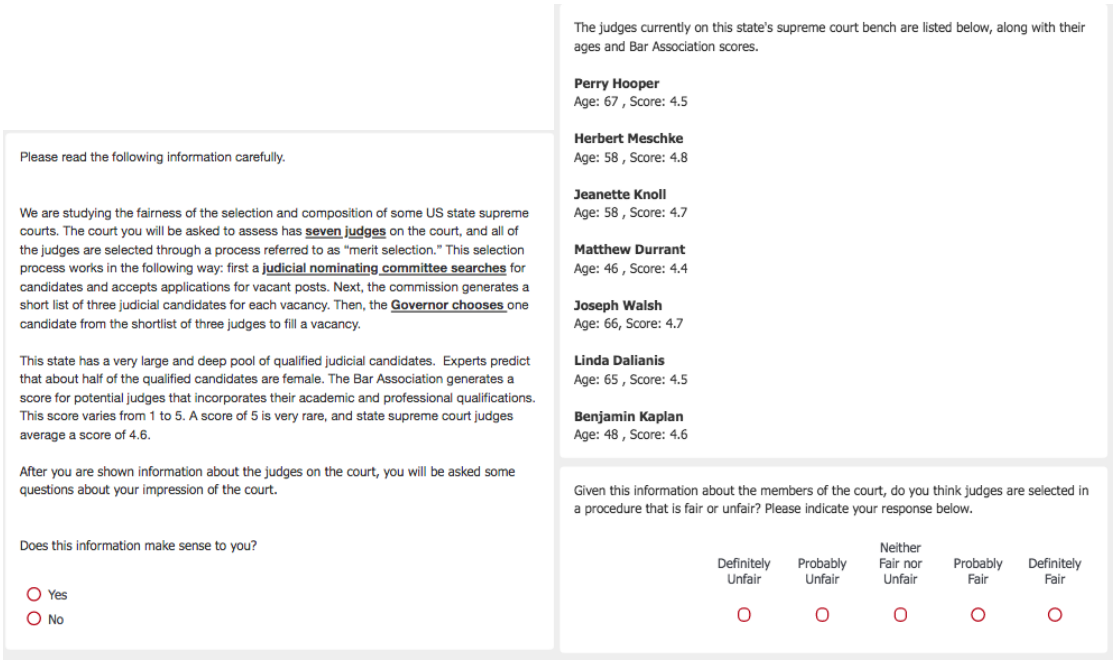
Summary characterisitcs for the MTurk survey respondents.

trary bar association score. There were two treatment variables. First, the number of women's names listed varied from zero to three. The names were drawn from a list of past and current state supreme court judges in which gender-ambiguous names were removed.[3] Second, half the respondents were told that judges were selected by the governor. The other half of respondents were told that judges were selected through a merit procedure in which a commission generated a short list of three names from which the governor chose one judge to fill a vacancy. Respondents were not given any information about the composition of the short-lists generated by the commission.

Based on the information given, respondents were asked if the selection of judges appeared fair or not using a five-point scale: Definitely Unfair, Probably Unfair, Neither Fair nor Unfair, Probably Fair, and Definitely Fair. Figure 4.1 shows the survey instrument for the merit selection and two women justices treatment group.

[3] I removed gender-ambiguous names from a list of all state supreme court judges selected between 1960 and 2010 from the State High Court Justice Database (*https* : *//www.lsu.edu/faculty/bratton/research.htm*) and then randomly drew male and female names from the list.

Figure 4.1: Example Survey Instrument



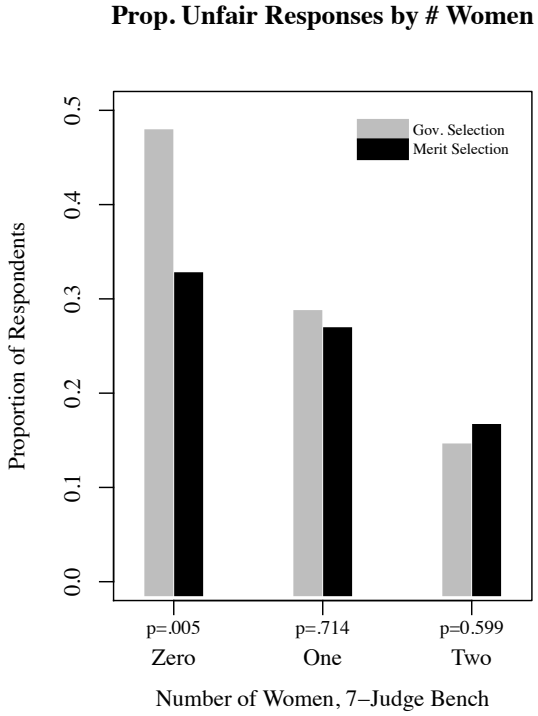
Example survey instruments for the merit selection and two women treatment group.

4.1.1 Results: institutions and the interpretation of information

If selection institutions have no effect on how respondents interpret information, perceptions of bias across different levels of gender diversity will be the same across both institutional treatments. Figure 4.2 shows the proportion of respondents who indicated that the process seemed unfair across the number of women judges *and* the selection institution type.

There is one notable difference in the interpretation of information across institution type: when there were no women on the bench, almost half of respondents in the gubernatorial selection group interpreted the gender disparity as evidence of bias. In contrast, only a third of respondents in the merit selection group interpreted the all-male court as evidence of bias. The p-value for this difference in perceptions of bias across institution type when there is one woman is small (p=0.005) and passes the Bonferroni correction. Conducting multiple tests using the same sample, as I do

Figure 4.2: Survey Results, Perceptions of Unfairness Across Institutions Type and Diversity



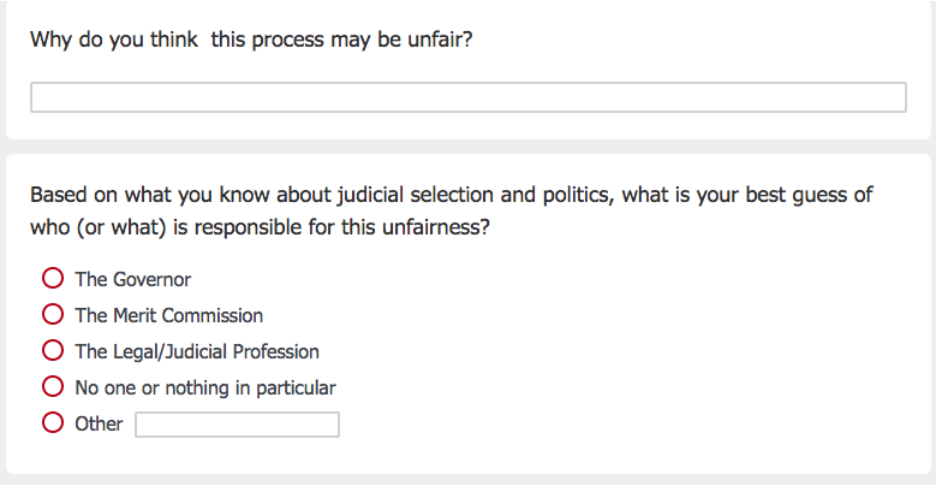
Perceptions of unfairness across levels of gender diversity and institution type. The bars show the proportion of respondents in each gender diversity treatment group who indicated that the process appeared "definitely unfair" or "probably unfair" disaggregated by institution type. The grey bars show the responses for those who saw gubernatorial selection institutions. The black bars show responses for those who saw merit selection procedures. When there were no women on the court, more respondents in the gubernatorial group perceived the process as biased realtive to those in the merit group (p=.005).

here, affects the probability of committing a type I error. That is, making multiple comparisons using the same sample increases the probability that two groups differ due to chance rather than a treatment effect. Using the Bonferroni Correction in which one rejects a null hypothesis when $p < \frac{\alpha}{m}$ and m= the number of comparisons, this difference in proportions is statistically significant. The adjusted p-value=.015.

4.1.2 Institutions and Blame Attribution

To determine whether selection institutions affect the attribution of blame, respondents who indicated that the process appeared biased where then asked, "Based on what you know about judicial selection and politics, what is your best guess of who (or what) is responsible for this unfairness?" Respondents were given a multiple-choice selection of responses, including an "other" response in which they could fill in an actor not included. Figure 4.3 shows the survey instrument for this portion of the study.

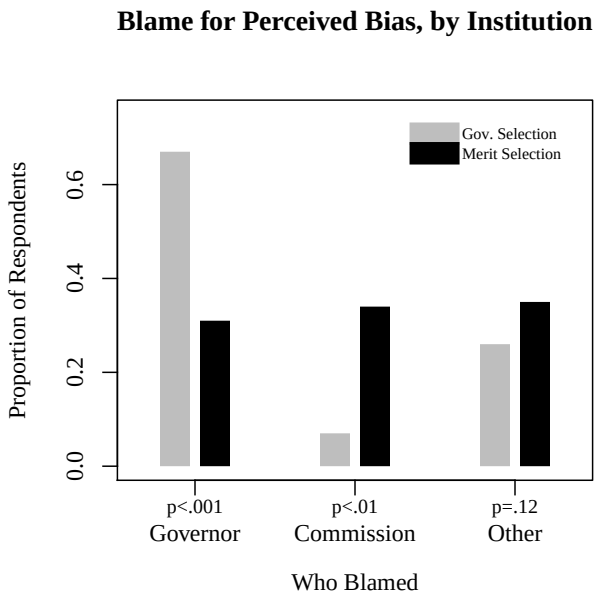
Figure 4.3: Survey Instrument, Blame Attribution



Those who were shown gubernatorial selection procedures most often blamed the governor. Those who saw merit selection were split evenly between blaming the governor and the commission. Figure 4.4 shows the blame responses across institutional treatment group.[4] What is interesting here is that respondents in the merit selection treatment group were not given any information about the gender composition of the short-list passed to the governor. Regardless of whether the governor or the merit

[4]A few respondents in the gubernatorial selection group blamed the merit commission even though respondents in that group were not given information about a merit commission. I attribute these responses to respondents who guessed or who were not paying attention (but still managed to pass the attention check). I provided "the merit commission" as a response for the gubernatorial treatment group to allow for direct comparisons across the two groups.

Figure 4.4: Results: Blame Attribution



The proportion of respondents in each treatment group who blamed the governor, the commission, or something else. As expected, under gubernatorial selection (grey bars), blame was concentrated on the governor. Under merit selection (black bars) blame was spread across the categories.

commission was responsible for disparate selections, then, the two actors shared the blame for perceived bias.

Chapter 3 showed that respondents believed, a priori, that merit selection systems were more fair than gubernatorial selection procedures due to the presence of multiple actors involved in the process. This chapter shows, first, selection institutions affect how observers make inferences about bias. Specifically, prior beliefs in the fairness of merit selection procedures lead observers to be more accepting of gender disparity under a merit selection system than under a gubernatorial selection system. Second, this chapter shows that the presence of multiple actors in selection obscures blame attribution. Under merit selection, blame is shared between the commission and the governor whereas all blame is concentrated on the governor under gubernatorial selection.

Taken together, these findings suggest that merit selection – or selection procedures in which blame is shared between multiple actors rather than a unitary actor – may dampen or slow the process of diversification.

4.2 Observational Empirical Analysis

Chapter 3 and the previous sections of this chapter indicate that merit selection procedures or procedures in which multiple actors are involved in judicial selection may undermine the process of gender diversification through prior beliefs in the fairness of selection, which serves to increase the amount of disparity required for making inferences about bias and through blame attribution.

In this section, I use observational, cross-national data to assess whether institutional change *away from unitary selection* dampens the process of gender diversification world wide. In the survey experimental work, I focused on "merit selection" in particular. In this section, I should note, I focus on non-unitary selection. While there is abundant variation in selection institutions cross-nationally, few countries use traditional "merit selection" as understood in the U.S. context.

Using a unique cross-national data set on the selection procedures for constitutional court judges collected for the Varieties of Democracy project and in collaboration with the Comparative Constitutions Project, I employ a matching design and signed-rank statistic to test for the effect of institutional change on the timing of diversification (Arrington et al., 2018).

If the presence of multiple actors undermines gender diversification, we should expect countries with multiple actors to have less gender diversity, on average, than other countries. Importantly, institutions and outcomes can be endogenous: expectations about the ability of a woman to successfully become a peak court justice under a specific selection system will shape whether she decides to pursue the qualifications necessary to hold the post, for example. To isolate the effect of institutions, therefore,

I focus on institutional change. Assuming constitutions are not changed with the *goal* of increasing gender diversity on the bench, I can treat changes as exogenous.[5] Specifically, I am interested in the gender diversity outcomes for countries that experienced a constitutional change that changed the peak court judicial selection systems from unitary selection to selection with multiple actors.

Once I identify the countries that experienced institutional changes expected to dampen diversification, I match those countries to control countries that also experienced a change, but not a change theorized to slow diversification. I match to countries that also experienced a change to their peak court selection institutions because Arrington et al. (2018) show that institutional changes – in general – to judicial selection process accelerate the process of gender diversification. I match the countries that experienced a particular change to countries that experienced different changes to avoid inaccurately attributing the effect of institutional change in general to the effect of a specific institution.

Defining the Treatment Variable As stated, the treatment variable for this analysis is a particular institutional change: going from having a unitary appointer to having more than one actor involved in the process. Table 4.2 shows examples of institutional changes that qualify as treatment for this analysis. Here, "unitary" actor refers to an institutional actor, so the legislature counts as one actor. A process that has the president appoint candidates and then the legislature approve the appointment counts as two actors. Treating legislatures as a unitary actor is a function of the coding protocol and is not ideal for this test because legislatures will obscure the processes of blame attribution and accountability. Classifying legislatures as unitary actors makes this test a conservative one.

[5] There is no evidence that the countries included in this analysis changed their judicial institutions with the goal of increasing gender diversity. Most changes to the judiciary coincide with broad constitutional changes.

Table 4.2: Examples of Treatment

Country	Year	Summary of Change
Afghanistan	2004	From king appoints to president appoints and house endorses
Albania	1976	From Assembly election to presidential appointment and assembly approval
Angola	1992	From Presidential appointment to three pathways to selection: National assembly elects some, President appoints some, and the Supreme Court chooses some judges
Belarus	1996	From elected by the Supreme Council to the President appoints half of justices and the Supreme Council elects half of justices
Benin	1984	From the Presidential Council of the Republic appoints to the Executive Council appoints and the National Assembly approves
Bulgaria	1991	From the National Assembly elects to three pathways: the National Assembly appoints four, the President appoints four, and a joint meeting of the Supreme court of Appeals and Administrative court appoints four
Croatia	2010	From parliament elects to the Committee of the Croatian parliament nominates and then the Croatian parliament elects by a 2/3rds majority
Dem. Rep. Congo	2005	From president appoints to three pathways: Parliament appoints three, the High Council of the Judiciary appoints three, and the President appoints three
El Salvador	1994	From the Legislative Assembly chooses judges to the National Council of the Judicature forms a list and the Legislative Assembly chooses judges from the list
Ethiopia	1987	Form the Emperor appoints to President nominates and Shengo (legislature) appoints.
Ghana	1979	From the President appoints to the President appoints and Parliament confirms

Examples of countries that changed from unitary selection to having multiple actors involved. Here, "unitary" means one institution, so the legislature, for example, counts as a unitary actor.

Dependent Variable I use as my dependent variable the time until the selection of the first woman judges rather than the total number of women on the court due to data limitations. It is much easier to accurately identify the date the first woman was selected than to identify the total number of women on the court over time. Attempts to collect the number of women on the court over time remain incomplete.[6] I therefore address the timing of the first woman judge rather than total levels of gender diversity over time.

Hypothesis: Countries that experienced an institutional change away *from unitary selection will take longer to have their first woman justice relative to countries that experienced a change to rules that did not change from unitary selection.*

Matching I match treated countries that experienced an institutional change from unitary selection to multiple actors to control countries that *also experienced a change,* but not a change from unitary selection. I match countries on the decade of change and the percentage of women in the lower house of the legislature.[7] Matching on decade controls for time and matching on percentage of women in the legislature controls for pressure for the selection of women to important political posts (Hoekstra, Kittilson and Bond, 2014). In addition to data limitations in the dependent variable, there are a limited number of countries that have experienced institutional changes. Given the data limitations on both the right-hand and left-hand sides, there are six matched pairs for which I have dependent variable data. Table 4.3 lists the six matched pairs used in this analysis.

[6] A team of researchers at Emory, lead by Leeann Bass, attempted to collect the total number of women on the bench over time, but substantial data missing-ness threaten the accuracy of the data. However, researchers were able to identify the year of the first woman more accurately.

[7] Using "optmatch" in r.

Table 4.3: Matched Pairs with DV Data

Treatment Country	Year Δ, T	Year 1st Wom., T	Control Country	Year Δ, C	Year 1st Wom., C
Afghanistan	2004	None	Tunisia	2014	None
Montenegro	2007	2008	Comoros	2001	2007
Nicaragua	1987	1995	Chile	1980	1989
Seychelles	1993	2011	Chad	1996	1998
Fiji	1997	2012	Ecuador	1995	1997
Romania	1991	2004	Cent. Afr. Rep.	1994	2005

Matched pairs used for the analysis. The left-most column shows "treated" countries (change from unitary selection). The next columns show the year of the institutional change and the year of the first woman justice on the peak court. The fourth column shows the matched control country. The fifth and sixth columns show the year of the constitutional change and the year the first woman judge selected to the peak court for the control country.

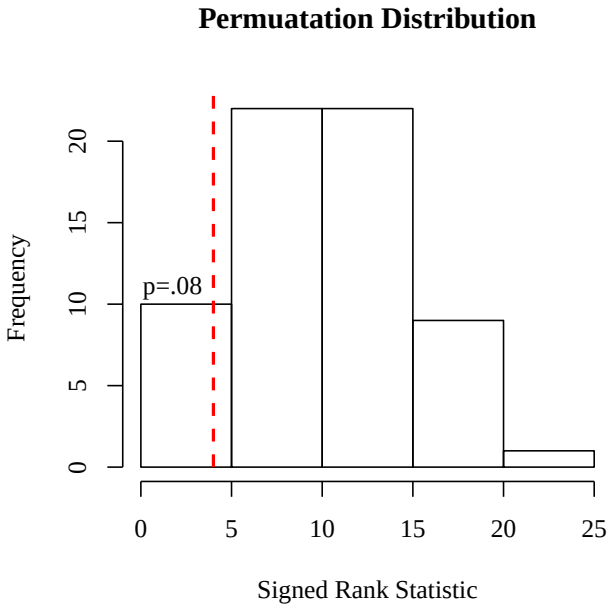
Analysis and Results Among the six matched pairs, the average number of years from an institutional change to the first woman peak court justice for treated countries (those that moved away from unitary selection) is 11.3 years. In contrast, the average number of years from institutional change until the first woman peak court justice for control countries was 5.5 years.

To test if an institutional change away from unitary selection affects the timing of the first woman justice, I use a sign-rank statistic (see Glynn and Ichino (2015)). For all treatment and control countries, I calculate the number of years between treatment and the selection of the first woman to the peak court.[8] Then, for each pair, I calculate the difference in the number of years until the selection of the first woman between control and treated countries.[9] Next, I rank the pairs by the size of the difference in years (the absolute value difference) from smallest to largest difference. That is, the pair with the smallest difference in years until the first woman is assigned a rank of 1 and the pair with the greatest difference in the number of years until the first woman is assigned the highest rank. The signed rank statistic is then calculated as

[8] I drop countries from the sample that had women prior to the institutional change.

[9] That is: (year first woman for a control country - treatment year for a control country) - (year first woman for a treated country - year of institutional change for the treated country)

Figure 4.5: Signed Rank Statistic: Time to First Woman



Among the six matched pairs in this data set, the control countries– on average – had fewer years between institutional change and the selction of the first woman justice (5.5 years) than treatment countries (11.3 years). The probability that the signed rank statistic would be smaller than the observed statistic if treatment was randomly assinged among matched pairs is .08.

the sum of the ranks between treatment and control countries for which the treated country "won" by selecting a woman to the court before the control country. For example, if there were five pairs and in each pair the treated country selected a woman in fewer years than the control country, the signed-rank statistic would be equal to 1+2+3+4+5. This statistic is larger when more of the pairs have a "winning" treated country and/or when the pairs in which the treated country "wins" have the greatest difference (the highest ranks).

To determine whether the observed signed rank statistic is unusual or not, I generate a reference distribution by permuting assignment of treatment and re-calculating the singed rank statistic for each permuted sample. This procedure results in a distribution that shows all possible signed rank statistics and the frequency of observing the signed rank statistic among all permutations. Then, I compare the observed signed

rank statistic to the permutation distribution to determine if the observed statistic is in the tails of the distribution and, therefore, unlikely due to chance. Figure 4.5 shows the permutation distribution and the observed sign rank statistic. The signed-rank test shows a negative (that is, control countries got a woman before treated countries) but statistically insignificant (p=.08) effect of a change away from unitary selection on the time until the selection of the first woman.

While we can not reject the null hypothesis of no relationship, the evidence here is suggestive that a movement away from unitary selection may indeed slow the process of diversification, as theorized.

4.3 Conclusions

In chapter 3, I used evidence from survey experiments to show that observers believe – prior to observing outcomes – that merit selection in which a commission nominates judges and the executive selects a judge is more fair than gubernatorial selection in which a unitary actor has all the power to select judges. Qualitative responses from respondents suggested that preferences for merit selection stem from the belief that having more actors involved in the process that limit the power of each other is more likely to result in "good" outcomes. In section 4.1, I showed that observers needed to observe *more* disparity before concluding that the process was biased under merit selection relative to gubernatorial selection. In section 4.1.2, I showed that the presence of multiple actors in merit selection can obscure blame attribution. Under merit selection, blame was shared between the commission and the governor (absent any information about who was actually to blame) whereas blame was concentrated on the governor under gubernatorial selection.

The evidence from the survey experiments in sections 4 and 4.1.2 suggest that the presence of multiple actors in selection systems like merit selection may undermine the process of gender diversification via two mechanisms: updating beliefs and blame

attribution. To test whether the presence of multiple actors is associated with delayed gender diversification cross-nationally, I employed a signed-rank statistic on matched data in which treatment is defined as a movement away from a unitary actor. There is a negative but statistically insignificant (p=.08) relationship between an institutional change away from unitary selection and the timing of the selection of the first woman justice to the peak court.

Chapter 5 Empirical Evidence: Accountability[1]

The final step in the framework of diversification outlined in chapter 2 is accountability. Once observers have inferred that bias in selection is occurring and identified who is to blame, they must hold those responsible for bias accountable. In this chapter, I test an observable implication of accountability. Specifically, if elites tasked with selecting judges expect to be held accountable for homogenous selections, should avoid gender homogeneity. Indeed, Bratton and Spill (2002) show that women are most likely to be selected to diversify all-male courts. Once a court has a woman justice, the probability another woman is selected decreases. Once diversified, however, those selectors who expect to be held accountable for homogeneity should work to prevent reversion to all-male or less diverse courts. An obvious way to prevent reversion to a less diverse or all-male court is for elites to seek out women justice to replace vacating women justices. In this chapter, I address patterns of gender and judicial replacement. I find that women are disproportionately selected to fill vacancies made by women, which is consistent with the expectation that selectors feel pressure to maintain levels of diversity on the bench.

5.1 Diversity, Quotas, and Gendered Patterns of Replacement

Despite the importance of gender diversity in office, the use of gender as a selection criterion is controversial, particularly in the United States.[2] While many countries

[1] This chapter is excerpted from Arrington (2018).

[2] "Gender" is often conceptualized as a spectrum of identity. In this project, I use gender to refer to how individuals present in the public sphere, either male or female. For practical purposes – data availability and model simplicity – I measure gender with a binary variable.

have turned to formal and informal gender quotas for office (Dahlerup, 2008), gendered selection criteria in the U.S. remain contentious (Baldez, 2006; Krook, 2006*a*).[3] Opponents criticize quotas for dismantling merit selection and for favoring certain groups of people over others; they argue that descriptive characteristics should not be – and, presumably, absent these policies are not – salient selection criteria. This argument has been particularly forceful in the judicial context, where gender is viewed as inconsequential to ones interpretation of the law: in the words of Minnesota Supreme Court Justice Jeanne Coyne, for example, "A wise old man and a wise old woman reach the same conclusion" (Margolick, 1991).[4] It is possible, however, that descriptive characteristics are important features of the selection process even in the absence of quotas, especially if people are attentive to and there is pressure for diversification (see Goelzhauser, 2011, p.776).

Indeed, studies indicate that gender is especially relevant for diversifying all-male courts (Bratton and Spill, 2002), but anecdotal evidence suggests that gender matters for judicial selection on courts that have already diversified too. Specifically, there are many instances of women judges retiring and being replaced by other woman judges. If gender is irrelevant in the selection process, we would only rarely observe women judges replacing women judges: given slow turnover and the historical dearth of women on state supreme courts, women replacing women by chance would be uncommon. In contrast, a pattern of women judges systematically replacing women judges would suggest a pattern of implicit reserved seats in which women replace women even though there is no formal rule requiring it.

If gender *is* a relevant selection criterion and women judges are being selected to replace women judges, we should consider both the causes and effects: why might judicial selectors choose or promote women judges to replace women judges? Where

[3] I use "gendered" to refer to phenomena in which "gender is present in the [process]," either implicitly or explicitly (Acker, 1992, p. 567).

[4] Some scholarship refutes this claim that gender and experiences of men and women are inconsequential in judging. See, for example (Martin, Reynolds and Keith, 2002) and (Martin, 1991).

state supreme court judges are elected, why might women choose to run to fill vacancies by women judges? And, has the pattern of women replacing women on the bench increased the presence of women judges, as traditional quotas aim to do, or have these patterns restricted the presence of women judges by limiting women to just one or a few seats?

There are at least two explanations for a gendered pattern of replacement on courts. For one, replacing women judges with women judges could be a tool for the continued exclusion or under-representation of women. From this perspective, gendered patterns of replacement create and perpetuate tokenism on the bench. By allowing but limiting the presence of woman judges to one or a few specific seats, "tokenism is ... symbolic equality" (Greene, 2013, p.82) that outwardly demonstrates a commitment to equality without addressing the underlying social and political treatment of historically marginalized groups (Laws, 1975, p.51). By limiting women judicial candidates to vacancies made by women candidates, a token seat or seats excludes women from other vacancies and limits diversity on the bench.

Alternatively, patterns of women replacing women may promote diversity on the bench. Rather than limiting women to one or a few seats, gendered patterns of replacement may ensure diversity by promoting the selection of women when courts might otherwise revert back to all-male or less diverse. This explanation for gendered replacement would be particularly beneficial for diversity when there are few women in the traditional candidate pools. The historical exclusion of women and minority individuals from higher education and posts that serve as informal qualifications for judicial office has limited and continues to limit the diversity of the candidate pool. In this setting of limited availability, patterns of gendered replacement might encourage judicial selectors to seek out those women judges who *are* qualified when they might otherwise select male judges.

For seats in which judges are elected, the same forces can be at play. If party

officials and donors only seek out and support women candidates for judicial races to fill vacancies by women – or if they discourage women from running for vacancies made by men – they could depress the presence of women on the bench. In contrast, if party elites and donors actively seek out women to run for a vacancy made by a woman when another woman might not run otherwise, they would be promoting diversity on the bench.

From the outset, it is unclear of gendered patterns of replacement have been beneficial for overall levels of diversity or not. Even if patterns of gendered replacement have been net positive for gender diversity on state supreme courts, these patterns may still have some pernicious outcomes for women (or minority) judges individually. For example, these patterns of women replacing women are consistent with concepts of "tracking" in which "indirect" bias funnels women and minority officials into particular posts or types of posts (Reingold, 2018). Women judges selected to replace women judges may also be subject to pressures of tokenism or to backlash for perceived gender-based favoritism. It is also important to note that the use of gender as a selection criterion can be explicit or implicit. Judicial selectors may actively and knowingly seek out women judges to replace women judges, but they may also do so unintentionally. Uncovering the internal motivations of judicial selectors and the experiences of individual women judges on the bench is beyond the scope of this project. Instead, I test whether patterns of gendered replacement are systematic and whether those patterns have suppressed overall gender diversity on the bench relative to a counterfactual in which gender is not a selection criterion.[5]

In this chapter I test whether women judges are more likely to fill vacancies made by women relative to vacancies made by men. Because broad structural forces such as the women's movement have resulted in the diversification of many professions over time and are correlated with the gender of both departing and replacement judges

[5]Given the same candidate pool.

across the U.S. states,[6] I adopt a matching design to ensure credible comparisons of judicial turnover over time and across all U.S. states.

I find that when female judges retire, a much greater proportion of replacement judges are women relative to when a male judge retires, which means that gender is a relevant selection criterion for state supreme courts. To interpret this result, I compare rates of selection to the gender composition of lawyers over time. In the aggregate, women are selected to state supreme courts about as often as expected given the composition of the candidate pool. This is not to say that women judges and judicial candidates do not face implicit or explicit biases that hinder the acquisition of prestigious judicial posts or burden their experiences once in office. Instead, the evidence from this analysis suggests that at the final selection stage, women are neither systematically excluded from state supreme court benches nor unfairly advantaged.[7] In the next sections, I describe the research design used to identify the pattern of women judges replacing women judges and compare those results to patterns of diversification in the candidate pool.

5.2 Data and Matching

Assessing whether the gender of vacating and replacement judges are independent presents a few methodological challenges. First, the gender of the retiring judge is by no means the *only* explanation for the gender of the replacement judge, so we must isolate the effect of gender, all else equal. Second, more women have been able to accumulate the qualifications for office over time. As the feminist movement took hold and played out, citizens came to accept women on courts and – for some – expect

[6] Specifically, these broad structural forces that affect both the independent and dependent variables would invalidate an empirical strategy in which the gender of a departing judge is used simply to predict the gender of the replacement

[7] It is worth noting that *presence* of women judges is not the same as *equality* among men and women judges. Likewise, the absence of exclusion is not the same as the absence of discrimination. This project addresses the presence of women judges; it does not speak to the experiences of women candidates and judges.

women on courts.[8] These over-time pressures mean that women are more likely to be selected to a vacancy over time, regardless of the gender of the retiring judge. To accurately identify the effect of gender, we need to account for time trends. Finally, to determine the effect of the gender of the vacating judge, there must be vacating judges *who are female*. Assessing the effect of the gender of the vacating judge has only recently been possible as more women judges have been selected to and have left state supreme court benches, so we need to manage inconsistent data availability over time.

To address these methodological concerns, I do two things. First, I use non-parametric matching to generate a data set of treatment (a woman retires) and control (a man retires) cases that share theoretically and empirically important characteristics. This allows me to isolate the effect of the gender of retiring judges on the gender of replacement judges by comparing the outcome across the treatment and control.[9] I match on time, the number of women, court size, the number of vacancies, and selection institutions to ensure plausible comparisons across treatment and control units. Second, I use a Cochran-Mantel-Haenszel (CMH) test of proportions to determine if the gender of the retiring judge is independent of the gender of the replacement judge. The CHM test is flexible to homogenous effects across time and other covariates and can accommodate differences in sample size across strata, which occurs in the data because more women have vacated the bench in recent years. Before describing the results of the CMH test, I detail the data and the matching procedures used.

[8] See, for example, Andrews (2016) and Bakalar (2016), two newspaper opinion pieces advocating the selection of a woman judge to the Alaska Supreme Court.

[9] The ideal way to identify and isolate the causal effect of a vacating judge's gender on the gender of the replacement judge would be to randomly assign gender to judges. Then, we could compare the gender of vacating judges to the gender of replacement judges knowing that all other traits are equal. Clearly this is not a feasible design. Instead, matching observations increases the balance across cases in which women and men retire, which helps isolate the causal effect of vacating judge's gender on replacement gender (see Boyd, Epstein and Martin (2010) for a more detailed discussion of matching and causal inference).

5.2.1 Data

The gender of judges retiring from and selected to state supreme courts comes from Kathleen Bratton's State High Court and Justice Database.[10] The data set includes judges selected to all 50 state supreme courts between 1960 and 2010 and describes how justices were selected, when they were selected, when they retired or vacated the court, and their gender.[11] I updated the data set to include judges who retired or were selected between 2010 and 2016 with information from Ballotopedia.[12] I restrict the data to the years between 1970 and 2016 to avoid data missingness in the early years.

The data are re-shaped so that the unit of analysis is state-years for which a judge vacates. While it is most common for a court to have only one vacancy at a time, there are many courts and years with multiple vacancies (see Table 5.1). Aggregating to state-year rather than treating each vacancy as the unit of analysis avoids an independence problem in cases with multiple vacancies: when two or more judges are replaced at the same time, the characteristics of one replacement judge might affect the probability the other replacement judge holds certain characteristics as well. Furthermore, evidence from the legislative arena suggests that the selection of multiple candidates at once, such as on a party list, is associated with increased gender diversity (Salmond, 2006; Paxton, 1997; Kenworthy and Malami, 1999). By aggregating to state-year and then matching on the number of vacancies, I avoid potential bias from

[10] http://www.lsu.edu/faculty/bratton/research.htm

[11] Texas and Oklahoma each have two high courts, one for civil cases and one for criminal cases. The criminal courts are not included here.

[12] Of the 1407 judges selected between 1960 and 2016 in this combined data set, there are 243 judges identified as women; 85 identified as Black or African American; 28 identified as Latinx, and 19 identified as Asian-American of Pacific Islander. Of the 243 women judges, 21 are Black women, 8 are Latina, and 8 are Asian-American/Pacific Islander women. Indicator variables for gender and Black judges are included in the State High Court and Justice Database. Indicator variables for Latinx and Asian American/Pacific Islander judges were generated through searches of NALEO directories and the Asian American and Pacific Islander Almanacs. In addition, keyword searches of Asian American, Pacific Islander, and Latino/a judges were used to augment missing directory and almanac years. Finally, when photographs of judges suggested a racial or ethnic identification, electronic newspaper searches or biographies were used to confirm race or ethnicity.

the interdependence of vacancies and I control for potential incentives to select women when there are multiple vacancies.

I link vacancies to replacements by time. If a vacancy and a replacement occur in the same year, those two judges are linked as a vacancy/replacement pair. Importantly, retirements and replacements do not always occur in the same calendar year. For example, a judge may retire in one year but her replacement may not be selected until the next year. In these cases when judges vacate or are selected in different years, I aggregate across two years to link the vacancy and replacement (see the Appendix for a description of the coding rules for aggregating two years). After cleaning and aggregating the data, I am left with 671 units that correspond to state-year(s)-vacancy(s) observations.[13]

The main independent variable (the "treatment" variable) is a dummy variable indicating if the vacating judge is female, and the dependent variable is a dummy variable indicating the gender of the judge selected to fill the vacancy. In cases where there is more than one vacancy and replacement, the dummy variable indicates if any of the vacating or replacing judges are female.[14] Table 5.1 summarizes the frequency of the treatment and dependent variables for the 671 units prior to matching. There are 109 "treated" cases in which a woman vacates the bench. In the next section I describe how I match those treated cases to control cases and then test for a relationship between the gender of the vacating judge and the gender of the replacement judge.

[13]To avoid counting a judge as his or her own replacement, I remove 44 judges who were selected and removed in the same year. When possible, I connect these brief-tenure juges's predecessors to their replacements; that is, I just skip over judges who served for less than a year. Additionally, there are 107 state-year-vacancies in which the number of vacancies and the number or replacements do not match up. These observations do not fit the criteria for two-year aggregation described in the appendix. For 49 of these observations, there is a vacancy but no corresponding appointment. For 58 of these observations there is at least one vacancy and one replacement, but the number of vacancies and replacements do not correspond. I remove these observations to prevent incorreclty attributing replacements to vacancies.

[14]To be clear, I match observations on the number of vacancies, so courts with two vacancies are only compared to other courts with two vacancies.

Table 5.1: Frequencies: Treatment and Dependent Variable

	Male Only	Female	Total
Vacating Judge (TV)	562	109	= 671
Replacement Judge (DV)	476	195	= 671

The gender of retiring and replacement judges in the cleaned but un-matched data set. The majority of retiring and replacement judges are male. There are 109 "treated" units in which a woman judge vacates.

5.2.2 Matching

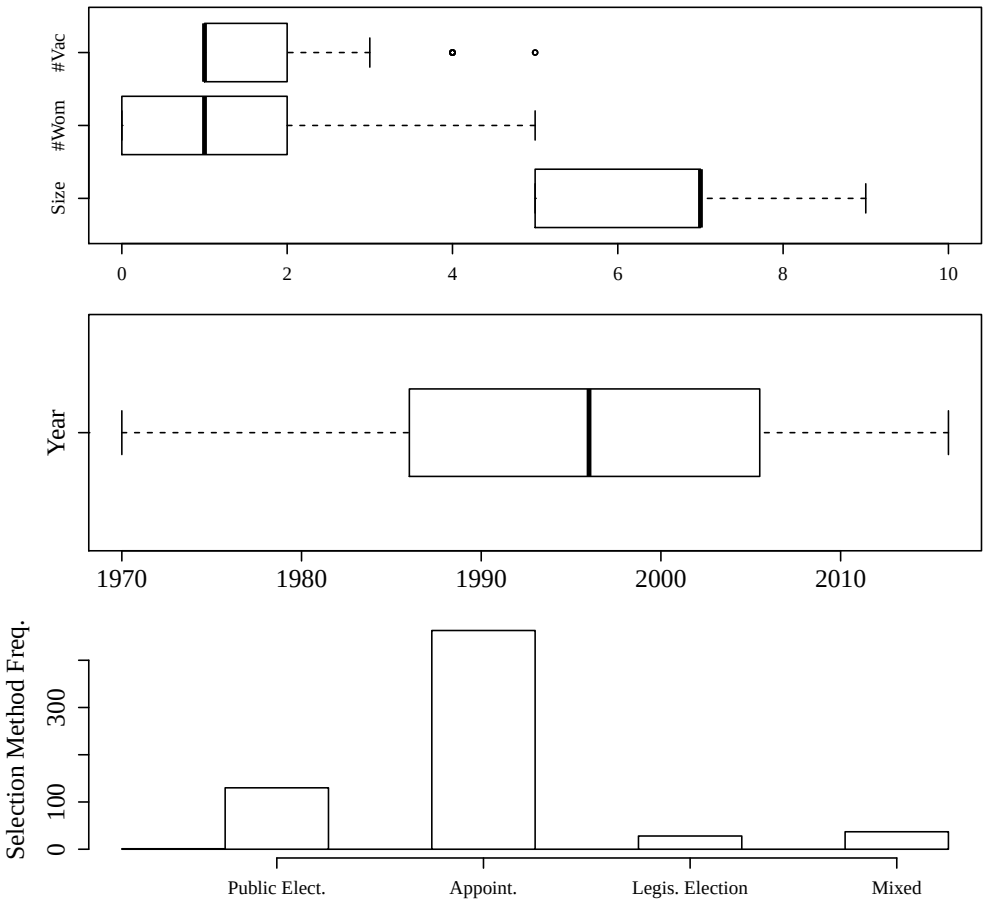
In order to better approximate an experimental study, I employ a matching design to minimize imbalance, which in turn reduces model dependence and bias (King and Zeng, 2006; Iacus, King and Porro, 2011). Cases are matched based on characteristics that affect the probability that the replacement and/or retirement judge is female. I match treatment and control cases on the (1) number of women on the court, (2) the size of the court, (3) time, (4) the number of vacancies, and (5) the selection method.

Matching on the number of women on the court is important for two reasons. First, the greater the number of women on the bench the greater the probability that any given vacancy is made by a woman. Second, as Bratton and Spill (2002) show, the number of women on the bench is negatively associated with the probability a woman is selected. Importantly, because it is only possible for a woman judge to retire from a bench on which there is at least one woman, all matched data have at least one woman on the bench.[15]

The size of the court is an important matching variable for three reasons. First, court size affects the probability that there is a vacancy. All else equal, the more judges there are on a bench the greater the opportunities for a vacancy. Second, the court size affects our interpretation of gender diversity. One additional woman has a greater effect on the gender composition of a 5 person court than a 9 person court. Third, extant literature suggests that women are more likely to be selected to larger

[15] This count includes the gender of the vacating judge.

Figure 5.1: Data Summary, Unmatched Data



The top panel shows distributions for the number of vacancies, the number women on the court, and the court size for the full, unmtched data-set. The middle panel shows the distribution of observations across time, and the bottom panel shows the distribution of selection methods. For the selection method, "public election" refers to both partisan and non-partisan elections. "Appointment" refers to gubernatorial appointment and merit selection, and "mixed" refers to years in which multiple judges were selected and the judges were selected in different ways.

courts (Williams and Thames, 2008; Cook, 1984).

Matching on time controls for the relationship between the presence of women on courts over time. More women have been selected to and have retired from state supreme courts in more recent years. In addition, matching on time controls for variation in pressure to diversify courts over time.

Matching on the number of vacancies is important for two reasons. First, the probability a woman vacates or is selected increases in the number of vacancies. Second, if the incentives for selecting women candidates in the legislative context apply to the judicial context, women may be more likely to be selected when there are multiple vacancies (Salmond, 2006; Paxton, 1997; Kenworthy and Malami, 1999).

Finally, observations are matched on selection method. Selection methods are grouped into three categories: popular election (both partisan and non-partisan), selection by elites (gubernatorial selection or merit selection), and legislative election.[16] While these groupings of selection procedures are broad and obscure variation within groupings, these categories capture important variation in opportunities for accountability over judicial selections and in the amount of coordination required to select judges. In popular election systems, accountability for the composition of the court is very diffuse, which means that any sanctions for perceived exclusion will be diluted, and credit-claiming for diversifying (Valdini and Shortell, 2016) will be less lucrative. In addition, coordination requirements to mobilize for the selection of a woman judge are high in the electoral context. In contrast, when only one or a few elites choose judges, accountability for a homogenous court is more easily attributed to those responsible and fewer people must coordinate in order to intentionally select a woman judge. Matching on these broadly-defined selection methods ensures that general patterns of accountability and coordination are held constant across treated

[16] In cases in which there were multiple vacancies and the replacement judges were selected through different procedures, the selection process was classified as "mixed," however none of these were successfully matched, which is not surprising given the few observations with mixed selections.

and control groups without seriously restricting the ability to successfully match. Figure 5.1 summarizes the matching variables for the full, unmatched data set.

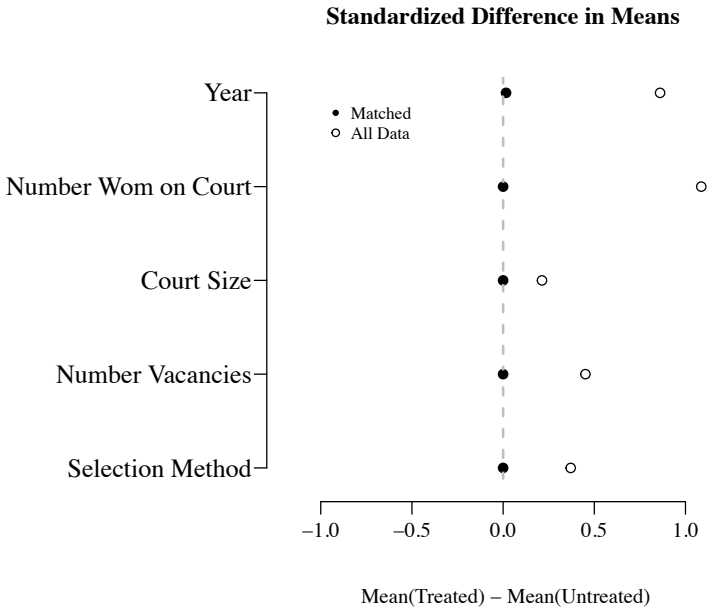
Treatment and control cases are matched using Coarsened Exact Matching procedures (Iacus, King and Porro, 2011) in the "MatchIt" package in R. Observations are *exact* matched on the court size, the number of women on the bench, and the three-category selection method. For these variables, the difference between a value of one and two, for example, is substantively different, so treated cases should only be matched to control cases that share the exact values for those variables.

For the year variable, cases should be matched within the same social and political context but need not be matched in the exact year. The difference between a vacancy in 1994, for example, and 1995 is not substantively meaningful. Moreover, vacancies on state supreme courts are relatively rare: only about half of the states in any given year have a vacancy on the court. Matching exactly on year seriously restricts the number of matched pairs. For these two reasons, the year variable bins are coarsened according the CEM coarsening algorithm, which matches treatment cases to control cases within five-year spans.

Figure 5.2 shows a balance plot that summarizes how treated units compare to control units, both for the full, unmatched data set (open circles) and for the matched data set (black dots).[17] The farther from zero the standardized difference of means is, the greater the difference between treated and untreated observations. There are substantial differences across the treated and untreated groups in the unmatched data: treated units occur later (greater value for year); there are slight differences in the frequency of selection method; treated units have more women on the bench, have more vacancies, and have a larger number of seats. If treatment and control units are matched appropriately, balance should improve and the differences in means of the

[17]The standardized difference in means are: [(mean of treated units - mean of untreated units)/ the standard deviation]. The standardized means of matched units are calculated with the CEM weights.

Figure 5.2: Balance Plot



This plot shows the standardized difference in means between the treated and untreated data for the full data set and for the matched data set. because each treatment case can be matched to multiple control cases, standardized differences in means are weighted.

matched data should be closer to zero than the differences for the full data set. As the black dots in Figure 5.2 show, balance is greatly improved.

5.3 Results

The Coarsened Exact Matching procedure groups treated and control cases into strata; each strata contains at least one treated unit and all the matched control units, of which there can be many. Table 5.2 shows a contingency table for the treatment and outcome variables for the matched data across all strata. The rows correspond to the gender of the vacating judges, and the columns show the proportion and number of vacancies filled by male and female judges. In this matched data set, when a male judge retired, more than three quarters of replacement judges were likewise men. Only 24.8% of vacancies made by male judges were replaced by female judges. In contrast, when a woman judge retired, almost half (44.8%) of her replacements were

also women.

Table 5.2: Matched Data, Contingency Table

		Replacement Judge		
		Male	Female	Total
Vacating Judge	Male	.75 (100)	.25 (33)	=1 (133)
	Female	.55 (32)	.45 (26)	= 1 (58)
C-M-H	χ^2=6.8	p=.01		

This contingency table summarizes how many vacancies were made by men and women and how many of those vacancies were filled by men and women. Proportions and total number of cases (in parentheses) are reported. The Cochran-Mantel-Haenszel test (C-M-H) generates a Chi-squared statistic of 6.8 with a corresponding p-value of .009, which means that we can reject the null hypothesis that the gender of the replacement judges is independent of the gender of the vacating judge.

To determine if the proportions of female judges selected as replacements are sufficiently different across male and female vacancies, I use a Cochran - Mantel - Haenszel test (CMH).[18] The null hypothesis of the CMH test is that there is no association between treatment and outcome variables. Under the null hypothesis, the proportion of women selected to fill a vacancy is independent of the gender of the vacating judge. For these data, the CMH test produces a χ^2 statistic of 6.8 with one degree of freedom.[19] The proportions of male and female judges selected to fill vacancies made by men and made by women are sufficiently different to reject the null hypothesis that the gender of the replacement judge is independent of the gender of the vacating judge (p-value=.009). This test indicates that the gender of the retiring judge does affect the gender of the judge selected to fill the vacancy. Vacancies made by women judges are filled by women judges at a greater rate than vacancies made by

[18]The CMH test is appropriate for binary treatment and outcome characteristics across differently sized strata (McDonald, 2014, p94-100).

[19]Calculated with the "mantelhaen.test" package in r.

men, and vacancies made by men are filled by men at a greater rate than vacancies made by women.

While this analysis demonstrates that gender is relevant in the selection process, it does not distinguish between a positive or negative outcome for judicial diversity overall. In the next section, I compare patterns of judicial selection to the gender composition of lawyers to determine if patterns of gendered replacement have suppressed or promoted the selection of women judges to state supreme courts.

5.4 Gender, Judicial Replacement, and Diversification

If the pattern of women judges replacing women judges requires judicial appointers or party elites to actively seek out women candidates when they otherwise would not, then this pattern may promote gender diversity. In contrast, if the pattern of women replacing women *limits* opportunities for women judges by restricting them to one or a few seats, then this pattern could suppress diversification. To investigate whether current patterns of selection promote or suppress the selection of women to state supreme courts, I compare the observed rates of selection to a counterfactual in which patterns of selection do not depend on the gender of vacating judges.[20] I use the gender composition of lawyers as a proxy for the candidate pool for state supreme court judges.[21] Then, I estimate the rate at which women *would* be selected to state supreme courts if gender were not a relevant selection criterion under existing and broadly defined standards of what it means to be qualified.

It is important to note that there is no agreed upon, ideal level of gender diversity on state supreme courts. In some contexts, such as the legislative context, standards

[20] I am not estimating what the rate of selection of women judges *should* or *ought* to be. For a discussion about the *right* or *just* rate of selection of women judges see, for example, Kenney (2013) or Mansbridge (1999).

[21] Cook (1978, 1984); Alozie (1996); Kenney (2012); Solberg and Bratton (2005) and others have also used the gender composition of lawyers as an estimate of or variable affecting the candidate pool for judicial posts.

of descriptive representation and diversity are compared to the composition of the population. The assumption, sometimes implicit, is that if political institutions are open and available to all, then the characteristics of the representatives should generally mirror the characteristics of the population. In the context of high courts, though, the vast majority of the electorate is not formally *eligible* for office under existing rules. At least 38 of the 50 states, for example, require state supreme court judges to be lawyers.[22] Requirements include practicing law for a certain number of years, being a member of the state bar, being a licensed attorney, or "being learned in the law."

Given how many states require judges to be lawyers, I choose the gender composition of the lawyer profession as a proxy for the candidate pool. This proxy is imperfect. First, using the composition of the legal profession as a proxy for the composition of the qualified candidate pool is generous. More women have been attending law school and becoming lawyers over time.[23] The gender balance among young lawyers is more equal than the gender balance of older lawyers, and it is the older, more experienced lawyers who are generally more qualified for prestigious judicial posts.

Second, using lawyers as the proxy for the candidate pool overlooks gender discrimination in informal qualifications.[24] To the extent that women are subject to pressures of the "leaky pipeline," fewer female lawyers may possess the informal qualifications that increase a judge's possibility of gaining a seat on the court (Cook, 1984; Epstein, Knight and Martin, 2003). In addition, gender differences in whether and how candidates perceive themselves as qualified shape decisions to run for office. Studies suggest that women may be less likely to view themselves as qualified (Richard L. Fox, 2004; Lawless and Fox, 2010), which in turn means that women who

[22]Based on the author's research with research assistance from Madeline Brown.

[23]Using a "generous" proxy for the composition of the candidate pool raises the bar for gender diversity, which in turn lowers the bar for observing patterns of exclusion.

[24]See Williams (2006, 2008) and Martin (1981) for discussions of gender variation in formal and informal qualifications for judicial offices.

do decide to pursue higher office may be *more* qualified than their male peers (Pearson and McGhee, 2013; Milyo and Schosberg, 2000). Treating all members of the candidate pool as equally qualified discounts gender differences in the accumulation and perception of qualifications.

Despite these limitations, the composition of the legal profession should give us a plausible estimate of the proportion of qualified candidates who are women. If the pattern of women judges replacing women judges systematically suppress diversity on the bench by limiting women to one or a few seats, there would be fewer women state supreme court judges than expected given the candidate pool (all else equal). In contrast, if patterns of women replacing women promotes women judges, we would see more women selected than expected.

5.4.1 Judicial Selection and the Candidate Pool

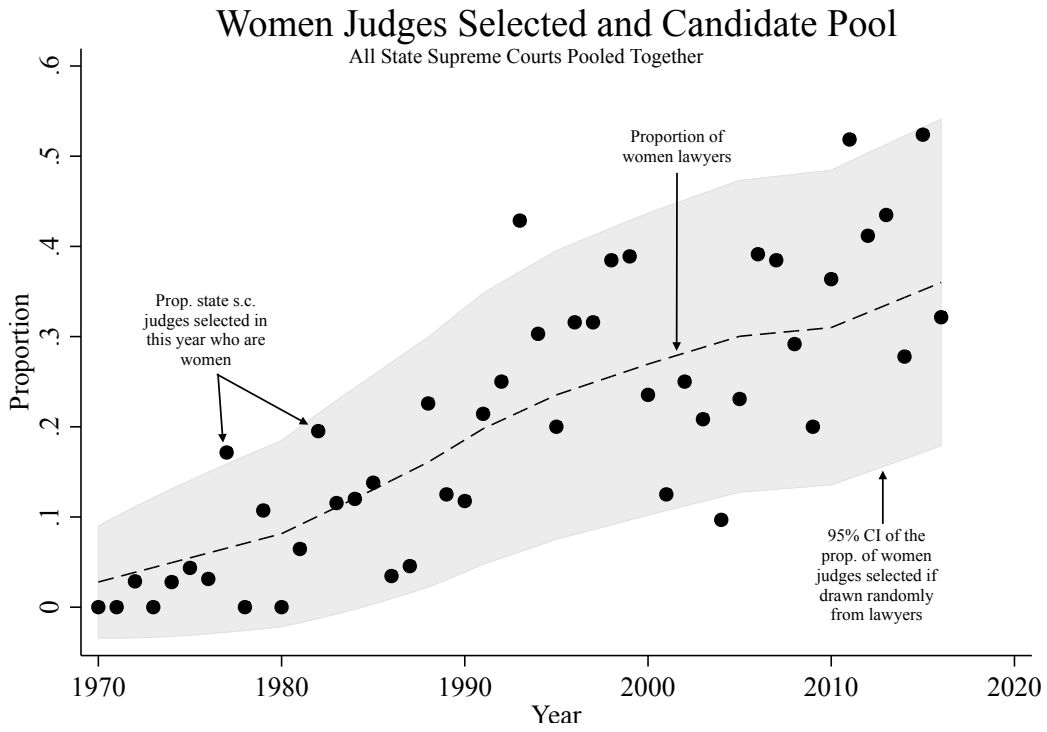
Figure 5.3 shows the average proportion of U.S. lawyers who are women (the dashed line).[25] The grey shaded region shows the 95% confidence interval around the composition of the candidate pool.[26] The dots show the actual proportion of vacancies filled by women.

We can see that the proportion of women judges selected to state supreme courts almost always falls within the expected range given the composition of the candidate pool. Of the 46 years in this data, 22 years see a greater proportion of women selected to state supreme courts than the proportion of women lawyers, and 25 years have

[25] The Lawyer Statistical Report provides state-level data for years 1952, 1963, 1966, 1970, 1980, 1985, 1988, 1991, 1994, 2000, and 2005. These data are aggregated for national comparisons. Data for 2010 and 2016 are only available at the national level. Missing years are estimated with linear interpolation.

[26] That is, if we randomly selected judges from the population of lawyers, 95% of the time the proportion of judges selected that are female would fall within the grey shaded region. The 95% confidence interval is calculated using the following formula for upper and lower bounds: $p \pm \sqrt{\frac{p*(1-p)}{n}} * 1.96$ where p is equal to the proportion of women lawyers for a given year and n is the average number of judges selected each year, 27. I use the average number of vacancies to smooth the confidence intervals for ease of interpretation. A plot that shows the confidence intervals calculated with the specific number of vacancies each year is provided in the appendix. Patterns are the same.

Figure 5.3: Lawyers as the Candidate Pool and Selections to State Supreme Courts



This plot shows the average percent women in the legal profession (dashed line), the 95% confidence intervals around the proportion of women lawyers (shaded region), and the proportion of women actually selected to state supreme courts (dots) in each year. In all but four years the proportion of women selected to state supreme courts is within the expected range, and there of the four out of range are above *the range.*

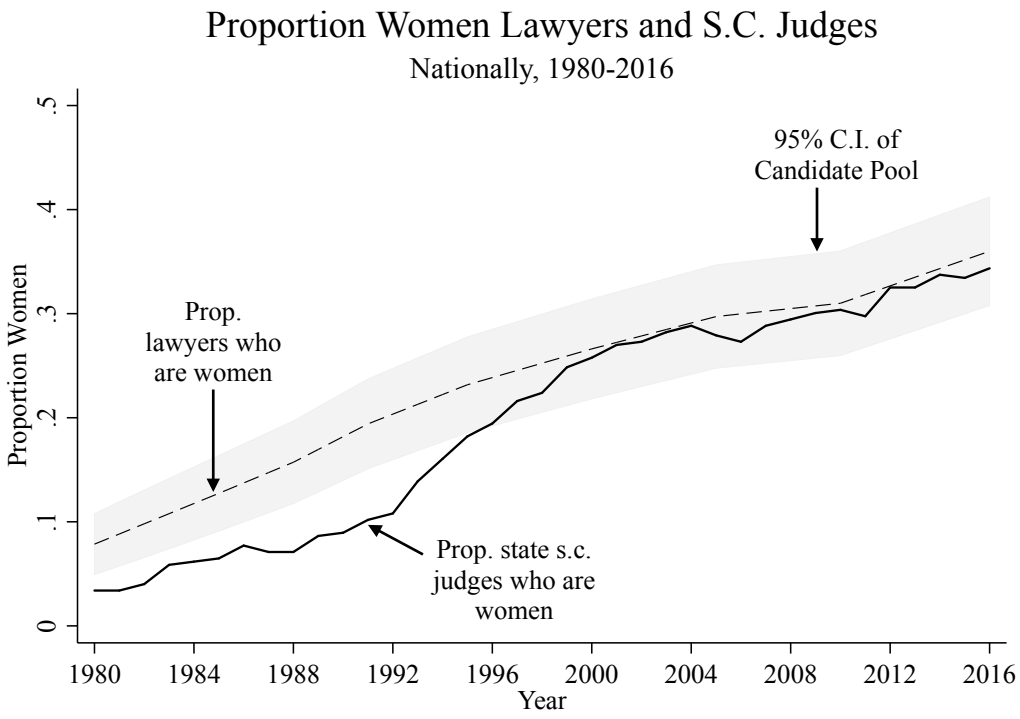
fewer women judges selected than expected. There are only four observations that fall outside the bounds of the confidence intervals, and three of those observations fall *above* the upper bound. This plot shows that the selection of women judges to state supreme courts mirrors a pattern in which judges are randomly selected from the population of attorneys.

5.4.2 Overall Judicial Diversity and the Candidate Pool

Comparing the proportion of judges *selected* to the composition of the qualified candidate pool is more generous than comparing the *overall* proportion of women judges on the bench to the candidate pool: low turnover on courts may depress the overall presence of women on a given bench even if the proportion of vacancies filled by women does reflect the candidate pool. Figure 5.4 shows the national proportion of lawyers that are women (dashed line) and the national, aggregate proportion of state supreme court judges that are women (solid line). The grey shaded region shows the 95% confidence interval around the proportion women lawyers, which represents variation from the candidate pool that might stem from randomness rather than bias or exclusion. Confirming Cook's (1984) finding that the gender diversity of state supreme court judges lagged behind the candidate pool, the proportion of women judges was below the confidence interval pool prior to 1996. Although the proportion of women state supreme court lawyers has not yet been greater than the proportion women lawyers, the proportion of women judges has been within the expected range for the last 20 years.

Using the gender composition of lawyers over the last 46 years as a proxy for the gender diversity of the candidate pool, women judges have been selected to state supreme court benches as often as expected if gender were not a relevant criterion. While there was a lag in the overall gender diversity of state supreme courts, on average courts have been about as diverse as expected since 1996.

Figure 5.4: The Candidate Pool and Supreme Court Justices Over Time



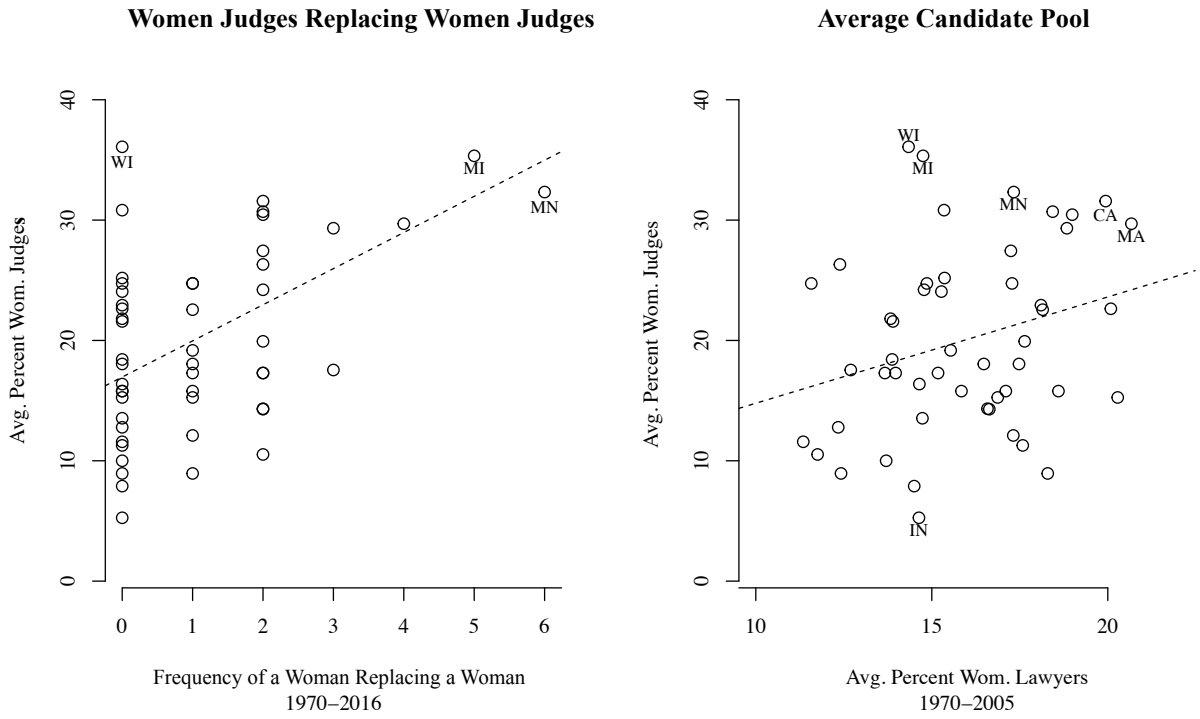
The proportion of women lawyers and the total proportion state supreme court judges on the bench who are women between 1980 and 2016. Data and the number of women selected in each year are more complete in early years than data on the overall composition of state supreme courts. To avoid missing data for the overall composition in early years, I restrict this analysis to 1980

5.4.3 Gendered Turnover and Diversity

These aggregate patterns of selection and diversity cannot rule out the possibility that gendered patterns of judicial selection limit opportunities for women judges. It could be, for example, that the states that most consistently conform to a pattern of gendered turnover are the states with only one or a few women judges. Empirically, though, that is not the case. The y-axis of the left panel of Figure 5.5 shows the average percentage of women on a state's supreme court between 1970 and 2016. The x-axis shows the number of instances in which a woman judge replaced a woman judge from 1970 to 2016.

The two states with the highest frequency of gendered turnover –Michigan and Minnesota – also are among the most gender diverse. Of course, the relationship

Figure 5.5: Frequency of Gendered Replacement and Average Proportion of Women Lawyers



The left panel shows the relationship between the number of of instances in which a woman judge replaced a woman judge by state between 1970 and 2016. The right panel shows the average proportion women judges and the average proportion women lawyers from 1970 to 2005 (2005 was the last year the Lawyer Statistical Report published data about the gender composition of lawyers at the state level).

between gendered turnover and gender diversity is endogenous: the more women on the court, the more chances there are for women to retire and be replaced by women. Likewise, there are gender diverse courts that do not have patterns of gendered replacement. What is important here is that the patterns of gendered replacement are not limited to courts with minimal diversity, which suggests that this pattern of women replacing women has not been systematically used to limit or exclude women from the bench. The right panel of Figure 5.5 shows the average percentage of women judges from 1970 to 2016 and the average proportion of women lawyers from 1970 to 2005.[27] While the relationship is positive, cross-state variation in the gender com-

[27]State level data on the gender composition of lawyers is only available until 2005.

position of lawyers does not account for all variation in the gender diversity on the bench. Neither gendered patterns of replacement nor a diverse candidate pool will *necessarily* lead to a diverse supreme court bench, but both factors are positively associated with greater gender diversity.

5.5 Discussion and Conclusions

This chapter demonstrates that gender is a relevant selection criterion for state supreme courts. Even though there are no formal rules or quotas requiring women to replace women, women judges are more likely to be selected to fill vacancies made by women than vacancies made by men. This pattern is consistent with the idea that those tasked with selecting judges are accountable to observers' preferences for gender diversity put forward in chapter 2. If elites tasked with selecting judges to state supreme courts believe they will be punished for reversion to all male courts, we should observe elites attempting to maintain a minimum level of gender diversity on the bench.

Of course, the selection of women to replace women could serve to suppress the gender diversity on the bench. However, using the gender composition of lawyers as a proxy for the composition of the qualified candidate pool for state supreme court benches, we see that women are selected to state supreme court benches as often as expected given the candidate pool. These two findings have a few important implications.

First, these findings suggest that the use of gender as a selection criterion has not systematically suppressed the diversification of state supreme court benches at the selection stage. While women are more likely to replace women judges, they do not exclusively replace women.[28] In addition, patterns of gendered judicial replacement are not limited to courts with token levels of diversity, and there is a positive

[28] See (Kenney, 2012) for an summary of how the replacement of women judges with men decreases overall diversity on the bench in the short term and can be a hurdle to diversification.

correlation between instances of women replacing women judges and levels of gender diversity on state supreme court benches. Patterns of women replacing women have not resulted in clear patterns of tokenism in which *only* one woman holds a seat on the bench at a time.

While there is no evidence here that gendered patterns of turnover have systematically promoted tokenism at the selection stage, this study does not speak to the experiences of women judges who were selected – or were selected at a specific time – to replace a woman judge. To the extent that people believe women judges are selected in whole or part because of their gender, women judges may be perceived as less qualified than their male peers. Evidence from non-judicial contexts suggests a "discounting principle" in which people perceive beneficiaries of affirmative action policies as less competent (Summers, 1991; Heilman, Block and Lucas, 1992). Notably, experimental evidence shows that perceptions of incompetence can be offset by unambiguous evidence that beneficiaries of affirmative action policies *are* competent (Heilman, Block and Stathatos, 1997). In the judicial context, formal and informal qualification requirements may provide those unambiguous signals of competency; the impressive resumes of women state supreme court justices may preclude potential for observers to "discount" competency. Moreover, Heilman et al. (1998); Evans (2003) find that stigmatization decreases as affirmative action policies become more moderate. The fact that the pattern of women judges replacing women judges is informal and incomplete – not every woman judge who retires is replaced by a woman judge – may serve to alleviate potential discounting. Assessing whether gendered patterns of replacement affect perceptions of the competency of women judges or the experiences of women judges will be a fruitful extension of the current project.

Second, gendered patterns of replacement have not resulted in women getting systematically *more* seats than expected given the gender composition of lawyers. Only three times in 30 years has the proportion of women selected been greater than the

bounds of the 95% confidence interval, and in those instances the observed proportion of women judges selected was only slightly greater than the expected proportion. If the gender composition of lawyers is an accurate proxy of the candidate pool, then there is no evidence that the use of gender in the selection process systematically favors women judges over men. For advocates of equal gender representation on the bench, however, the selection of women at rates commensurate with the composition of the candidate pool may be viewed as the bare minimum. From this perspective, the selection of women at rates less than 50% is unjust because it reinforces the expectation that women need not be or should not be equally represented in positions of power and provides tacit approval for the current standards for qualification that favor men over women.

Third, this project highlights the importance of continued efforts to diversify the judicial candidate pool and state supreme court benches. While there is no settled standard of gender diversity on state supreme courts, advocates of descriptive representation argue that political bodies ought to mirror the descriptive characteristics of the population. Even though women have been selected as often as expected given the composition of the candidate pool, women do not make up 50% of lawyers or supreme court justices. Women have faced – and continue to face – barriers to the accumulation of formal and informal requirements for office (Cook, 1984; Guinier et al., 1994; Drachman, 2001; Epstein, Knight and Martin, 2003; Redfield, 2009; Rikleen, 2015). These barriers to accessing qualifications, in turn, suppress the diversity of the eligibility pool. Removing the barriers for women to accumulate the necessary qualifications, or more critically, re-defining what it means for one to be "qualified," may help lead to more diverse state supreme court benches over time.

Encouragingly, law school graduation rates and the composition of the legal profession are much more diverse now than even 20 years ago. In fact, there were more women law school graduates than men law school graduates in 2015 and 2016. As

these new lawyers mature and acquire informal qualifications, the candidate pool for high office will come to increasingly mirror the gender composition of the electorate. While a more diverse candidate pool may will not necessarily lead to a more diverse judiciary (Cook, 1984), a more diverse candidate pool should facilitate efforts to further diversify the judiciary.[29]

Finally, although the aggregate patterns of judicial selection conform to expectations, there is cross-state variation in the timing and consistency of diversification and in the use of gendered patterns of replacement.[30] This project has not addressed why some states chose to seek out women to replace women while other states did not. What is it about the social, political, or judicial context in Minnesota and Michigan that accounts for the frequent replacement of women judges with women judges? Of course, having more women on the bench provides more opportunities for gendered patterns of turnover, and Minnesota is unique: its seven judge court had a four woman majority in 1991 (Margolick, 1991). To put that in context, Alaska selected its *first* woman judge, Dana Fabe, in 1996 (Alaska Judicial Council, 2006), and New Hampshire did not select its first woman judge, Linda Stewart Dalianis, until 2000 (New Hampshire Judicial Branch, 2017). Furthermore, within a given state, not all women are replaced by women. Under what conditions are governors and political elites most likely to seek out – either implicitly or explicitly – women judges to replace women judges?

Future research ought to address cross-state and over-time variation in the use of gendered patterns of judicial turnover. It is possible that patterns of women replacing women developed out of efforts to diversify the bench but could turn into a ceiling that limits the openness of any seat to a woman judge. Future research should continue to

[29] In addition to continued diversification of the candidate pool, other scholars suggest, for example, that diversifying merit committees (Esterling and Andersen, 1999), having women governors (Kenney and Windett, 2012), or having more women in legislative office (Hoekstra, Kittilson and Bond, 2014) may also lead to increased gender diversity on courts.

[30] See the Appendix for a state-by-state comparison of supreme court diversity and candidate pool diversity.

explore whether the judiciary – the branch tasked with ensuring equal justice under the law – is selected through fair, equal, and non-discriminatory practices.

Chapter 6 An Alternative? Slate Selection and Gender Diversity on the Bench

In chapter 2, I presented a framework for diversity that linked observer inferences about bias and fairness to accountability for diversity. In chapters 3 and 4 I presented evidence suggesting that selection procedures with multiple actors like merit selection may serve to undermine opportunities for gender diversity on the bench. In this chapter, I present an institutional alternative to unitary selection versus multiple actor selection that may promote gender diversity in the judiciary: slate selection.

In the legislative context, empirical researchers have identified a relationship between party list proportional electoral systems (PR) and greater gender diversity in office relative to single member district electoral systems (SMD), on average (Kenworthy and Malami, 1999; Salmond, 2006; Rule, 1987; Yoon, 2004). There are at least three mechanisms that might explain this relationship between party list PR and greater gender diversity: information, balanced lists, and party competition.[1]

First, party list ballots provide more information to voters about the gender composition of candidates than ballots in SMD elections, which allows observers to better identify gender disparity under party list systems. Second, the presence of several candidates on a party list means that no one candidate has to appeal to such large group of voters. Instead, traits of each candidate can be viewed as compliments or substitutes to traits of other candidates on the list, which allows for a more diverse set of candidates to be electable. Finally, because PR systems allow for smaller viable parties, PR systems facilitate the inclusion of new ideas and expectations. If smaller, niche parties popularize gender balanced party lists, larger parties have an incen-

[1]These are, of course, not the *only* three mechanisms. Others include centralized candidate nomination (Norris, 1993) decreased emphasis on incumbency and greater turnover (**?**), and ease of implementing party quotas (Caul, 2001; Krook, 2006*b*).

tive to also balance their lists to remain competitive, thereby increasing the gender diversity of candidates and elected officials (Matland and Studlar, 1996).

These three mechanisms (information, balanced lists, and party competition) help explain the consistent empirical relationship between PR systems and more women in office. However, discussions and tests of these mechanisms have been limited to the legislative context. In this paper, I apply the first two mechanisms of the party list PR system logic to a feature of judicial selection: the selection of multiple judges simultaneously rather than on a rolling, one-by-one basis. I argue that the first two mechanisms that explain gender diversity in party list PR systems – information and balanced lists – apply to the group selection of judges and should result in greater diversity in the judiciary relative to systems in which judges are selected one-by-one.

After elaborating on the mechanisms through which group selection facilitates gender diversity, I present evidence from a survey experiment that supports the expectation that group selection facilitates gender diversity. Then, I summarize an observational analysis of the selection of judges to peak courts cross-nationally. Using a novel new data set, I isolate changes to formal selection methods that result in group selection of judges. Using a matched pair design and a signed rank statistic, I show that there is a positive but statistically insignificant relationship between an institutional change that results in group selection and the selection of the first woman to the country's peak court. I conclude with recommendations for future research.

6.1 The Role of Party Lists PR in Legislative Diversity

While there is substantial variation within PR systems (Schmidt, 2009) that may affect the ease and opportunity for women candidates to win seats (district magnitude, position requirements, and vote thresholds, for example), we can generally identify three main mechanisms that facilitate the inclusion of women candidates in office.

The first mechanism is information. When voters are presented with lists of several candidates on a PR ballot rather than just two or three candidates per district on an SMD ballot, they have more information about the gender balance of candidates for office. It is much easier for voters to make inferences about gender bias when confronted with several all-male party lists than when confronted with two or three male candidates in an SMD system. Particularly as district magnitude increases, the "exclusion of women from the party's list of candidates becomes increasingly obvious and increases the danger of a negative reaction from voter"(Matland, 1993, p.738).

Second, in an SMD election, voters vote for just one candidate, so that one candidate – alone – must earn sufficient votes to win. S/he must appeal to a broad audience. In contrast, voters under list PR systems vote for a group of candidates, which means that the *combination of traits* of several candidates must be sufficiently appealing to win votes. Any individual candidate, therefore, need not appeal to such a wide set of voters.[2] Having a list of candidates allows the traits of individual candidates to be viewed as complimentary or substitutable to traits of others on the same list, which means that a more diverse set of candidates may be viable under a PR system. Indeed, if traits of those in a group are viewed as complimentary, the presence of a list may *encourage* diversity by encouraging the selection of different types of candidates to produce a "balanced" list.

Finally, PR systems facilitate the inclusion of smaller, niche parties. If the small parties champion a popular issue, larger parties have an incentive to absorb these issues to prevent loosing seats to the smaller party. If small parties champion gender diversity by including women on their party lists, other parties, then, may feel pressure to also include women on their lists to avoid perceptions of sexism or exclusion (Matland and Studlar, 1996).

[2]This depends, however, on whether lists are open or closed. When lists are open and candidates face intra-party competition, there is an incentive to develop personal reputations to garner votes to determine their position on the list. See Carey and Shugart (1995) for a discussion of intra-party competition and the cultivation of personal vote-seeking.

The first two of these mechanisms – list selection providing more information and allowing for balanced lists – are not necessarily unique to the PR system. Instead, these mechanisms should apply to institutions in which candidates are selected as a group rather than individually. In the U.S. context, we see some features of these two mechanisms at the presidential level when Presidential candidates attempt to broaden their popular appeal by selecting Vice Presidential candidates with different characteristics (Nelson, 1988; Baumgartner, 2012). In the next section, I describe how these two mechanisms linking group selection to greater gender diversity in office can be applied to judicial selection cross-nationally.

6.2 Group Selection and Diversity in the Judiciary

A growing literature addresses gender diversity in the judiciary. The size and prestige of a court (Williams and Thames, 2008), norm diffusion across space and institution (Williams and Thames, 2008; Goelzhauser, 2011; Hoekstra, Kittilson and Bond, 2014), the legal culture (Remiche, 2015), and features of the common law and civil law systems (Schultz and Shaw, 2013) are expected to affect prospects for gender diversity on the bench. When it comes to the role of selection institutions in gender diversity in the judiciary, however, debate still exists. Some find the concentration of accountability on a unitary selector is associated with greater diversity (Carbon, Houlden and Berkson, 1982; Bratton and Spill, 2002; Williams and Thames, 2008). In contrast, Goelzhauser (2011) finds that the presence of a merit commission is associated with the earlier selection of the first woman on state supreme courts. However more recent work finds that the merit commission stage disadvantages women judicial candidates (Goelzhauser, 2018). Others still find no or little relationship between selection institutions and diversity (Alozie, 1988, 1990; Hoekstra, Kittilson and Bond, 2014).

Despite the many explanations for variation in gender diversity in the judiciary,

no one[3] has addressed the role of group or slate selection for judicial gender diversity. Cross-nationally, there is substantial variation in the selection procedures for peak court judges,[4] including variation in the timing and turnover of judges. Several countries have judicial selection institutions that result in (or require) the selection of multiple judges at once.

For example, in contrast the U.S. system in which Supreme Court judges are selected on a rolling, one-by-one basis, in Albania, the nine justices of the constitutional court are selected for nine year terms, and three justices are replaced every three years. Applying the first the mechanisms of party list PR systems – information – to the judicial context, citizens in Albania have more information about the gender composition of newly selected judges. While citizens are not *voting* on the slates of candidates, they can still make inferences about whether or not the selection of judges is fair. If citizens decide that the process is unfair due to observing persistent homogeneity, they can exert pressure on the elites tasked with selecting judges.

Indeed, Valdini and Shortell (2016) find that political elites tasked with selecting judges will select women to the bench when doing so is electorally beneficial. Specifically, elites who are subject to institutional "exposure" – that is, when they are electorally vulnerable for their selections – are the most likely to claim credit for choosing women judges. Elites "sheltered" from electoral accountability are less-likely to select women judges because they do not benefit from claiming credit for the selection of women. Valdini and Shortell's work suggests, then, that observers of judicial selection can affect prospects for gender diversity on the bench even when judges are not elected. Having judges selected as a group provides observers more evidence about the gender composition of newly selected candidates which, in turn, facilitates observers' abilities to pressure elites for homogeneous selections.

[3] To my knowledge, at least

[4] By peak court judge, I mean the constitutional court or the highest ordinary court in countries in which there is no constitutional court.

The second mechanism through which party list PR facilitates the election of women candidates is through balanced lists. By allowing multiple candidates to be presented to voters and elected at once, no one candidate must appeal to all voters. Instead, voters and elites can evaluate the traits of the candidates taken as a whole.[5] This mechanism can apply to the judicial context as well. Judges selected to the U.S. Supreme Court are selected as individuals and must appeal to a majority of the Senate to be confirmed. In contrast, if judges are selected as a group, elites must agree on the slate as a whole rather than each individual judge.[6] Likewise, observers can make decisions about the qualifications and characteristics of judges as a group. By considering groups of judges rather than individual characteristics, there may be more flexibility for the selection of non-traditional or "outsider" candidates such as women. If slate selection facilitates the inclusion of historical outsiders, we should observe greater and earlier gender diversification, on average, in countries in which judges are selected as a slate.

Hypothesis 1 (Information): Observers will be more critical of gender disparity when they have more information about judges selected to the bench.

Hypothesis 2 (Balanced Lists): Given the same information, observers will be more critical of gender disparity when judges are selected as a slate rather than one-by-one.

Hypothesis 3 (Diverse Benches): There will be greater gender diversity in countries in which judges are selected as a slate or in pairs relative to when judges are selected

[5]If voters have expectations about how many candidates will be elected from a particular party list, they can asses the traits of the top X candidates they expect to actually make it into office rather than the whole list.

[6]The extent to which individual characteristics of judges are relevant to selection depends on actual rules. In some cases, each member of the slate may be confirmed/selected separately. In other cases, the slate as a whole may be confirmed. Either way, we should expect this mechanism to hold: when judges are selected at the same time –regardless of whether they need to be confirmed separately – the confirmation of one judge ought to affect prospects for the others. That is, the individual confirmation of judges in the same short time frame should not be independent.

on a rolling, one-by-one basis.

6.3 Research Design

To test whether group selection facilitates the inclusion of women in the judiciary, I conduct two empirical analyses. The first, a survey experiment, tests the information and balanced list hypotheses. I find that observers are more critical of homogeneous courts when judges are selected as a slate rather than one-by-one, even when they observe the same exact judges selected to the court. The second study uses observational, cross-national data to test whether slate selection increases gender diversity. Due to data limitation, I test for the timing of the first woman selected to a peak court rather than levels gender diversity. This analysis leverages a new and unique data set on the formal selection procedures of peak court judges cross-nationally to identify instances of institutional change. Using non-parametric matching, I isolate the effect of implementing slate selection. I find a positive, but insignificant effect of slate selection on the timing of the first peak court woman judge.

6.3.1 Survey Experimental Evidence

To test whether slate selection affects observers' responses to gender disparity, I fielded a survey experiment in the summer of 2017. The survey was designed and hosted on Qualtrics and respondents were recruited through Amazon's Mechanical Turk (MTurk) service. As is well documented, MTurk survey respondents are not randomly drawn from the population. Instead, MTurk workers tend to be younger and more politically liberal than randomly selected respondents. Even so, research suggests that MTurk samples are more representative than in person convenience samples and student samples (Berinsky, Huber and Lenz, 2012). Respondents were paid $0.35 for taking an approximately 3 minute survey. For the analysis presented here, survey respondents must have indicated that the lived in the United States or

taken the survey from within the U.S.[7] In addition, respondents who did not pass a simple attention check were removed from the analysis. Table 6.1 shows summary characteristics for the MTurk respondents used in this analysis.

Table 6.1: Characteristics of Survey Respondents

	Male	Female			
Gender	.49 (351)	.51 (369)			
	<25	25-34	35-49	50+	
Age	.1 (69)	.43 (311)	.30 (214)	.18 (128)	
	Very Cons.	Smwht Cons.	Moderate	Smwht Lib.	Very Lib.
Ideology	.05 (33)	.25 (171)	.23 (165)	.37 (269)	.12 (84)
	High School	Some Cllge	BA/BS	Masters	Doctorate
Edu.	.08 (59)	.35 (251)	.40 (287)	.14 (99)	.04 (25)

Summary characteristics for the MTurk respondents used in the survey experimental analysis.

In this experiment, I manipulate two variables to create four treatment conditions. I manipulate the amount of information observers receive to test how the role of information affects perceptions of bias, and I manipulate the selection institutions to test how slate selection might encourage observers to asses judicial candidates as a "balanced" slate. Table 6.2 describes the treatment characteristics of the four groups.

All groups were given general information about a hypothetical, five-judge court, and they were told some basic information about the judges: their age, gender, the prestige of their law school, and their years of judicial experience. Groups A, B, and C were told judges were selected to five year terms, and each year one judge retired and one judge was selected to the bench. Respondents in group A were only given information about the one judge selected to the bench in the current year. Group B

[7] Data on longitude/latitude and IP addresses were used to verify location of respondents

Table 6.2: Treatment Groups

		Selection Process:	
		Rolling,	Slate,
	One male judge	Group A	
Information:	Five judges, all-male	Group B	Group D
	Five judges, mixed gender	Group C	

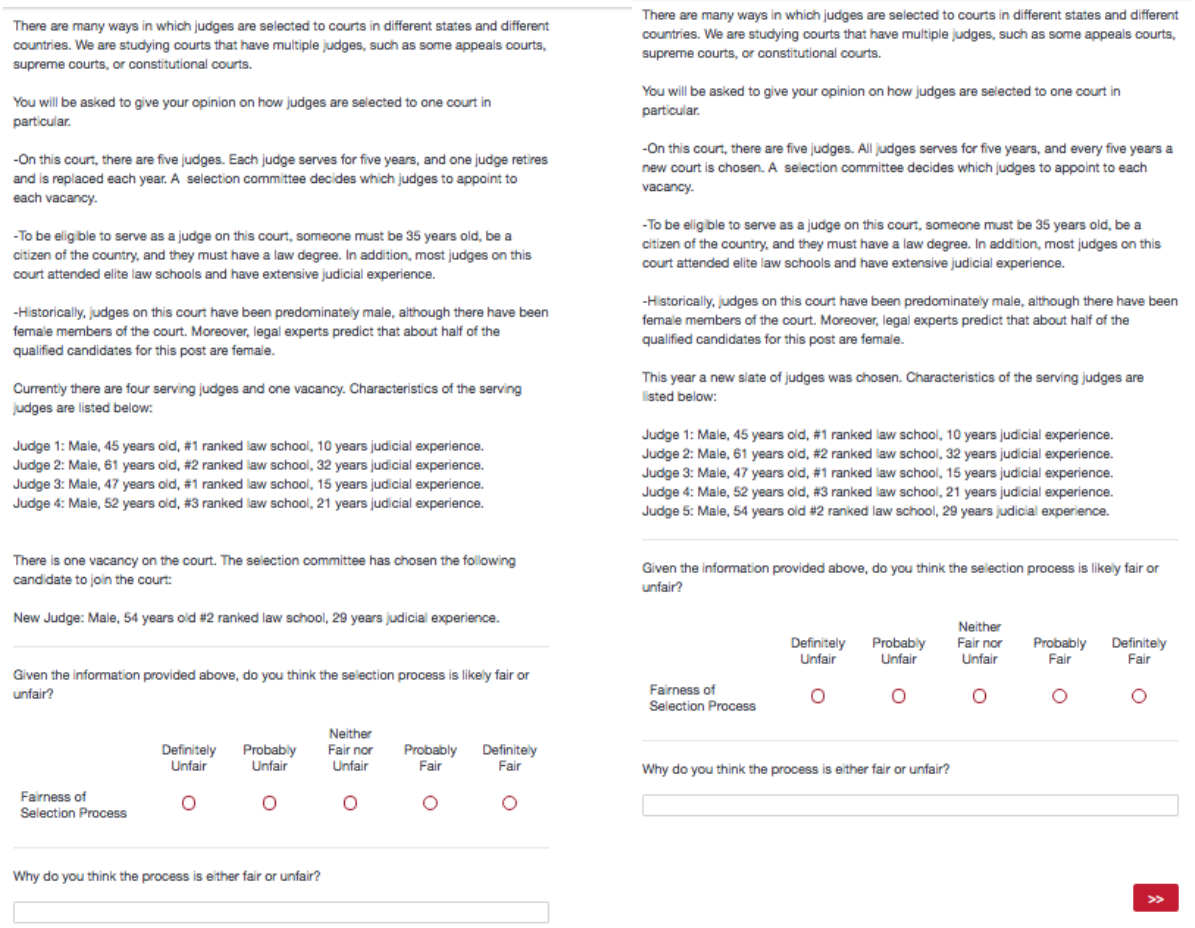
The treatment groups; two treatment variables were manipulated. Group A is compared to group B to test the information hypothesis. Group B is compared to grpup D to test the balnced lists hypothesis. Group B is compared to group C to test whether observers perceive a gender balanced court as more fair.

was shown four male judges currently serving on the bench and one male judge newly selected to the bench. For group C, two of the currently serving judges were women. Those in group D were told that judges were selected to five year terms, but every five years all five judges retired and were replaced. Those in group D saw the same information about the same judges as those in group B.

After reading information about the court and the members of the court, respondents were asked, "Given the information provided above, do you think the selection process is likely fair or unfair?" Respondents indicated their responses on a five-point scale: definitely unfair, probably unfair, neither fair nor unfair, probably fair, or definitely fair.

To test how information affects observers' inferences about bias (H1), I compare group A (one male judge, one-by-one selection) to group B (five male judges, one-by-one selection). To test how slate selection might encourage observers to view selections as "balanced lists," I compare group B (five male judges, one-by-one selection) to group D (five male judges, slate selection). As a check to see if observers are more accepting – that is, less likely indicate the process is unfair – when courts are more

Figure 6.1: Survey Instrument: One-by-One and Slate



Two of the three survey instruments. The figures above show the text and questions asked of those in the one-by-one (all male) treatment group and the slate (all male) treatment group.

gender balanced, I compare group B (five male judges, one-by-one) to group C (two female and three male judges, one-by-one). Figure 6.1 shows the survey instruments for Groups B (five men, one-by-one) and D (five men, slate).

As a test to see if observers recognize and care about gender diversity – a necessary assumption of the mechanisms outlined above – I added an additional control group in which judges are selected on a rolling, one-by-one basis but some of the judges are female. If respondents do not notice or care about gender diversity, responses should be the same across the mixed gender, one-by-one group and the all-male, one-by-one

group. Table 6.2 describes the treatment characteristics of the three group.

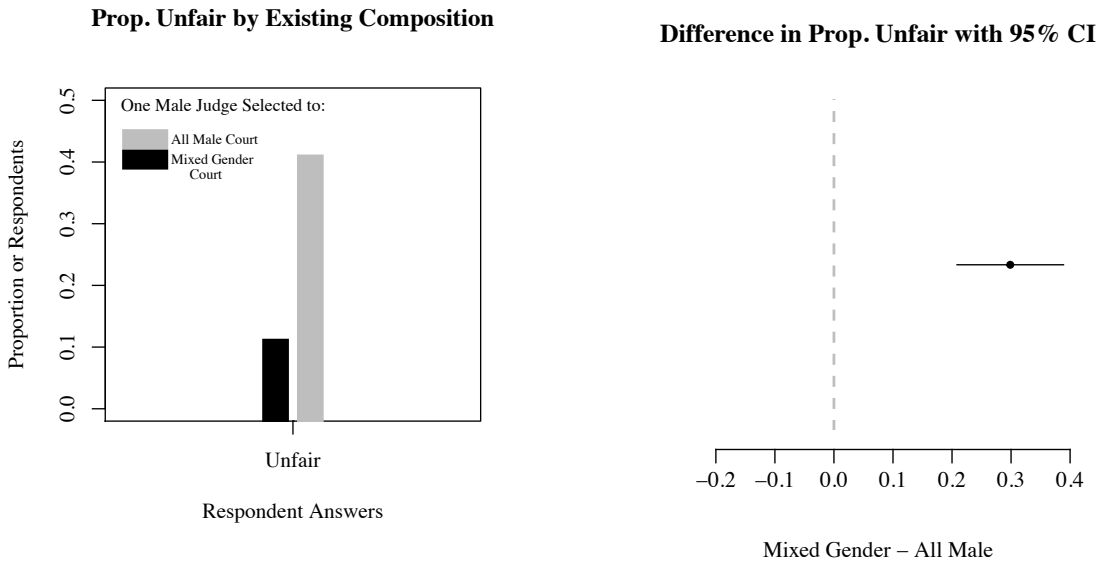
Results

Do respondents care about the gender composition of the court? As a first test to check whether respondents noticed gender parity or disparity whether disparity affects perceptions of bias, I compared Group B (those who observed one man selected to an all-male court) to group C (those who observed one man selected to a gender-equal court). If respondents do not care about gender diversity on the bench, responses about fairness or bias should be the same across both groups as the only difference between these two treatment groups is the gender of two judges. If, however, respondents do care about the gender diversity of courts then more respondents should indicate that the process appears unfair for the all-male court.

The left panel of Figure 6.2 shows the proportion of respondents in each group who indicated that the process was either "definitely unfair" or "probably unfair." The black bar shows the responses for the mixed-gender court; the grey bar shows the responses for the all-male court. As you can see, more respondents who saw an all-male court interpreted the selection of an additional male judge as evidence of bias than respondents who saw a mixed-gender court. Moreover, the difference in the proportion of respondents who indicated that the process appeared biased is statistically significant ($p < .001$), which suggests that this pattern is not due to chance. The right panel plots the point estimates and 95% confidence interval for a difference in proportions test.

The comparison of these two groups demonstrate that respondents are sensitive to the gender composition of courts and that some respondents do interpret homogeneous courts as evidence of a biased selection process. In other words, this comparison shows that how respondents interpret the fairness of the selection of a new judge to the bench depends on the current composition of the court.

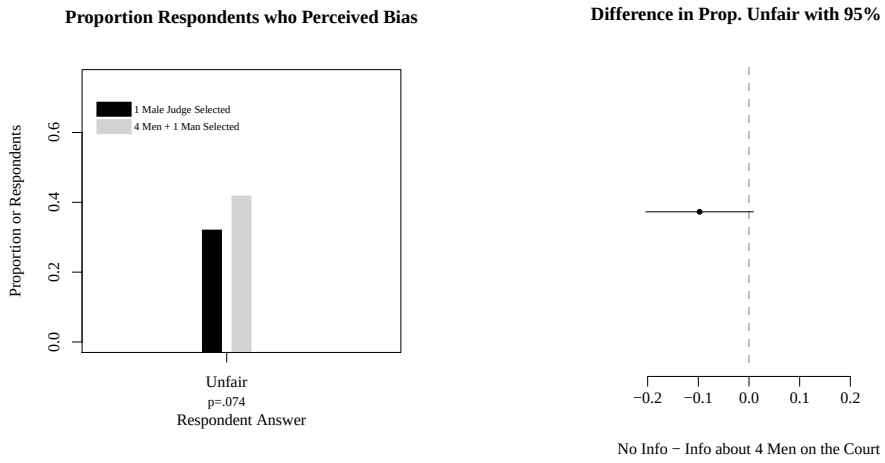
Figure 6.2: Information and Perceptions of Unfairness



The left panel shows the proportion of respondents who indicated that the process appeared "definitely unfair" or "probably unfair." The grey bar shows respondents who saw the selection of a male judge to an all-male court. The black bar shows answers from respondents who saw a man selected to a mixed-gender court. The right panel shows the point estimate and 95% confidence interval for a difference in proportions test. The p-value for this test is $p < .001$

Hypothesis 1: Does the Amount of Information Affect Perceptions of Unfairness? The information hypothesis predicts that having more information about judges and the gender composition of judges will facilitate observers' abilities to make accurate inferences about bias. This hypothesis follows form the idea that in the legislative context, party list systems provide more information to observers about the gender composition of candidates than SMD systems. To test how information about judges affects observers's ability to make inferences about bias, I compare group A (one male judge, one-by-one selection) to group B (five male judges, one-by-one selection). Both of these groups were told the same information about the selection process, and both were told about one judge that is newly selected to the bench. Group B, however, is told about four currently serving male judges. If information about the judges affects perceptions of bias, those in group B should be more critical of the selection process than those in group A. Figure 6.3 shoes the proportion of

Figure 6.3: Amount of Information and Perceptions of Unfairness



The left shows the proportion of respondents who answered "definitely unfair" or "probably unfair." The black bars show the answers from respondents who saw 1 male judge selected to a court without any information about the existing judges. The grey bars show the responses for those where were given information about a male judge selected to a bench with four existing male judges. The right panel shows the difference in proportions. The p-value for the difference in proportions test (which follows a χ^2 distribution) of "unfair" responses is p= .07 which indicates the probability of observing a difference at least as great as the difference observed is 7%. The 95% confidence interval includes 0, or no difference between the two proportions.

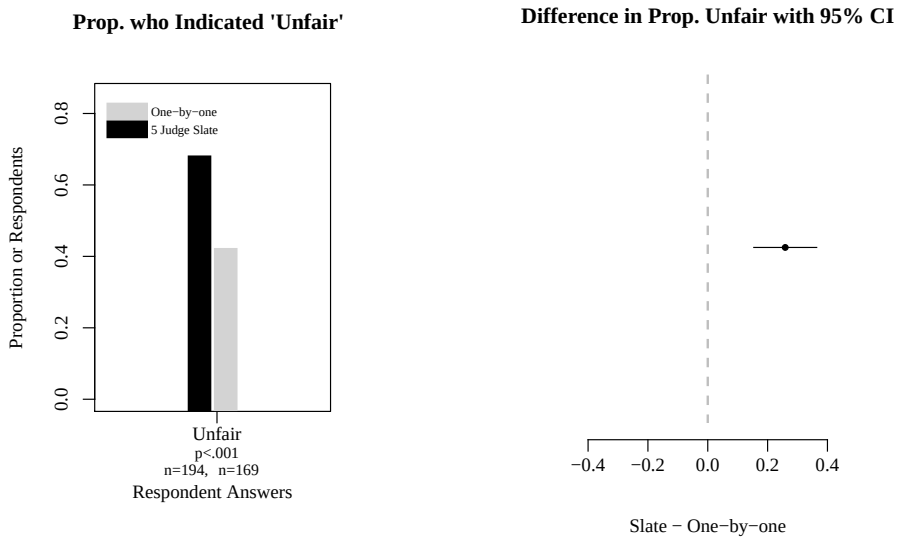
respondents across the two groups who indicated that the process was either "definitely unfair" or "probably unfair." More respondents in Group B (five male judges) indicated that the process appeared unfair than those in group A (one male judge). The p-value for a difference in proportions test is p=.074.

This comparison demonstrates that information likely does have a small effect on how and whether observers make inferences about gender bias in judicial selection: observers were more critical when they had more information about homogeneity on the bench.

Hypothesis 2: Does Slate Selection affect Perceptions of Unfairness? To determine whether slate selection affects perceptions of unfairness relative to rolling, one-by-one selection, I compare responses of group B (five male judges, on-by-one) to group D (five male judges, slate). It is important to note that respondents saw information about the same five judges. The only difference in the timing of selection

of the judges. If selecting judges as a group rather than one by one has no effect on perceptions of unfairness, then the responses across the two groups should be the same. The left panel of Figure 6.4 shows the proportion of respondents who indicated that the process appeared either "definitely unfair" or "probably unfair" across the two treatment groups. Respondents who saw five men selected as a slate were much more likely to indicate that the process appeared unfair, and the difference in the proportion of respondents who indicated the process appeared unfair is statistically significant ($p < .001$).

Figure 6.4: Slate Selection and Perceptions of Unfairness

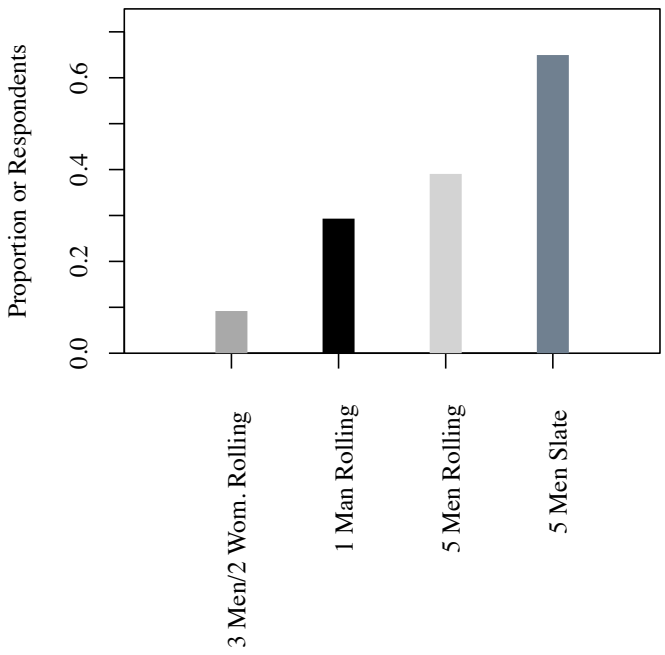


The left panel shows the proportion of respondents who indicated the process seemed "definitely unfair" or "probably unfair." The grey bar shows respondents who saw the selection of one man to a court with four men . The black bar shows the respondents who saw a slate of five male judges selected. The right panel shows the point estimate and 95% confidence interval for the difference in proportions test. The p-value for this test is $p < .001$.

The evidence from the survey experiments indicates that observers are more critical of homogeneity when judges are selected as a group rather than one by one. While some of the difference can be accredited to differences in information, the greater difference is between those who saw five male judges selected at once versus five male judges selected on a rolling basis. Figure 6.5 shows the proportion of respondents

who indicated the process was either "definitely unfair" or "probably unfair" across all treatment groups. The difference between those in group D (five men, slate) and group B (five men, rolling) is much greater than the difference between group B (five men, rolling) and group A (one man, rolling), which suggests that the information mechanism is not the sole explanation for the theorized relationship between group selection and gender diversity. Instead, it appears that the selection of judges as a group rather than one-by-one leads observers to make different inferences despite having the same amount of information. This is consistent with the idea that observers assess candidates differently when they are selected as a group – the balanced list hypothesis. Observers concluded that excluding a woman from a group of five judges was more problematic than not selecting a woman for five binary choices.

Figure 6.5: Responses, All Groups



The proportion of respondents who indicated "definitely unfair" or "probably unfair" across all treatment groups.

To elucidate why those in the slate group were more critical of gender disparity

Table 6.3: Summary of Gender in Qualitative Responses

Institution	Number of Qual. Responses	Percent who Noted Gender	Of those who Noted Gender, Percent Unfair
Slate	247	59.5%	88.4%
One-by-one	238	37.0%	81.8%

than those who saw the same information in the one-by-one group, I turn to qualitative explanations for respondents' choices. In addition to indicating their perception of fairness on a five-point scale, respondents were asked to provide an explanation for their decision. Among those who provided a qualitative explanation for their response, about 60% in the slate selection treatment group noted the gender of the justices in their responses. In contrast, among those in the one-by-one control group, only 37% noted gender in their qualitative responses. Interestingly, among those who noticed the gender of the judges, the proportion who concluded that the process was either "definitely unfair" or "probably unfair" across both groups is similar: 88% among slate selection and 81% among one-by-one selection. Table 6.3 summarizes the number of respondents who noted gender in their qualitative responses and Table 6.4 lists examples of qualitative responses.

These responses suggest that respondents were more likely to notice the gender of judges when judges were selected simultaneously as a slate than on a rolling, one-by-one basis. Among respondents who noticed the gender imbalances, the proportion who inferred that the process was unfair and justifications for their responses were similar across institution type. It appears, then, that the mechanism linking slate selection to perceptions of gender bias among observers is through encouraging observers to *notice* gender imbalances given the same amount of information.

Table 6.4: Examples of Qualitative Responses

Institution	**Response**	**Explanation for Response**
One-by-one	Definitely Unfair	"Half of the qualified candidates are female, so it would make MUCH more sense to appoint a female for this panel of judges as she would certainly bring a different perspective!"
One-by-one	Probably Unfair	"I think that these judges are definitely qualified, but I would like to see at least one woman serving"
One-by-one	Neither Fair nor Unfair	"While it may seem a little male centric, the rules for selection are fair. Age, schooling, and experience are all things both genders can accomplish equally."
One-by-one	Probably Fair	"They all have extensive experience. Its not like an unqualified male is being picked instead of an unqualified female. Sometimes results like these aren't the sexist results people are looking for but instead just a matter of merit."
Slate	Definitely Unfair	"Because if over half of qualified candidates are female then WHY aren't they being chosen."
Slate	Probably Unfair	"[O]nly men are on a 5 person panel which should have at least one or two women to be truly representative of the population."
Slate	Neither Fair nor Unfair	"Just because they are all male doesn't make it unfair. They have a lot of experience, so I think it is fair."
Slate	Probably Fair	"Although it lacks diversity, it appears the judges chosen are highly qualified which should be the primary criteria of this sort decision."

Examples of respondents' explanations for their responses about the fairness/unfairness of the selection process. Answers are similar across institution type.

6.3.2 Slate Selection: Observational Evidence

The evidence from the survey experiments indicates that respondents are more likely to notice homogeneity when judges are selected as a group rather than one-by-one. If observers' are better able to identify and be critical of disparity under slate selection, elites tasked with appointing judges should feel more pressure to select women when selection institutions require the selection of multiple judges at once relative to systems in which judges are selected on a one-by-one basis. In turn, we ought to observe greater diversity in the judiciary under slate selection systems.

In this section, I combine a unique cross-national data set on the selection procedures for constitutional court judges with a research design for identifying the effects on institutional change on gender diversity in the judiciary first described in Arrington et al. (2018). The data, collected for the Varieties of Democracy project and in collaboration with the Comparative Constitutions Project, identifies the selection and removal procedures for peak court justices cross-nationally and over time. I have identified processes in which judges are selected as groups (or pairs) rather than individually.[8] Importantly, however, this data is limited to constitutional rules. Some states may have statues – rather than constitutional rules – that dictate group selection. Those states (if they exist) are not included here.

Institutions are, of course, endogenous to outcomes. Expectations about the ability of a woman to successfully become a peak court justice under a specific selection system will shape whether she decides to pursue the qualifications necessary to hold the post. If women are excluded from the bench, potential women candidates have less incentive to acquire the necessary qualifications, for example. In order to account for this endogeneity, I focus on constitutional changes to selection institutions. Assuming constitutions are not changed with the *goal* of increasing gender diversity

[8]Most courts with slate selection select two or three judges at once, rather than the whole court. A common method of slate selection is for three judges on a nine-judge court to be selected every three years. Full courts can be selected at once, though, such as when a court is newly implemented.

Table 6.5: Constitutional Examples of Slate Selection

Country	**Year**	**Text from the Constitution Describing Slate Selection**
Chile	1986	"The [seven] members of the Court shall serve eight years, be partially replaced every four years, and must not be removed."
Niger	2009	(Translated from French by the Vdem Judiciary team) The President of the Republic designates five members to the court; the President of the Assembly proposes 2 members to the court, and the President of the Senate proposes two members. The nine judges serve for one six year, non-renewable term.
Romania	1991	"The Constitutional Court is composed of nine justices, appointed for a nine-year term, which cannot be extended or renewed...Every three years, one-third of the members of the Constitutional Court are replaced, under the conditions stipulated by the statutory law of the Court."
Spain	1978	"The Constitutional Court is composed of twelve members...The members of the Constitutional Court shall be appointed for a period of nine years and shall be renewed by thirds every three years."

Examples from countries' constitutions that indicate slate selection. Examples come from the Varieties of Democracy Judiciary data collected by a team at Emory University using constitutions organized through the Comparative Constitutions Project.

on the bench, I can treat changes as exogenous.[9] After identifying the countries and years in which a judicial selection system changes to slate selection, I match those treated unites to control units. Then I compare the time from treatment until the first woman in selected to the court. If slate selection does facilitate the selection of women, we should observe women being selected – on average – sooner in treated countries relative to control countries.

I use as my dependent variable the time until the selection of the first woman judges rather than the total number of women on the court due to data limitations. It is much easier to accurately identify the date the first woman was selected than to identify the total number of women on the court over time. Attempts to collect the number of women on the court over time remain incomplete.[10] I therefore address the timing of the first woman judge rather than total levels of gender diversity over time.

Matching

I match countries that experienced an institutional change that resulted in slate selection to countries that also had an institutional change but *not* a change to slate selection. As Arrington et al. (2018) show, institutional changes to judicial selection procedures are associated with decreased time to the selection of a woman. Because any change is associated with decreased time to the selection of a woman, matching countries with a change to slate selection to countries that had no change would overestimate the effect of slate selection by attributing effects of institutional change, generally, to slate selection, specifically. Instead, I match treated countries that changed to slate selection to countries that experienced a change to their judicial selection institutions that did not affect whether judges were selected one-by-one or

[9] There is no evidence that the countries included in this analysis changed their judicial institutions with the goal of increasing gender diversity. Most changes to the judiciary coincide with broad constitutional changes.

[10] A team of researchers at Emory, lead by Leeann Bass, attempted to collect the total number of women on the bench over time, but substantial data missing-ness threaten the accuracy of the data. However, researchers were able to identify the year of the first woman more accurately.

Table 6.6: Treatment and Control Countries

Treated Country & Treatment Year	**Year 1st Wom. (t)**	**Control Country & Year**	**Year 1st Wom. (c)**	**Diff. (t-c)**
Chile, 1986	1989	Portugal, 1982	1989	-4
Ecuador, 1983	1997	Iran, 1989	1998	5
Mauritania, 1991	2017	Tanzania, 1995	2004	17
Niger, 2009	2013	Cent. Afr. Rep., 2004	2005	3
Panama, 1983	1998	Afghanistan, 1980	2017	-22
Romania, 1991	2004	Cent. Afr. Rep., 1994	2005	2
Senegal, 1992	1993	Kazakhstan, 1995	2002	-6
Spain, 1978	1980	Madagascar, 1970	1991	-19
Thailand, 1997	1998	Malawi, 1994	1997	-2
Togo, 1992	2007	Malaysia, 1994	2001	8
Tunisia, 2014	2017	Mozambique, 2004	2017	-10
Cape Verde, 1992	.	Lithuania, 1992	1993	.
Dem. Rep. Congo, 2005	.	Niger, 2010	2013	.
Morocco, 1972	.	Benin, 1970	1972	.

The matched pairs. Treated Countries are those that had an institutional change that resulted in slate or pair selection. Treated countries are matched to control countries that also experienced a change to their selection institutions, but those changes did not include a movement to slate selection. Pairs were matched within decades on the percent women in the lower house of the legislature.

as a slate. In addition, I match on decade and the percentage of women in the lower house.[11] Matching on decade controls for time, and matching on the percentage of women in the legislature controls for pressure for the selection of women to important political posts (Hoekstra, Kittilson and Bond, 2014). Table 6.6 lists the treated countries and their matched pairs.

Analysis

To determine the effect of an institutional change to slate selection, I use a sign-rank statistic (see Glynn and Ichino (2015)). For all treatment and control countries, I calculate the number of years between treatment and the selection of the first

[11] Using "optmatch" in r.

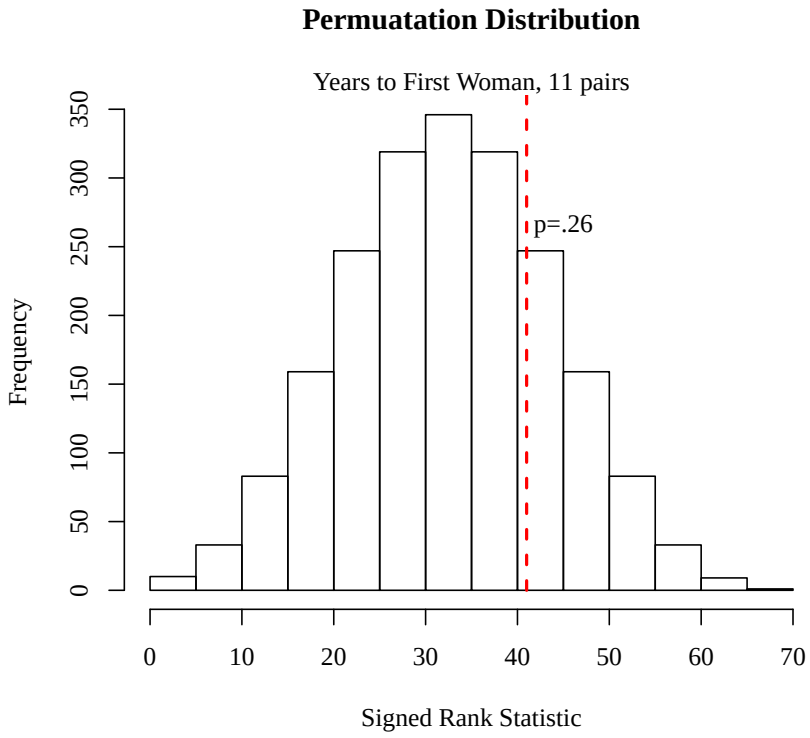
woman to the peak court.[12] Then, for each pair, I calculate the difference in the number of years until the selection of the first woman between treatment and control countries.[13] Next, I rank the pairs by the size of the difference in years (the absolute value difference) from smallest to largest difference. That is, the pair with the smallest difference in years until the first woman is assigned a rank of 1 and the pair with the greatest difference in the number of years until the first woman is assigned the highest rank. The signed rank statistic is then calculated as the sum of the ranks between treatment and control countries for which the treated country "won" by selecting a woman to the court before the control country. For example, if there were five pairs and in each pair the treated country selected a woman in fewer years than the control country, the signed-rank statistic would be equal to 1+2+3+4+5. This statistic is larger when more of the pairs have a "winning" treated country and/or when the pairs in which the treated country "wins" have the greatest difference (the highest ranks).

To determine whether the observed signed rank statistic is unusual or not, I generate a reference distribution by permuting assignment of treatment and re-calculating the singed rank statistic for each permuted sample. This procedure results in a distribution that shows all possible signed rank statistics and the frequency of observing the signed rank statistic among all permutations. Then, I compare the observed signed rank statistic to the permutation distribution to determine if the observed statistic is in the tails of the distribution and, therefore, unlikely due to chance. Figure 6.6 shows the permutation distribution and the observed sign rank statistic. It shows that the observed signed rank statistic indicates a positive but statistically insignificant relationship between slate treatment and the timing of the selection of the first woman to the peak court. The observational evidence presented here are not sufficient to reject

[12]Countries that had women on the court prior to the institutional change were dropped from the sample.

[13]That is: (year first woman for a treated country - treatment year for a treated country) - (year first woman for a control country - year of institutional change for the control country)

Figure 6.6: Permutation Distribution and Observed Statistic



This graph shows the permutation distribution of all possible assignments of treatment to matched pairs. The red dashed line shows the observed sign rank statistic. 11 pairs.

the null hypothesis of no relationship between slate selection and gender diversity on the bench.

It is worth noting, however, that this study is conservative. Institutions should only affect gender diversity under particular circumstances. Specifically, in countries and time periods when observers do not expect (or, even, do not want) gender diversity on the bench, the ability of observers to more easily identify gender disparity under slate selection should not affect outcomes. We should, therefore, expect institutions to have the largest affect under contexts of moderate pressure for gender diversification in politics (Arrington et al., 2018). Future research should to address the role of slate selection procedures in contexts where there is moderate social pressure for gender diversity.

6.4 Conclusions

The logic explaining why party list PR systems are associated with greater gender diversity in the legislative context is not – I argue – limited to the legislative context. In this project, I applied the logic of list PR systems to the selection of peak court judges. I hypothesized that selecting judges as a slate rather than on a rolling, one-by-one basis should facilitate gender diversity by providing more information to respondents and by encouraging respondents to asses candidates as a "balanced" group rather than individually. Evidence from survey experiments shows that respondents are more critical of gender disparity when judges were selected as a slate rather than one-by-one, even when respondents across the two groups saw the same information. Based on respondents' qualitative explanations for their choices, it appears that those in the slate selection group were more likely to notice gender disparity than those in the on-by-one group.

Observational evidence is less conclusive. Using a matching design for casual inference, I find a positive but statistically insignificant effect of slate selection on the timing of gender diversification cross-nationally. I note, however, that the analysis presented here is conservative: the role of institutions in affecting gender diversity depends on the social context. In settings where communities do not expect or even want women serving on constitutional courts, there should be little effect of institutions. Future research should situate institutional change within a broader social/political context.

Chapter 7
Intersectionality: Minority Women State Supreme Court Judges[1]

This dissertation, so far, has focused on gender diversity in the judiciary. However it is worth noting two details. First, the theoretical framework outlined here is not specific to women judges; this framework can be applied to the selection of many different types of groups. The complication, though, is that the fewer the members of the group – that is, the as a group holds a smaller and smaller proportion of the candidate pool – the harder it will be for observers to make accurate inferences about the fairness of selection institutions. Specifically, for some small groups, the frequent absence of representatives from that group in office will be consistent with a *fair* selection process.

Second, by focusing on women as a group, the project thus far has largely overlooked diversity *among* women. In this chapter, I leverage a new data set on the race/ethnicity of U.S. state supreme court justices in order to shed light on the increasing presence of minority women state supreme court judges. In this chapter, I describe characteristics of judges and compare minority women to minority men and white women. I show that minority women are – on average– more similar to minority men and white women on the characteristics included here.

Until very recently, any assessment of minority women on the bench was constrained by the marked absence of female judges of color.[2] At the state supreme court

[1]This work is excerpted from my chapter in *The Politics of Race Gender and the Judiciary*, Eds. Sharon Navarro and Samantha Hernandez, forthcoming.

[2]I identify the minority judges who describe themselves as members of minority racial/ethnic groups or judges who have been identified as members of minority groups by other scholars or news media. When I refer to the "first" members of a racial/ethnic group on the bench, I mean the first member of that racial or ethnic group that I was able to identify.

level, the first minority woman justice was a Latina judge named Dorothy Comstock Riley selected to the Michigan supreme court in 1982.[3] Juanita Kidd Stout was selected to the Pennsylvania supreme court in 1988, becoming the first Black female state supreme court justice.[4] A year later Joyce Luther Kennard became the first Asian American/Pacific Islander state supreme court judge when she was selected to the California supreme court.

Since the early 1980s, 36 women of color have been selected to state supreme court benches. The inclusion of women of color on the highest state courts coincided with increases in the presence of white women and men of color as well. Figure 7.1 shows the proportion of state supreme court judges who are white men, white women, minority men, and minority women over time.[5]

Given the newness and relative rarity of minority women to state supreme court benches, little is known about the characteristics of minority women judges and the circumstances under which they are selected to states' highest courts. It may be that minority women judicial candidates or nominees are held to a particularly high standard and have to overcome additional hurdles.[6] For example, Goldman et al. (2000) find that among President Clinton's district court judicial appointments, the time between nomination and a Judiciary Committee hearing for women and minority judicial nominees took, on average, 43 days longer than the time for white male judges (120 days versus 77, respectively). Insights from the intersectionality paradigm

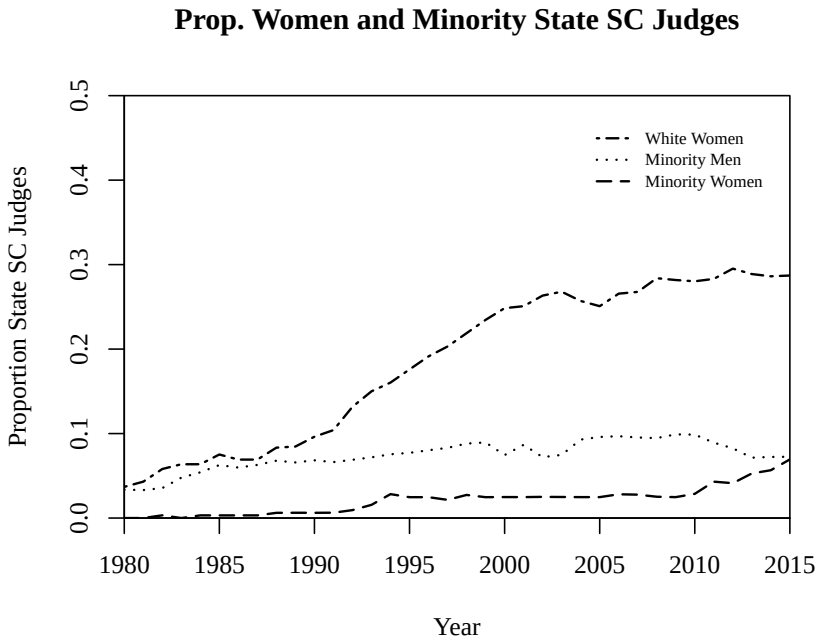
[3]The Michigan Supreme Court Historical Society identifies Judge Riley as Hispanic, although Martin and Pyle (1999) describe her as "a white Repubican"(p. 1207).

[4]Julia Cooper Mack, a Black woman judge, was appointed to the D.C. Court of Appeals – the DC equivalent of a state supreme court – by President Ford in 1975 (Blackburne-Rigsby, 2009).

[5]Data on the gender of judges and whether judges are Black/African American from 1960 to 2010 are from the State High Court and Justice Database (Bratton, 2017). Data on the gender of judges selected since 2010 were collected by the author from Ballotopedia and internet searches. Indicator variables for Latinx and Asian American/Pacific Islander judges were generated through searches of NALEO directories and the Asian American and Pacific Islander Almanacs. In addition, keyword searches of Asian American, Pacific Islander, and Latino/a judges were used to augment missing directory and almanac years. Finally, when photographs of judges suggested a racial or ethnic identification, electronic newspaper searches or biographies were used to confirm race or ethnicity.

[6]In the legislative context, see Lawless and Fox (2005); Milyo and Schosberg (2000); Anzia and Berry (2011).

Figure 7.1: Women, Minority Men, and Minority Women State Supreme Court Judges Over Time



The proportion of judges selected to state supreme courts over time who are minority women, minority men, and white women. Most judges are white men, and much of the diversity on the bench gained in recent decades is due to the increased presence of white women.

(Crenshaw, 1989; Hancock, 2007; Hawkesworth, 2003) suggest that minority women may have experiences that are distinct from both white women and minority men. If minority women judges face hurdles related to both gender and race, they may be held to a particularly high standard for selection.

This chapter uses a new data set on the racial/ethnic identification of state supreme court judges to describe ways in which characteristics of minority-women state supreme justices are similar to or different from characteristics of women justices and minority men justices. There are myriad ways in which minority women may have different experiences in the accumulation of qualifications, in selection, and in retirement compared to white women or minority men. The characteristics addressed here are by no means exhaustive. This analysis provides a preliminary look at the ways in which minority women are similar to and district from their white and male counterparts on state supreme court benches.

The next sections outline arguments for why the inclusion of women, minority, and minority women judges is important, then compare several characteristics of state supreme court judges across race and gender. These characteristics are organized into sections on qualifications, selection, and retirement. Because multiple comparisons with one limited sample increases the probability of finding a relationship where one does not exist (type I errors), findings should be considered preliminary. That said, I find that for the characteristics assessed here, minority women state supreme court justices are more similar to minority men than to white women. Across all comparisons, minority women judges are more similar to their white and male counterparts than they are distinct.

7.1 The Importance of Race and Gender Diversity on the Bench

Although political representation is often addressed in the legislative context, the implications of descriptive, symbolic, and substantive representation are particularly important in the context of courts because "judges have a more direct and irrevocable impact in the lives of many Americans than local or even national legislators." (Ifill, 2000, p. 407-8). Indeed, existing research suggests that a diverse judiciary can have important consequences on judicial behavior and outcomes, judicial legitimacy, and judicial discourse.

Latina Justice Sonia Sotomayor, prior to her appointment to the U.S. Supreme Court, said, "Whether born from experience of inherent physiological or cultural differences... our gender and national origins may and will make a difference in our judging" (Sotomayor, 2002, p. 92). Similarly, Judge Anna Blackburne-Rigsby, a Black women currently serving as the Chief Judge of the District of Columbia Court of Appeals, wrote, "I have seen that being both black and female brings an important additional voice to the deliberative process, but that voice is varied because there is no singular 'black woman' perspective" (Blackburne-Rigsby, 2009, p. 689).

Although empirical evidence of race and gender differences in judging is mixed, there is some evidence that race and gender identity and experiences may affect judging. For example, Welch, Combs and Gruhl (1988) find racial differences in decisions to incarcerate and in sentencing decisions at the trial court level. Collins, Manning and Carp (2010) find that at the district court level, women judges make different decisions than their male counterparts on civil rights and liberties cases and criminal justice cases. Similarly, they find that minority judges decide more liberally than white judges across all issue areas except labor and economics. Gruhl, Spohn and Welch (1981) find that women judges are more likely to sentence women defendants to prison than male judges, although on other measures men and women judges convict

and sentence similarly. Allen and Wall (1993) find that women justices on state supreme courts are more likely to decide pro-woman on women's issues; Songer and Crews-Meyer (2000) find that women state supreme court judges voted more liberally than male judges in death penalty and obscenity cases, and McCall and McCall (2007) find that women state supreme court justices decided more liberally than male judges on search and seizure cases after 1991. Others, however, find no race or gender differences in judging or find that differences have been overstated (Westergren, 2003; Dixon, 2009; Kenney, 2008; Walker and Barrow, 1985; Segal, 2000).

Evidence suggests that diversity on the bench can affect the behavior or male judges as well. For example, the presence of women judges on a three judge U.S. appellate panel affects the behavior of male judges: Farhang and Wawro (2004) They find that male judges voted more liberally in anti-discrimination cases when there was a woman serving on the panel. Similarly, Boyd, Epstein and Martin (2010) find that male appellate judges are more likely to decide in favor of plaintiffs in sex discrimination cases when a woman is serving on a panel alongside them.

Finally, diversity in the life experiences of those on the bench can alter the judicial discourse and the deliberative process (Ifill, 2000) and lead to "structural impartiality" (Ifill, 1998, p. 99). Ifill (2000, p. 455) writes, "we should value racial diversity if it brings alternative perspectives and analysis to the process and enriches the legal decision-making."

Regardless of the extent to which women or minority judges make different decisions than their white male counterparts, add new perspectives to the deliberative process, or alter the trajectory of judicial decision-making, the presence of diverse judges has important symbolic consequences (Scherer and Curry, 2010) and signals that powerful state institutions are open to women, minority, and minority-women individuals. Even in the *absence* of racial or gender differences in behavior, Kenney (2013) argues that the presence of women on judicial benches is important because

the presence of women in the judiciary "normalizes women's authority and power" (p. 175). Understanding characteristics of minority women judges and the conditions under which they are selected to state supreme courts is an important next step in understanding the role of diversity in the judiciary, but the study of minority-women judges has been minimal.

7.2 Characteristics of Minority, Women, and Minority Women Judges

Using new data on the race and gender of state supreme court judges, I identified 98 minority men and 36 minority women state supreme court justices selected between 1970 and 2016. Table 7.1 lists the number of men and women who are Black/African American, Asian American or Pacific Islander, Latinx, Native American, or white.[7] Of Black judges, 25% are female. Of Asian judges, 37.5% are female. The only Native American judge is female, and 18.4% of white judges are female. Of all male judges, just over 9% are men of color. Among women judges, almost 15% are women of color. Table 7.2 lists the 36 identified minority women state supreme court justices.

In the following sections, I compare characteristics of minority women, minority men, white women, and white men justices of state supreme courts. Reported p-values should be interpreted with caution: because several comparisons are made from one sample, the probability of committing a type I error – inferring a pattern where none exists – increases. The comparisons are of characteristics associated with qualification, selection, and retirement.

[7]The "white" category iuncludes all judges not identifed as either Black/African American, Latino, Asian American/Pacific Islander, or Native American. Judges described as Irish-American or Greek-American, for example, were classified as'white.'

Table 7.1: Judges Selected 1970-2016

	Male	**Female**	**(%Female)**
Black/African American	63	21	(25.0%)
Asian Am./Pacific Is.	10	6	(37.5%)
Latinx	20	9	(31.0%)
Native American		1	(100%)
White/Not Identified	935	211	(18.4%)
(% Minority)	(9.1%)	(14.9%)	

Gender and racial/ethnic summary of judges selected between 1970 and 2016, with the exception of Lorna Lockwood, a white, female Justice of the Arizona Supreme Court selected in 1961 and the first woman selected to a state supreme court. She is included here for an exhaustive count of women judges selected through 2016.

Table 7.2: Minority Women State Supeme Court Judges

Name	State	Year Joined	Race/ Ethnicity
Riley, Dorothy Comstock	MI	1982	Latina
Stout, Juanita Kidd	PA	1988	Black/African Am.
Kennard, Joyce Luther	CA	1989	Asian Am/Pcf. Is.
Sears, Leah Ward	GA	1992	Black/African Am.
Ciparick, Carmen Beauchamp	NY	1993	Latina
Nakayama, Paula Aikko	HI	1993	Asian Am/Pcf. Is.
Johnson, Bernette Joshua	LA	1994	Black/African Am.
Roaf, Andree Layton	AR	1994	Black/African Am.
Selby, Myra Consetta	IN	1994	Black/African Am.
Toney-Williams, Felicia	LA	1994	Black/African Am.
Brown, Janice Rogers	CA	1996	Black/African Am.
Maes, Petra Jimenez	NM	1998	Latina
Quince, Peggy A.	FL	1998	Black/African Am.
Baldwin, Cynthia Ackron	PA	2005	Black/African Am.
Timmons-Goodson, Patricia	NC	2006	Black/African Am.
Guzman, Eva	TX	2009	Latina
Marquez, Monica M.	CO	2010	Latina
Brown, Yvette McGee	OH	2011	Black/African Am.
Cantil-Sakauye, Tani Gorre	CA	2011	Asian Am/Pcf. Is.
Duffly, Fernande R.V.	MA	2011	Asian Am/Pcf. Is.
McKenna, Sabrina S.	HI	2011	Asian Am/Pcf. Is.
Powell, Cleo E.	VA	2011	Black/African Am.
Beasley, Cheri	NC	2012	Black/African Am.
Wright, Wilhelmina	MN	2012	Black/African Am.
Abdus-Salaam, Sheila	NY	2013	Black/African Am.
Espinosa, Carmen	CT	2013	Latina
Rivera, Jenny	NY	2013	Latina
Watts, Shirley Marie	MD	2013	Black/African Am.
Hines, Geraldine	MA	2014	Black/African Am.
Yu, Mary	WA	2014	Latina & Asian Am.
Hotten, Michele	MD	2015	Black/African Am.
Hudson, Natalie	MN	2015	Black/African Am.
Kruger, Leondra	CA	2015	Black/African Am.
Budd, Kimberly	MA	2016	Black/African Am.
Makamoto, Lynn	OR	2016	Asian Am/Pcf. Is.
McKeig, Anne	MN	2016	Native Am.

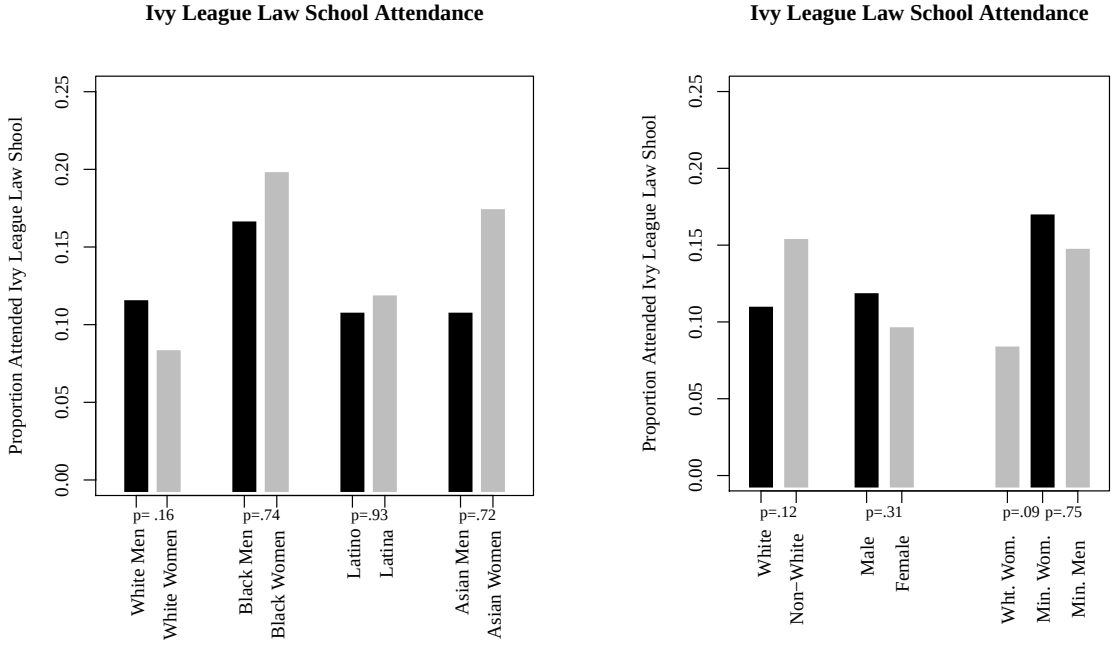
Minority-Women state supreme court judges that have been identified as female and minority.

7.2.1 Qualification Characteristics

If women, minority, or minority women are held to a higher standard for qualification, then we may see discrepancies in qualification characteristics across race and gender. To test whether minority women are held to a higher standard of qualification, I compare the rates of ivy league law school attendance, prior judicial experience, and age at selection. If white women, minority men, or minority women are held to a higher standard of qualification, more women and minority judges may have attended ivy league law schools; more women and minority judges may have had prior judicial experience, and minority and women judges may have been older at the time of selection to allow for the accumulation of qualifications.

Ivy League Law School Attendance Figure 7.2 shows that there are no significant gender differences in ivy league law school attendance among co-ethnics, although greater proportion of minority women attended ivy league law schools than white women (p=.09). There are no clear differences across race and gender for ivy league law school attendance. When we measure elite education as ivy league law school attendance, minority women are not held to a higher standard of education.

Figure 7.2: Ivy League Law School Attendence

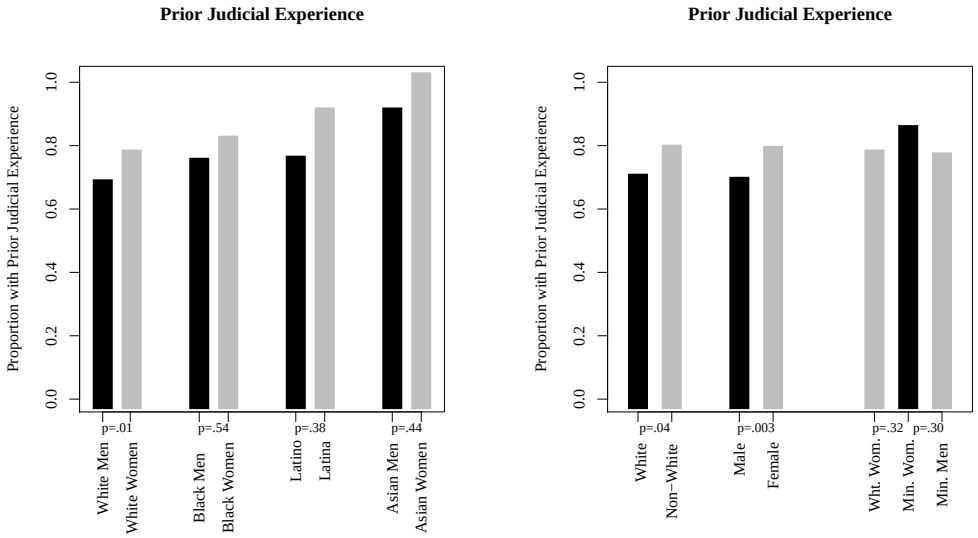


The proportion of judges who attended Ivy League law schools by identity groups. Small numbers of Latino, Black, and Asian/Pacific Islander judges limits statistical significance.

Prior Judicial Experience Epstein, Knight and Martin (2003) argue that the norm of prior judicial experience for U.S. Supreme Court justices may serve to restrict professional, gender, and racial diversity on the Supreme Court. They argue that the disparity of women and minority judges on federal benches restricts prospects for diversity on the Supreme Court. Martin (1987) acknowledges that in the selection of federal judges, both Reagan and Carter may "have applied a double standard in demanding more judicial experience from women" (p. 141). If women, minority, and minority women are held to a higher standard for qualification, we may observe a greater proportion of non-white and non-male judges with prior judicial experience.

Figure 7.3 shows that across all race/ethnic groups, a greater proportion of women had prior judicial experience than men, however the gender difference among co-ethnics is only statistically significant among white judges (p=.01). Aggregating to compare white and minority judges, a greater proportion of minority judges had prior judicial experience than white judges (p=.04) and a greater proportion of women than men had prior judicial experience (p=.003). However, there are no differences between minority women and white women or minority men.

Figure 7.3: Judicial Experience

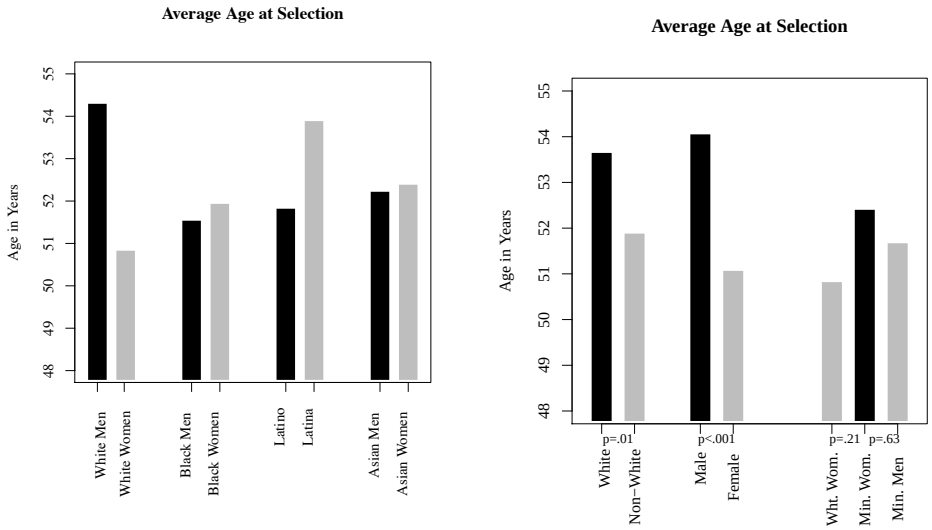


Proportion of judges with prior judicial experience. The only within ethnicity gender difference is among white judges (p=.01). Aggregating to all white versus all non-white judges, the p-value for the difference in proportions is p=.04, and the p-value for the difference in proportion between all men and all women is p=.003. These comparisons indicate that a greater proportion of women state supreme court judges have prior judicial experience relative to men.

Age at Selection In the legislative context, female candidates (Burrell, 1992) and representatives tend to be older than their male counterparts (Dubeck, 1976; Carroll, 1983; Moncrief and Thompson, 1992). Traditional explanations for this discrepancy suggested that women delayed running for office until after their children were born or were older. Carroll (1983) found, however, that among state legislators, patterns of delaying office due to the number and age of children were consistent between men and women, suggesting that childbirth and rearing was not the explanation for the age difference between men and women legislators. An alternative explanation focuses on gender differences in political ambition among younger potential candidates; in surveys of potential legislative candidates, there was a substantial gender gap among men and women under 40 in responses to questions about whether they had considered running for office, had seriously thought about a political career, or who had discussed the prospect of running for office with party leaders (Lawless and Fox, 2005). Delaying

candidacy also allows for more time to accumulate qualifications.

Figure 7.4: Age at Selection



Average age at time of selection. White judges are older at the time of selection than non-white judges (p=.01) and men are older than women (p<.001). The gender difference in age is driven by the gender difference among white judges (p<.001) as there are no gender differences among Back, Latina, or Asian American judges.

In the context of the judiciary, the appointment of (some) judges rather than election may diminish the effect of disparate political ambition among men and women because appointed judges do not need to organize and undertake a campaign.[8] In addition, differences in the pool of judicial candidates versus legislative candidates may also affect gender differences in aging across context: most state have a formal or informal requirement that judges hold a law degree or are members of the bar. Since graduation rates of women from law schools have increased substantially in the last several years, the candidate pool for female state supreme court judges may be younger than the candidate pool for male supreme court justices, which could serve to cancel out differences in political ambition or time to accumulate qualifications between male, female, white, and minority judicial candidates. Indeed Yoon (2003)

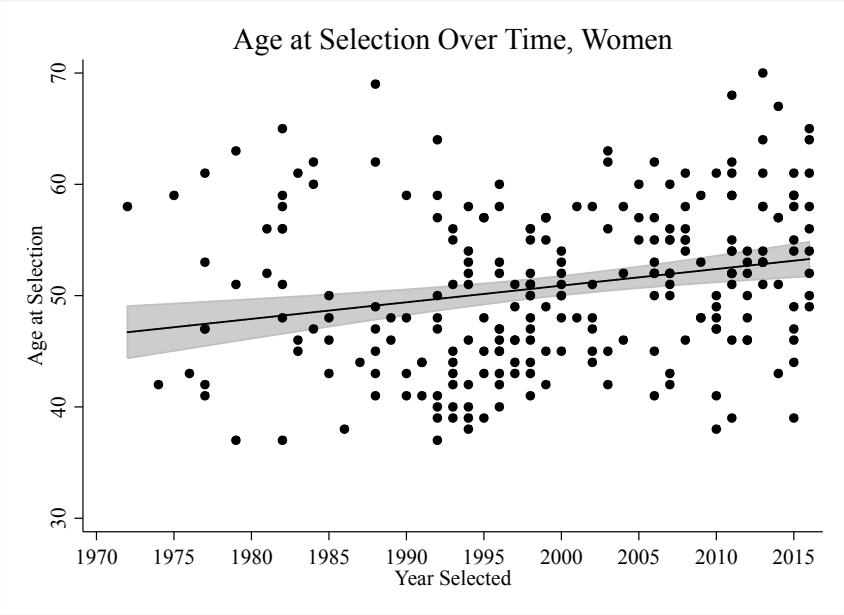
[8]This is not to say that appointed judges are not or have not been involved in politics

finds that women and minority judges on the federal bench between 1945 and 2000 began their judgeships at a younger age than their white male peers. Black judges were, on average, 49.2 years old when then joined the bench. Women were 47.2 years old while white and male judges averaged 51.3 years.

Figure 7.4 shows the average age of judges at the time of selection. White men are, on average, the oldest at selection (54.1) and white women are the youngest (50.6). On average, white judges are older at selection than non-white judges ($p=.01$) and men are older than women ($p<.001$). The gender difference among all judges is due to the gender difference between white men and white women ($p<.001$), as there are no gender differences in age at selection among minority women and men.

If the age difference between men and women stems from an age difference in the candidate pool because women have only relatively recently been graduating from law school at rates commensurate with men, then the age difference ought to decrease over time as the candidate pool of women lawyers for judgeships has gotten larger and has been able to age. Figure 7.5 shows a scatter plot of the ages of women judges selected over time. The black line is the regression of age by year. The grey shaded region is the 95% confidence interval. Woman have been getting older at selection over time, but the change is slight, suggesting that it may not just be differences in the maturity of the candidate pool that accounts for the age difference between men and women judges.

Figure 7.5: Age at Selection Overtime



The average age at selection for women over time. The average age has increased slightly *over time. A one year increase in the appointment year is associated with a .15 increase in average age at selection (p<.001)*

Summary: Similarities and Differences in Qualifications

The qualification characteristics of white men, white women, minority men, and minority women are similar in some ways and distinct in others. There are no differences across gender or race/ethnicity in Ivy League law school attendance. When it comes to prior judicial experience, across all ethnic categories, a greater proportion of women had prior judicial experience, but the only gender difference among co-ethnics is among white judges. For age at time of selection, white judges and male judges are older – on average – than women and minority judges. The gender discrepancy in age at selection between men and women judges is driven by the age difference among white men and women judges (p<.00). Among racial minority judges, there are no statistically significant differences in age.

7.2.2 Selection Characteristics

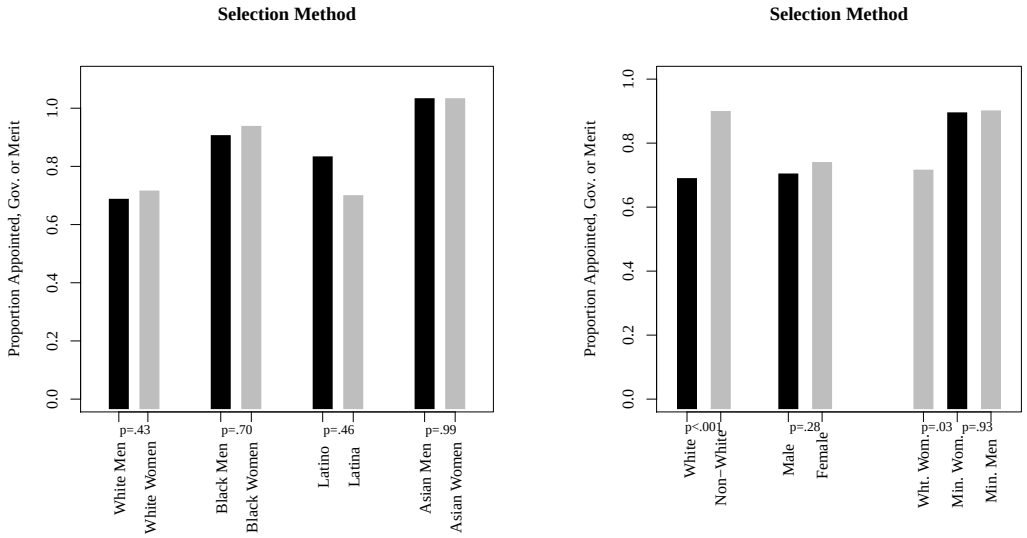
Moving from qualification characteristics to characteristics of judicial selection, in this section I compare party identification and the method of selection across race and gender.

Appointment versus Election Given the importance of minority and women judges on the bench, scholars have studied the conditions under which minority and women judges are selected to courts. Many explanations focus on differences in selection institutions. Findings on the effect of different selection methods – executive appointment, elections, or merit selection – on the appointment of women and minority judges are mixed. Some find that the concentration of accountability on elite selectors such as a governor (Bratton and Spill, 2002; Carbon, Houlden and Berkson, 1982) or merit commission (Goelzhauser, 2011; Esterling and Andersen, 1999) leads to greater diversification.[9] However, many scholars find no relationship between selection method and diversity on the bench (Hurwitz and Lanier, 2003; Alozie, 1988, 1990, 1996; Hurwitz and Lanier, 2001).

Figure 7.6 shows the proportion of judges appointed by elites, either through gubernatorial selection or merit selection. There are no gender differences among co-ethnics or in the aggregate, but there does appear to be a difference between white judges and minority judges; a greater proportion of minority judges were appointed relative to white judges ($p<.001$). Among women, a greater proportion of minority women were appointed than white women.

[9]Likewise, Hall (2001) find some – albeit weak ($p=.06$) – evidence that minority judges fare worse in non partisan elections than their peers. Goelzhauser (2011) finds that when citizen liberalism is relatively high, partisan election is associated with state supreme courts having their first Black justice selected sooner. Glick and Emmert (1986) find that merit selection systems are the least likely to select religious minorities to state supreme court benches.

Figure 7.6: Selection Method



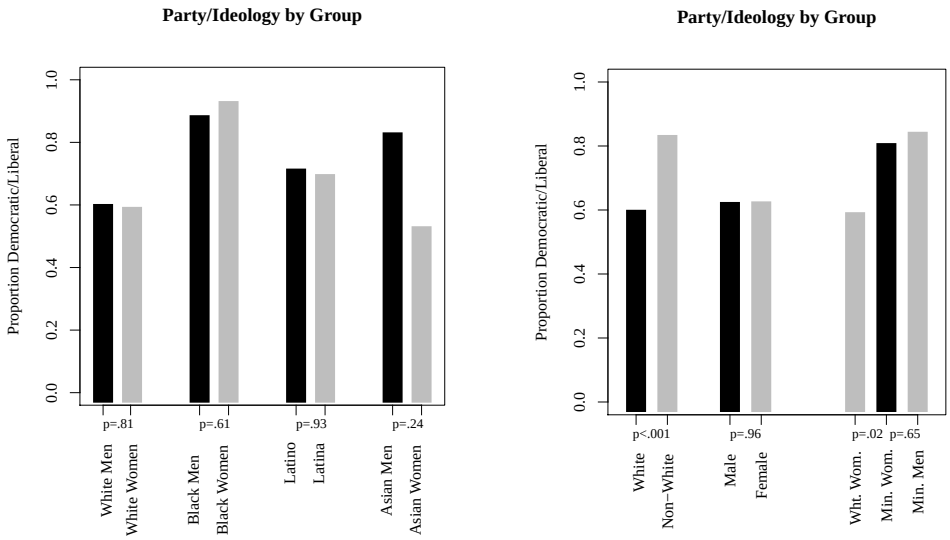
Proportion of judges selected through Gubernatorial selection or Merit Selection. Includes judges selected by the Governor to fill a vacancy in between elections.

Party In addition to selection institutions, the partisanship of those tasked with selecting judges may shape judicial diversity: liberal selectors are expected to select more women and minority judges.(Bratton and Spill, 2002), and liberal constituencies may be more likely to elect women and minority candidates both because diversity has become an issue associated with the Democratic party[10] and because women and minority judges may be more likely to run as Democrats in partisan elections due to gender and racial partisan gaps (e.g. Kaufmann, 2002; Gay and Tate, 1998). Figure 7.7 shows that there are no gender differences among co-ethnics in party identification, but there is a difference (p<.00) between white and minority judges.[11]

[10]For exampple, *The 2016 Democratic Platform* states, "Above all, Democrats are the party of inclusion. We know that diversity is not our problem – it is our promise" DNC (2016).

[11]Party identification information comes from Bratton's original data set for judges selected prior to 2010. For judges selected after 2010, partisanship comes from (1) the candidate's party in partisan elections, (2) party affiliation described by newspapers in descriptions of the candidates, and (3) party self-identification of candidates in newspaper articles or biographies. If party attribution could not be gleaned from the above three rules, the party of the appointing governor is used as a proxy (for appointed judges only).

Figure 7.7: Party/Ideology



Proportion of Judges who have been identified as liberal or Democratic.

Summary: Selection Characterisitcs

Among selection characteristics, two patterns emerge. First, a greater proportion of minority judges aren appointed by elites than white judges. Second, a greater proportion of minority judges are identified as liberal or Democratic than white judges. Both of these findings are resilient to p-value corrections for multiple comparisons.

7.2.3 Retirement characteristics

In this final section I compare characteristics of retirement across race and gender. Specifically, I compare length of tenure and reasons for vacating the bench, and then I describe patterns of replacement for minority women judges.

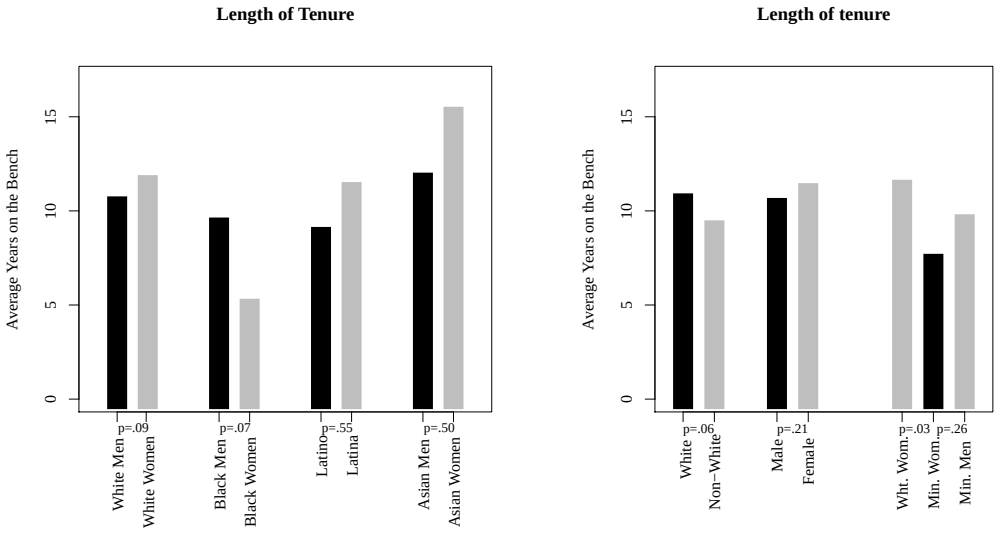
Length of Tenure There are a few reasons why we might expect women, minority, or minority women to serve on the bench for a different length of time than white or male judges. Women, minority, and minority women may be appealing candidates for the federal bench, particularly for presidents who aim to diversify the federal judiciary. If women and minority judges are often nominated to the federal judiciary,

their tenures on state supreme court benches may be brief. In addition, experiences of sexism or racism while serving on the bench may lead female and minority judges to retire earlier than their white male colleagues. Finally, differences in the likelihood of winning retention elections or facing electoral opposition may cause discrepancies in tenure length (although Luskin et al. (1993) find no racial differences in retention election vote share). In contrast, limitations in moving up the career ladder outside of the judiciary – that is, less opportunity in the private or academic sectors – may leave women, minority, and minority women delaying retirement and maintaining their seats longer. Yoon (2003) finds that for judges on federal courts between 1945 and 2000, on average, women served 8.5 years less than their male counterparts while Black judges served 3.9 years less than white judges.

Figure 7.8 shows racial and gender differences in length of tenure on state supreme court benches. White women may serve longer than white men (p=.09); Black men may serve longer than Black women (p=.07), white judges may serve slightly longer than non-white judges (p=.06), and minority women serve for a shorter length of time than white women due to short average length of tenure among Black women (p=.03). However, the multiple comparisons made from on sample with few observations for minority women, it should be noted that the probability of committing a type-I error is greater than the reported p-values.

Reasons for Leaving the Bench While length of tenure does not show clear discrepancies across race or gender, there may be patterns across race and gender for the reasons judges leave the bench. Table 7.3 shows the number and percentage of judges who vacated the bench for different reasons. A greater percentage of minority women are selected to temporary, interim appointments than white and male judges, and a greater percentage of minority women are nominated to the federal judiciary than white and male judges. Table **??** in the appendix shows the results of several

Figure 7.8: Length of Tenure



The average length of tenure of judges on State Supreme Courts.

linear probability models that estimate race and gender based on reasons for vacating the bench. The only differences that remain statistically significant once multiple comparisons are accounted for are differences in interim appointments: both women and minority judges are more likely to be appointed on an interim basis that white and male judges. There are no other statistically significant differences in reasons for vacating state supreme court benches across race and gender.

Table 7.3: Reasons for Vacancies

Reason for Leaving the Bench	**White Men**	**White Women**	**Minority Men**	**Minority Women**
Mandatory Retirement	147 22.7%	17 13.8%	15 20.5%	3 18.6%
Defeated, Removed, Not re-nominated	66 10.2%	18 14.6%	13 17.8%	2 12.5%
Retired or Resigned	341 52.7%	68 55.%3	33 45.2%	5 31.3%
Nominated to Federal Judiciary	35 5.4%	9 7.3%	2 2.3%	2 12.5%
Interim Appointment	9 1.4%	6 4.9%	5 6.8%	3 18.8%
Died in Office	35 5.4%	5 4.1%	4 5.5%	1 6.3%
Appointed/Elected to National office	14 2.2%		1 1.4%	
Total, non-missing	647	123	73	16

Reasons for vacating the bench by race and gender. There have only been 15 judges identified as minority women to serve and leave the bench. Dorothy Comstock Riley of Michigan is counted twice in this table: in 1983 she was removed from office after she was appointed to fill a vacancy by Governor Milliken right before his term ended. In 1984 she was elected to the Supreme Court and served until she retired in 1997. There is substantial missing data for white men (n=111) that is not included in the calculation of the percentages for "All Men."

Replacement Patterns A final analysis addresses the predecessors and successors of minority women state supreme court judges. Arrington (2018) shows that the gender of vacating judges and replacement judges are interdependent: when a woman judge vacates the bench she is more often replaced by another woman. When men vacate, they are more often replaced by men than women. Replacing a minority woman justice with a white man justice will decrease both the gender and racial diversity on the bench. Therefore, judicial selectors may face additional pressure to replace a vacating female judge of color with a woman, minority, or minority woman in order to maintain some diversity on the bench.

Of the 15 minority women who have left the bench, 13 have clear successors.[12] Of those 13 minority women judges, six were replaced by white women, four were replaced by minority women, one was replaced by a minority man, and two were replaced by white men.

Of the 36 identified minority women selected to a state supreme court, 35 have clear predecessors.[13] Of the 35 minority women justices selected to the bench, 18 replaced white men, eight replaced minority men, five replaced white women, and four replaced minority women. In other words, 48.6% of minority women were selected to fill seats vacated by non-white-men. Table 7.4 reports the number and percentage of white men, minority men, white women, and minority women who were the predecessors and successors of minority women alongside the percentage of judges selected in total between 1970-2016 and 2000-2016 as a comparison. Given the high overall proportion of judges who are white men, minority women have low rates of being replaced by white men. Although there are very few minority women who have vacated the bench, only 15% of their successors were white men. In contrast, between 2000 and 2016, 61% of judges selected to the bench were white men. The pattern of

[12] I have not confirmed the successors of Cynthia Baldwin and Felicia Toney-Williams. Both justices were selected on an interim basis.

[13] Bernette Joshua Johnson of Louisiana was selected to fill a new seat on the bench.

Table 7.4: Predecessors and Successors

	Predecessors of Minority Women	**Successors of Minority Women**	**All Judges Selected, 1970-2016**	**All Judges Selected, 2000-2016**
White Men	51.4% (18)	15.4% (2)	73.3% (934)	61% (239)
Minority Men	22.9% (8)	7.7% (1)	7.3% (93)	8.2% (32)
White Women	11.3% (5)	46.2% (6)	16.5% (201)	25% (98)
Minority Women	11.4% (4)	30.8% (4)	2.9% (35)	5.9% (23)

Race and gender characteristics of those who preceded and succeeded minority women state supreme court justices. Due to the vast number of white judges, I do not have specific information on who succeeded and replaced each white judge.

vacancies made by minority women being predominately filled by other women and minorities suggests that selectors face strong incentives to maintain some diversity on the bench when women of color vacate.

7.3 Conclusion: Are Minority Women Judges Distinct?

Using data on the racial and gender identification of women, minority, and minority-women state supreme court judges, the goal of this project was to uncover potential ways in which women, minority, and especially minority-women are similar to and different from white and male judges. Although statistical significance should be interpreted with caution due to the increased probability of committing type I errors with multiple comparisons, the evidence presented here indicates that there may be a few race and gender differences in characteristics of state supreme court justices.

Minority women are more likely to have attended ivy league law school than white women but are no more or less likely to have prior judicial experience than white women or minority men. Minority women are not discernibly older or younger than white women or minority men at selection, while white men are, on average, the oldest at time of selection. Minority women are less likely to be elected than white women, but there are no differences in selection method between minority men and minority women. Minority men and women are more likely to be liberal or Democratic than white men or white women, and minority women have shorter average length of tenure than white women. Minority women have the highest rate of interim appointment, and they also have the highest rate of nominations to the federal judiciary. Although there have only been 15 minority women who have left the bench, their vacancies are rarely filled by white men. Instead, vacancies by minority women are most often filled by white women and other minority women.

Across the six categories where minority women can be directly compared to

minority men and white women (ivy league education, prior judicial experience, age, selection method, party, and length of tenure), minority women are indistinguishable from minority men in all six categories. In contrast, minority women differ from white women across four categories (ivy league education, selection method, ideology, and tenure). The comparisons presented here suggest that minority women state supreme court justices may be more similar to minority men than white women.

Chapter 8 Conclusion: Accountability for Diversity on the Bench

Diversity on the bench can have important substantive effects on the discourse, outcomes, and legitimacy of the judiciary (Scherer and Curry, 2010; Farhang and Wawro, 2004; Boyd, Epstein and Martin, 2010; Collins, Manning and Carp, 2010; Kenney, 2012). Despite the importance of gender diversity in the judiciary, there is substantial variation in the timing and level of diversity, both within the United States and cross-nationally. Scholars have posited several potential explanations: culture and gender roles (Norris, 1987; Siaroff, 2000; Inglehart and Norris, 2003); political socialization and the willingness of women to participate in political office (Burns, Schlozman and Verba, 2001; Verba, Burns and Schlozman, 1997; Chhibber, 2002; Richard L. Fox, 2004; Lawless and Fox, 2010); norm diffusion across space and institutions (Hoekstra, Kittilson and Bond, 2014; Williams and Thames, 2008); prestige (Williams and Thames, 2008); the nature of the judicial system (Remiche, 2015; Schultz and Shaw, 2013); and selection institutions (Williams and Thames, 2008; Carbon, Houlden and Berkson, 1982; Bratton and Spill, 2002; Gill, 2012; Alozie, 1988, 1990; Slotnick, 1984). However, empirical evidence is inconsistent.

This project presented and tested a holistic framework of judicial diversification in which judicial diversity is the outcome of citizen inferences about bias and accountability. For citizens to hold selectors accountable for homogeneous courts, citizens must be able to (1) make accurate inferences about whether or not bias is occurring in selection, (2) accurately attribute blame for perceived bias, and (3) hold those responsible for bias accountable to induce diverse selections. Each of these three steps is mediated by the specific institutional arrangements of the selection process. Institutional features that solicit trust will shape prior beliefs in the fairness of institutions,

which, in turn, shapes how observers interpret information. Institutions that affect the size of the court and turnover shape how much data citizens have to update beliefs, and the presence of multiple actors/steps in selection obscures blame attribution. Finally, institutions that shelter selectors from sanctions undermine accountability for diversity.

Chapter 3 showed how institutions can and do affect prior beliefs in the fairness of selection institutions. Respondents overwhelmingly perceived merit selection procedures as more fair than gubernatorial selection. While the term "merit" in merit selection had a small, positive effect on perceptions of fairness, the preference for the two-step process over gubernatorial selection persisted when it was labeled "commission assisted," which suggests that most respondents' preferences for the institution were based on the procedure rather than the name. Specifically, in the qualitative explanations for their responses, respondents indicated that it was the presence of multiple actors who can limit each other's power that explained their preferences for the commission based process over gubernatorial selection.

Chapter 4 addressed how institutions affect inferences about fairness in judicial selection. I found that institutions do affect perceptions of fairness. Respondents who were told a homogenous court was the outcome of merit selection were largely content with the all-male bench. In contrast, those who were told the process was the outcome of gubernatorial selection were more critical. In section 4.1.2, I showed that the presence of multiple actors in merit selection can obscure blame attribution. Under merit selection, blame was shared between the commission and the governor (absent any information about who was actually to blame) whereas blame was concentrated on the governor under gubernatorial selection.

The evidence from the survey experiment in this chapter suggests that the presence of multiple actors in selection systems (like merit selection) may undermine the process of gender diversification via two mechanisms: (1) prior beliefs and the inter-

pretation of information and (2) blame attribution. If true, we should observe slower and lower levels of gender diversity where judges are selected through a process with multiple actors relative to a process with a unitary selector. To test whether an institution requiring multiple actors to select judges is associated with delayed gender diversification cross-nationally, I employed a signed-rank statistic on matched data in which treatment is defined as a movement away from a unitary actor. I focused on institutional changes to better isolate the causal effect of institutions. There is a negative but statistically insignificant (p=.08) relationship between an institutional change away from unitary selection and the timing of the selection of the first woman justice to the peak court. On average, countries that changed *from* unitary selection were slower to select their first female justice.

In chapter 5, I turned to the question of selector behavior in light of expectations about accountability. If elites tasked with selecting judges to state supreme courts expect to be punished for reversion to all-male or less diverse courts, elites should work to maintain existing levels of diversity on the bench. One way of doing so is to seek out and choose women judges to replace vacating women judges. Indeed, the empirical data in this chapter showed that judicial replacement is gendered. Women judges are more likely to be selected to fill vacancies made by women than vacancies made by men. This pattern is consistent with the idea that those tasked with selecting judges are accountable to observers' preferences for gender diversity put forward in chapter 2.

Chapter 6 addressed an alternative institutional feature of judicial selection that might increase diversification. In this chapter, I applied the logic of list PR systems to the selection of peak court judges. I hypothesized that selecting judges as a slate rather than on a rolling, one-by-one basis should facilitate gender diversity by providing more information to respondents and by encouraging respondents to asses candidates as a "balanced" group rather than individually. Evidence from the sur-

vey experiment showed that respondents were more critical of gender disparity when judges were selected as a slate rather than one-by-one, even when respondents across the two groups saw the same information. Based on respondents' qualitative explanations for their choices, it appeared that those in the slate selection group were more likely to notice gender disparity than those in the one-by-one group. Observational evidence was less conclusive. Using a matching design for casual inference, I found a positive but statistically insignificant effect of slate selection on the timing of gender diversification cross-nationally.

Chapter 7 turned to diversity among women justices on state supreme courts. Using data on the racial and gender identification of women, minority, and minority-women state supreme court judges, I compared characteristics of minority women state supreme court judges to their male and white colleagues. Across the six categories where minority women were directly compared to minority men and white women (ivy league education, prior judicial experience, age, selection method, party, and length of tenure), minority women were indistinguishable from minority men in all six categories. In contrast, minority women differed from white women across four categories (ivy league education, selection method, ideology, and tenure). The comparisons presented in this chapter suggested that minority women state supreme court justices may be more similar to minority men than white women.

The evidence presented in chapters 3 to 6 demonstrates that institutions do matter for diversity; institutions shape how citizens make inferences about bias and hold those responsible for bias accountable. In particular, the evidence from chapter 3 and 4 suggest that merit selection procedures – a topic of substantial debate in the U.S. – may undermine the process of diversification on the bench through two mechanisms: (1) prior beliefs in the fairness of institutions that, in turn, require observers to see *more* disparity before concluding the process is biased and (2) the obfuscation of blame attribution caused by having multiple actors involved in the process.

Do these findings imply that merit selection is "bad" for diversity? Not necessarily. First, it could be that merit selection systems select more women candidates despite the opportunities for diversity to be undermined. Empirical evidence about this question is mixed, and the study I conducted showed a negative but statistically insignificant relationship between mutli-actor selection and the speed with which a court selects its first woman. I focused on the presence of multiple actors, which I expected to obscure inference and accountability. However, the presence of multiple actors is not the only institutional feature of merit selection that distinguishes the institution from unitary or executive selection. It is quite possible that selection committees are more willing or better able to seek out and identify female candidates than a governor; in that case, one institutional feature that promotes diversity (candidate identification) might cancel out or even overcome the institutional features that may undermine diversity (the presence of multiple actors).

Second, it could be that merit selection procedures favor one gender over another but produce better qualified candidates. If observers care more about the improved quality than about dampened diversity, they may still perceive merit selection systems as "good" and better than the alternatives. Future research should consider how different institutional features of selection systems and their implications work together as a bundle. For example, if merit selection is associated with dampened opportunities for diversification relative to gubernatorial selection and slate selection is associated with greater opportunities for diversity than one-be-one selection, which is better for diversity: merit selection combined with slate selection or gubernatorial selection combined with one-by-one selection?

In addition, this dissertation project only addressed diversity among women briefly and in the context of the United States. Future work should acknowledge and celebrate the rich diversity among women, minority, minority women, and white men judges. Similarly, this project addressed bias in the *selection* of women judges and

does not speak to discrimination or bias in the training for, recruitment to, and experience on the bench.

Finally, the focus of this project has been selection institutions, but selection institutions are not the only institutional features of the judiciary that affect prospects for diversification. For one, I treated the qualified candidate pools as fixed in this project, but these pools vary over time and across space. The standards by which judges are deemed qualified can have a substantial effect on the gender composition of the pool. Similarly, institutions that shape the recruitment or identification of judges can be narrow (i.e., supreme court justices must have prior appeals court experience) or broad (i.e., in Niger, where justices come from different sectors of the population). To the extent that professions or social sectors are gendered, the breadth of recruitment networks can seriously affect the gender composition of the candidate pool.

As another example, procedures for *removing* judges may also have an effect on diversity. Removal institutions can determine the frequency with which judges are removed from the bench and the length of their tenures. If removal from the court is very rare so tenures are long, judges have the opportunity to shape the outcomes of more cases. This, combined with the evidence that indicates women judicial candidates are more successful on less prestigious courts (Remiche, 2015; Williams and Thames, 2008) suggests that difficulty in removing judges may decrease opportunities for women on the bench.

In sum, this dissertation project has provided a framework for thinking about how institutions affect diversification, but the project leaves many additional implications untested. Future research will both expand on and deconstruct the implications and institutions tested here.

Appendix A Appendix

Coding Rules, Aggregating Two Years for Chapter ?? Because vacancies and selections do not always occur in the same year, I aggregated two years when vacancies and selections did not occur in the same year. Specifically, courts must meet one of three requirements for two years to be aggregated. I implement these rules in order; two years can only be aggregated under rule two, for example, if the two years are not aggregated under rule one. The rules are:

1. If the number of vacancies matched the number of selections in a given year, those selections were paired to those vacancies. Of the 671 state-year-vacancy(s) observations in the unmatched data set, 89 observations were paired under this rule.

2. If there is a vacancy in year t, no new judge selected in year t, no vacancy in year $t+1$, but there is a judge selected in year $t+1$, I aggregate the two years so that the judge selected in year $t+1$ is counted as the replacement for the judge who retired in year t. Of the 671 state-year-vacancy(s) observations in the unmatched data set, 497 observations were paired under this rule.

3. In the Bratton data set, judges who take office early in a year but were selected in the previous year are listed as selected in the previous year. Therefore, if a judge retires in year t, no judge is appointed in year t, no judge is selected in year $t+1$, no judge vacated in year $t-1$, but a judge was selected in year $t-1$, I count the judge selected in year $t-1$ as the replacement to the judge who retired in year t. Of the 671 state-year-vacancy(s) observations in the unmatched data set, 16 observations were paired under this rule.

4. Finally, I aggregate two years when there is a discrepancy in the number of

vacancies and selections in one year, but not over two years. For example, if one judge retires in year t, no judge is selected in year t, one judge retires in year $t+1$ and two judges are selected in year $t+1$, I aggregate the two years so that the two retiring judges are matched with the two replacement judges. In this case, I treat the aggregated two years as one year with two vacancies. Of the 671 state-year-vacancy(s) observations in the unmatched data set, 69 observations were paired under this rule.

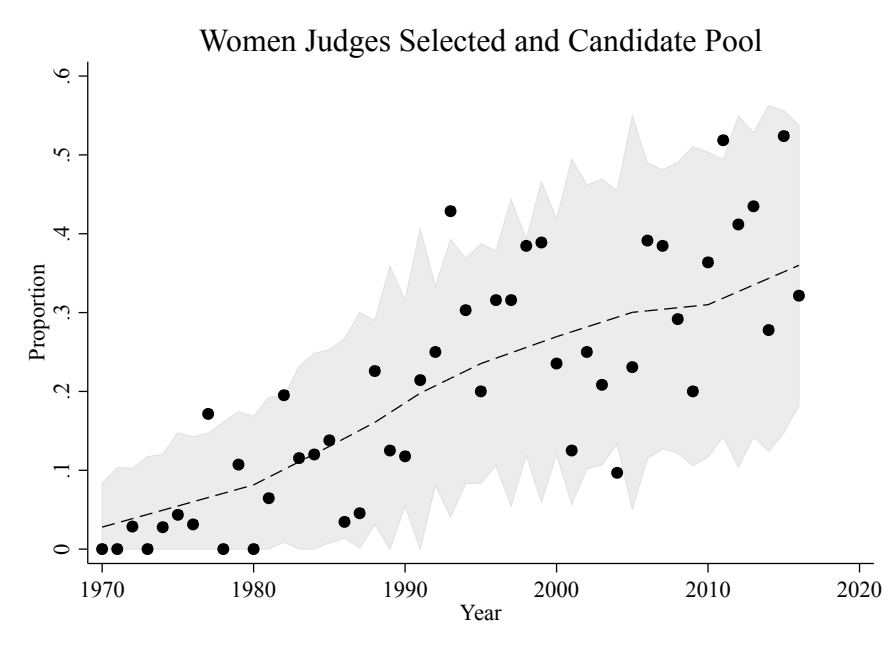
Disaggregated CMH tests To determine if the gender of vacating and replacement judges is independent across different conditions, I disaggregate the CMH test across various covariates. Disaggregating across covariates reduces the power of each test. Importantly, all observations are matched on the criteria listed inTable A.1. The CMH test is flexible to variation across strata, so the aggregated analysis reported in the main text is valid for overall patterns of judicial replacement.

Table A.1: C-M-H, disaggregated groups

	Group	# Treated units	χ^2	p
Number Vacancies	Only one Vacancy	51	6.86	0.01
	$>$ 1 Vaancy	7	.15	0.70
# Women on Court	1 Womant	18	3.35	0.07
	$>$ 1 Woman	40	2.82	0.09
	$>$ 2 Women	14	0.32	0.57
Court Size	5 Judge Court	15	.739	0.39
	7 Judge Court	41	4.78	0.03
	9 Judge Court	2	Not enough	data
Selection Method	Popular Election	9	.50	0.48
	App or Merit	48	6.94	0.01
	Legis. Election	1	Not enough	data

Candidate Pool Confidence Intervals, Actual Figure 5.3 plots the candidate pool of lawyers with confidence intervals calculated with the average number of vacancies each year. Figure 7 shows the same plot but with confidence intervals calculated with the actual number of vacancies each year. Patterns are the same.

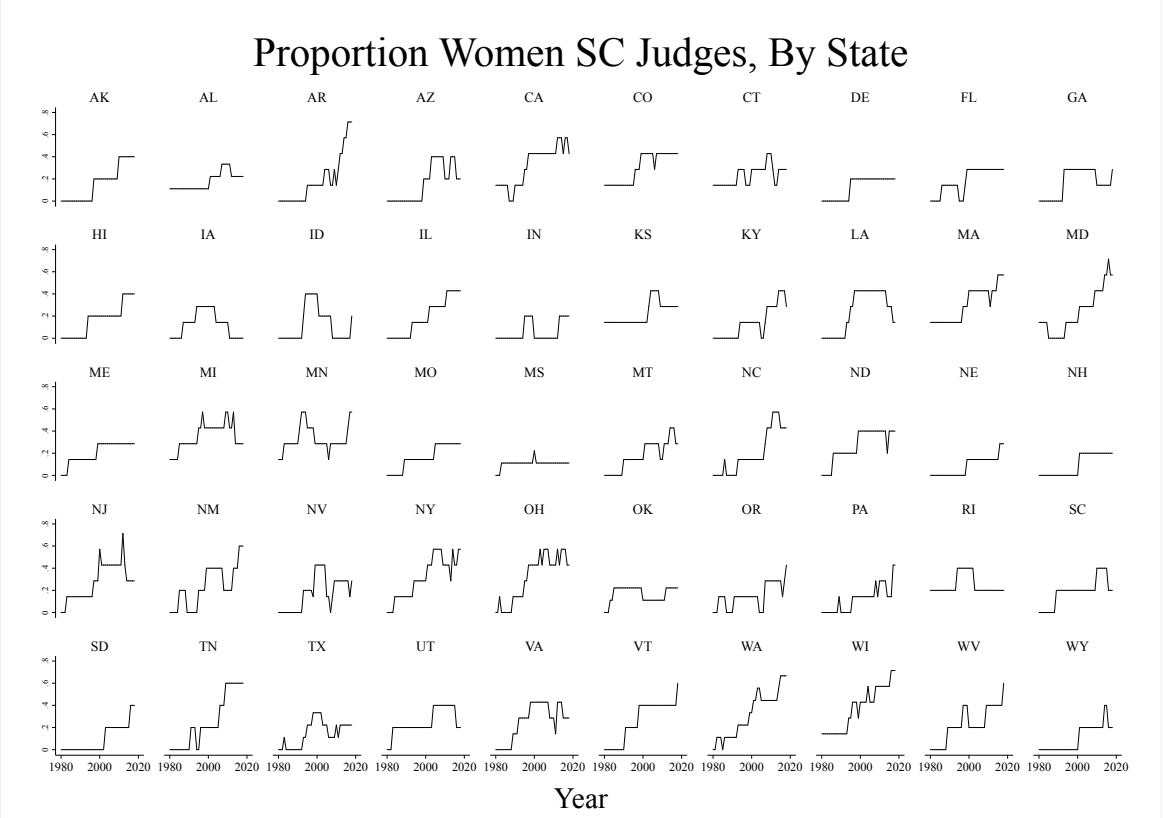
Figure A.1: Candidate Pool as Female Lawyers



The dashed line shows the proportion of lawyers who are women over time. The black dots show the proportion of judges selected each year who are women. The shaded region shows the 95% confidence interval for the expected proportion of women selected each year. In this plot the confidence intervals are calculated with the actual number of vacancies each year

State-by-state Variation In addition to obscuring over-time stickiness in diversification, aggregate patterns of selection may obscure variation across state in the selection of women judges. Figure A.2 shows how the proportion of women supreme court judges varies by state.

Figure A.2: Proportion of Women State Supreme Court Justices, Over Time and by State



The proportion of women judges over time and by state. Note: truncating the data at 1980 obscures the trajectories of gender diversity in states where women were selected to the bench earlier: Florence Allen served on the Ohio supreme court from 1923 to 1934; Anne Alpern served on the Pennsylvania Supreme Court in 1961; Lorna Lockwood served on the Arizona state supreme court from 1961 to 1975, and Elsijane Trimble Roy served on the Arkansas sate supreme court from 1975 to 1977.

Bibliography

Abraham, H.J. 1992. *Justices and Presidents: A Political history of Appointments to the Supreme Court*. New York: Oxford University Press.

Acker, Joan. 1992. "From Sex Roles to Gendered Institutions." *Contemporary Sociology* 21(5):565–569.

Alaska Judicial Council. 2006. "Dana Fabe.".
URL: *https://web.archive.org/web/20060117150532/http://www.ajc.state.ak.us/Judges/FabeDa*

Allen, David W. and Diane E. Wall. 1993. "Role Orientations and Women State Supreme Court Justices." *Judicature* 77(3).

Alozie, Nicholas O. 1988. "Black Representation on State Judiciaries." *Social Science Quarterly* 69:979–986.

Alozie, Nicholas O. 1990. "Distribution of Women and Minority Judges: The Effects of Judicial Selection Methods." *Social Science Quarterly* 71(2):315–325.

Alozie, NIcholas O. 1996. "Selection Methods and the Recruitment of Women to State Courts of Last Resort." *Social Science Quarterly (University of Texas Press)* 77(1):110 – 126.

Anderson, Christopher. 1995. *Blaming the government: Citizens and the economy in five European democracies*. New York: ME Sharpe.

Anderson, Christopher J. 2007. The Interaction of Structures and Voter Behavior. In *The Oxford Handbook of Political Behaviour*, ed. Russell Dalton and Hans Dieter Klingemann. Oxford: Oxford University Press pp. 589–609.

Andrews, Elaine. 2016. "Many Good Reasons for Gov. Walker to Appoint a Aoman to the Alaska Supreme Court." *Alaska Dispatch News* .

URL: *https://www.adn.com/commentary/article/many-good-reasons-gov-walker-appoint-woman-alaska-supreme-court/2016/05/03/*

Anzia, Sarah F and Christopher R Berry. 2011. "The Jackie (and Jill) Robinson effect: Why do congresswomen outperform congressmen?" *American Journal of Political Science* 55(3):478–493.

Arrington, Nancy B. 2018. "Gender and Judicial Replacment: The Case of US State Supreme Courts." *Journal of Law and Courts* 6(2).

Arrington, Nancy, Leeann Bass, Adam Glynn, Jeffrey Staton, Brian Delgado and Staffan Lindberg. 2018. "Gender Diversity on High Courts." *Varieties of Democracy Working Paper Series* .

URL: *https://www.v-dem.net/en/news-publications/working-papers/*

Atkeson, Lonna Rae. 2003. "Not All Cues Are Created Equal: The Conditional Impact of Female Candidates on Political Engagmenet." *The Journal of Politics* 65(4):1040–1061.

Bakalar, Linda. 2016. "Wanting One Woman on Alaska Supreme Court is not 'Reverse Sexism'." *Alaska Dispatch News* .

URL: *https://www.adn.com/commentary/article/wanting-woman-alaska-supreme-court-not-reverse-sexism/2016/05/05/*

Baldez, Lisa. 2006. "The Pros and Cons of Gender Quota Laws: What Happens When You Kick Men Out and Let Women In?" *Politics and Gender* 2(1):102–109.

Baumgartner, Jody. 2012. "Vice Presidential Selection in the Convention Era: Experience or Electoral Advantage?" *Congress and the Presidency* 39(3):297–315.

Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3):351–368.
URL: *http://www.jstor.org/stable/23260322*

Berkson, Larry C, Susan Carbon and Alan Neff. 1979. "A Study of the US Circuit Judge Nominating Commission: Findings, Conclusions and Recommendations." *Judicature* 63:104.

Blackburne-Rigsby, Anna. 2009. "Black Women Judges: The Historical Journey of Black Women to the Nation's Highest Courts." *Howard LJ* 53:645.

Boyd, Christina L., Lee Epstein and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54(2):389–411.

Bratton, K. A. and Rorie L. Spill. 2002. "Existing Diversity and Judicial Selection: The Role of the Appointment Method in Establishing Gender Diversity in State Supreme Courts." *Social Science Quarterly* 83(2):504–518.

Bratton, K.A. 2017. "State High Court Justice Database.".
URL: *https://www.lsu.edu/faculty/bratton/research.htm*

Burns, Nancy, Kay Lehman Schlozman and Sidney Verba. 2001. *The private roots of public action*. Harvard University Press.

Burrell, Barbara. 1992. "Women Candidates in Open-Seat Primaries for the U. S. House: 1968-1990." *Legislative Studies Quarterly* 17(4):493–508.

Cahill-O'Callaghan, Rachel. 2015. "Reframing the Judicial Diversity Debate: Personal Values and Tacit Diversity." *Legal Studies* 35(1):1–29.

Carbon, Susan, Pauline Houlden and Larry Berkson. 1982. "Women on the State

Bench: Their Characteristics and Attitudes about Judicial Selection." *Judicature* 65(6):294–305.

Carey, John M and Matthew Soberg Shugart. 1995. "Incentives to cultivate a personal vote: A rank ordering of electoral formulas." *Electoral Studies* 14(4):417–439.

Carroll, Kathleen M. 1983. "The Age Difference Between Men and Women Politicians." *Social Science Quarterly (University of Texas Press)* 64(2):332 – 339.

Caul, Miki. 2001. "Political parties and the adoption of candidate gender quotas: A cross–national analysis." *Journal of Politics* 63(4):1214–1229.

Chhibber, Pradeep. 2002. "Why are some women politically active? The household, public space, and political participation in India." *International Journal of Comparative Sociology* 43(3-5):409–429.

Collins, Paul M., Jr., Kenneth L. Manning and Roberta A. Carp. 2010. "Gender, Critical Mass, and Judicial Decision Making." *Law and Policy* 32(2):260–281.

Cook, Beverly B. 1984. "Women Judges: A Preface to Their History." *Golden Gate UL Rev.* 14:573.

Cook, Beverly Blair. 1978. Women Judges: The End of Tokenism. In *Women in the Courts*, ed. Winifred L. Hepperle and Laura Crites. Williamsburg: National Center for State Courts.

Crenshaw, Kimberle. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum* 1989(1):139–167.

Dahlerup, Drude. 2008. "Gender Quotas – Controversial But Trendy." *International Feminist Journal of Politics* 10(3):322–328.

Dixon, Rosalind. 2009. "Female Justices, Feminism and the Politics of Judicial Appointment: A Re-Examination." *University of Chicago, Public Law and Legal Theory Working Papers* .

DNC. 2016. "The 2016 Democratic Platrform.".
URL: *https://www.democrats.org/party-platform*

Drachman, Virginia G. 2001. *Sisters in Law: Women Lawyers in Modern American History*. Harvard University Press.

Dubeck, Paula J. 1976. "Women and access to political office: A comparison of female and male state legislators." *The Sociological Quarterly* 17(1):42–52.

Epstein, Lee, Jack Knight and Andrew D. Martin. 2003. "The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court." *California Law Review* 91:903–965.

Esterling, Kevin M and Seth S Andersen. 1999. "Diversity and the judicial merit selection process: a statistical report." *Research on Judicial Selection, The American Judicature Society* 2.

Evans, David C. 2003. "A Comparison of the Other-Directed Stigmatization Produced by Legal and Illegal Forms of Affirmative Action." *Journal of Applied Psychology* 88(1):121.

Farhang, Sean and Gregory Wawro. 2004. "Institutional Dynamics on the U.S Court of Appeals: Minority Representation Under Panel Decision Making." *The Journal of Law, Economics, and Organization* 20(2).

Feenan, Dermot. 2008. "Women Judges: Gendering Judging, Justifying Diversity." *Journal of Law and Society* 35(4):490–519.

Fox, Richard L. and Jennifer L. Lawless. 2010. "If Only They'd Ask: Gender, Recruitment, and Political Ambition." *Journal of Politics* 72(2):310 – 326.

Fox, Richard L and Zoe M Oxley. 2003. "Gender stereotyping in state executive elections: Candidate selection and success." *Journal of Politics* 65(3):833–850.

G. Bingham Powell, Guy D. Whitten. 1993. "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context." *American Journal of Political Science* 37(2):391–414.

Gay, Claudine and Katherine Tate. 1998. "Doubly bound: The impact of gender and race on the politics of black women." *Political Psychology* 19(1):169–184.

Gill, Rebecca. 2012. "A Framework for Comparative Judicial Selection Research With an Application to Gender Diversity on High Courts." *Conference Paper* .

Glick, Henry R and Craig F Emmert. 1986. "Selection Systems and Judicial Charateristics: The Recruitment State Supreme Court Judges." *Judicature* 70:228.

Glynn, Adam N and Nahomi Ichino. 2015. "Using qualitative information to improve causal inference." *American Journal of Political Science* 59(4):1055–1071.

Goelzhauser, Greg. 2011. "Diversifying State Supreme Courts." *Law and Society Review* 45(3):761–781.

Goelzhauser, Greg. 2018. "Does Merit Selection Work?: Evidence from Commission and Gubernatorial Choices." *Journal of Law and Courts* 1(6):155–187.

Goldman, Sheldon, Elliot Slotnick, Gerard Gryski and Gary Zuk. 2000. "Clinton's judges: Summing up the legacy." *Judicature* 84:228.

Greene, Linda S. 2013. "Tokens, Role Models, and Pedagogical Politics: Lamentations of an African American Female Law Professor." *Berkeley Journal of Gender, Law and Justice* 6(1):81–92.

Griffin, John D. and Michael Keane. 2006. "Descriptive Representation and the Composition of African American Turnout." *American Journal of Political Science* 50(4):998–1012.

Gruhl, John, Cassia Spohn and Susan Welch. 1981. "Women as policymakers: The case of trial judges." *American Journal of Political Science* pp. 308–322.

Guinier, Lani, Michelle Fine, Jane Balin, Ann Bartow and Deborah Lee Stachel. 1994. "Becoming Gentlemen: Women's Exxperiences at One Ivy League Law School." *University of Pennsylvania Law Review* 143(1):1–110.

Hall, Melinda Gann. 2001. "State supreme courts in American democracy: Probing the myths of judicial reform." *American Political Science Review* 95(2):315–330.

Hancock, Ange-Marie. 2007. "When multiplication doesn't equal quick addition: Examining intersectionality as a research paradigm." *Perspectives on politics* 5(1):63–79.

Hawkesworth, Mary. 2003. "Congressional enactments of race–gender: Toward a theory of raced–gendered institutions." *American Political Science Review* 97(4):529–550.

Heilman, Madeline E, Caryn J Block and Jonathan A Lucas. 1992. "Presumed incompetent? Stigmatization and affirmative action efforts." *Journal of Applied Psychology* 77(4):536.

Heilman, Madeline E, Caryn J Block and Peter Stathatos. 1997. "The affirmative action stigma of incompetence: Effects of performance information ambiguity." *Academy of Management Journal* 40(3):603–625.

Heilman, Madeline E, William S Battle, Chris E Keller and R Andrew Lee. 1998.

"Type of affirmative action policy: A determinant of reactions to sex-based preferential selection?" *Journal of Applied Psychology* 83(2):190.

Hoekstra, Valerie, Miki Caul Kittilson and Elizabeth Andrews Bond. 2014. Gender, High Courts, and Ideas about Representation in Western Europe. In *Representation: The Case of Women*, ed. Maria C. Escobar-Lemmon and Michelle M. Taylor-Robinson. New York: Oxford University Press book section 3, pp. 103–117.

Hurwitz, Mark S and Drew Noble Lanier. 2001. "Women and minorities on state and federal appellate benches, 1985 and 1999." *Judicature* 85:84.

Hurwitz, Mark S and Drew Noble Lanier. 2003. "Explaining judicial diversity: The differential ability of women and minorities to attain seats on state supreme and appellate courts." *State Politics & Policy Quarterly* 3(4):329–352.

Iacus, Stefano M, Gary King and Giuseppe Porro. 2011. "Causal inference without balance checking: Coarsened exact matching." *Political Analysis* .

Ifill, Sherrilyn. 1998. "Judging the Judges: Racial Diversity, Impartiality, and Representation on State Trial Courts." *Boston College Law Review* 39(1).

Ifill, Sherrilyn A. 2000. "Racial diversity on the bench: Beyond role models and public confidence." *Wash. & Lee L. Rev.* 57:405.

Inglehart, Robert and Pippa Norris. 2003. *Rising Tide: Gender Equality and Cultural Change around the World*. New York: Cambridge University Press.

Jones, Brian Christopher. 2012. "Drafting Proper Short Bill Titles: Do States Have the Answer Feature." *Stanford Law and Policy Review* 23:455.

Kalantry, Sital. 2012. "Women in Robes." *Americas Quarterly* Summer:82–88.

Kaufmann, Karen M. 2002. "Culture wars, secular realignment, and the gender gap in party identification." *Political Behavior* 24(3):283–307.

Kenney, Sally J. 2008. "Thinking about gender and judging." *International Journal of the Legfal Profession* 15(1-2):87–110.

Kenney, Sally J. 2012. "Choosing Judges: A Bumpy Road to Women's Equality and a Long Way to Go." *Mich. St. L. Rev.* p. 1499.

Kenney, Sally J. and Jason Windett. 2012. "Diffusion of Innovation or State Political Culture? Explaining the First Woman State SUpreme Court Justices." *Working Paper, Cited with Permission* .

Kenney, Sally Jane. 2013. *Gender and Justice: Why Women in the Judiciary Really Matter*. Routledge.

Kenworthy, Lane and Melissa Malami. 1999. "Gender Inequality in Political Representation: A Worldwide Comparative Analysis." *Social Forces* 78(1):235–269.

Kevin M. Leyden, Stephen A. Borrelli. 1995. "The Effect of State Economic Conditions on Gubernatorial Elections: Does Unified Government Make a Difference?" *Political Research Quarterly* 48(2):275–290.

King, Gary and Langche Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14:131–159.

Kittilson, Miki Caul. 2006. *Challenging Parties, Changing arliaments: Women and Elected Office in Contemporary Western Europe*. Ohio State University Press.

Krook, Mona Lena. 2006*a*. "GenderQuotas, Norms, and Politics." *Politics and Gender* 2(1):110–118.

Krook, Mona Lena. 2006*b*. "Reforming Representation: The Diffusion of Candidate Gender Quotas Worldwide." *Politics and Gender* 2:303–327.

Lawless, Jennifer L. 2004. "Women, war, and winning elections: Gender stereotyping in the post-September 11th era." *Political Research Quarterly* 57(3):479–490.

Lawless, Jennifer L and Richard L Fox. 2005. *It takes a candidate: Why women don't run for office*. Cambridge University Press.

Lawless, Jennifer L. and Richard L. Fox. 2010. *It Still Takes a Candidate: Why Women Don't Run for Office*. New York: Cambridge University Press.

Laws, Judith Long. 1975. "The Psychology of Tokenism: An Analysis." *Sex Roles* 1(1):51–67.

Lovenduski, Joni and Pippa Norris. 1993. *Gender and Party Politics*. London: Sage Publications.

Luskin, Robert C, Christopher N Bratcher, Christopher G Jordan and Tracy K Renner. 1993. "How minority judges fare in retention elections." *Judicature* 77:316.

Mansbridge, Jane. 1999. "Should Blacks Represent Blacks and Women Represent Women? A Contingent 'Yes.'." *Journal of Politics* 61(3):628.

Margolick, David. 1991. "Women's Milestone: Majority on Minnesota Court." *The New York Times* .

URL: *http://www.nytimes.com/1991/02/22/news/women-s-milestone-majority-on-minnesota-court.html?pagewanted=all*

Marsh, Michael and James Tilley. 2010. "The Attribution of Credit and Blame to Governments and Its Impact on Vote Choice." *British Journal of Political Science* 40(1):115–134.

Martin, Elaine. 1981. "Women on the Federal Bench: A Comparative Profile." *Judicature* 65:306–313.

Martin, Elaine. 1987. "Gender and Judicial Selection: A Comparison of the Reagan and Carter Administration." *Judicature* 71:136.

Martin, Elaine. 1991. *Judicial Gender and Judicial Choices.* Center for the American Woman and Politics: The Impact of Women in Public Office Rutgers.

Martin, Elaine and Barry Pyle. 1999. "Gender, race, and partisanship on the Michigan Supreme Court." *Albany Law Review* 63:1205.

Martin, Patricia Yancy, John R. Reynolds and Shelley Keith. 2002. "Gender Bias and Feminist Consciousness among Judges and Attorneys: A Standpoint Theory Analysis." *Signs* 27(3):665–701.

Matland, Richard E. 1993. "Institutional Variables Affecting Female Representation in National Legislatures: The Case of Norway." *The Journal of Politics* 55(3):737–755.

URL: *http://www.jstor.org/stable/2131998*

Matland, Richard E. and Deborah Dwight Brown. 1992. "District Magnitude's Effect on Female Representation in U.S. State Legislatures." *Legislative Studies Quarterly* 17(4):469–492.

Matland, Richard E and Donley T Studlar. 1996. "The contagion of women candidates in single-member district and proportional representation electoral systems: Canada and Norway." *The journal of politics* 58(03):707–733.

McCall, Madhavi and Michael A McCall. 2007. "How far does the gender gap extend?: decision making on state Supreme Courts in Fourth Amendment Cases, 1980–2000." *The Social Science Journal* 44(1):67–82.

McDonald, John H. 2014. *Handbook of Biological Statistics.* 3rd ed. Baltimore: Sparky House Publishing.

Milyo, Jeffrey and Samantha Schosberg. 2000. "Gender bias and selection bias in House elections." *Public Choice* 105(1-2):41–59.

Moncrief, Gary F and Joel A Thompson. 1992. *Changing patterns in state legislative careers*. University of Michigan Press.

Mullarky, Mary J. 2012. "A Brief History of Women of the Colorado Supreme Court." *The Colorado Lawyer* 41(27).

Murray, Rainbow. 2010. *Cracking the Highest Glass Ceiling: A Global Comparison of Women's Campaigns for Executive Office*. Women and Minorities in Politics Santa Barbara: Praeger.

Nelson, Michael. 1988. "Choosing the vice president." *PS: Political Science & Politics* 21(4):858–868.

New Hampshire Judicial Branch. 2017. "Supreme Court - Chief Justice Linda Stewart Dalianis.".
URL: *https://www.courts.state.nh.us/supreme/meetdalian.htm*

Norris, Pippa. 1987. *Politics and Sexual Equality, the comparative Postion of Women in Western Democracies*. Boulder: Lynne Rienner Publishers, Inc.

Norris, Pippa. 1993. *Conclusions: Comparing Legislative Recruitment*. London: Sage.

Paxton, Pamela. 1997. "Women in National Legislatures: A Cross-National Analysis." *Social Science Research* 26(4):442–464.

Pearson, Kathryn and Eric McGhee. 2013. "What It Takes to Win: Questioning 'Gender Neutral' Outcomes in U.S. House Elections." *Politics and Gender* 9(04):439–462.

Phillips, Anne. 1995. *The Politics of Presence*. Oxford Political Theory Oxford: Clarendon Press.

Pitkin, Hanna Fenichel. 1967. *The Concept of Representation*. New York: Atherton Press, Inc.

Redfield, Sarah E. 2009. "The Educational Pipeline to Law School: Too Broken and Too Narrow to Provide Diversity." *Pierce Law Review* 8.

Reingold, Beth. 2018. "Who Leads What? Legislative Leadership and Intersections of Gender, Race, and Ethnicity in the American States." *Working Paper, Emory University* .

Reingold, Beth and Jessica Harrell. 2010. "The Impact of Descriptive Representation on Women's Political Engagement: Does Party Matter?" *Political Research Quarterly* 63(2):280–294.

Remiche, Adelaide. 2015. "When Judging is Power." *Journal of Law and Courts* Spring:95–114.

Richard L. Fox, Jennifer L. Lawless. 2004. "Entering the Arena? Gender and the Decision to Run for Office." *American Journal of Political Science* 48(2):264–280.

Rikleen, Luaren Stiller. 2015. "Women Lawyers Continue to Lag Behind Male Colleagues: Report on the Ninth Annual NAWL National Survey on Retention and Promotion of Women in Law Firms." *National Association of Women Lawyers* .

Robert C. Lowry, James E. Alt, Karen E. Ferree. 1998. "Fiscal Policy Outcomes and Electoral Accountability in American States." *The American Political Science Review* 92(4):759–774.

Rule, Wilma. 1987. "Electoral Systems, Contextual Factors and Women's Opportunities for Election to Parliament in 23 Democracies." *Western Political Quarterly* 40:477–498.

Salmond, Rob. 2006. "Proportional Representation and Female Parliamentarians." *Legislative Studies Quarterly* 31(2):175–204.

Scherer, Nancy and Brett Curry. 2010. "Does Descriptive Race Representation Enhance Institutional Legitimacy? The Case of the U.S. Courts." *The Journal of Politics* 72(1):90–104.

Schmidt, Gregory D. 2009. "The election of women in list PR systems: Testing the conventional wisdom." *Electoral Studies* 28(2):190 – 203.

Schultz, Ulrike and Gisela Shaw. 2013. Introduction: Gender and Judging: Overiview and Synthesis. In *Gender and Judging*, ed. Ulrike Schultz and Gisela Shaw. Bloomsbury.

Segal, Jennifer A. 2000. "Representative decision making on the federal bench: Clinton's district court appointees." *Political Research Quarterly* 53(1):137–150.

Siaroff, Alan. 2000. "Women's Representation in Legislatures and Cabinets in Industrial Democracies." *International Political Science Review* 21(2):197–215.

Slotnick, Elliot E. 1984. "Judicial Selection Systems and Nomination Outcomes Does the Process Make a Difference?" *American Politics Quarterly* 12(2):225–240.

Solberg, R. L. Spill and K. A. Bratton. 2005. "Diversifying the Federal Bench: Presidential Patterns." *The Justice System Journal* 26(2):119–133.

Songer, Donald R and Kelley A Crews-Meyer. 2000. "Does judge gender matter? Decision making in state supreme courts." *Social Science Quarterly* pp. 750–762.

Sotomayor, Sonia. 2002. "A Latina Judge's Voice." *Berkeley La Raza LJ* 87.

Summers, Russel J. 1991. "The Influence of Affirmative Action on Perceptions of a Beneficiary's Qualifications1." *Journal of Applied Social Psychology* 21(15):1265–1276.

Toobin, Jeffrey. 2014. "The Obama Brief: The President Considers his Judicial Legacy." *The New Yorker* .
URL: *http://www.newyorker.com/magazine/2014/10/27/obama-brief*

Valdini, Melody E. and Christopher Shortell. 2016. "Women's Representation in the Highest Courts: A Comparative Analysis of the Appointment of Female Justices." *Political Research Quarterly* 69(4):865–876.

Verba, Sidney, Nancy Burns and Kay Lehman Schlozman. 1997. "Knowing and caring about politics: Gender and political engagement." *The Journal of Politics* 59(04):1051–1072.

Walker, Peter. 2015. "Don't rush gender equality in UK judiciary, says supreme court judge.".
URL: *http://www.theguardian.com/law/2015/sep/22/gender-equality-warning-uk-legal-profession-supreme-court-judge-jonathan-sumption*

Walker, Thomas G and Deborah J Barrow. 1985. "The diversification of the federal bench: Policy and process ramifications." *The Journal of Politics* 47(2):596–617.

Weaver, Kent R. 1986. "The Politics of Blame." *Journal of Public Policy* 6(4).

Welch, Susan, Michael Combs and John Gruhl. 1988. "Do black judges make a difference?" *American Journal of Political Science* pp. 126–136.

Westergren, Sarah. 2003. "Gender Effects in the Courts of Appeals Revisited: The Data Since 1994." *Geo. LJ* 92:689.

Williams, Margaret S. 2006. "In a Different Path: The Process of Becoming a Judge for Women and Men." *Judicature* 90(3):104–113.

Williams, Margaret S. 2008. "Ambition, Gender, and the Judiciary." *Political Research Quarterly* 61(1):68–78.

Williams, Margaret S. and Frank C. Thames. 2008. "Women's Representation on High Courts in Advanced Industrialized Countries." *Politics and Gender* 4(03).

Yoon, Albert. 2003. "Love's Labor's Lost-Judicial Tenure among Federal Court Judges: 1945-2000." *Cal. L. Rev.* 91:1029.

Yoon, Mi Yung. 2004. "Explaining Women's Legislative Representation in Sub-Saharan Africa." *Legislative Studies Quarterly* 29(3):447–468.