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 on courts by making it easier – or at least more likely – for observers to notice and be critical of gender disparity in judicial selection. Evidence from a survey experiment fielded in the United States demonstrates that observers are more likely to notice and are more critical of gender homogeneity when judges are selected as a group rather than one-by-one. 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.

Keywords: Descriptive Representation, Gender Diversity, Judicial Diversity, Slate Selection, Balanced-Lists, Information

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), improved policy outcomes for historically excluded communities (Saint-Germain 1989; Phillips 1995; O'Regan 2000; Reingold 2000; Celis 2006; Reingold 2008; Wangnerud 2009), and – even absent observed differences between historically excluded populations and the status quo – the normalization of historically excluded groups in positions of power (Kenney 2013). Despite the benefits of diversity, there is still substantial cross-national 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, only three countries had lower or single houses with at least 50% women in 2019 (Rwanda, Cuba, and Bolivia), and only 15 countries had at least 40% women (Inter-ParliamentaryUnion 2019). In the judiciary, Rwanda (50%) and Serbia (67%) were the only two countries with high courts with least 50% women justices in 2011, according to UN data (Turquet et al. 2011), and in a study on peak courts in democracies, the only two democracies with at least 50% women in 2012 were Latvia (57.1%) and Slovenia (55.6%) (Valdini and Shortell 2016).

In this project, I apply lessons from existing literature on gender diversity in the legislative context to the judicial context. More specifically, 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 – by showing

---

[1]This 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).

[2]Most of this research 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.

[3]These 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 electoral institutions because these two institutions – I argue – parallel features of judicial selection institutions.

[4]Other 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). These mechanisms can – and probably do – manifest in the judicial context as well, but empirical evi-

voters full lists of many candidates for each party rather than just one or a few candidates per party – provide more information to voters about the gender composition of candidates than ballots in SMD elections. By showing voters more candidates, party-list systems allow observers to better identify gender disparity relative to SMD 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 and provides incentives for gatekeepers to choose balanced lists (Norris 2000; Caul 1999).

These two mechanisms – information and balanced lists – help explain the consistent empirical relationship between electoral systems and variation in the 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 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 likely to notice and are more critical of gender homogeneity when judges are selected simultaneously as a slate rather than one-by-one, which suggests that the logic of gender diversity under party-list PR systems does apply to non-legislative contexts such as the judiciary. More generally, the survey experiment confirms expectations that descriptive representation is important for perceptions of institutional legitimacy: respondents were more likely to indicate that the selection process was fair when the outcome of selection included both male and female judges. 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 that one way

dence is thus far limited. Term limits or mandatory retirement ages would increase turnover, which should increase the amount of information observers receive about judicial selections. Quotas in the Judicial context are rare, but do exist (see the Appendix for examples and Malleson (2006) for a discussion of affirmative action policies and judicial selection). Likewise, centralized candidate nomination procedures may serve to concentrate accountability for homogeneous selections, and existing evidence demonstrates that executives are rewarded for diverse selections (or punished for its absence) (Valdini and Shortell 2016; Badas and Stauffer 2019).

to facilitate diversification in the judiciary is to alter the timing of judicial turnover so that judges are selected in pairs or as a group rather than on a rolling, one-by-one basis.

The Role of party-list 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). Political culture (Inglehart and Norris 2003; Norris and Inglehart 2001), district magnitude (Engstrom 1987), gender quotas (Tripp and Kang 2008), ballot structure (Rule and Shugart 1995; Stegmaier, Tosun and Vlachová 2014), and intra-party policies such as list position for women (Compaoré 2005) or recruitment and selection procedures (Vandeleene 2014; Kunovich and Paxton 2005; Caul 1999), for example, can all shape the extent to which PR systems facilitate the inclusion of women politicians. Despite this within PR-system variation, there are at least two mechanisms through which PR systems.

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, voters have more information about the gender balance of candidates for office. It is much easier for voters to make inferences about the possibility of gender bias in selection when confronted with several all-male party lists than when confronted with two or three male candidates in an SMD system. As Matland (1993) explains, 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 voters" (p.738). Simply put: voters gain more 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* must earn sufficient votes to win. SMD elections, therefore, require candidates to have broad appeal. To the extent that women candidates are less likely to appeal to a wide audience, SMD elections can "exacerbate voter reluctance to select women candidates" (Goetz 2003, p. 55).

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] Instead, the traits of individual candidates can be viewed as complementary or substitutable to traits of others on the same list, which alleviates the potential electoral risk of possessing or reflecting traits that diverge from the historically "typical" candidate. Indeed, if traits of those in a group are viewed as complementary, the presence of a list may *promote* diversity by encouraging the selection of different types of candidates to produce a "balanced" list (Valdini 2012; 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 other 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, parties, or coalitions (Bäck, Debus and Müller 2016) intentionally select gender-balanced and/or racially-diverse cabinets.[6] Canadian Prime Minister Justin Trudeau, for example, received international attention for his choice to select a gender-balanced cabinet in 2015 (Chartrand 2016), a choice that is increasingly common. Indeed, in 2019, nine countries had cabinets with at least 50% women, an all-time high,[7] which speaks to the increased salience of the presence of women in political office.[8]

The judiciary poses an interesting case with which to assess the extension of party-list PR logic. Cross-nationally, judicial selection procedures reflect robust institutional variation that allows for

---

[5]This depends, however, on whether lists are open, closed, or flexible. When lists are open or flexible and candidates face intra-party competition, there is an incentive to develop personal reputations to garner votes for election and/or future list position (Carey and Shugart 1995; Crisp et al. 2013; André et al. 2017).

[6]See Matland and Brown (1992) for an application of the balanced-list logic to multi-member districts for US State Legislatures.

[7]Data from the International Parliamentary Union (https://www.ipu.org/resources/publications/infographics) reports that Spain (64.7%), Nicaragua (55.6%), Sweden (54.4%), Albania (53.3%), Colombia (52.9%), Costa Rica (51.9%), Rwanda (51/9%), Canada (50%), and France (50%) had 50% or more of their cabinet seats held by women.

[8]Interestingly, more countries have gender balanced cabinets (9) than gender balanced legislatures (3), perhaps because a Prime Minister, party, or coalition tasked with setting a cabinet can overcome the coordination problem that plagues multiple parties and voters when selecting candidates to the parliament.

the assessment of myriad institutional effects on diversity. In addition, judiciaries are becoming increasingly powerful globally (Hirschl 2008; Tate and Vallinder 1995). Factors that shape the composition of courts, therefore, have the capacity to increasingly shape a country's political context. In the next sections, I outline how the logic of party-list PR systems and the selection of women applies to the judicial context as well.

Group Selection and Diversity in the Judiciary

A growing literature addresses the conditions under which courts diversify: 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), the creation of new courts (Escobar-Lemmon et al. 2019), institutional change (Arrington et al. 2018), and legal systems (Dawuni and Kang 2015; Schultz and Shaw 2013) all affect prospects for gender diversity on the bench.

Among those who assess how selection institutions shape gender diversity in the judiciary, some find that the concentration of accountability on a unitary selector such as a president (or, at the sub-national level, a governor) leads to greater diversity (Williams and Thames 2008; Bratton and Spill 2002; Carbon, Houlden and Berkson 1982). Recent work in the United States on commission-assisted or "merit" selection procedures in which a commission generates a short list from which the executive selects a judges suggests – contrary to expectation (i.e. Krivosha 1987) – that this selection procedure does not facilitate the selection of women (Reddick, Nelson and Caufield 2010; Goelzhauser 2011; 2018; 2019; Arrington 2019).[9] Others still find no or little relationship between

---

[9]Goelzhauser (2011) does find that merit selection in the US context is associated with the earlier selection of a state's first Black judge.

selection institutions and diversity (Hoekstra, Kittilson and Bond 2014; Hurwitz and Lanier 2003; Alozie 1990; 1988).[10]

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. Anecdotal evidence suggests that selecting judges as groups should affect the selection of women. For example, Judge Mabel Van Camp – the first woman selected to the Ontario Supreme Court[11]– recounts, "there was pressure" to select women to public office; "they were going to add five [judges] and it was difficult to add five and not have one of them a woman at that time" (Mossman 2013, p.61). Van Camp identifies the increased pressure to select a woman in a context in which several judges were to be selected at the same time. Importantly, the selection of judges as a group a slate is not limited to special instances; 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) the selection of multiple judges at once. Table 1 shows examples of selection procedures for peak court[12] judges that involve slate/group selection.

Slate selection ought to lead to increased gender diversity on the bench through the two mechanisms addressed above: information and balanced lists. First, selecting candidates as a group rather than one-by-one provides observers greater information about the gender composition of selected judges and, subsequently, the judiciary more generally. Take, for example, an observer living in a country with a highly male dominated judiciary. Assume that this observer has little background information about the judiciary and is not particularly aware of the gender disparity on the bench. If she reads a newspaper article about a newly selected supreme court justice who happens to be a man, that one piece of information is unlikely to lead our observer to question gender disparity in the judiciary. In contrast, if that same observer lives in a country where judges are selected as a slate and reads an article about five new supreme court justices – all of whom appear in the

---

[10]See 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.

[11]Subsequent legislation – including the Courts of Justice Amendment Act, 1989 – re-named and restructured the Ontario Supreme Court. See JudgesLibrary (2016) for more detail.

[12]Peak 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.

Table 1: 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 two 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. Of the constitutions and constitutional amendments recorded by this team, 34 countries have or had peak court selection processes that either implicitly or explicitly indicated group selection. Data on the gender composition of the court in 2011 is from Turquet et al. (2011).

photo as male or have names common for men – our observer may notice the absence of a woman and become more skeptical about the extent to which women are present in the judiciary. In other words, selecting judges as a slate can provide more information to observers in a way that makes the absence of women more stark, which, in turn, increases the probability that observers will perceive gender disparity to be problematic. To the extent that elites tasked with selecting judges will subsequently face pressure to rectify the disparity, slate selection should result in more diverse judiciaries.

Second, slate selection allows observers to assess new judges as a balanced list. When a judge is selected independently, he/she must appeal to the minimum winning coalition among the selectorate. Outsiders such as women or minorities may be more likely to alienate some portion of that coalition, which makes their selection less likely. In contrast, when judges are selected as a group, traits of one candidate can be treated as substitutes or compliments to traits of other candidates,[13]

---

[13]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

Figure 1: Path Diagram
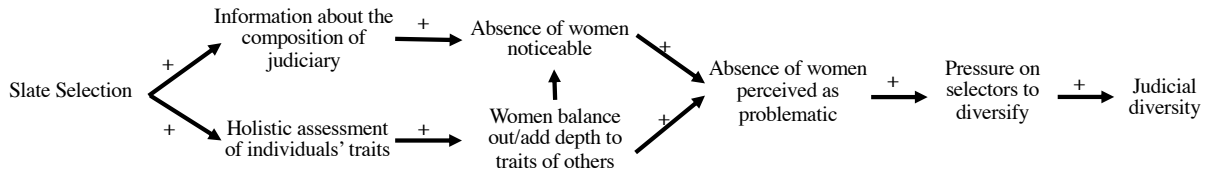
Two mechanisms – information and balanced lists – link slate selection to increased diversity in the judiciary.

which minimizes attention to unusual or "outsider" traits of individual judges. Someone skeptical of the ability of a woman judge may be less skeptical if a woman is one of multiple judges selected at the same time. Indeed, the presence of historically excluded traits may be advantageous to the reception of a slate of candidates as a way of adding depth and balance to a group. By adding a woman judge to a slate, selectors can build coalitions of those who want diverse traits represented in the judiciary. If many observers do prefer increased diversity in the judiciary, the absence of a woman among a slate of candidates may result in increased public pressure on selectors to diversify the bench. By allowing the holistic assessment of a group of candidates, slate selection not only de-emphasizes the potential electoral risk of selecting historical outsiders, but promotes the selection of a diverse group of candidates to produce a balanced list with wide appeal.

The path diagram in figure 1 summarises the two mechanisms through which slate selection ought to increase gender diversity in the judiciary. Both the information and balanced list mechanisms are theorised to increase the extend to which the absence of women is perceived as problematic. In turn, facilitating the extent to which gender disparity is perceived of as problematic should result in increased pressure on judicial selectors to chose qualified women candidates.

---

other cases, the slate as a whole may be confirmed or rejected. 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 a short time frame should not treated as independent.

It is important to note that both the information and balanced lists mechanisms depend on observers noticing gender disparity and pressuring selectors to choose judges who increase the diversity of the judiciary.[14] Valdini and Shortell (2016) find that elites are more likely to 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 and, presumably, are sheltered from sanctions for the continued selection of homogeneous courts. Valdini and Shortell's work confirms that observers of judicial selection can affect prospects for gender diversity on the bench indirectly, even when judges are selected by elites. In other words, what observers think about selections to the court can affect who is selected to the court, even when observers are not directly involved in selection.

Survey Experimental Evidence

Observing a relationship between slate selection and increased gender diversity in the judiciary presents several empirical challenges. 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 will not affect prospects for diversification. Second, institutions are endogenous. Not only are institutions chosen with favored outcomes in mind, but one's beliefs about the potential for selection under a given institutional regime will determine her willingness to pursue office 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; isolating the *independent* effect of slate selection in the observational setting is, therefore, unlikely. Given the complications to observational analysis, this study instead focuses on identifying and testing the micro-foundations of the mechanisms through which slate selection can lead to increased gender diversity with evidence from a

---

[14]That is, these mechanisms depend on the willingness and ability of the public to exert "vertical accountability" (Goetz 2003; O'Donnell 1998).

survey experiment fielded in the summer of 2017. The survey experiment is designed to isolate the information and balanced lists mechanisms and test how each shapes whether and the extent to which observers are critical of gender homogeneity. More precisely, I test the following two hypotheses:

Hypothesis 1 (Information): *Holding the timing of selection fixed,* observers will be more critical of gender disparity when they have more information about homogeneous judges selected to the bench.

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

Survey Instrument

The survey experiment 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 (See, for example, Casey et al. 2017; McCredie and Morey 2018).Even so, research suggests that MTurk samples are more representative than in-person convenience samples and student samples (Berinsky, Huber and Lenz 2012) and can be useful for political science research (Huff and Tingley 2015). Respondents were randomly assigned into treatment and control groups and were paid $0.35 for taking an approximately three minute survey. For the analysis presented here (n=722), survey respondents must have indicated that they lived in the United States or have taken the survey from within the US.[15] In addition, respondents who did not pass a simple attention check were removed from the analysis. Table 2 shows summary characteristics for the MTurk respondents used in this analysis, and figure 2 plots descriptive characteristics by treatment group to show sample balance.[16]

---

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

[16]Although respondents were randomly assigned into treatment and control groups, it is possible for confounding characteristics to be over-represented in one group relative to another. Simple tests of differences in characteristics across treatment groups are all statistically insignificant, indicating that the variance across groups is insubstantial: Gender, p=.41 (χ^2 test), Age, p=.63 (F statistic, OLS); Ideology, p=.83 (χ^2 test), and Education, p=.57 (χ^2 test).

Table 2: 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 (n=722). Two respondents did not report their sex/gender and one respondent did not report their education level.

In this experiment, I manipulate two variables to create four treatment conditions. I manipulate the amount of information observers receive to test how information affects perceptions of bias, and I manipulate the selection institutions to test how slate selection might encourage observers to assess judicial candidates as a "balanced" slate. All groups were given general information about a hypothetical, five-judge court. Respondents were informed that a selection committee chooses judges,[17] and were told some basic information about the judges: their age, gender, the prestige of their law school, and their years of judicial experience. All respondents were primed in the same way with respect to gender; respondents were informed that "legal experts predict that about half of the qualified candidates for this post are female." While this prime increases the probability that respondents notice gender (thereby reducing the external validity of the study), it ensures that all respondents have the same background information with which to make inferences about fairness or bias (which serves to increase internal validity).[18]

---

[17]While executive selection or – in the context of the United States, commission assisted selection or elections – are more common judicial selection procedures, committee selection was used here to avoid priming respondents to project their approval of an executive onto the outcome of the process. As such, this choice limits external validity but promotes internal validity.

[18]The goal of this experiment is to test how selection institutions shape respondents' inferences about fairness or bias based on observed gender (dis)parity on the bench. If a respondent observes the selection of all men to a court, however, she could plausibly conclude either (1) that the process overlooks qualified women or (2) that there must not be any qualified women candidates. The observer's inference in the latter case is about the pool of candidates and not the fairness of the selection process. Without fixing one of these two beliefs – either about the composition of the

Figure 2: Descriptive Characteristics by Treatment Group
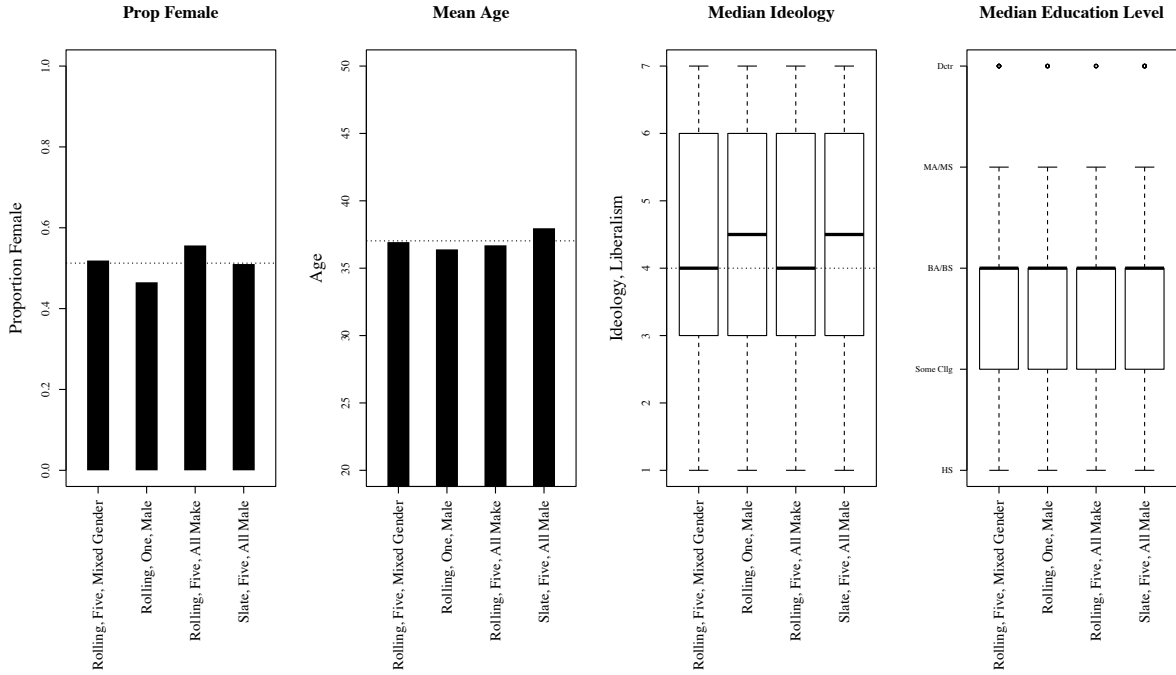
Dotted horizontal lines indicate the overall means/median. These four characteristics are reported because, plausibly, they may be associated with variation in the dependent variable. Women, young people, those who are more liberal, and those who have spent more time in formal institutions of higher education may all be more likely to notice gender disparity and infer that gender disparity is the outcome of institutional processes. Differences in these characteristics across treatment groups reflect random variation and are statistically insignificant: gender, p=.41 (χ^2 test), age, p=.63 (F statistic, OLS); ideology, p=.83 (χ^2 test), and education, p=.57 (χ^2 test)

Table 3 outlines the four experimental groups. Groups A, B, and C were told judges were selected to five year terms, and each year one judge retired and one new 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

---

candidate pool or about the fairness of selection – some respondents may update beliefs and make inferences about the candidate pool and some respondents may update beliefs and make inferences about the fairness of selection institutions. I fix beliefs about the composition of the candidate pool by telling all respondents that the candidate pool is 50/50 male/female so that information about the selection of men and women can be used to make inferences about the fairness of the selection institution. While this prime reduces the external validity of the experiment – observers may not and probably do not know the precise gender composition of the candidate pool – it increases the internal validity of the experiment by fixing a potential confounding variable to isolate how observers learn about the fairness of the selection process.

Table 3: 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 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.

and one male judge newly selected to the bench. Group C also observed the selection of one new male judge, but this judge was selected to a bench with two male and two female judges currently serving. 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 (five men, rolling); the only difference between groups B and D is the timing of selection.

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 responses from group A (one male judge, one-by-one selection) to responses from 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 responses from 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 – Group C includes both male and female judges serving on a court with rolling, one-by-one selection. If respondents do not notice or care about gender diversity, responses about the extent to which the process appears fair should be the same across the mixed

Figure 3: 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	O	O	O	O	O

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	O	O	O	O	O

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

>>

Two of the four 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, one-by-one group (Group C) and the all-male, one-by-one group (Group B). The three comparisons are detailed below.

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 a court with four currently serving men) to group C (those who observed one man selected to a court with two current female and two current male judges). The only difference between these two treatment groups is the gender of two sitting judges. If respondents do not care about gender diversity on the bench, responses about fairness or bias should be the same across both groups as. 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.

The left panel of figure 4 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 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. The right panel of 4 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 diversity in office and legitimacy.
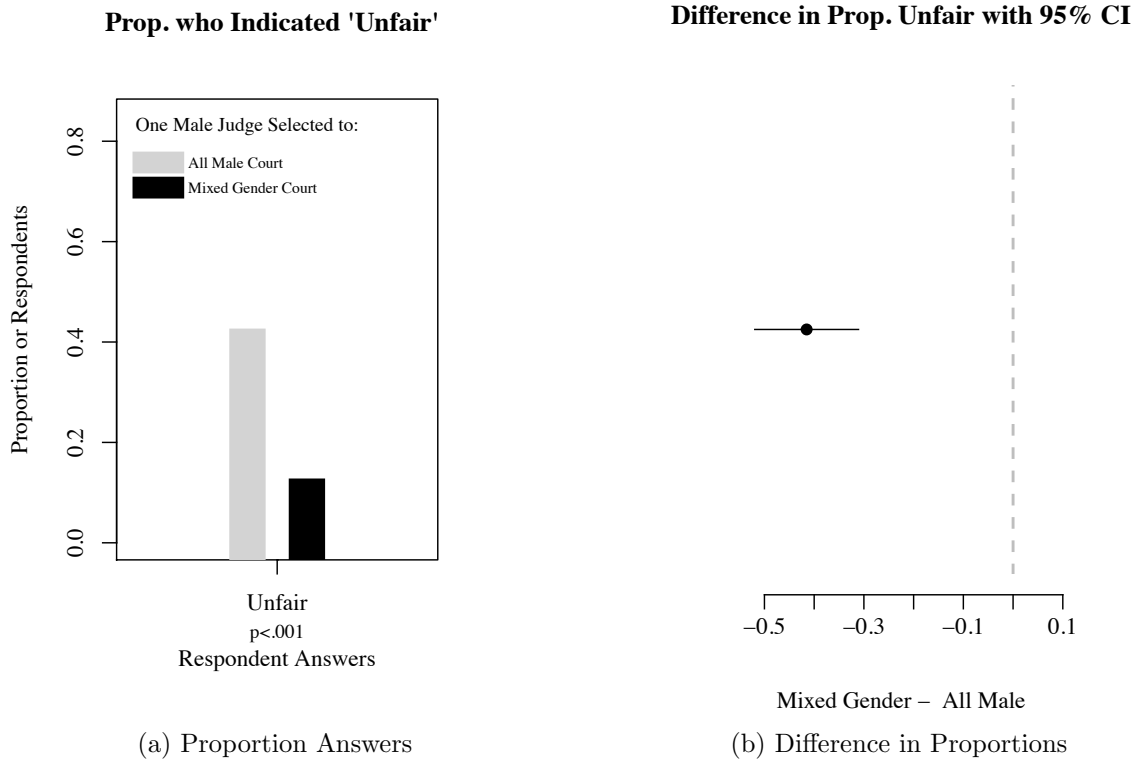
Prop. who Indicated 'Unfair'

Difference in Prop. Unfair with 95% CI

(a) Proportion Answers

(b) Difference in Proportions

Figure 4: *The left panel 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. The right panel shows the point estimate and 95% confidence interval for a differences in proportions test. The p-value for this test is $p < .001$.*

This finding supports an "integrated model" of representation (Schwindt-Bayer and Mishler 2005) in the judicial context by demonstrating how descriptive representation – that is, representation that occurs when those who are representing reflect traits of the represented– can shape perceptions of the procedures used to select and authorize officials to act (i.e., "formal" representation, a conception of representation based on the "formal arrangements which proceed and intitiate" representation (Pitkin 1967, p. 11)). Scholarship that addresses the link between formal and descriptive representation of women typically addresses how formal rules shape prospects for the selection of women. The finding presented above – that observers are more likely to perceive a selection process as fair when women are present in office – confirms that descriptive representation can shape beliefs about formal representation (Schwindt-Bayer and Mishler 2005) in the context of the judiciary.

Does the Amount of Information Affect Perceptions of Unfairness? The information hypothesis (H1) predicts that having more information about the gender disparity amon 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 one male judge who is newly selected to the bench. Group B, however, was told about four currently serving male judges. If increased information about gender disparity among judges affects perceptions of bias, those in group B should be more critical of the selection process than those in group A.

The left panel of figure 5 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 may lead more observers to be critical of the selection process. Despite a ten percentage point difference across groups, the difference in proportions test fails to reject the null hypothesis at the traditional,

Figure 5: Amount of Information and Perceptions of Unfairness

Prop. who Indicated 'Unfair'

Difference in Prop. Unfair with 95% CI
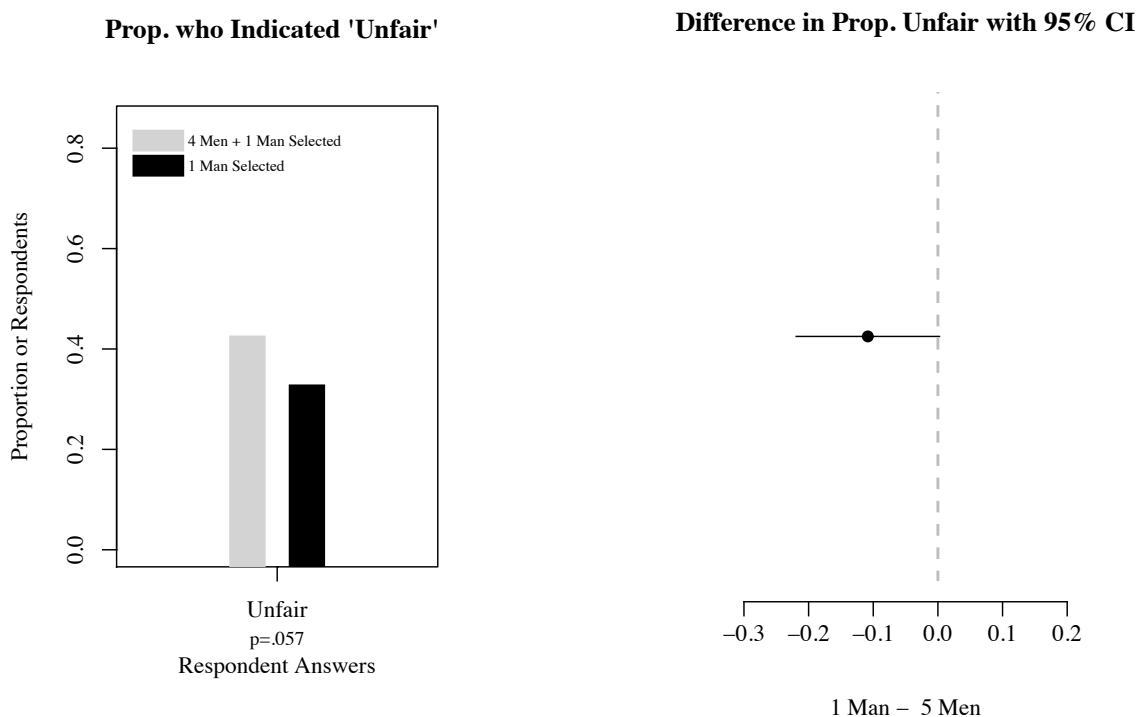
The left panel shows the proportion of respondents who answered either "definitely unfair" or "probably unfair." The black bars show the answers from respondents who saw 1 male judge selected to a court without any information about the existing judges. The grey bars show the responses for those where were given information about a male judge selected to a bench with four existing male judges. The right panel shows the point estimate and 95% confidence interval for a difference in proportions test. The p-value for the difference in proportions test (which follows a χ^2 distribution) of "unfair" responses is p=.057.

95% confidence level (p=.057), so this observed difference across treatment groups may be due to chance. These results, therefore, are at best suggestive.

Does Slate Selection affect Perceptions of Unfairness when Information is Held Constant?   To determine whether slate selection affects perceptions of unfairness relative to rolling, one-by-one selection via the balanced-list mechanism, 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 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 (un)fairness, then the responses about perceived unfairness across the two groups should be the same.

The left panel of figure 6 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). Of those who saw the five male judges selected but were told those judges were selected one-by-one, just shy of 40% indicated that the process was unfair. Among those who saw the *same* five male judges but were told the five judges were selected as a slate, 65% indicated that the process appeared either probably or definitely unfair, a difference from the one-by-one group that is statistically significant (p<.001). The right panel of figure 6 shows the point estimate and 95% confidence interval of the difference in means.

This difference in the perception of unfairness across treatment groups shows that more respondents are critical of gender homogeneity when judges are selected as a group rather than on a rolling, one-by-one basis. Because both groups were given the same information about the same five judges, this finding supports the balanced list hypothesis and the expectation that observers assess candidates differently when candidates are selected as a group versus individually.

Explanations for Differences, Qualitative Responses   The evidence presented in the three tests above and summarized in figure 7 suggests that respondents care about gender diversity on
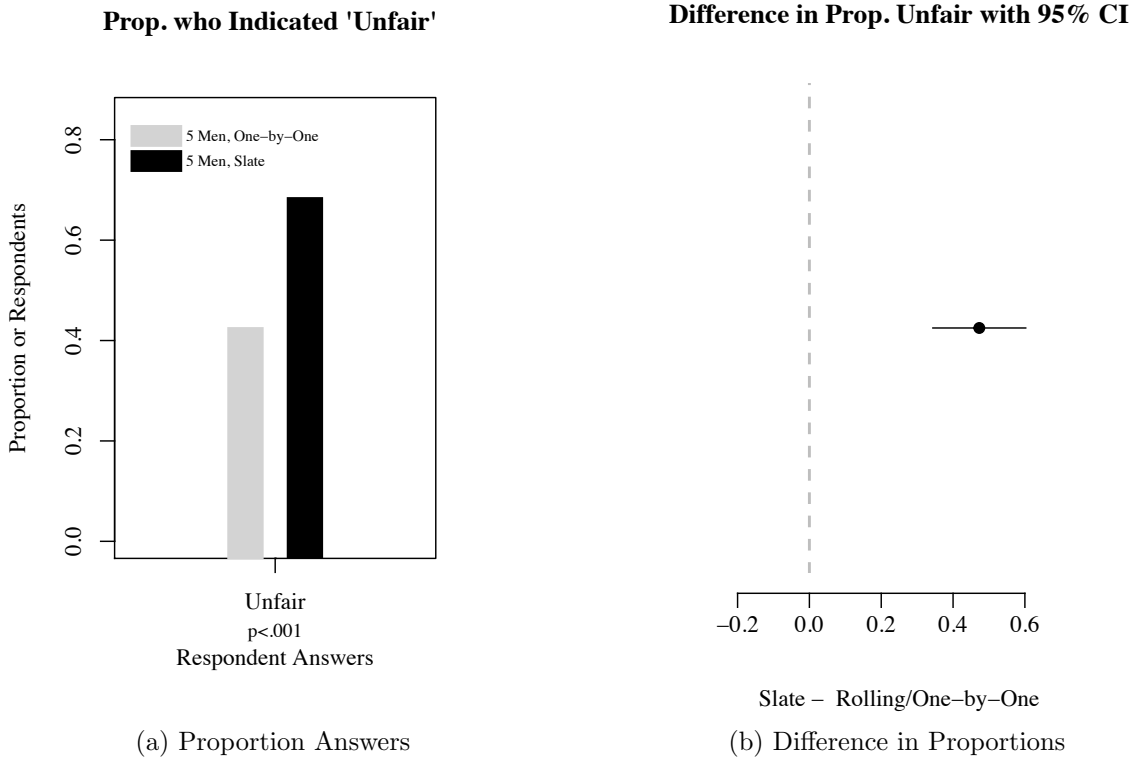
| **Prop. who Indicated 'Unfair'** | **Difference in Prop. Unfair with 95% CI** |

(a) Proportion Answers   (b) Difference in Proportions

Figure 6: *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 associated with this test is $p < .001$.*

courts; respondents are less likely to indicate that the selection process is unfair when there are men and women on a court compared to when there are just men. In addition, the amount of information observers receive about gender disparity *may* affect perceptions of unfairness: more respondents indicated that the process was unfair when they were given information about five male judges than when respondents were given information about just one male judge. However, this difference is not statistically significant (p=.057) at traditional levels of confidence, so may stem from chance. Third, more respondents were critical of gender homogeneity when judges are selected as a slate than when they are selected on a rolling basis, even when the selected judges are identical (p<.001). This difference in perceptions of unfairness is consistent with the balanced list hypothesis in which observers assess candidates differently and more holistically when candidates are selected as a group rather than individually.

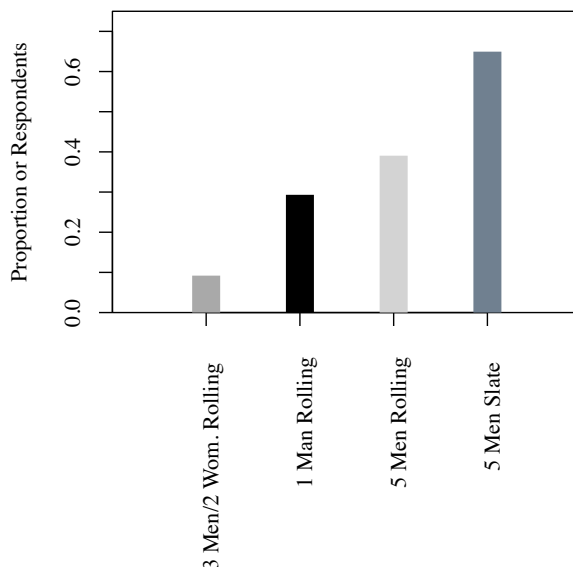**Prop. 'Unfair' across Information and Institutions**

Figure 7: *The proportion of respondents who indicated "definitely unfair" or "probably unfair" across all treatment groups.*

Table 4: Summary of Gender in Qualitative Responses across Treatment Groups

Institution	# of Qual. Responses	% Who Noted Gender	Of those who Noted Gen. % Unfair
Rolling, 2 Wom/3Men	181	33%	17%
Rolling, 1 Man	173	39%	63%
Rolling, 5 Men	168	46%	79%
Slate, 5 Men	189	67%	91%

This table reports the number of qualitative responses by treatment group. The percentage of respondents who noted gender in their qualitative explanations for why they thought the process was either fair or unfair is listed by treatment group in column 3. Among those who noted gender in their responses, column 4 reports the percentage who indicated that the process was either "probably unfair" or "definitely unfair." Comparing the rolling, five men group to the slate, five men group (the balanced list hypothesis), respondents in the slate group were 13 percentage points more likely to notice gender (p<.001), and among those who noted gender there was a 12 percentage point difference in the proportion of respondents who concluded that the process was either definitely or probably unfair (p=.016). For the information hypothesis (comparing one man, rolling to five men, rolling, there is a seven percentage point difference in the rate at which respondents noted gender (p=.15) and a 16 percentage point difference in the proportion of respondents who indicated the process was unfair (p=.0251).

To elucidate why those in the slate group were more critical of gender disparity than those who saw the same information in the rolling/one-by-one group, I turn to qualitative explanations for respondents' choices. In addition to indicating their perception of fairness or unfairness on a five-point scale, respondents were asked to provide an explanation for their decision. Table 5 provides examples of typical responses. As table 4 shows, among those who provided a qualitative explanation for their response, 67% in the slate selection treatment group noted the gender of the justices in their responses.[19] In contrast, among those in the rolling/one-by-one with five men group, only 46% noted gender in their qualitative responses, a difference that is statistically significant (p <.001). Among those who noticed the gender of the judges, those in the slate group were also more likely to indicate that the process was either "definitely unfair" or "probably unfair" than those in the rolling, five men group (p=.016). The qualitative responses suggest that respondents were more likely to both *notice* and *be critical of* the gender of judges when judges were selected simultaneously as a slate than on a rolling, one-by-one basis. To the extent that those tasked with selecting officials are responsive to increased criticism, slate selection procedures should lead to increased gender diversity in office.

---

[19]Any qualitative response that included the following terms were classified as noting gender: men, women, man, woman, male, female, gender.

Table 5: Examples of Qualitative Responses

Institu-tion	Response	Noted Gender?	Explanation for Response
Rolling, Five Men	Probably Unfair	Yes	"I think that these judges are definitely qualified, but I would like to see at least one woman serving"
Rolling, Five Men	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."
Rolling, Five Men	Probably Fair	No	"Candidates are judged based on age, experience and education, so it seems fair."
Slate, Five Men	Definitely Unfair	Yes	"Because if over half of qualified candidates are female then WHY aren't they being chosen."
Slate, Five Men	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, Five Men	Probably Fair	No	"I think it is fair because they all have many years of experience and graduated from top ranked law schools."

Examples of respondents' explanations for their responses about the fairness/unfairness of the selection process.

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 the survey experiments offer a few lessons. First, respondents care about judicial diversity; respondents were more likely to indicate that the selection process was fair when the outcome of the selection process included both men and women. Respondents confronted with an all-male court were critical of the process that resulted in gender homogeneity, which confirms expectations that gender diversity shapes perceptions of institutional fairness and legitimacy.

Second, I theorised that one mechanism through which slate selection may increase the extent to which the absence of women is perceived as problematic is through increased information about

disparity. Holding selection procedures fixed, I manipulated how many male judges respondents observed. Results here are suggestive: respondents who saw five male judges were more critical than those who saw only one male judge, but this difference is not statistically significant at the traditional thresholds (p=.057).

Third, given the same information about the composition of the court, respondents were more critical of gender disparity when judges were selected as a slate rather than one-by-one (p<.001). This finding indicates that observers assess the characteristics of officials differently when officials are assessed as a group rather than individually, in line with the balanced list hypothesis. Moreover, qualitative responses demonstrate that respondents are not only more critical of gender disparity but are also more likely to even note gender when judges are selected as a group rather than one-by-one. In this study, observers applied stricter standards for gender diversity to officials selected as a group.

Fourth, taken together, this study indicates that one way to increase *accountability* for gender diversity in office – and, subsequently, increase gender diversity *among* those holding office – is to alter the timing of selection so that multiple officials are selected simultaneously.[20] Unlike quotas or other institutional features designed to explicitly address the inclusion of one or a few specific identities, altering the timing of selection may facilitate diversification without sparking the resistance associated with explicit gender quotas (Krook 2016). Moreover, because the timing of selection promotes diversity by altering the way in which observers assess traits generally, selecting judges as a group/slate may facilitate the inclusion of both women judges and judges who hold myriad other characteristics that have been historically excluded from positions of power. Finally, this paper shows how lessons from one institutional context – legislative elections – can shed light on other institutional contexts such as the judiciary. As such, this paper contributes to the development of a more general theory of institutions and diversification.

---

[20]The findings in this project are limited to a specific survey experimental context designed to promote *internal* validity; the extent to which the patterns observed in this project would be reflected in the observational setting depends on many other social and political factors not tested in this project.

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.

André, Audrey, Sam Depauw, Matthew S Shugart and Roman Chytilek. 2017. "Party nomination strategies in flexible-list systems: Do preference votes matter?" *Party Politics* 23(5):589–600.

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. 2019. "Judicial Merit Selection: Beliefs about Fairness and the Undermining of Gender Diversity on the Bench." *Paper presented at the 2019 Meeting of the American Political Science Association, Washington DC* .

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 Engagement." *The Journal of Politics* 65(4):1040–1061.

Bäck, Hanna, Marc Debus and Wolfgang C Müller. 2016. "Intra-party diversity and ministerial selection in coalition governments." *Public Choice* 166(3-4):355–378.

Badas, Alex and Katelyn E. Stauffer. 2019. "'Scoring Points?' Diversity in Judicial Nominations and Perceptions of Presidential Accomplishment." *Paper presented at the Annual 2019 Meeting of the American Political Science Association, Washington D.C.* .

BasicDocuments. N.d. "Statute of the International Court of Justice,.". Accessed: Oct 11, 2016.

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.

Casey, Logan S, Jesse Chandler, Adam Seth Levine, Andrew Proctor and Dara Z Strolovitch. 2017. "Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection." *SAGE Open* 7(2):2158244017712774.

Caul, Miki. 1999. "Women's representation in parliament: The role of political parties." *Party politics* 5(1):79–98.

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/*

Compaoré, Nestorine. 2005. "Burkina Faso: Recruiting women for legislative elections." *Women in Parliament: Beyond Numbers* pp. 132–140.

Constitute. N.d. "Central African Republic 2013.". Accessed October 11, 2016.
URL: $https://www.constituteproject.org/constitution/Central_African_Republic_2013?lang = en$

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.

Crisp, Brian F, Santiago Olivella, Michael Malecki and Mindy Sher. 2013. "Vote-earning strategies in flexible list systems: Seats at the price of unity." *Electoral Studies* 32(4):658–669.

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.

Dawuni, Josephine and Alice Kang. 2015. "Her ladyship chief justice: the rise of female leaders in the judiciary in Africa." *Africa Today* 62(2):45–69.

Engstrom, Richard L. 1987. "District magnitudes and the election of women to the Irish Dáil." *Electoral Studies* 6(2):123–132.

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.

Escobar-Lemmon, Maria C, Valerie Hoekstra, Alice Kang and Miki Caul Kittilson. 2019. "Breaking the Judicial Glass Ceiling: The Appointment of Women to High Courts Worldwide." *Paper presented at the 2019 European Conference on Politics and Gender, Amsterdam* .

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.

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.

Goelzhauser, Greg. 2019. *Judicial Merit Selection: Institutional Design and Performance for State Courts.* Temple University Press.

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 Representation of Women." *PS: Political Science and Politics* 28(4):711–713.

Hirschl, Ran. 2008. The Judicialization of Politics. In *The Oxford handbook of political science.*

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.

Huff, Connor and Dustin Tingley. 2015. ""Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2(3):2053168015604648.

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.

Inter-ParliamentaryUnion. 2019. "Women In Politics Infographics." *United Nations Entity for Gender Equality and the Empowerment of Women* .
   URL: *https://www.ipu.org/resources/publications/infographics*

JudgesLibrary. 2016. "Superior Court of Justice: History of the Court.".

  URL: *http://www.ontariocourts.ca/scj/about/history/*

Junn, Jane. 1997. Assimilating or Coloring Participation? Gender, Race, and Democratic Political Participation. In *Women Transformng 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.

Krivosha, Norman. 1987. "Acquiring Judges by the Merit Selection Method: The Case for Adopting Such a Method." *Southwestern Law Journal* 40:15.

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

Krook, Mona Lena. 2016. "Contesting gender quotas: dynamics of resistance." *Politics, Groups, and Identities* 4(2):268–283.

Kunovich, Sheri and Pamela Paxton. 2005. "Pathways to power: The role of political parties in women's national political representation." *American Journal of Sociology* 111(2):505–552.

Malleson, Kate. 2006. "Rethinking the merit principle in judicial selection." *Journal of Law and Society* 33(1):126–140.

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 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. "Thee 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.

McCredie, Morgan N and Leslie C Morey. 2018. "Who are the Turkers? A characterization of MTurk workers using the personality assessment inventory." *Assessment* p. 1073191118760709.

Mossman, Mary Jane. 2013. Becoming the First Woman Judges in Ontario: Women Lawyers, Gender and the Politics of Appointment. In *Gender and Judging*, ed. Ulrike Schultz and Gisela Shaw. The Oñati Institute for the Sociology of Law Portland, Oregon: Hart Publishing pp. 51–67.

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.

Norris, Pippa and Ronald Inglehart. 2001. "Women and democracy: Cultural obstacles to equal representation." *Journal of democracy* 12(3):126–140.

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.

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

Reddick, Malia, Michael Nelson and Rachel Caufield. 2010. "Examining diversity on state courts: How does the judicial selection environment advance–and inhibit–judicial diversity." *American Judicature Society. Retrieved February* 21:2012.

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.

Rule, Wilma and Matthew Shugart. 1995. "The preference vote and the election of women." *Voting and Democracy Report* .

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: Overiview 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.

Stegmaier, Mary, Jale Tosun and Klára Vlachová. 2014. "Women's parliamentary representation in the Czech Republic: does preference voting matter?" *East European Politics and Societies* 28(1):187–204.

Tate, C Neal and Torbjorn Vallinder. 1995. *The global expansion of judicial power*. NYU Press.

The Constitution of the Republic of South Africa: December 16, 1996 (as Amended to March 26, 2009). N.d. Accessed via The Varieties of Democracy Judiciary Project and the Comparative Constitutions Project.

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.

Valdini, Melody Ellis. 2012. "A deterrent to diversity: The conditional effect of electoral rules on the nomination of women candidates." *Electoral Studies* 31(4):740–749.

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.

1 Appendix

1.1 Gender Quotas on Courts

While less common than legislative quotas, there are examples of quotas for courts. The 2013 interim constitution for the Central African Republic had an explicit gender quota for their constitutional court: "The Constitutional Court of the Transition is composed of nine (9) members of which at least four (4) are women, who will hold the title of Constitutional Judges" (Constitute N.d., Article 80). Other courts have more nebulous goals of diversity.

While not a numeric quota, the 2009 South African Constitution states the need for descriptively representative courts: 'The need for the judiciary to reflect broadly the racial and gender composition of South Africa must be considered when judicial officers are appointed"(*The Constitution of the Republic of South Africa: December 16, 1996 (as Amended to March 26, 2009)* N.d., Chapter 8, Article 174).

Article 9 of the Statute of the International Court of Justice states, "At every election, the electors shall bear in mind not only that the persons to be elected should individually possess the qualifications required, but also that in the body as a whole the representation of the main forms of civilization and of the principal legal systems of the world should be assured." (BasicDocuments N.d.).