

Group Selection and Opportunities for Gender Diversity in the Judiciary

Nancy B. Arrington*

Abstract

In the legislative context, there is a persistent empirical relationship between party-list proportional electoral systems and higher levels of gender diversity in office. I apply the logic of party-list selection to the judicial context and argue that the selection of judges as pairs or in groups may facilitate the process of gender diversification in office by making it easier – or at least more likely – for observers to notice gender disparity in judicial selection. Evidence from a survey experiment demonstrates that observers are more critical of gender homogeneity when judges are selected as a group rather than one-by-one. Indeed, respondents were more likely to even note gender disparity when judges were selected as a group rather than on a one-by-one basis. These micro-foundations demonstrate that the logic of party-list PR in the legislative context may also apply to the judicial context, suggesting that one way to improve prospects of judicial diversity could be re-structuring the timing of judicial turnover.

Introduction

The idea that political offices should – at least to some extent – mirror the descriptive characteristics of the populations they serve has increasingly gained popular traction across political contexts, including the judiciary (see Grossman et al. (2016); Ifill (1998); Scherer and Curry (2010), for example). Descriptively representative institutions, in turn, are associated with increased political engagement and participation (Junn 1997; Gay 2002; Atkeson 2003; Campbell and Wolbrecht 2006; Reingold and Harrell 2010), increased perceptions of legitimacy (Scherer and Curry 2010), increased acceptance of policy decisions (Arnesen and Peters 2018), and improved policy outcomes for historically excluded communities (Saint-Germain 1989; Phillips 1995; O'Regan 2000; Reingold 2000; Celis 2006; Reingold 2008; Wangnerud 2009). Even so, there is still substantial cross-national

*Assistant Professor at Cal Poly State University, naarring@calpoly.edu. Prepared for the European Conference on Politics and Gender 2019. An earlier version of this paper was presented at the 2018 Annual Meeting of the Midwest Political Science Association. Gracious thanks to Beth Reingold and Jeffrey Staton for constructive comments.

variation in the extent to which historical outsiders such as women and ethnic minorities (or women who are also ethnic minorities) are included in political offices.¹

In the legislative context,² scholars 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) (Kenworthy and Malami 1999; Salmond 2006; Yoon 2004; McAllister and Studlar 2002; Norris 2000; Rule 1987; Norris 1985).³ There are at least two mechanisms that might explain the relationship between party-list PR and greater gender diversity: information and balanced lists.⁴ 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 must appeal to a majority 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 (Norris 2000).

These two mechanisms – information and balanced lists – help explain the consistent empirical relationship between PR systems and a greater presence of women in office. However, discussions and tests of these mechanisms have been limited to the legislative context. In this article, I apply these two mechanisms of the party-list PR system logic to a feature of judicial selection. I argue that the selection of judges in pairs or as a group – that is, as a slate – rather than one-by-one ought to increase levels of gender diversity in the judiciary through the mechanisms of increased information and balanced lists. Evidence from a survey experiment confirms that respondents are more critical of gender homogeneity when judges are selected simultaneously, which suggests that

¹This project studies the selection of women as a group, but it is important to emphasize that women are a diverse group and hold myriad intersecting identities (Crenshaw 1989).

²Most of this work focuses on industrialized democracies. Matland (1998) finds that effect of PR electoral systems on women’s representation is limited to industrialized democracies and does not extend to less economically developed democracies.

³These two systems are not the only electoral institutions that might shape the success of women candidates in the legislative setting. The single transferable vote (STV), for example, is also studied in relation to legislative gender diversity (Hirczy 1995). I focus on PR and SMD because these two institutions – I argue – parallel features of judicial selection institutions.

⁴Other mechanisms include party competition/contagion (Matland and Studlar 1996), centralized candidate nomination (Norris 1993), decreased emphasis on incumbency and greater turnover (Darcy and Choike 1986), and ease of implementing party quotas (Caul 2001; Krook 2006).

the logic of gender diversity under party-list PR systems does apply to non-legislative contexts such as the judiciary. In the next sections I outline the logic of gender diversity under party-list PR systems, apply the logic to the judiciary, and then describe the survey experiment used to test the two mechanisms. I conclude with recommendations for future research.

The Role of party-lists PR in Legislative Diversity

There is substantial institutional and social variation that affects the ease and opportunity for women to run for and win seats in office even among countries with party-list PR electoral systems (Schmidt 2009). Gender quotas (Tripp and Kang 2008), district magnitude, position requirements, party candidate selection requirements (Vandeleene 2014), vote thresholds, economic development, and political culture (Inglehart and Norris 2003), for example, can all shape the extent to which PR systems facilitate the inclusion of women politicians. Despite this variation, there are at least two mechanisms through which PR systems facilitate the election of women that might also apply to judicial selection.

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). Simply put: voters more easily gain information about the gender composition of candidates running for office under party-list PR than SMD systems.

Second, in an SMD election, voters vote for just one candidate, so that one candidate – alone – must earn sufficient votes to win, which means 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.⁵ Having a list of candidates allows the traits of individual candidates to be viewed as complementary or substitutable to traits of others on the same list, which means that a more diverse set of candidates may be more viable under a PR system. Indeed, if traits of those in a group are viewed as complementary, the presence of a list may *encourage* diversity by encouraging the selection of different types of candidates to produce a “balanced” list (Norris 2000; Matland and Studlar 1996).

These two features of party-list PR systems – 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. For example, we see features of the balanced list mechanism in US Presidential elections when Presidential candidates attempt to broaden their popular appeal by selecting Vice Presidential candidates with different characteristics (Baumgartner 2012; Nelson 1988) or when Prime Ministers intentionally select gender-balanced and/or racially-diverse cabinets. Canadian Prime Minister Justin Trudeau received international attention for his choice to select a gender balanced cabinet in 2015 Chartrand (2016), a choice that is increasingly common. As table 1 shows, 2019 saw the greatest number of countries with at least 50% women ministers, which speaks to the increased salience of gender parity in political office. Interestingly, more countries have gender balanced cabinets than gender balanced legislatures, perhaps because a Prime Minister setting a cabinet can overcome the coordination problem that plagues multiple parties and voters when selecting candidates to the parliament.

⁵This 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.

Table 1: Gender Balanced Cabinets and Legislatures. 2008-2019

Year	# 50%+ Cabinets	Country, %	# 50%+ Legislatures	Country, %
2008	3	Finland (57.9%) Norway (55.6%) Grenada (50.0%)	0	
2010	4	Finland (63.2%) Cape Verde (53.3%) Spain (52.9%) Norway (52.6%)	1	Rwanda (56.3%)
2012	4	Norway (52.6%) Sweden (52.2%) Finland (50.0%) Iceland (50.0%)	2	Rwanda (56.3%) Andorra (50.9%)
2014	3	Nicaragua (57.1%) Sweden (56.5%) Finland (50.0%)	2	Rwanda (63.8%) Andorra (50.0%)
2015	5	Finland (62.5%) Cabo Verde (52.9%) Sweden (52.2%) France (50.0%) Leichtenstein (50.0%)	3	Rwanda (63.8%) Bolivia (53.1%) Andorra (50.0%)
2017	6	Bulgaria (52.9%) France (52.9%) Nicaragua (52.9%) Sweden (52.2%) Canada (51.7%) Slovenia (50.0%)	2	Rwanda (61.3 %) Bolivia (53.1%)
2019	9	Spain (64.7%) Nicaragua (55.6%) Sweden (54.5%) Albania (53.3%) Colombia (52.9%) Costa Rica (51.9%) Rwanda (51.9%) Canada (50%) France (50%)	3	Rwanda (61.3%) Cuba (53.2%) Bolivia (53.1%)

The number of countries with cabinets and legislatures with at least 50% of seats held by women for the years 2008-2019. Data from the International Parliamentary Union <https://www.ipu.org/resources/publications/infographics>.

Group Selection and Diversity in the Judiciary

A growing extant literature addresses gender diversity in the judiciary. The size and prestige of a court (Williams and Thames 2008), norm diffusion across space and institution (Hoekstra, Kittilson and Bond 2014; Goelzhauser 2011; Williams and Thames 2008); the existing or previous gender diversity on a court (Arrington 2018; Bratton and Spill 2002); the legal culture (Remiche 2015), the media (Escobar-Lemmon et al. 2016; Kenney 2008), quotas (Hoekstra 2010), 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 shaping gender diversity, debate still exists. Some find the concentration of accountability on a unitary selector such as a president (or, at the subnational level, a governor) leads to greater diversity (Williams and Thames 2008; Bratton and Spill 2002; Carbon, Houlden and Berkson 1982). Reddick, Nelson and Caufield (2009) find that the “merit” selection process in the United States in which a commission generates a short list from which the executive selects a nominee is associated with the selection of more women judges. Similarly, Goelzhauser (2011) finds that the presence of a merit commission is associated with the earlier selection woman state supreme courts justices in the United States. 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 (Hoekstra, Kittilson and Bond 2014; Hurwitz and Lanier 2003; Alozie 1990; 1988).⁶

Despite the many explanations for variation in gender diversity in the judiciary, the role of group or slate selection for judicial gender diversity has not been addressed. Cross-nationally, there is substantial variation in the selection procedures for judges, including variation in the timing and turnover of judges. Some countries have judicial selection institutions that result in (or require)

⁶See Frederick and Streb (2008) for a discussion of women judges running for election. They find that women are no less likely to win judicial elections for intermediate state appellate courts than men, and may in fact receive a slight boost.

Table 2: Constitutional Examples of Slate Selection

Country	Year	Text from the Constitution Describing Slate Selection	% Women, 2011
Chile	1986	“The [seven] members of the Court shall serve eight years, be partially replaced every four years, and must not be removed.”	10%
Niger	2009	(Translated from French) 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.	0%
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.”	22%
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.”	17%

Examples from countries’ constitutions that indicate slate selection. Examples come from the Varieties of Democracy Judiciary data collected by a team at Emory using constitutions organized through the Comparative Constitutions Project. Data on the gender composition of the court is from Turquet et al. (2011).

the selection of multiple judges at once. Table 2 shows examples of selection procedures for peak court judges that involve slate selection.⁷

For example, in contrast to the US 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. In nine years, observers in Albania will observe the selection of nine judges. In comparison, only four judges have been selected to the US Supreme Court in the nine year span of 2009-2018.⁸ Applying the first mechanism 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

⁷Peak courts refer to the constitutional court with the power of judicial review or the highest ordinary court in countries in which there is no constitutional court.

⁸Sotomayor, 2009; Kagan, 2010; Gorsuch, 2017; Kavanaugh, 2018

pressure – that is, “vertical accountability” (Goetz 2003; O’Donnell 1998) – on the elites tasked with selecting judges.

Indeed, Valdini and Shortell (2016) find that elites will select women to the bench when doing so is electorally beneficial. Specifically, elites who are subject to institutional “exposure” (p. 865) – 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 (p. 865). Valdini and Shortell’s work indicates, then, that observers of judicial selection can affect prospects for gender diversity on the bench indirectly *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 to encourage diverse selections.

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.⁹ This mechanism can apply to the judicial context as well. Using the US Supreme Court as an example, nominees to the 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.¹⁰ 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

⁹Or, in part. If voters have expectations about how many candidates will be elected from a particular party list, they should assess the traits of the top X candidates that they expect will actually make it into office.

¹⁰The 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.

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 homogenous 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.*

Survey Experimental Evidence

Observing a relationship between slate selection and increased gender diversity in the judiciary presents several empirical problems. First, both the information and balanced lists mechanisms depend on a population that cares about descriptive gender diversity. If observers are indifferent to the exclusion of women, the ability to more easily notice exclusion is unimportant. Second, institutions are endogenous. Not only are institutions chosen with favored outcomes in mind, but one's beliefs about prospects for success under one institutional regime will determine her willingness to participate by, for example, accumulating the necessary qualifications for office. Finally, selecting judges in pairs or groups is just one of many institutional features that may affect prospects for diversification. Given the complications to observational analysis, this study aims to identify the micro-foundations of the mechanisms through which slate selection would lead to increased gender diversity.

To test how slate-selection affects observer's responses to gender disparity through the mechanism of information and balanced lists, 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). 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 three minute survey. For

the analysis presented here, survey respondents must have indicated that they lived in the United States or have taken the survey from within the US.¹¹ In addition, respondents who did not pass a simple attention check were removed from the analysis. Table 3 shows summary characteristics for the MTurk respondents used in this analysis, and figure 1 plots descriptive characteristics by treatment group to show sample balance.

Table 3: 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 4 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 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

¹¹Data on longitude/latitude and IP addresses were used to verify location of respondents

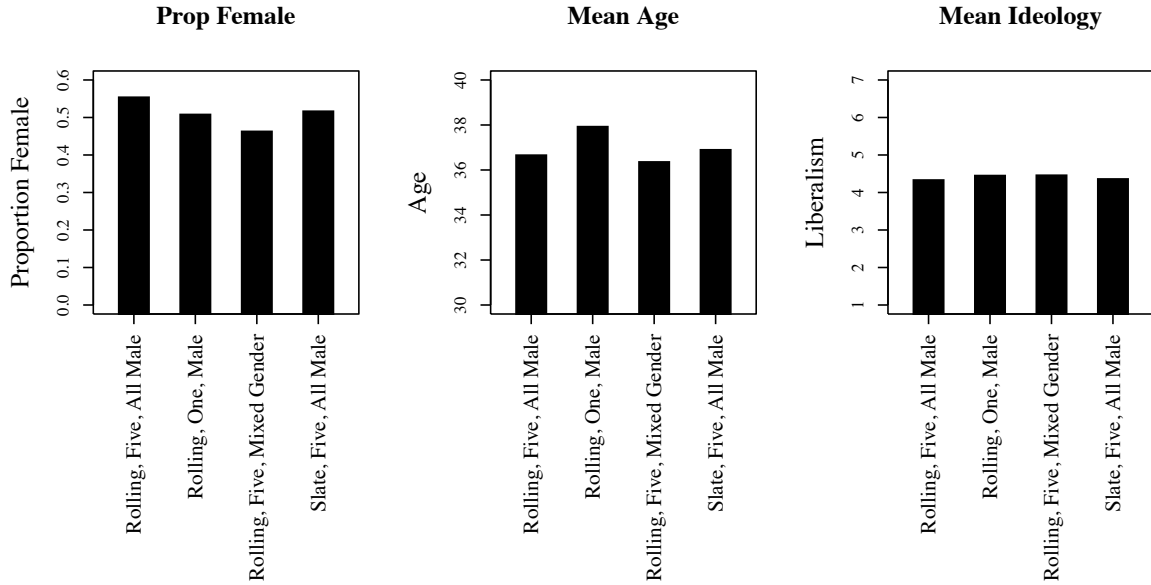


Figure 1: *Descriptive characteristics by treatment group*

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.

Table 4

		Selection Process:	
Information:	One male judge	Rolling, Group A	Slate, Group D
	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 group D to test the balanced lists hypothesis. Group B is compared to group C to test whether observers perceive a gender balanced court as more fair.

Figure 2: Survey Instrument: One-by-One and Slate

There are many ways in which judges are selected to courts in different states and different countries. We are studying courts that have multiple judges, such as some appeals courts, supreme courts, or constitutional courts.

You will be asked to give your opinion on how judges are selected to one court in particular.

-On this court, there are five judges. Each judge serves for five years, and one judge retires and is replaced each year. A selection committee decides which judges to appoint to each vacancy.

-To be eligible to serve as a judge on this court, someone must be 35 years old, be a citizen of the country, and they must have a law degree. In addition, most judges on this court attended elite law schools and have extensive judicial experience.

-Historically, judges on this court have been predominately male, although there have been female members of the court. Moreover, legal experts predict that about half of the qualified candidates for this post are female.

Currently there are four serving judges and one vacancy. Characteristics of the serving judges are listed below:

Judge 1: Male, 45 years old, #1 ranked law school, 10 years judicial experience.
 Judge 2: Male, 61 years old, #2 ranked law school, 32 years judicial experience.
 Judge 3: Male, 47 years old, #1 ranked law school, 15 years judicial experience.
 Judge 4: Male, 52 years old, #3 ranked law school, 21 years judicial experience.

There is one vacancy on the court. The selection committee has chosen the following candidate to join the court:

New Judge: Male, 54 years old #2 ranked law school, 29 years judicial experience.

Given the information provided above, do you think the selection process is likely fair or unfair?

	Definitely Unfair	Probably Unfair	Neither Fair nor Unfair	Probably Fair	Definitely Fair
Fairness of Selection Process	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Why do you think the process is either fair or unfair?

There are many ways in which judges are selected to courts in different states and different countries. We are studying courts that have multiple judges, such as some appeals courts, supreme courts, or constitutional courts.

You will be asked to give your opinion on how judges are selected to one court in particular.

-On this court, there are five judges. All judges serves for five years, and every five years a new court is chosen. A selection committee decides which judges to appoint to each vacancy.

-To be eligible to serve as a judge on this court, someone must be 35 years old, be a citizen of the country, and they must have a law degree. In addition, most judges on this court attended elite law schools and have extensive judicial experience.

-Historically, judges on this court have been predominately male, although there have been female members of the court. Moreover, legal experts predict that about half of the qualified candidates for this post are female.

This year a new slate of judges was chosen. Characteristics of the serving judges are listed below:

Judge 1: Male, 45 years old, #1 ranked law school, 10 years judicial experience.
 Judge 2: Male, 61 years old, #2 ranked law school, 32 years judicial experience.
 Judge 3: Male, 47 years old, #1 ranked law school, 15 years judicial experience.
 Judge 4: Male, 52 years old, #3 ranked law school, 21 years judicial experience.
 Judge 5: Male, 54 years old #2 ranked law school, 29 years judicial experience.

Given the information provided above, do you think the selection process is likely fair or unfair?

	Definitely Unfair	Probably Unfair	Neither Fair nor Unfair	Probably Fair	Definitely Fair
Fairness of Selection Process	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Why do you think the process is either fair or unfair?

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.

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 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 (Group C) 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 (Group C) and the all-male, one-by-one group (Group B). Table 4 describes the treatment characteristics of the three groups.

Do respondents care about the gender composition of the court? As a first test to check whether respondents noticed gender parity or disparity and whether disparity affected 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 sitting judges. If respondents do care about the gender diversity of courts, then more respondents should indicate that the process appears unfair for the all-male court.

Panel (a) of figure 3 shows the proportion of respondents in each group who indicated that the process was either “definitely unfair” or “probably unfair.” The black bar show the responses for the mixed-gender court; the grey bars show the responses for the all-male court. 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. Panel (b) plots the point estimates and 95% confidence interval for a differences in proportions test.

The comparison of these two groups demonstrates 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 the current gender composition of the court shapes how respondents interpret the fairness of the selection of a new judge to the bench; homogeneity undermines perceptions of the fairness of the institution. To the extent that perceptions of fairness reflect perceptions of legitimacy, this finding confirms both theoretical expectations (Kenney 2013; Mansbridge 1999) and existing empirical scholarship (Scherer and Curry 2010) on the relationship between descriptive representation and legitimacy. Interestingly, the idea that homogeneity on the court leads respondents to perceive of the selection process as unfair suggests that descriptive representation can affect perceptions of procedural fairness or,

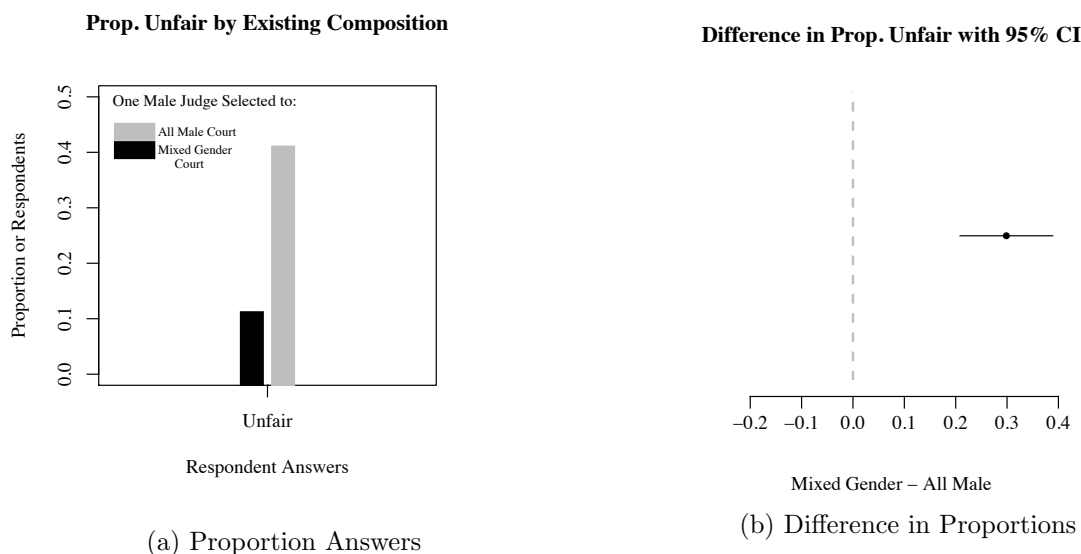
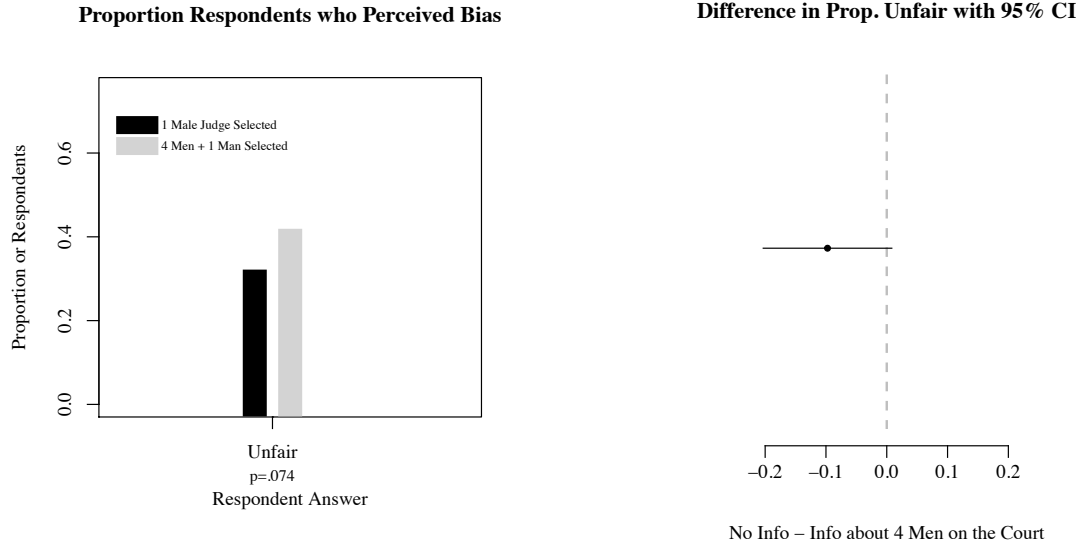


Figure 3: Panel (a) shows the proportion of respondents who indicated that the process appeared “definitely unfair” or “probably unfair.” The grey bars show respondents who saw the selection of a male judge to an all-male court. The black bars show answers from respondents who saw a man selected to a mixed-gender court. Panel (b) shows the point estimate and 95% confidence interval for a differences in proportions test. The p -value for this test is $p < .001$

rather, perceptions of the legitimacy of processes of formal representation. Scholarship addressing the link between formal and descriptive representation typically address formal representation as something that is fixed and affects levels of descriptive representation (Schwindt-Bayer and Mishler 2005). This finding that homogeneity affects perceptions of fairness of the selection procedures flips the relationship and addresses how descriptive representation affects perceptions of the processes of formal representation.

Does the Amount of Information Affect Perceptions of Unfairness? The information hypothesis (H1) 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 from 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’ abilities 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

Figure 4: 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 is the probability of observing a difference at least as great as the difference observed in the study. The 95% confidence interval includes 0, or no difference between the two proportions.

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 4 shows the proportion of 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), suggesting that increased information about gender homogeneity on the bench does lead observers to be more critical of the selection process. However, a difference in proportions test does not reflect traditional standards for a two-tailed test of statistical significance ($p = .074$). In sum, the evidence suggests a weak effect of the information mechanism of group selection.

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.

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, one-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 is what respondents were told about the timing of the selection of the judges. Some were told that the judges one judge was selected every year (one-by-one), other respondents were told that five judges were selected every five years (slate). 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. Panel (a) of figure 5 shows the proportion of respondents who indicated that the process appeared either “definitely unfair” or “probably unfair” across the two treatment groups. The black bar shows the responses of group D (five male judges, slate) and the grey bar shows responses for group B (five male judges, one-by-one). 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 ($p < .001$). This demonstrates that respondents are more critical of gender homogeneity when judges are selected as a group rather than one-by-one, which is consistent with the balanced list mechanism.

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 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). This suggests that the information mechanism is not the sole explanation for the theorized relationship between group selection and gender diversity. Instead,

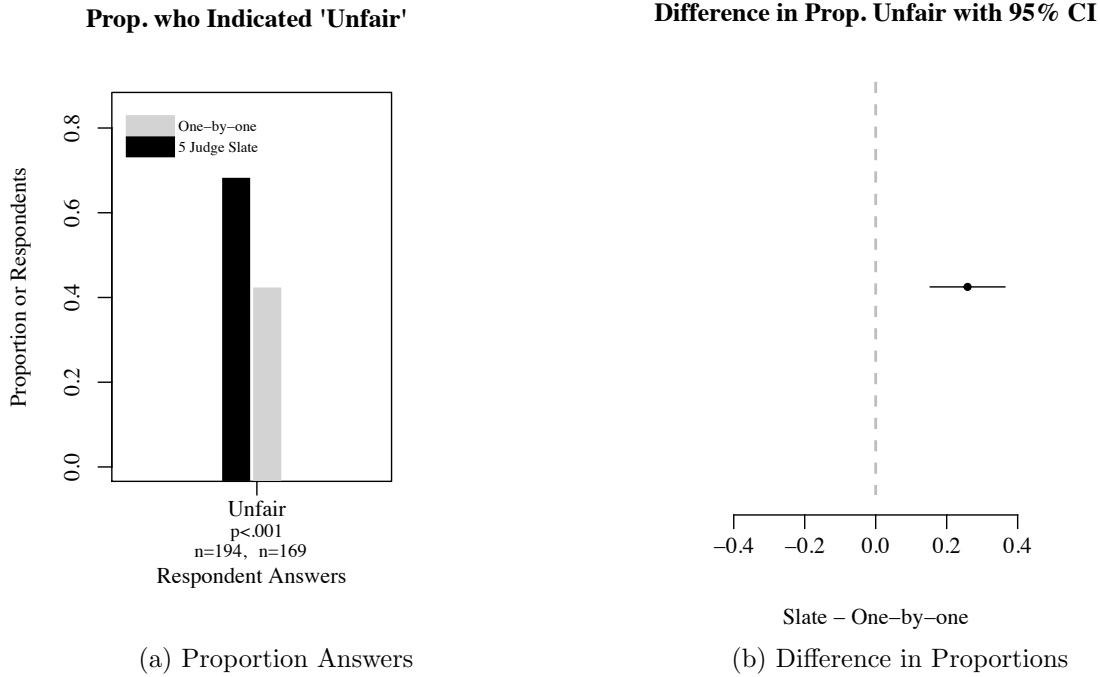


Figure 5: Panel (a) shows the proportion of respondents who indicated the process seemed “definitely unfair” or “probably unfair.” The grey bar shows respondents who saw the selection of one man to a court with four men. The black bar shows the respondents who saw a slate of five male judges selected. Panel (b) shows the point estimate and 95% confidence interval for the difference in proportions test. The p -value associated with this test is $p < .001$.

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 (i.e., 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.

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

Prop. 'Unfair' across Information and Institutions

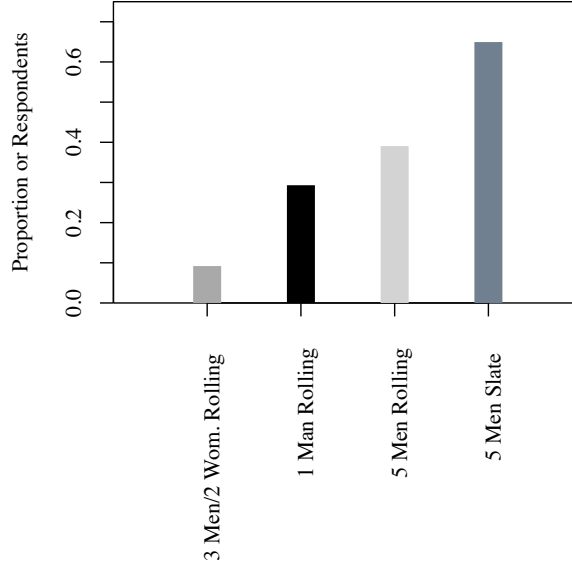


Figure 6: *The proportion of respondents who indicated “definitely unfair” or “probably unfair” across all treatment groups.*

Table 5: 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%

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 5 summarises the number of respondents who noted gender in their qualitative responses and table 6 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 was similar across institution type, as were the justifications for their responses. It appears, then, that the mechanism linking slate selection to perceptions of gender bias among observers is through

Table 6: Examples of Qualitative Responses

Institu- tion	Response	Noted Gender?	Explanation for Response
One-by-one	Definitely Unfair	Yes	“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	Yes	“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	No	“[H]e seems qualified but without knowing the details of the other candidates [I can’t] say how fair or unfair it is.”
One-by-one	Definitely Fair	No	“His credentials and age are comparable to the other judges.”
Slate	Definitely Unfair	Yes	“Because if over half of qualified candidates are female then WHY aren’t they being chosen.”
Slate	Neither Fair nor Unfair	Yes	“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	Yes	“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.

encouraging observers to *notice* gender imbalances, given the same amount of information. In other words, group selection increases respondent sensitivity to gender imbalances in office, which in turn allows more people to be critical of homogeneity. If those tasked with selecting officials are responsive to this increased criticism, slate selection procedures should lead to increased gender diversity in office.

Conclusions

The logic explaining why list-PR systems are associated with greater gender diversity in the legislative context is not – I argue – limited to legislatures. In this project, I applied the logic of party-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 assess candidates as a “balanced” group rather than individually. Evidence from a survey experiment shows that respondents

are more critical of gender disparity when judges are selected as a slate rather than one-by-one, even when respondents across the two groups saw the same information. While both mechanisms affect respondents' perceptions of fairness, the balanced-list mechanism has a greater effect than the information mechanism. 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 one-by-one group.

These findings teach us a few lessons. First, the respondents in this survey experiment care about judicial diversity; respondents were more likely to indicate that selection processes were fair when the outcome of the selection process included men and women. This finding confirms expectations that descriptive representation is important for perceptions of institutional legitimacy. Second, increased information about gender homogeneity on the bench lead respondents to be more likely to perceive the selection process to be unfair ($p=.074$), which indicates that consciousness raising campaigns tasked with educating voters about the dearth of women in politics could be useful. Third, observers are more critical of gender disparity when officials are selected as a group rather than one-by-one. Under conditions in which judicial selectors are responsive to popular criticism, group selection should promote gender diversity by increasing the costs to the selectors of excluding women. Finally, selecting officials as a group rather than one-by-one makes it more likely that observers will *notice* gender homogeneity. Even though two groups of respondents were given information about the *same* five judges, how respondents perceived the representativeness of those judges depended on the timing of judges' selection. In other words, observers apply different standards to groups selected as a slate, and in this study, observers applied stricter standards for gender diversity to officials selected as a group.

Taken together, this study indicates that one way to increase accountability for gender diversity in office is to alter the timing of selection so that multiple officials are selected simultaneously. Future research will test how slate selection manifests in different observational contexts, such as the selection of judges,¹² boards, commissions, or other groups of officials.

¹²see Appendix 1 for a preliminary observational test of how institutional changes to slate selection are weakly associated with faster gender diversification on national peak courts.

References

- 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.
- Arnesen, Sveinung and Yvette Peters. 2018. "The legitimacy of representation: how descriptive, formal, and responsiveness representation affect the acceptability of political decisions." *Comparative Political Studies* 51(7):868–899.
- Arrington, Nancy. 2018. "Gender and Judicial Replacement: The Case of US State Supreme Courts." *Journal of Law and Courts* 2(6).
- 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 Engagemenet." *The Journal of Politics* 65(4):1040–1061.
- Baumgartner, Jody. 2012. "Vice Presidential Selection in the Convention Era: Experience or Electoral Advantage?" *Congress and the Presidency* 39:297–315.
- 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.
- 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.

- Campbell, David E. and Christina Wolbrecht. 2006. "See Jane Run: Women Politicians as Role Models for Adolescents." *The Journal of Politics* 68(2):233–247.
- 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.
- Caul, Miki. 2001. "Political parties and the adoption of candidate gender quotas: A cross-national analysis." *Journal of Politics* 63(4):1214–1229.
- Celis, Karen. 2006. "Substantive Representation of Women: The Representation of Women's Interests and the Impact of Descriptive Representation in the Belgian Parliament (1900-1979)." *Journal of Women, Politics and Policy* 28(2):85–114.
- Chartrand, Fred. 2016. "Trudeau's 'Because it's 2015' Retort Draws International Attention."
URL: <https://www.theglobeandmail.com/news/politics/trudeaus-because-its-2015-retort-draws-international-cheers/article27119856/>
- 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* pp. 57–80.
- Darcy, Robert and James R Choike. 1986. "A formal analysis of legislative turnover: Women candidates and legislative representation." *American Journal of Political Science* pp. 237–255.
- Escobar-Lemmon, Maria C, Valerie Hoekstra, Alice Kang and Miki Caul Kittilson. 2016. "Just the Facts? Media Coverage of Female and Male High Court Appointees in Five Democracies." *Politics and Gender* 12(2):254–274.

- Frederick, Brian and Matthew J. Streb. 2008. "Women Running for Judge: The Impact of Sex on Candidate Success in State Intermediate Appellate Court Elections." *Social Science Quarterly* 89(4):937–954.
- Gay, Claudine. 2002. "Spirals of Trust? The Effect of Descriptive Representation on the Relationship between Citizens and Their Government." *American Journal of Political Science* 46(4):717–732.
- 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.
- Goetz, Anne Marie. 2003. Gender and Accountability. In *Women Making Constitutions*, ed. Alexandra Dobrowolsky and Vivien Hart. Springer pp. 52–67.
- Grossman, Guy, Oren Gazal-Ayal, Samuel D Pimentel and Jeremy M Weinstein. 2016. "Descriptive representation and judicial outcomes in multiethnic societies." *American Journal of Political Science* 60(1):44–69.
- Hirczy, Wolfgang. 1995. "STV and the Representaiton of Women." *PS: Political Science and Politics* 28(4):711–713.
- Hoekstra, Valerie. 2010. "Increasing the gender diversity of High Courts: A comparative view." *Politics and Gender* 6(3):474–484.
- 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. 2003. “Explaining Judicial Diversity: The Differential Ability of Women and Minorities to Attain Seats on State Supreme and Appellate Courts.” *State Politics and Policy Quarterly* 3(4):329–352.
- Ifill, Sherrilyn. 1998. “Judging the Judges: Racial Diversity, Impartiality, and Representation on State Trial Courts.” *Boston College Law Review* 39(1).
- Inglehart, Robert and Pippa Norris. 2003. *Rising Tide: Gender Equality and Cultural Change around the World*. New York: Cambridge University Press.
- Junn, Jane. 1997. Assimilating or Coloring Participation? Gender, Race, and Democratic Political Participation. In *Women Transforming Politics: An Alternative Reader*, ed. Cathy J. Cohen, Kathleen B. Jones and Joan C. Tronto. New York: New York University Press book section 22, pp. 387–397.
- Kenney, Sally J. 2008. “Gender on the agenda: how the paucity of women judges became an issue.” *The Journal of Politics* 70(3):717–735.
- 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.
- Krook, Mona Lena. 2006. “Reforming Representation: The Diffusion of Candidate Gender Quotas Worldwide.” *Politics and Gender* 2:303–327.
- Mansbridge, Jane. 1999. “Should Blacks Represent Blacks and Women Represent Women? A Contingent ‘Yes.’” *Journal of Politics* 61(3):628.

- 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. 1998. "Women's Representation in National Legislatures: Developed and Developing Countries." *Legislative Studies Quarterly* 23(1):109–125.
- URL:** <http://www.jstor.org/stable/440217>
- Matland, Richard E and Donley T Studlar. 1996. "The Contagion of Women Candidates in Single-Member District and Proportional Representation Electoral Systems." *The Journal of Politics* 58(03):707–733.
- McAllister, Ian and Donley T Studlar. 2002. "Electoral systems and women's representation: a long-term perspective." *Representation* 39(1):3–14.
- Nelson, Michael. 1988. "Choosing the vice president." *PS: Political Science & Politics* 21(4):858–868.
- Norris, Pippa. 1985. "Women's Legislative Participation in Western Europe." *West European Politics* 8(4):90–101.
- Norris, Pippa. 1993. Conclusions: Comparing Legislative Recruitment. In *Gender and Party Politics*, ed. Joni Lovenduski and Pippa Norris. London: Sage.
- Norris, Pippa. 2000. Women's representation and electoral systems. In *The International Encyclopedia of Elections*, ed. Richard Rose. Washington DC: CQ Press.
- O'Donnell, Guillermo A. 1998. "Horizontal accountability in new democracies." *Journal of democracy* 9(3):112–126.
- O'Regan, Valerie R. 2000. *Gender Matters: Female Policymakers' Influence in Industrialized Nations*. Greenwood Publishing Group.

- Phillips, Anne. 1995. *The Politics of Presence*. Oxford Political Theory Oxford: Clarendon Press.
- Reddick, Malia, Michael J. Nelson and Rachel Raine Caufield. 2009. "Racail and Gender Diversity on State Courts: An AJS Study." *Judges' Journal* 48(3):28–32.
- Reingold, Beth. 2000. *Representing Women: Sex, Gender, and Legislative Behavior in Arizona and California*. Chapel Hill: The University of North Carolina Press.
- Reingold, Beth. 2008. Women as Office Holders: Linking Descriptive and Substantive Representation. In *Political Women and American Democracy*, ed. Christina Wolbrecht, Karen Beckwith and Lisa Baldez. New York: Cambridge University Press book section 9.
- 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.
- Rule, Wilma. 1987. "Electoral Systems, Contextual Factors and Women's Opportunities for Election to Parliament in 23 Democracies." *Western Political Quarterly* 40:477–498.
- Saint-Germain. 1989. "Does Their Difference Make a Differnece? The Impact of Women on Public Policy in the Arizon Legislature." *Social Science Quarterly* 70(4):956–968.
- 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: Overview and Synthesis. In *Gender and Judging*, ed. Ulrike Schultz and Gisela Shaw. Bloomsbury.
- Schwindt-Bayer, Leslie A. and William Mishler. 2005. "An Integrated Model of Women's Representation." *Journal of Politics* 67(2):407–428.
- Tripp, Aili Mari and Alice Kang. 2008. "The global impact of quotas: On the fast track to increased female legislative representation." *Comparative Political Studies* 41(3):338–361.
- Turquet, Laura, Papa Seck, Ginette Azcona, Roshni Menon, Caitlin Boyce, Nicole Pierron and Emma Harbour. 2011. "Progress of the World's Women 2011-2012: In Pursuit of Justice." *The United Nations Entity for Gender Equality and the Empowerment of Women* .
- 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.
- Vandeleene, Audrey. 2014. "Gender quotas and 'women-friendly' candidate selection: Evidence from Belgium." *Representation* 50(3):337–349.
- Wangnerud, Lena. 2009. "Women in Parliaments: Descriptive and Substantive Representation." *Annual Review of Political Science* 12:51–69.
- Williams, Margaret S. and Frank C. Thames. 2008. "Women's Representation on High Courts in Advanced Industrialized Countries." *Politics and Gender* 4(03).
- Yoon, Mi Yung. 2004. "Explaining Women's Legislative Representation in Sub-Saharan Africa." *Legislative Studies Quarterly* 29(3):447–468.

Appendix 1: 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 Appendix, 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.¹³ Importantly, though, this data is limited to constitutional rules. It is possible, therefore, that some states have statutes dictating group selection. If there is no evidence of group selection in the constitution, a process is not counted as having group selection.

Because institutions outcomes are endogenous to institutions – 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 – I focus on constitutional changes to selection institutions. Assuming constitutions are not changed with the *goal* of increasing gender diversity on the bench, I can treat institutional changes as exogenous.¹⁴

¹³Most 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.

¹⁴There 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.

After identifying the countries and years in which a judicial selection system changes to slate selection, I match those treated units to control units. Then, I compare the time from treatment until the first woman is 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 judge 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.¹⁵ 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 as a slate. In addition, I match on decade and the percent women in the lower house.¹⁶ Matching on decade controls for time and matching

¹⁵A team of researchers at Emory, lead by Leeann Bass, attempted to collect the total number of women on the bench over time, but substantial missing data threaten the accuracy of the data. However, researchers were able to identify the year of the first woman more accurately.

¹⁶Using 'optmatch' in R.

Table 7: Treatment and Control Countries

Treated Country	Year First Woman (t)	Control Country	Year First Woman (c)	Diff.*
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. The year of the first woman peak court justice is missing for Cape Verde, the Democratic Republic of the Congo, and Morocco.

**The “Diff.” column refers to the difference between the treated and control countries in how long it took from the time of the institutional change until the selection of the first woman judge. That is, $[YearFirstWoman_{Treated} - YearChange_{Treated}] - [YearFirstWoman_{Control} - YearChange_{Control}]$. Negative numbers in the “Diff.” column indicate that it took fewer years from the time of the institutional change to the selection of the first woman in treated countries.*

on percent women in the legislature controls for pressure for the selection of women to important political posts (Hoekstra, Kittilson and Bond 2014). Table 7 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 woman to the peak court.¹⁷ 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.¹⁸ 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 signed 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 7 shows the permutation distribution and the observed signed 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 the null hypothesis of no relationship between slate selection and gender diversity on the bench.

¹⁷I drop countries from the sample that had women prior to the institutional change.

¹⁸That 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)

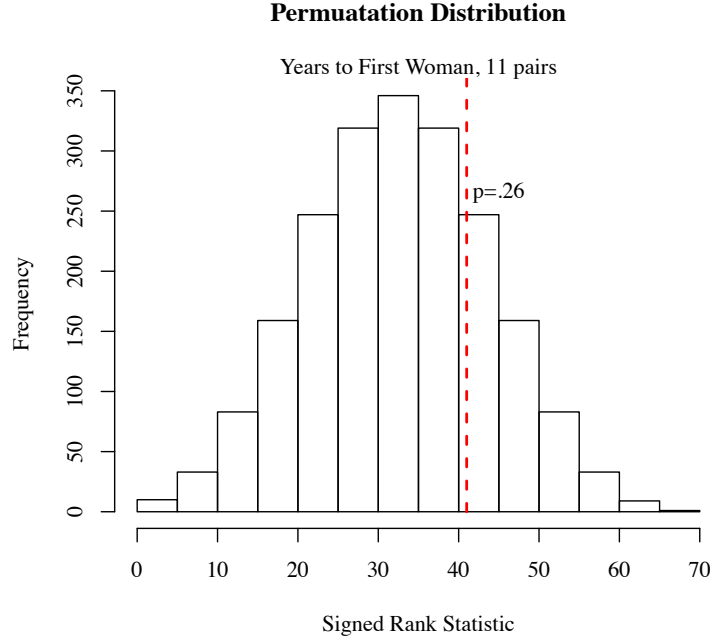


Figure 7: *This graph shows the permutation distribution of all possible assignments of treatment to matched pairs. The red dashed line shows the observed sign rank statistic. 11 pairs.*

It is worth noting, however, that this study is conservative. Institutions should only affect gender diversity under particular circumstances. For example, 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 ought to address the role of slate selection procedures in contexts where there is moderate social pressure for gender diversity.