

Nominating Commissions, Judicial Retention, and Forward-Looking Behavior on State Supreme Courts: An Empirical Examination of Selection and Retention Methods

State Politics & Policy Quarterly
2015, Vol. 15(2) 211–238
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1532440014567858
sppq.sagepub.com



Ryan J. Owens¹, Alexander Tahk¹,
Patrick C. Wohlfarth², and Amanda C. Bryan³

Abstract

High-profile advocates are pushing states to move away from judicial elections and toward a “merit” method because it purportedly produces the best quality judges. Quality, however, is difficult to measure empirically. Rather than attempt to measure quality, we examine whether certain types of state supreme courts are more forward-looking than others. States are likely to desire forward-looking behavior among judges because it can protect judicial legitimacy, help states to control policy, and could be more efficient than myopic behavior. Using a recent innovation in matching called covariate-balancing propensity scores, we find that the U.S. Supreme Court is equally likely to review and reverse decisions by judges regardless of their selection or retention methods. These results suggest that state supreme court justices, no matter their paths of getting to (and staying on) their courts, are roughly equal in terms of forward-looking behavior.

Keywords

judicial elections, judicial politics, judicial selection, Missouri plan, judicial reforms

¹University of Wisconsin, Madison, WI, USA

²University of Maryland, College Park, MD, USA

³Loyola University Chicago, Chicago, IL, USA

Corresponding Author:

Ryan J. Owens, Department of Political Science, University of Wisconsin, 214 North Hall, 1050 Bascom Mall, Madison, WI 53706, USA.

Email: rjowens@wisc.edu

In recent years, various controversies arising from judicial elections have led to renewed calls for states to abolish contested judicial elections and reform their judicial selection methods. Perhaps chief among judicial reform proselytizers is former Supreme Court Justice Sandra Day O'Connor, who declares that judicial elections are "awful" (Podgers 2009) and that her "first piece of advice" for states with elected judiciaries is to switch to the "Missouri Plan" (O'Connor 2009, 490, discussed in Bonneau and Hall 2009). O'Connor believes that nominating commissions—not voters or elected officials—should identify suitable individuals to be state supreme court justices. And she is not alone. Justice Ruth Bader Ginsburg has also labeled judicial elections as "dreadful" and backed the Missouri Plan (Hardin 2013). Indeed, there are widespread calls today to replace elected judiciaries with the Missouri Plan selection method (but see Bonneau and Hall 2009; M. G. Hall 2014b). Often, though not always, the purported reason to change is that the Missouri Plan generates better quality judges and legal opinions (Hardin 2013).

The problem, however, is that the "quality" is elusive to measure, which means that the debate over judicial selection methods can reduce to hunches and grousing. Rather than engage in a largely esoteric debate over quality, we elect, instead, to examine a concept that *can* be measured empirically and, therefore, can contribute usefully to the debate over judicial selection. And it is a concept with practical consequences for states and judges alike. We examine the forward-looking behavior of state high courts. As we see it, forward-looking behavior is valuable to states because it likely protects judicial legitimacy, can help states retain more control over their judicial policies, and is efficient. In other words, we ask the following question: Are judges selected and retained by some methods more forward-looking than others?

To determine whether some types of state supreme courts are more forward-looking than others, we examine whether the U.S. Supreme Court is more likely to review or reverse state supreme courts' judicial opinions. In terms of the broader debate, we ask whether judges selected under the merit plan are more forward-looking than other judges. We employ the covariate-balancing propensity-score (CBPS; i.e., matching) approach recently devised by Imai and Ratkovic (2014) to determine the effects of selection and retention methods on upper court review and reversal. We first analyze whether the Supreme Court is more likely to review some state courts over others. We then look at whether the Court is likely to reverse some state supreme courts more than others.

Our results are consistent and clear. The U.S. Supreme Court does not disproportionately review or reverse the decisions of state judges selected or retained by any particular method. It treats decisions by Missouri Plan judges *the same* as decisions by elected and appointed judges. What is more, neither the use of nominating commissions nor retention elections—two major features touted by judicial reformers—lead to greater forward-looking behavior and insulation. In short, judges selected under the merit plan are just as forward-looking as other judges.

These results are important for a number of reasons. First, the policy implications are manifest, as they question whether a move away from one selection and retention method to another will lead to a more forward-looking judiciary. If we believe

forward-looking behavior is tied to the protection of legitimacy, state control over policy, and efficiency, all of these things can be achieved (and in fact have been achieved) without changing selection methods. Indeed, a number of scholars have challenged the arguments of reformers on empirical grounds (Bonneau and Hall 2009; Comparato and McClurg 2007; Gibson 2012; M. G. Hall 2014b), arguing that elected judiciaries have benefits that merit plan judiciaries do not—and that the purported benefits of the plan are overestimated. These results add to such counterclaims. Second, although other scholars have compared judges' characteristics across selection methods, they have done so on different dimensions (see, for example, Glick and Emmert 1987; Goelzhauser and Cann 2014). Ours is the first large-scale, systematic examination of state judicial selection mechanisms that focuses on both the Supreme Court's agenda *and* merits stages. Furthermore, unlike previous studies, we separately investigate nominating commissions and retention elections, which allows us to scrutinize more precisely institutional features of selection methods that may be relevant. Third, and not to be overlooked, the matching approach we employ allows us to make much stronger inferences about the effects of selection methods on forward-looking behavior than standard parametric measures. And, as far as we can tell, we are among the first to employ the approach.

Debates over Judicial Selection Methods

Because the consequences of judicial selection are profound, states have adjusted their selection practices for over 200 years. During the early nineteenth century, states used appointment systems to select their judges. In the Jacksonian era, however, many abandoned appointments in favor of elected judiciaries. By electing judges, states could remove judicial ties to governors and legislatures, giving them an independent base of authority and, presumably, greater decisional independence (McLeod 2007). Soon, though, observers began to question the propriety of judicial elections, criticizing that they sacrificed too much judicial independence for accountability. Another concern focused on whether judges should raise campaign contributions (Pound 1906). In response, reformers proposed a plan they hoped would take politics (and other perceived negatives) out of the judicial selection equation. Governors would choose state supreme court judges from a pool of candidates put forth by judicial merit commissions, and these judges would subsequently sit in retention elections (Krivosha 1990, 130). In 1940, Missouri became the first state to adopt this "merit plan" (Fitzpatrick 2009), which consequently earned the label, the "Missouri Plan."

Today, the Missouri Plan is one of many methods states employ to select their supreme court justices and judges. Some states still select (and retain) their justices using partisan elections, with justices running as Democrats or Republicans. Others employ nonpartisan elections, in which justices run against opponents, but none clearly identify with a party—at least on the ballot. Other states employ legislative or gubernatorial appointive systems, subject to varying kinds of tenure and retention methods. The prototypical Missouri Plan system, though, has two unique features—nominating commissions and retention elections.

The first unique feature of the Missouri Plan is its use of nominating commissions that provide governors with names of potential judicial nominees, one of whom he or she must select to fill a vacancy on the court. More specifically, a judicial nominating commission—made up of lawyers, politicians, and sometimes members of the public—compiles a list of qualified individuals for a judicial vacancy. The commission forwards this list to the governor who then must select someone from the list for the judicial vacancy.¹ These commissions were created to place judicial selection decisions in the hands of so-called experts (i.e., state bars) rather than potentially misinformed voters or politicians who might have their own agendas in mind when making judicial selections (Bierman 2002, 678). The goal, in other words, was to create a commission that would filter out unqualified applicants (who might otherwise run for, and win, a judicial election) and select only the most qualified.

Some reformers claim that these nomination commissions produce better state supreme court justices than other selection methods. This is the case because the nominating commissions consist of legal experts who pick only the most qualified individuals to serve as judges. As Caufield (2010, 772) explains, “widespread acceptance of merit selection . . . [has come] to rest on the notion that the individuals chosen through merit selection would be different—and, on the whole, better—than those who were chosen through electoral processes.” Indeed, a number of scholars and policy makers have made this point. Crompton (2002) argues that commissions limit the role of politics in judicial selection. Harrison, Swisher, and Grabel (2007, 256) claim that selection by commission “is almost entirely transparent, exacting, and virtually devoid of political influence.” The American Judicature Society (n.d., 1) declared that the merit selection process is “the best way to choose the best judges” because nominating commissions will choose applicants “on the basis of their qualifications, not on the basis of political and social connections.”

The second unique feature of the plan is its use of retention elections—and the independence they arguably afford justices. After some period of time on the bench, a justice under the Missouri Plan will be subject to a retention election. Voters cast their ballots simply to retain or not retain the justice (i.e., without an opponent). If a majority votes to retain the justice, he or she continues to serve. If a majority votes not to retain the judge, the position becomes vacant and must be refilled.

Advocates claim that retention elections enhance judicial independence while retaining a small degree of democratic accountability. That is, by making the judge difficult to remove—by taking away an electoral opponent—the judge has more independence to make sound legal decisions. Independence is desirable because judges “cannot make the hard decisions . . . unless they are truly independent” (Webster 1995, 9). Chemerinsky (1988, 1988) argues that the “entire concept of the rule of law *requires* that judges decide cases based on their views of the legal merits, not based on what will please voters” (emphasis in original). Missouri Plan judges and their opinions are theoretically better, then, because they enjoy decisional independence.

The scholarship, however, is mixed regarding which methods are best. Some scholarship suggests that judges selected by different methods sometimes manifest different behavior, but whether this means one selection and retention method is superior to

another is unclear. For example, Tabarrok and Helland (1999) find that partisan elected judges intentionally generate higher tort awards (for reelection purposes) than other types of judges (see also Helland and Tabarrok 2002). Huber and Gordon (2004) find that judges who are elected issue harsher sentences as they approach reelection. Iaryczower, Lewis, and Shum (2013) evaluate the trade-offs between bureaucrats and elected (compared with unelected) state supreme court justices and find that in criminal rights cases, “justices that are shielded from voters’ evaluations on average have higher quality information than justices that face either reelection or retention elections” and thus make fewer mistakes.² Cann (2007) finds that judges in states that use gubernatorial appointment or the Missouri Plan rate their state judiciaries higher in terms of quality than judges in states with electoral systems. Choi, Gulati, and Posner (2010) find that opinions written by partisan elected judges (especially in large states) are cited less often than opinions written by appointed judges. And, looking to judicial ethics, Reddick (n.d.) finds that judges disciplined for ethical lapses are more likely to hail from elected states.

Other studies resurrect elected courts, especially partisan elected courts. Caldarone, Canes-Wrone, and Clark (2009) and Canes-Wrone, Clark, and Park (2012) suggest that state court judges who are selected through the Missouri Plan or nonpartisan elections actually are *more* beholden to popular opinion than judges elected through partisan elections. Choi, Gulati, and Posner (2010) find that partisan elected judges are just as independent as appointed judges—and are more productive in terms of opinions published. Glick and Emmert (1987), focusing on where state supreme court justices obtained their law degrees and how much previous legal experience they enjoyed before going on the bench, find no differences among judges based on selection methods.

Similar studies focus on the positive impacts of judicial elections. For example, Gibson (2012) finds that the benefits of judicial elections offset the costs from campaigning. That is, even though respondents have some misgivings about judges taking campaign contributions to run in elections, the benefits of holding an election overall outweigh those negative reactions. In other words, he argues, the “new style” of judicial campaigning is not as ruinous as reformers believe (see also Comparato and McClurg 2007). Other scholars agree. Bonneau and Hall (2009) find that connecting voters to the judiciary offers needed accountability among the judiciary and provokes more legitimacy. In other words, if voters value accountability, judicial elections are worthwhile.

As this brief summary no doubt signals, it is nearly impossible to determine whether one type of approach is “better” than another, or generates better quality judges or opinions than others. In large part, this is because the notion of quality is so broad that it must be defined and measured in a specific sense to be of any use. And, unfortunately, this is not often done. So, rather than participate in the debate over judicial quality, we look instead at a related, but empirically measurable, concept—forward-looking behavior.

State Selection Systems and Forward-Looking Behavior

Although states are likely to vary in their preferences over a great many things (e.g., Owens and Wohlfarth 2014), surely they should all be interested in having courts that

are seen as legitimate, that keep their states in control of their policies, and that are efficient. Of course, judges should also value these things. We contend that forward-looking judicial behavior can help them accomplish these goals. That is, by deciding cases in a manner that reduces the likelihood of Supreme Court review and reversal, state supreme courts can enhance their legitimacy, keep control of their policies, and minimize costs. We address each of these concerns in turn.

Protecting Legitimacy through Forward-Looking Behavior

A forward-looking court might protect its legitimacy more than a court that does not anticipate upper court behavior. In general, a court's legitimacy is the bedrock of its support. Justice Frankfurter, referring to the U.S. Supreme Court, once claimed, "The Court's authority . . . rests on sustained public confidence in its moral sanction" (Caldeira 1986, 1209). While state courts are different insofar as many of them have a closer electoral tie to voters than the Supreme Court, the principle nevertheless remains: courts are strongest when their institutional support floats at high tide. And, a court that is reviewed and reversed regularly is likely to be seen as less legitimate than a court that avoids review and reversal.

A host of empirical scholarship examining the ebbs and flows of state court legitimacy lends support for the expectation that there is a link between state supreme court legitimacy and the U.S. Supreme Court's review and reversal of state court decisions. While there is a debate in the literature over whether a state's judicial selection mechanism can directly affect perceptions of legitimacy (see Cann and Yates 2008; Wenzel, Bowler, and Lanoue 2003), Benesh (2003) argues that the foundation of legitimacy lies in perceptions of procedural fairness. State courts, she argues, have the highest perceived legitimacy when their procedures are seen as fair and accurate. One of the primary reasons the U.S. Supreme Court reviews and reverses decisions is to alter lower court rules and procedures. Thus, Supreme Court attention can send a powerful signal calling into question the procedural fairness of state high courts.³ When fairness is called into question, legitimacy is challenged.

Moreover, the Supreme Court has a unique effect on public opinion. Decisions by the U.S. Supreme Court tend to legitimize legal rules, leading the public to accept them as fair, correct, and just even when the Court's decisions do not comport with the public's preferences (Gibson 1989; Johnson and Martin 1998; Tyler and Rasinski 1991). When the Supreme Court reviews and reverses state courts, the public's likely reaction is to view the lower court decision as inaccurate or even unfair. Thus, forward-looking state court justices would want to avoid attention from the U.S. Supreme Court which might produce costly hits to their legitimacy.

Finally, a number of scholars have suggested that noncompliance and nonimplementation of court decisions can harm judicial legitimacy (Clark 2009; Dahl 1957; Eskridge 1991; M. E. K. Hall 2014a). Kosaki (2003) argues that state courts are particularly cognizant of the impact nonimplementation can have on their legitimacy and strategically alter their behavior accordingly. At the state court level, Supreme Court review and reversal has the same outcome as legislative reversal: nonimplementation

of state court decisions. Thus, in the same way, justices on state courts expect legislative overrides to harm their legitimacy, so might reversal by the Supreme Court.

Controlling Policy through Forward-Looking Behavior

What is more, forward-looking behavior is useful for states because it allows them to retain control over their policies. When a state supreme court decision gets appealed to the U.S. Supreme Court, that Court has the power to undo the decision and its policies. If states are concerned about creating and controlling their own policies—and we have every reason to believe they are—they should want to avoid such a state of affairs. By strategically anticipating High Court reactions, state court justices can reduce the likelihood of review and reversal. As Comparato and McClurg (2007, 730) state, judges (at least those elected) should aim to avoid upper court monitoring to enhance their “latitude.” They can avoid such monitoring by not “engaging in protracted contact with the Supreme Court . . .” (Comparato and McClurg 2007, 730). Thus, state judges who are better able to avoid review and reversal by the U.S. Supreme Court are likely to minimize incurring the long-term costs associated with establishing an unfavorable Court precedent or attracting greater future monitoring of state policies.

Indeed, a host of studies show how lower courts and agencies try to insulate their policy decisions from upper court review, something Tiller and Spiller (1999) call “strategic instrumentalism.” When faced with judicial scrutiny, agencies often employ high-cost approaches (e.g., adjudication) rather than low-cost approaches (e.g., rule-making) to make it more difficult for reviewing courts to change their policies (Tiller and Spiller 1999). As the preferences of the agency and reviewing court diverge, the agency makes policy by adjudication rather than broad rulemaking because it is more costly for reviewing courts to supervise specific agency adjudications. Likewise, Smith and Tiller (2002) find that circuit courts review agency decisions with an eye toward obstructing Supreme Court review. That is, circuit judges strategically base their decisions on legal grounds that make Supreme Court review more costly. As the circuit court becomes increasingly distant ideologically from the agency decision, the probability that the circuit court utilizes factual “reasoning process” review (looking at the *facts* of cases) rather than statutory interpretation increases significantly.⁴

Enhancing Efficiency through Forward-Looking Behavior

A forward-looking court can help states avoid the legal instability and expense associated with Supreme Court reversals. When the Supreme Court grants review of a lower court opinion, it confuses the law in the case. Often, the Court puts a stay on the lower court decision, and the law is unsettled while the Court decides. For the litigants, this legal intermission is problematic. And for states that must regulate in the interim, it can be remarkably frustrating. Frequent occurrences only exacerbate this problem. At the same time, Supreme Court review of state high court decisions can be costly.

To be sure, many state supreme court decisions are litigated, and then appealed, by private parties who bear the most direct costs of review. But many state supreme court

decisions that make their way to the U.S. Supreme Court involve issues litigated by state Attorney General offices. And on appeal, these attorneys often must devote their entire attention to the Supreme Court case. These are, of course, resources that could be devoted elsewhere. In addition, if the High Court remands the decision, the state court must decide it again, exerting yet more attention that could be devoted elsewhere. With large workloads and little time, deciding a remanded case is costly for the courts. In other words, although it would be nice for the Supreme Court to affirm a lower court decision, the likelihood of reversal plus the costs necessary to litigate (or redo) the case make forward-looking behavior (i.e., insulation) attractive.

Judge-Level Benefits of Forward-Looking Behavior

Of course, state supreme court justices are likely to want to avoid review and reversal as well, and for similar reasons (Comparato and McClurg 2007). They likely want to keep control over legal policy rather than turning it over to the U.S. Supreme Court. They will also want to avoid having to rehear cases, as it consumes their busy time. And for those state supreme court justices who have aspirations to secure an appointment to the federal bench, a large reversal rate is likely to cast doubt on their qualifications. Indeed, nearly all court watchers use reversal rates to evaluate circuit judges nominated to the Supreme Court because “judicial reversal reflects professional criticism by other professionals” (Cass 1995, 984). Others have used reversal as a measure of performance as well. Cross and Lindquist (2009), for example, analyze Supreme Court reversal rates when measuring the quality of federal circuit court judges. In a study of antitrust litigation, Baye and Wright (2011) examine how formal training in economics can help judges avoid being reversed on appeal, as reversal has “potentially deleterious effects . . . that could damage [their] reputations(p. 4).” Even prominent federal circuit judge Richard Posner has suggested that reversal rates capture the concept of judicial performance (Posner 2005).⁵ Thus, judges who are likely to seek elevation should engage in forward-looking behavior.

Along the same lines, judges who are frequently reversed by the High Court may find their ability to persuade colleagues in other state courts across the country diminished (Klein and Morrisroe 1999). If frequent reversal takes a toll on a judge’s prestige, that judge’s ability to influence the law beyond his or her jurisdictional boundaries will be limited. However, making sound decisions that are not reversed can help. In other words, forward-looking behavior can help a judge’s reach across the judiciary.

Data and Measures

We are interested in examining whether some kinds of state high courts are better able to look forward and prevent review or reversal than others—whether the Supreme Court is more likely to review or reverse some kinds of courts more than others. As such, we analyze two broad models: a model of Supreme Court review and a model of Supreme Court reversal.⁶ One especially useful benefit of examining review and reversal is that they scale across states, allowing for large-scale empirical analyses that

might not be available with other approaches. That is, to compare judicial selection methods across states, the measure of interest must “scale.” For example, assume that we, instead, analyzed judicial discipline. If judges chosen by some judicial selection method are punished more than judges chosen by other methods, one might infer that the first method is “worse” than the second. Yet, there is not one common measure of discipline among them: the various states have different sets of ethical guidelines and different ethics commissions. However, state supreme courts are all constantly under the threat of U.S. Supreme Court review and reversal. This dynamic thus represents a single measure whereby all state courts are held accountable.

Supreme Court Review of State Court Decisions

We first analyze whether the Supreme Court grants certiorari to (i.e., opts to review) decisions by some kinds of state courts more than others. Of course, examining Supreme Court review of state court decisions is not a trivial data undertaking. Because the Supreme Court does not issue opinions on the petitions to which it denies review, the characteristics of denied cases are unobservable, making analysis of agenda setting difficult. To overcome this limitation, we draw on Harry Blackmun’s private agenda documents (Epstein, Segal, and Spaeth 2007), which include the Court’s docket sheets (detailing how each justice voted), as well as the internal “cert pool” memos prepared by the justices’ law clerks.⁷ These memos become available to the public only after a justice passes away and allows his or her estate to release them. Blackmun’s materials, which are highly reliable (Black and Owens 2009b), contain the Court’s agenda materials for all petitions for review filed during the 1986–93 terms. Our data thus provide one of the first systematic examinations of Supreme Court review of state high court decisions.

Our dependent variable, *State Court Reviewed*, captures whether the U.S. Supreme Court granted review to a cert petition that challenged a state supreme court decision. To construct this variable, we identified all paid certiorari petitions (i.e., excluding in forma pauperis petitions) asking the Court to review a case emanating from a state court of last resort that made the Court’s discuss list,⁸ from 1986 to 1993.⁹ *State Court Reviewed* thus equals 1 if the Supreme Court granted certiorari to the petition challenging a state supreme court decision, 0 otherwise.

Our main covariates identify the judicial selection system used to select the state supreme court justices whose decisions were under review. Table 1 shows the selection systems used in each state between 1960 and 2011. Based on the metric provided by the American Judicature Society,¹⁰ we specify dummy variables for each of the following selection systems—*Nonpartisan Election*, *Partisan Election*, *Appointment*, and *Merit System* (i.e., the Missouri Plan).¹¹ To be sure, not all states that employ the same broad approach use their methods identically. For example, Hawaii’s merit system is different from Iowa’s. Our grouping of states, however, fits within conventional literature on the topic and broadly within the American Judicature Society’s approach.¹² Each variable takes on a value of 1 if the state under review employed the particular selection system, 0 otherwise. In our models, we use the merit selection system as our

Table 1. Selection Systems for State Supreme Courts, 1960–2011.

State	Selection system	Years used
Alaska	Merit system	1960–2011
Alabama	Partisan election	1960–2011
Arkansas	Partisan election	1960–99
	Nonpartisan election	2000–11
Arizona	Partisan election	1960–73
	Merit system	1974–2011
California	Merit system	1960–2011
Colorado	Merit system	1960–2011
Connecticut	Appointment	1960–64
	Merit system	1965–2011
Delaware	Appointment	1960–76
	Merit system	1977–2011
Florida	Partisan election	1960–71
	Merit system	1972–2011
Georgia	Partisan election	1960–82
	Nonpartisan election	1983–2011
Hawaii	Merit system	1960–2011
Iowa	Partisan election	1960–61
	Merit system	1962–2011
Idaho	Nonpartisan election	1960–2011
Illinois	Partisan election	1960–2011
Indiana	Partisan election	1960–69
	Merit system	1970–2011
Kansas	Merit system	1960–2011
Kentucky	Partisan election	1960–74
	Merit system	1975–2011
Louisiana	Partisan election	1960–2011
Massachusetts	Appointment	1960–2011
Maryland	Nonpartisan election	1960–69
	Merit system	1970–2011
Maine	Appointment	1960–2011
Michigan	Nonpartisan election	1960–2011
Minnesota	Nonpartisan election	1960–2011
Missouri	Merit system	1960–2011
Mississippi	Partisan election	1960–93
	Nonpartisan election	1994–2011
Montana	Nonpartisan election	1960–2011
North Carolina	Partisan election	1960–2001
	Nonpartisan election	2002–11
North Dakota	Nonpartisan election	1960–2011
Nebraska	Partisan election	1960–61
	Merit system	1962–2011

(continued)

Table 1. (continued)

State	Selection system	Years used
New Hampshire	Appointment	1960–2011
New Jersey	Appointment	1960–2011
New Mexico	Partisan election	1960–2011
Nevada	Partisan election	1960–75
	Nonpartisan election	1976–2011
New York	Partisan election	1960–76
	Merit system	1977–2011
Ohio	Nonpartisan election	1960–2011
Oklahoma	Partisan election	1960–66
	Merit system	1967–2011
Oregon	Nonpartisan election	1960–2011
Pennsylvania	Partisan election	1960–2011
Rhode Island	Appointment	1960–93
	Merit system	1994–2011
South Carolina	Appointment	1960–2011
South Dakota	Nonpartisan election	1960–79
	Merit system	1980–2011
Tennessee	Partisan election	1960–70
	Merit system	1971–74
	Partisan election	1975–93
	Merit system	1994–2011
Texas	Partisan election	1960–2011
Utah	Nonpartisan election	1960–84
	Merit system	1985–2011
Virginia	Appointment	1960–2011
Vermont	Appointment	1960–73
	Merit system	1974–2011
Washington	Nonpartisan election	1960–2011
Wisconsin	Nonpartisan election	1960–2011
West Virginia	Partisan election	1960–2011
Wyoming	Partisan election	1960–71
	Merit system	1972–2011

baseline category for comparison. Thus, each state selection dummy captures the extent to which the selection method is more (or less) likely to be reviewed compared with a case originating from a Missouri Plan court.

As we are examining Supreme Court auditing of lower courts, we also control for the factors associated with Supreme Court review (Black and Owens 2009a; Caldeira and Wright 1988; Perry 1991). These factors include whether there was a dissenting vote in the state supreme court, whether the state court exercised judicial review,

whether the Solicitor General supported or opposed review, whether the petition involved a civil liberties issue, whether the case was legally salient, whether the petitioner alleged a conflict among the lower court, and whether there was weak, strong, or no conflict among the lower courts. We define our coding scheme in the online supplement.

Supreme Court Reversal of State Court Decisions

Our second model examines whether the U.S. Supreme Court reversed a state supreme court. To construct our dependent variable, *State Court Reversed*, we identified all cases decided by the U.S. Supreme Court between the 1960 and 2011 terms in which the Court reviewed a state supreme court decision.¹³ We coded *State Court Reversed* to reflect whether the U.S. Supreme Court's final disposition in each case reversed the state court. We code *State Court Reversed* as 1 if the Court reversed the state supreme court (on the merits), 0 otherwise.

Our main covariates of interest are the four judicial selection system dummy variables outlined above. We also control for additional features that might lead the Supreme Court to reverse a state court. As we explain more fully in the online supplement, we control for the degree of amicus curiae support for each litigant, the Supreme Court's ideological proclivity to reverse the state supreme court, whether the Solicitor General opposes the state court decision, whether there was dissent in the state court decision under review, the professionalism of the state court under review, and the issue area of the case.¹⁴

Method

To examine whether the Supreme Court reviews or reverses Missouri Plan courts less than other courts, we estimate treatment effects using a matching approach that employs CBPS (Imai and Ratkovic 2014).¹⁵ We employ a matching approach like CBPS because we are essentially asking a “but for” question—but for the presence of the selection method, would the U.S. Supreme Court have reviewed or reversed the state supreme court decision? Matching allows researchers to remove “imbalance” between the treatment and control groups that might otherwise lead to inappropriate inferences (Black and Owens 2012). That is, the matching process adjusts for differences between treatment and control groups so the remaining differences may be attributed to differences in treatment assignment (Boyd, Epstein, and Martin 2010; Epstein et al. 2005).

Our treatment is whether the court decision under review came from a state supreme court that employed one of our selection methods (the Missouri Plan, nonpartisan elections, partisan elections, or appointments). If we observe differences in the probability of review or reversal between the treatment groups (e.g., between a partisan elected court and a Missouri Plan court), we can infer that the selection method is responsible (assuming no omitted variable bias). Because there are more than two possible comparisons, however, the treatment we consider must be multivalued rather

than binary—a feature which CBPS allows.¹⁶ If we ignored this feature and performed matching separately when comparing different pairs of treatment groups, we would be comparing treatment effects for different populations, which would be problematic. It could, for instance, result in intransitivities in which System A has a higher average reversal rate than System B, Selection System B has a higher average reversal rate than System C, and System C has a higher reversal rate than System A.

We employ Imai and Ratkovic's (2014) recent CBPS approach.¹⁷ CBPS attempts to achieve balance on all covariates while giving more weight to those that appear most relevant to treatment assignment by combining the score conditions implied by maximum likelihood and explicit covariance balance conditions.¹⁸ If our estimated propensity-score model, $\hat{\pi}(\cdot)$, is properly specified, then the balance conditions are redundant and our estimated propensity scores and treatment effects will be asymptotically unbiased, as with a more naive approach to propensity-score estimation. Importantly, if the model is misspecified, the explicit inclusion of balance conditions for covariates that are separate from the propensity-score model limits the bias from misspecification. Indeed, this is precisely the situation where CBPS may be most useful. With CBPS, we can leverage a roughly correct propensity-score model without being sensitive to slight misspecification.

As with propensity scores estimated through other methods, we have several choices of treatment-effect estimators to use with these CBPS. We use the inverse propensity-score weighting (IPW) estimator of Hirano, Imbens, and Ridder (2003) to estimate the average treatment effect. In the context of multiple treatments and binary outcomes, IPW computes the risk difference between treatment a and treatment b as

$$\sum_{i:T_i=b} \pi_b(X_i) \sum_{i:T_i=b} \frac{Y_i}{\pi_b(X_i)} - \sum_{i:T_i=a} \pi_a(X_i) \sum_{i:T_i=a} \frac{Y_i}{\pi_a(X_i)}.$$

This estimator has been found to perform well for estimating risk differences (Austin 2010), as we are estimating here.

Many of our case-specific variables (e.g., the number of amici) are found *after* the treatment effect (the creation of the judicial selection method). Accordingly, it is possible that these variables could be affected by or determined by the treatment. Controlling for these posttreatment variables would eliminate any portion of the treatment effect that occurs through the effect of the selection system on these variables and might bias our estimates (Rosenbaum 1984). Thus, we estimate one set of models that match only on *State Court Professionalism* and *Issue Area*, which are determined well before a case reaches a state supreme court.

Other variables, however, may also be capturing pretreatment characteristics of the case or legal environment that might be *correlated* with the selection system but not *caused* by it. For example, *Amici* might reflect the importance of outside interest in the case independent of the state supreme court decision. If so, controlling for variables like this that serve as proxies for pretreatment variables is desirable. So, we estimate a second set of matched models that match on a full battery of available covariates. In our model of review, the full battery of matched covariates includes the following:

Table 2. Estimates of Risk Differences Using Covariate-Balancing Propensity Scores.

	Model 1	Model 2	Model 3	Model 4
Appointment	0.016 [-0.161, 0.193]	0.011 [-0.175, 0.197]	0.106* [0.005, 0.208]	0.078 [-0.025, 0.180]
Nonpartisan	0.115 [-0.025, 0.256]	0.100 [-0.049, 0.249]	0.035 [-0.061, 0.131]	0.012 [-0.084, 0.108]
Partisan	-0.011 [-0.157, 0.135]	0.068 [-0.116, 0.252]	0.029 [-0.048, 0.105]	-0.015 [-0.094, 0.064]
Joint hypothesis test p value	.345	.574	.236	.331
Observations	276	248	964	961
Dependant variable	State Court Reviewed		State Court Reversed	

Note. 95% confidence intervals in brackets. Estimates of risk differences of reversal between each selection system in comparison to merit.

* $p < .05$.

State Court Professionalism, Issue Area, Strong Conflict, Amici, Dissent Below, SG Amicus Support, Legal Salience, and Judicial Review Below. In our model of reversal, the full battery of matched covariates includes the following: *State Court Professionalism, Issue Area, Petitioner Amici, Respondent Amici, Congruence, Lower Court Disagreement, and US Opposes State.*¹⁹

Results

Supreme Court Review of Lower Court Decisions

Is the Supreme Court less likely to review cases decided by merit plan judges than cases decided by other types of judges? The answer is no. Models 1 and 2 in Table 2 analyze the Court's decision to review state court decisions from 1986 to 1993. The data suggest that the High Court treats petitions from Missouri Plan state courts the same as petitions from appointed, partisan, or nonpartisan elected state supreme courts. Across both matched models, the confidence intervals always include 0, and the coefficients on selection methods are nowhere near conventional levels of significance.²⁰

Looking, first, at Model 1—which matches only on *State Court Professionalism* and *Issue Area*—the Court is no more likely to review partisan elected, nonpartisan elected, or appointed state courts than it is Missouri Plan state courts. Model 2—which matches on the full battery of covariates—tells the same story. Once again, there are no differences among selection methods in terms of Supreme Court review. In both cases, a joint hypothesis test is also consistent with no difference in the risk of review between treatment groups.²¹

We present visual estimates of risk difference for review between each pair of selection systems in Figure 1. Here, because we are explicitly making every possible comparison, we adjust our confidence intervals for the number of comparisons using the Tukey–Kramer method to ensure that the probability that *all* six confidence

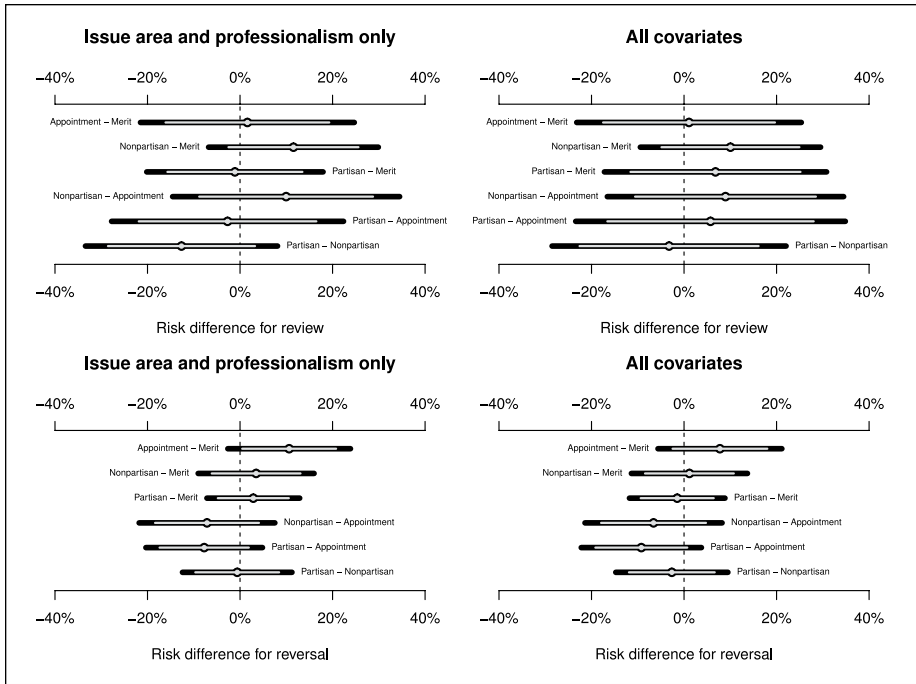


Figure 1. Estimated differences between selection systems in probability of review (top) or reversal (bottom) based on covariate-balancing propensity scores. Note. Longer black bars are 95% family-wise confidence intervals using the Tukey–Kramer method. Shorter gray bars are 95% confidence intervals without a multiple-comparison adjustment. Graphs correspond to Models 1, 2, 3, and 4 of Table 2. Graphs on the left are matched only on issue area and court professionalism. Graphs on the right are matched on all covariates.

intervals exclude the treatment effect is at least 95% (see Bretz, Hothorn, and Westfall 2010). Thus, these confidence intervals are larger than those in Table 2, which show 95% confidence intervals for risk differences from the merit system. (By not making this adjustment in Table 2, we are biasing our results *against* our finding of no difference across systems.) The top portion of Figure 1 highlights this nonsignificance. There is no relationship between High Court review and state judicial selection method. And for those readers who are skeptical of matching, we point out that standard (i.e., nonmatched) probit models (in the online appendix) tell a similar story.

To check the robustness of our findings, we next consider an alternative understanding of the treatment as one that occurs instead at the *state* level. Here, the treatment is the choice of selection system (or selection systems in the case of states that change system over the time of our data) used in a given state. The unobserved potential outcomes are the outcomes that would have occurred *had each state used a different selection system*. Thus, we are implicitly comparing the observed outcomes with those in counterfactual worlds in which states used different selection systems rather

than in counterfactual worlds in which the cases in our data set were decided by different state supreme courts. This formulation allows for the possibility that the choice of selection system affects the rate at which the Supreme Court reviews state supreme court decisions through its influence on the type of cases that arise—for example, the issue areas involved in the cases. It also allows for the possibility that unobserved similarities exist between cases that are decided by the same state supreme court.

We begin with a permutation test at the state level. If the selection system used by each state is unrelated to the rate of review, then the relationship between selection system and review should be no different than if we analyze our data after we randomly permute which states are recorded as using which systems. We can thus test the hypothesis that the selection system each state uses is independent of review by comparing a test statistic from our full data with the distribution of test statistics that result from randomly permuting state labels.²² Performing this test, we find no evidence of a relationship between judicial selection/retention and Supreme Court review ($p = .43$).²³ In short, there appears to be no link between Supreme Court review of state court decisions and state judicial selection method. In terms of avoiding review, all state court methods appear to be equally forward-looking.

Supreme Court Reversal of Lower Court Decisions

We next examine whether the Supreme Court is less likely to reverse decisions by some courts rather than others. Once again, the data say no. As the right half of Table 2 shows, the Supreme Court is no less likely to reverse decisions rendered by Missouri Plan justices than they are justices selected via other methods. Looking, first, at Model 3—which matches only on *State Court Professionalism* and *Issue Area*—the Court is no more likely to reverse elected or appointed states than it is Missouri Plan states. Model 4—which matches on the full battery of covariates—tells the same story. There are no differences among Missouri Plan and other methods in terms of Supreme Court reversals. And, in both cases, a joint hypothesis test once again reveals no difference in risk of reversal between treatment groups.

Furthermore, the estimates of risk difference for reversal between each pair of selection systems in Figure 1 confirm there is no relationship between reversal and state court selection method. The bottom portion of Figure 1 shows this nonsignificance. There is no relationship between High Court reversal and state judicial selection method. The 95% family-wise confidence intervals always contain zero.

What is more, another series of permutation tests once again reveal no evidence of a relationship between selection system and reversal when considering the treatment to be the choice of system used by each state ($p = .105$). Thus, regardless of how broad we wish to make our treatment effects—whether we wish to include effects that occur through factors determined after each case is decided, such as through the number of amicus briefs with the U.S. Supreme Court, or even indirect effects through the types of cases that arise in each state—we find no evidence of an effect of selection system on High Court reversal.²⁴

Of course, one might wonder whether review or reversal are fair measures of forward-looking behavior because the Supreme Court sometimes changes doctrinal

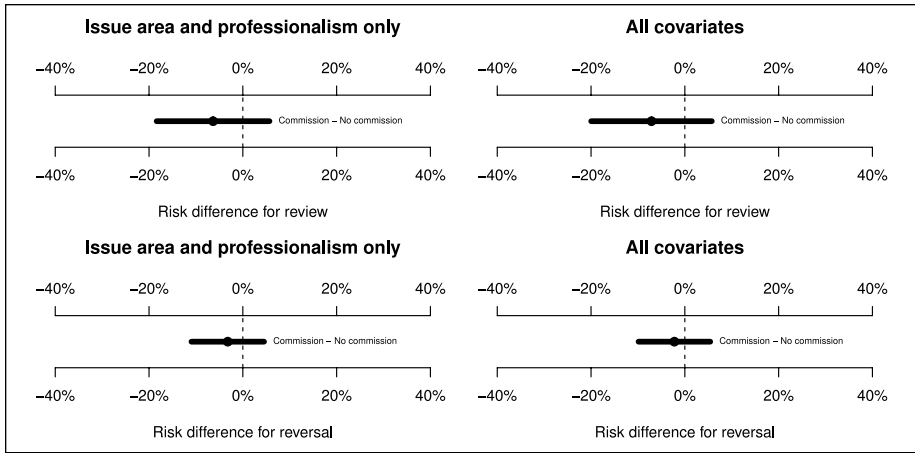


Figure 2. Estimated differences by nominating commission use in probability of review (top) or reversal (bottom) based on covariate-balancing propensity scores. Note. Bars are 95% confidence intervals. Graphs on the left are matched only on issue area and court professionalism. Graphs on the right are matched on all covariates.

course with little warning and might therefore gin up a state’s reversal or review rate unfairly. Although this occurrence is infrequent—and likely not correlated with any selection method—we nevertheless refit our models excluding those cases in which the Supreme Court overruled one of its precedents. We did so because those are likely the kinds of cases that blind side states and could lead us to inappropriate inferences. When we refit our models excluding cases where the Court overruled its own precedent, our results remain unchanged. Neither reversal nor review was related to a state’s selection method.²⁵

The Independent Role of Nominating Commissions

Given the null finding of no difference between Missouri Plan states and states that employ other selection methods, we decided to look deeper. After all, some states treat selection methods like a smorgasbord, picking and choosing aspects of different selection methods that they find most palatable. Some employ nominating commissions but not other aspects of the prototypical Missouri Plan. Perhaps nominating commissions or retention elections *independently* induce more forward-looking behavior. We begin with the role of the nominating commission.

We divide states into two camps: those that use nominating commissions to “select” their justices and those that do not. We examine whether Supreme Court review or reversal is less likely for cases in which state justices were selected by nominating commissions. We find no evidence that selection via nominating commissions affects the probability of review or reversal. As Figure 2 shows, the estimated decrease in probability of review or reversal is not significantly different from zero in any of the

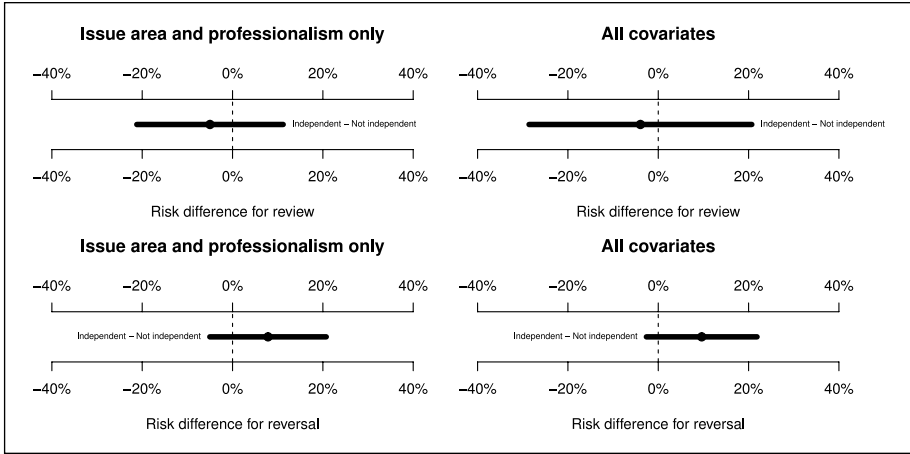


Figure 3. Estimated differences by bar control over its delegation to nominating commission in probability of review (top) or reversal (bottom) based on covariate-balancing propensity scores.
 Note. Bars are 95% confidence intervals. Graphs on the left are matched only on issue area and court professionalism. Graphs on the right are matched on all covariates.

four models that focus on nominating commissions. While the point estimates are all negative, none of the coefficients come close to conventional levels of significance. What is more, considering a state-level treatment as described above, a permutation test also reveals no evidence of an effect ($p = .29$ for review and $p = .26$ for reversal).

Still, not all nominating commissions are alike. In particular, some states grant their bar associations complete independence to pick the lawyer delegations to the nominating commission. In other states, the governor—in conjunction with the bar—selects the lawyers who will serve as the bar’s delegation to the nominating commission (alongside members of the public and political appointees who also serve on the commission). If the claims of commission superiority are correct—that experts are better at picking judges than the public and politicians—the Supreme Court would be less likely to review and reverse decisions by judges selected from such bar-empowered commissions. Once again, however, the data do not support such claims. As Figure 3 shows, the Court is just as likely to review or reverse state court decisions from justices selected by commissions with more bar control. In all four models, the 95% family-wise confidence intervals contain zero. Whether a state bar has total control over its delegation to the nominating commission does not matter. The Supreme Court is just as likely to review and reverse decision by justices selected with strong bar control as those by justices in other states.

Judicial Retention and Independence

We next examine forward-looking behavior and claims of judicial independence. Reformers claim that political accountability is problematic for judges and that they

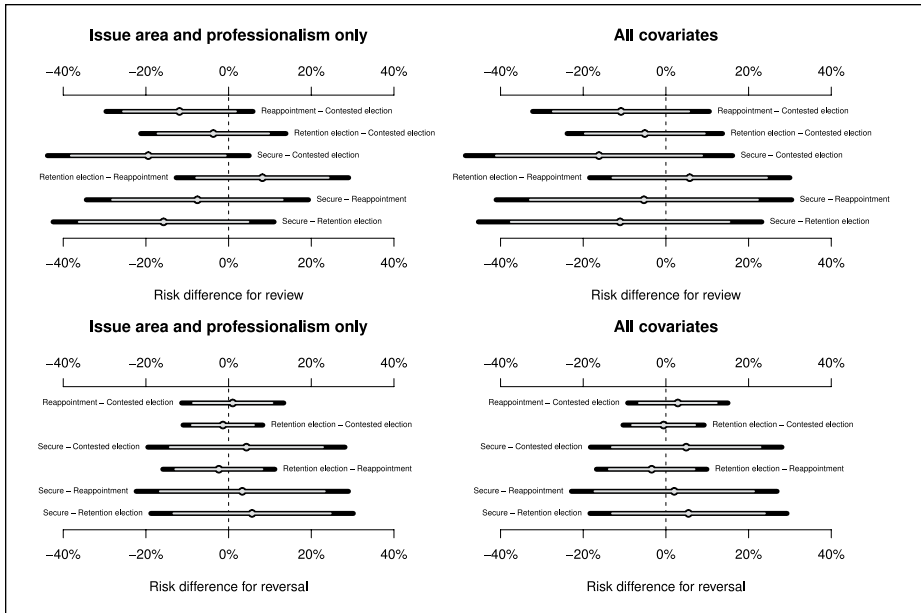


Figure 4. Estimated differences between retention systems in probability of review (top) or reversal (bottom) based on covariate-balancing propensity scores. Note. Longer black bars are 95% family-wise confidence intervals using the Tukey–Kramer method. Shorter gray bars are 95% confidence intervals without a multiple-comparison adjustment. Graphs on the left are matched only on issue area and court professionalism. Graphs on the right are matched on all covariates.

make better decisions when they are not politically accountable. Accordingly, we divide states into four camps depending on how they retain their justices. More specifically, and following the strategy of Goelzhauser and Cann (2014), we divide state courts into four groups: (1) states that use contested elections to retain justices (i.e., nonpartisan or partisan elections); (2) states that use reappointment by a governor, legislature, or commission to retain justices; (3) states whose justices enjoy lifetime tenure; and (4) states that use retention elections to retain their justices.

As Figure 4 shows, we observe no evidence that retention elections lower the probability of review. In addition, we find no evidence that secure lifetime tenure or retention elections lead to a lower probability of Supreme Court reversal. We observe weak evidence (from the state court professionalism and issue area match only model) that cases decided by justices with lifetime tenure are less likely to be reviewed by the Supreme Court than those decided by justices retained by competitive elections, but this statistically significant effect disappears once we adjust for the number of hypotheses being tested. It also disappears when we match on all covariates versus just the professionalism and issue area covariates only. At any rate, even this weak result does not conform to the claims of reform advocates. Finally, a permutation test, as

performed before, reveals no evidence of an effect of the retention system used ($p = .69$ for review and $p = .25$ for reversal).

Unanimous Reversals of State Court Decisions

Thus far, we have found no evidence that the Supreme is less likely to reverse (or review) Missouri Plan courts than other courts. Perhaps, though, not all reversals are alike. Unanimous reversals might better tap into poor forward-looking behavior. As Posner (2000, 716) states, "A lower court decision that the Court reverses unanimously, even after a full briefing and argument, is more likely to be just plain incorrect, rather than merely the reflection of political difference." So, we opted to refit all our reversal models and focus only on instances where the justices issued decisions by a unanimous vote. The results are nearly identical to what we report above.²⁶ In fact, when we focus only on unanimous U.S. Supreme Court decisions, we actually observe that the High Court is more likely to reverse decisions coming from states that empower their bar associations at the commission stage. This result, of course, conflicts with the claims of independent nominating commission supporters. Once again, the data provide no support for the claim that Missouri Plan judges are better able to avoid upper court review and reversal.

In sum, not a single model we examined showed support for the claim that Missouri Plan justices are better able to engage in forward-looking behavior than justices from other selection methods. The Supreme Court, on average, reviews and reverses them just as often as other justices. What is more, even the independent influence of nominating commissions, state bars, and retention elections did not appear to generate more favorable results. And all of these results hold when we limit our examinations to unanimous Supreme Court behavior. Put simply, the data show no differences among judges in terms of forward-looking behavior.

Conclusion

Institutional reform of state judicial selection methods has long been (and continues to be) the subject of great divisiveness. Policy makers routinely debate the merits of different approaches, including principles such as the desired level of accountability to the public, the importance of judicial independence, the role of campaigning, and voter information. Scholars of judicial institutions and state courts have also produced a rather extensive literature seeking to evaluate these institutions. These efforts have often sought to evaluate judges and their selection methods on the basis of their "quality." Yet, defining (and measuring) the notion of quality is often subjective and remains elusive, thus hindering the ability of scholars to foster a significant consensus.

We set out to contribute to this debate by evaluating how different judicial selection methods might produce judges who are more or less forward-looking toward potential review by the U.S. Supreme Court. We contend that forward-looking behavior is a concept with practical consequences for states and judges alike. It is valuable to states because it likely protects judicial legitimacy, can help states retain more control over

their policies, and is efficient. We evaluated particularly whether judges selected under the merit plan system are more forward-looking than other judges. Using an analytically rigorous CBPS (i.e., matching; Imai and Ratkovic 2014), our empirical results consistently suggest that the U.S. Supreme Court does not disproportionately review or reverse the decisions of state judges selected or retained by any particular method. That is, the High Court seems to treat decisions by Missouri Plan judges *the same* as decisions by elected and appointed judges. What is more, neither the use of nominating commissions nor retention elections independently leads to greater forward-looking behavior.

While there is no perfect measure of judicial quality—and we certainly do not think our approach is a measure of quality—we do believe that our approach can at least provide some insight into the debate over quality. As then-Justice Rehnquist (1984) once stated, “the most common reason members of our Court vote to grant certiorari is that they doubt the correctness of the decision of the lower court (p. 1027).” Chief Justice Hughes occasionally voted to grant review to cases “of no public importance whatever simply because the decision below was unjust, unreasonable, or plainly wrong” (McElwain 1949, 13). The same can be said about the decision to reverse a decision. Indeed, nearly all Court watchers use reversal rates to evaluate circuit judges nominated to the Supreme Court because “judicial reversal reflects professional criticism by other professionals” (Cass 1995, 984). Ultimately, even though the notion of quality is subjective, our hope is to have followed the tradition of recent scholars (Bonneau and Hall 2009; Gibson and Caldeira 2011; M. G. Hall 2014b) and focus on the testable.

Although we believe our analysis offers compelling results about differences (or lack thereof) in forward-looking behavior across state judicial selection systems, we should note that our inferences do not necessarily generalize beyond judges serving on state courts of last resort. That is, states select judges to serve on many lower trial and appellate courts that are further removed from the threat of review by the U.S. Supreme Court. Their more pressing considerations lie in the courts above them on the state level; these lower state judges might therefore be less inclined to consider the prospect of eventual review in federal court, making our approach less tractable for studying those courts. Moreover, because state trial courts often are courts of fact, rather than courts of policy making, the considerations that go into their decisions are likely very different than state supreme courts. That being said, most of the debate over judicial selection focuses on state high courts and thus we believe that, despite this limitation, our results still inform this debate.

The data suggest that different judicial selection methods do not lead to different forward-looking behavior. Yet, just like their varying tastes over the role of government in society, states and their citizens obviously have varied preferences for institutional design. These disagreements are born out of the policy makers’ desires to emphasize different attributes that various selection mechanisms foster. Partisan elections facilitate transparency over judicial preferences and hotly contested judicial campaigns can mobilize voters (M. G. Hall and Bonneau 2013), providing substantial benefits to voter engagement (Gibson and Caldeira 2011), but those campaigns may

(Cann and Yates 2008) or may not (Wenzel, Bowler, and Lanoue 2003) harm the legitimacy of the Courts. The Missouri Plan may not only shield judges and justices from undue external influences but may also distance judges from voters and make the courts less democratically responsive. Decisions over policy and institutions are ultimately decisions about values. And on the score, states can be laboratories of experimentation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. In some states, the governor may ask for a second slate of names, and in others, the governor is not bound by the commission's list of names (American Judicature Society 2011). In the "pure" version of the Missouri Plan, the governor must select someone from the commission's list, and the selection does not require legislative confirmation.
2. That study, however, examined only criminal cases from 1995 to 1998, and lumped partisan and nonpartisan states together into "Elected" states. As such, it cannot speak directly to our questions here.
3. Research on the U.S. Solicitor General (SG) before the Supreme Court also suggests that justices' decisions can signal a lack of credibility and legitimacy when the SG promotes an excessively political agenda (Wohlfarth 2009).
4. In a similar vein, Schanzenbach and Tiller (2007) find that judges depart from sentencing guidelines on factual rather than legal grounds to make it costly for reviewing courts to overturn the sentence modification. And while upper court review is not impossible, it becomes more costly and, therefore, less likely.
5. Scholars outside the United States have also used reversal as a measure of quality (Maitra and Smyth 2004). Indeed, Salzberger and Fenn (1999) found that English Court of Appeal judges who were reversed more often by the House of Lords (in its judicial capacity) were seen as less qualified for promotion and, therefore, were less likely to be elevated.
6. Another fruitful avenue of research for future scholarship would be to examine the content of judicial decisions, rather than their outcomes.
7. The cert pool was established in 1972 to minimize the time justices and their clerks had to spend reviewing an ever-increasing stream of petitions for review. Rather than each justice reading the thousands of petitions, one randomly selected clerk prepares a memo summarizing the facts and legal arguments in a case and makes a recommendation that is circulated to all participating chambers. Currently, Justice Alito is the only justice out of the cert pool. His clerks must read every petition filed with the Court.
8. We obtained the discuss lists from Justice Blackmun's files at the Library of Congress. We analyze cases from the discuss list rather than the population of all petitions because they are the cases most likely to receive consideration by the full Court and the least likely to be

merely frivolous claims (Perry 1991, 89). It takes, after all, only one justice to put a petition on the Court's discuss list. The same holds true for nonpaid (i.e., in forma pauperis) petitions, which, again, are mostly frivolous (Black and Owens 2009a).

9. Because the type of data our analysis requires on cases denied by the Supreme Court only exists in the justices' archives, data are not available later than 1993. This time period is admittedly limited given that the entirety of the time period occurs before the Supreme Court's decision in *Republican Party of Minnesota v. White* (2002) reinvigorated the controversy over judicial elections. That being said, while our analysis predates the controversy, it does not predate contested partisan elections in some of the states' high courts. Thus, if the systems did produce differentially forward-looking judges, we believe that would have likewise been true even before the controversy over judicial elections was salient. Moreover, our analysis in the second half of the article, covering reversals of state court decisions, covers nine years after *White* and likewise finds no adverse effect of judicial elections.
10. <http://www.judicialselection.us/>
11. We recognize that in many states, vacancies arise before a judge's term expires, and the governor can appoint a successor. As, however, we are focusing on the systems as a whole, we ignore such instances.
12. Following Nelson, Caufield, and Martin (2013), all subsequent empirical results and substantive inferences are robust to categorizing Ohio and Michigan as partisan election systems as opposed to nonpartisan election states.
13. We collect these data from the Supreme Court Database (available at <http://scdb.wustl.edu/>). Our dependent variable relies on the *partyWinning* variable in the database.
14. By looking at whether the Supreme Court reviews or reverses state supreme court decisions, we necessarily exclude a large number of state court decisions that involve *only* state issues. Our empirical results and inferences regarding the extent of forward-looking behavior are thus limited to those cases that have a potentially relevant federal question. But, we should also point out that even if a state court decided a case solely on a state issue—and ignored federal claims raised by a party—that party still would have the right to appeal to the U.S. Supreme Court, having preserved the federal issue in the record.
15. Following Stuart (2010) and others, we use the term “matching” for any method that attempts to achieve covariate balance between treatment and control groups.
16. There are six different comparison groups: Missouri Plan versus appointment, Missouri Plan versus nonpartisan election, Missouri Plan versus partisan election, appointment versus partisan election, appointment versus nonpartisan election, and partisan election versus nonpartisan election.
17. We use the CBPS package for R (Ratkovic, Imai, and Fong, n.d.) to estimate the covariate-balancing propensity scores used in this article.
18. Specifically, covariate-balancing propensity score (CBPS) estimates propensity scores, π_j , using the population moment condition

$$E \left[\frac{1_{T_i=b} f(X_i)}{\hat{\pi}_b(X_i)} - \frac{1_{T_i=a} f(X_i)}{\hat{\pi}_a(X_i)} \right] = 0 \forall a, b$$

for an appropriate choice of f , with estimation through generalized method of moments, as we will use in this article, or empirical likelihood. If $f(X_i) = \frac{\partial \pi_\beta(X_i)}{\partial \beta}$, this results in the usual maximum-likelihood estimator of β , from which we could derive propensity scores

specified by π_{β} assuming that $\pi_{\beta} \in C^2$, the set of functions with continuous second derivatives. However, if we augment f with covariates or functions of covariates (such as higher order moments), then we are also attempting to achieve balance on these covariates separate from their entry into the propensity-score model.

19. For a discussion of how we measured these covariates, see the online appendix.
20. For information on specific coefficients at the review stage, see the online appendix.
21. A joint hypothesis test allows us to test the hypothesis that there is no difference between any two systems without needing to perform multiple tests or to specify a hypothesis about which two systems are expected to differ. Although we can also perform separate hypothesis tests for each pair of systems, the probability of rejecting the null in at least one of these tests is greater than the nominal level of each test unless we adjust our tests for the number of comparisons. Nonetheless, as none of these separate hypothesis tests are significant even without such an adjustment, it is not surprising that a joint hypothesis test does not reveal a significant difference between systems.
22. We use Pearson's chi-square statistic as a test statistic. However, the distribution of the test statistic will not necessarily follow a chi-square distribution, as it would in Pearson's chi-square test.
23. Even if we were to see an effect of this sort, it is unclear that it would be relevant for understanding the quality of the decisions made by state supreme court justices. If the choice of selection system affects review (or reversal) only through its influence on the type of cases that arise in each state but does not affect the rate of review once these characteristics are determined, it would be difficult to attribute this affect to the quality of a state supreme court decision that was issued after the characteristics of the case had been determined. Thus, we believe this is less relevant for our inquiry than comparisons of decisions in similar cases that are issued by state supreme courts that used different selection systems. Nonetheless, the results still show little evidence of differences across treatment systems even when considering indirect effects of this sort.
24. Some may wonder whether there is a connection between the decision to grant review and reverse a case. We fit a Heckman selection model to examine this dynamic. The ρ statistic does not approach statistical significance. We present these results in the supplementary online appendix.
25. We present these results in the online appendix.
26. Figures on unanimous treatment by the Supreme Court appear in the online appendix.

References

- American Judicature Society. 2011. "Judicial Merit Selection: Current Status." http://www.judicialselection.com/uploads/Documents/Judicial_Merit_Charts_0FC20225EC6C2.pdf (accessed January 2, 2015).
- American Judicature Society. n.d. "Merit Selection: The Best Way to Choose the Best Judges." http://www.judicialselection.us/uploads/documents/ms_descrip_1185462202120.pdf (accessed January 12, 2015).
- Austin, Peter C. 2010. "The Performance of Different Propensity-Score Methods for Estimating Differences in Proportions (Risk Differences or Absolute Risk Reductions) in Observational Studies." *Statistics in Medicine* 29 (20): 2137–48.
- Baye, Michael R., and Joshua D. Wright. 2011. "Is Antitrust Too Complicated for Generalist Judges? The Impact of Economic Complexity and Judicial Training on Appeals." *Journal of Law & Economics* 54 (1): 1–24.

- Benesh, Sara C. 2003. "Understanding Public Confidence in American Courts." *Journal of Politics* 68 (3): 697–707.
- Bierman, Luke. 2002. "Beyond Merit Selection." *Fordham Urban Law Journal* 29 (3): 851–72.
- Black, Ryan C., and Ryan J. Owens. 2009a. "Agenda-Setting in the Supreme Court: The Collision of Policy and Jurisprudence." *Journal of Politics* 71 (3): 1062–75.
- Black, Ryan C., and Ryan J. Owens. 2009b. "Join-3 Votes and Supreme Court Agenda Setting." Unpublished manuscript on file with authors. <http://www.judgesonmerit.org/2009/05/12/justice-oconnor-has-strong-word-for-judicial-elections/>
- Black, Ryan C., and Ryan J. Owens. 2012. *The Solicitor General and the United States Supreme Court: Executive Influence and Judicial Decisions*. New York, NY: Cambridge University Press.
- Bonneau, Chris W., and Melinda Gann Hall. 2009. *In Defense of Judicial Elections*. New York, NY: Routledge.
- Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54 (2): 389–411.
- Bretz, Frank, Torsten Hothorn, and Peter Westfall. 2010. *Multiple Comparisons Using R*. Boca Raton, FL: Chapman & Hall/CRC. <http://books.google.com/books?id=U8Xc9zujgcsC> (accessed January 2, 2015).
- Caldarone, Richard P., Brandice Canes-Wrone, and Tom S. Clark. 2009. "Partisan Labels and Democratic Accountability: An Analysis of State Supreme Court Abortion Decisions." *Journal of Politics* 71 (2): 560–73.
- Caldeira, Gregory A. 1986. "Neither the Purse nor the Sword: Dynamics of Public Confidence in the Supreme Court." *American Political Science Review* 80 (4): 1209–26.
- Caldeira, Gregory A., and John R. Wright. 1988. "Organized Interests and Agenda Setting in the U.S. Supreme Court." *American Political Science Review* 82 (4): 1109–27.
- Canes-Wrone, Brandice, Tom S. Clark, and Jee-Kwang Park. 2012. "Judicial Independence and Retention Elections." *Journal of Law, Economics, & Organization* 28 (2): 211–34.
- Cann, Damon M. 2007. "Justice for Sale? Campaign Contributions and Judicial Decisionmaking." *State Politics & Policy Quarterly* 7 (3): 281–97.
- Cann, Damon M., and Jeff Yates. 2008. "Homegrown Institutional Legitimacy: Assessing Citizens' Diffuse Support for State Courts." *American Politics Research* 36 (2): 297–329.
- Cass, Ronald A. 1995. "Judging: Norms and Incentives of Retrospective Decision-Making." *Boston University Law Review* 75:941–96.
- Caufield, Rachel Paine. 2010. "What Makes Merit Selection Different?" *Roger Williams University Law Review* 15:765–92.
- Chemerinsky, Erwin. 1988. "Evaluating Judicial Candidates." *Southern California Law Review* 61:1985–94.
- Choi, Stephen J., G. Mitu Gulati, and Eric A. Posner. 2010. "Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather than Appointed Judiciary." *Journal of Law, Economics, & Organization* 26 (2): 290–336.
- Clark, Tom S. 2009. "The Separation of Powers, Court-Curbing and Judicial Legitimacy." *American Journal of Political Science* 53 (4): 971–89.
- Comparato, Scott A., and Scott D. McClurg. 2007. "A Neo-institutional Explanation of State Supreme Court Responses in Search and Seizure Cases." *American Politics Research* 35 (5): 726–54.
- Crompton, J. Andrew. 2002. "Pennsylvanians Should Adopt a Merit Selection System for State Appellate Court Judges." *Dickinson Law Review* 106 (4): 755–68.
- Cross, Frank B., and Stefanie A. Lindquist. 2009. "Judging the Judges." *Duke Law Journal* 58:1383–437.

- Dahl, Robert A. 1957. "Decision-Making in a Democracy: The Supreme Court as a National Policy-Maker." *Journal of Public Law* 6 (2): 279–95.
- Epstein, Lee, Daniel E. Ho, Gary King, and Jeffrey A. Segal. 2005. "The Supreme Court during Crisis: How War Affects Only Non-war Cases." *New York University Law Review* 80 (1): 1–116.
- Epstein, Lee, Jeffrey A. Segal, and Harold J. Spaeth. 2007. "Digital Archive of the Papers of Harry A. Blackmun." <http://epstein.wustl.edu/blackmun.php> (accessed January 12, 2015).
- Eskridge, William N., Jr. 1991. "Overriding Supreme Court Statutory Interpretation Decisions." *Yale Law Journal* 101 (2): 331–455.
- Fitzpatrick, Brian T. 2009. "The Politics of Merit Selection." *Missouri Law Review* 74:675–710.
- Gibson, James L. 1989. "Understandings of Justice: Institutional Legitimacy, Procedural Justice, and Political Tolerance." *Law & Society Review* 23 (3): 469–96.
- Gibson, James L. 2012. *Electing Judges: The Surprising Effects of Campaigning on Judicial Legitimacy*. Chicago, IL: University of Chicago Press.
- Gibson, James L., and Gregory A. Caldeira. 2011. "Has Legal Realism Damaged the Legitimacy of the U.S. Supreme Court?" *Law & Society Review* 45 (1): 195–219.
- Glick, Henry R., and Craig Emmert. 1987. "Selection Systems and Judicial Characteristics: The Recruitment of State Supreme Court Justices." *Judicature* 70:228–35.
- Goelzhauser, Greg, and Damon M. Cann. 2014. "Judicial Independence and Opinion Clarity on State Supreme Courts." *State Politics & Policy Quarterly* 14:123–41.
- Hall, Matthew E. K. 2014. "The Semiconstrained Court: Public Opinion, the Separation of Powers, and the U.S. Supreme Court's Fear of Nonimplementation." *American Journal of Political Science* 58 (2): 352–66.
- Hall, Melinda Gann. 2014. *Attacking Judges: How Campaign Advertising Influences State Supreme Court Elections*. Redwood City, CA: Stanford University Press.
- Hall, Melinda Gann, and Chris W. Bonneau. 2013. "Attack Advertising, the White Decision, and Voter Participation in State Supreme Court Elections." *Political Research Quarterly* 66:115–26.
- Hardin, Peter. 2013. "Backers Push for Merit Selection of Minnesota Judges Is Spotlighted." *Gavel Grab*, October 3. <http://www.gavelgrab.org/?p=63884> (accessed January 12, 2015).
- Harrison, Mark L., Keith Swisher, and Meghan H. Grabel. 2007. "On the Validity and Vitality of Arizona's Judicial Merit Selection System: Past, Present, and Future." *Fordham Urban Law Journal* 34 (1): 239–64.
- Helland, Eric, and Alex Tabarrok. 2002. "The Effect of Electoral Institutions on Tort Awards." *American Law and Economics Review* 4:341–70.
- Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.
- Huber, Gregory A., and Stanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind When It Runs for Office?" *American Journal of Political Science* 48 (2): 247–63.
- Iaryczower, Matias, Garrett Lewis, and Matthew Shum. 2013. "The Elect or Appoint? Bias, Information, and Responsiveness of Bureaucrats and Politicians." *Journal of Public Economics* 97:230–44.
- Imai, Kosuke, and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society: Series B* 76:243–63.
- Johnson, Timothy R., and Andrew D. Martin. 1998. "The Public's Conditional Response to Supreme Court Decisions." *American Political Science Review* 92 (2): 299–309.
- Klein, David, and Darby Morrisroe. 1999. "The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals." *The Journal of Legal Studies* 19 (4): 371–91.

- Kosaki, William S. 2003. "The Politics of Judicial Decision-Making in Educational Policy Reform Litigation." *Hastings Law Journal* 55:1077–233.
- Krivosha, Norman. 1990. "In Celebration of the 50th Anniversary of Merit Selection." *Judicature* 74 (3): 128–32.
- Maitra, Pushkar, and Russell Smyth. 2004. "Judicial Independence, Judicial Promotion and the Enforcement of Legislative Wealth Transfers—An Empirical Study of the New Zealand High Court." *European Journal of Law and Economics* 17:209–35.
- McElwain, Edwin. 1949. "The Business of the Supreme Court as Conducted by Chief Justice Hughes." *Harvard Law Review* 63 (1): 5–26.
- McLeod, Aman L. 2007. "A Comparison of the Criminal Appellate Decisions of Appointed State Supreme Courts: Insights, Questions, and Implications for Judicial Independence." *Fordham Urban Law Journal* 34 (1): 343–62.
- Nelson, Michael J., Rachel Paine Caufield, and Andrew D. Martin. 2013. "OH, MI: A Note on Empirical Examinations of Judicial Elections." *State Politics & Policy Quarterly* 13 (4): 495–511.
- O'Connor, Sandra Day. 2009. "The Essentials and Expendables of the Missouri Plan." *Missouri Law Review* 74:479–94.
- Owens, Ryan J., and Patrick C. Wohlfarth. 2014. "State Solicitors General, Appellate Expertise, and State Success before the U.S. Supreme Court." *Law & Society Review* 48 (3): 657–85.
- Perry, H. W., Jr. 1991. *Deciding to Decide: Agenda Setting in the United States Supreme Court*. Cambridge, MA: Harvard University Press.
- Podgers, James. 2009. "O'Connor on Judicial Elections: 'They're Awful. I Hate Them.'" *ABA Journal*, May 9.
- Posner, Richard A. 2000. "Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality." *The Journal of Legal Studies* 29 (2): 711–19.
- Posner, Richard A. 2005. "Judicial Behavior and Performance: An Economic Approach." *Florida State University Law Review* 32:1259–80.
- Pound, Roscoe. 1906. "The Causes of Popular Dissatisfaction with the Administration of Justice." *American Law Review* 40:729–49.
- Ratkovic, Marc, Kosuke Imai, and Christian Fong. N.d. "CBPS: R Package for Covariate Balancing Propensity Score." <http://CRAN.R-project.org/package=CBPS> (accessed January 2, 2015).
- Reddick, Malia. n.d. "Judging the Quality of Judicial Selection Methods: Merit Selection, Elections, and Judicial Discipline." Unpublished study. http://www.judicialselection.us/uploads/documents/Judging_the_Quality_of_Judicial_Sel_8EF0DC3806ED8.pdf
- Rehnquist, William H. 1984. "Oral Advocacy: A Disappearing Act." *Mercer Law Review* 35:1015–28.
- Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable that Has Been Affected by the Treatment." *Journal of the Royal Statistical Society: Series A (General)* 147 (5): 656–66.
- Salzberger, Eli, and Paul Fenn. 1999. "Judicial Independence: Some Evidence from the English Court of Appeal." *Journal of Law & Economics* 42 (3): 831–47.
- Schanzenbach, Max, and Emerson H. Tiller. 2007. "Strategic Judging under the United States Sentencing Guidelines: Positive Political Theory and Evidence." *Journal of Law, Economics, & Organization* 23 (1): 24–56.
- Smith, Joseph L., and Emerson H. Tiller. 2002. "The Strategy of Judging: Evidence from Administrative Law." *The Journal of Legal Studies* 31 (1): 61–82.

- Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25 (1): 1–21.
- Tabarrok, Alexander, and Eric Helland. 1999. "Court Politics: The Political Economy of Tort Awards." *Journal of Law & Economics* 42:157–88.
- Tiller, Emerson H., and Pablo T. Spiller. 1999. "Strategic Instruments: Legal Structure and Political Games in Administrative Law." *Journal of Law, Economics, & Organization* 15 (2): 349–77.
- Tyler, Tom R., and Kenneth Rasinski. 1991. "Procedural Justice, Institutional Legitimacy, and the Acceptance of Unpopular U.S. Supreme Court Decisions: A Reply to Gibson." *Law & Society Review* 25 (3): 621–30.
- Webster, Peter D. 1995. "Selection and Retention of Judges: Is There One 'Best' Method?" *Florida State University Law Review* 23:1–42.
- Wenzel, James P., Shaun Bowler, and David J. Lanoue. 2003. "The Sources of Public Confidence in State Courts: Experience and Institutions." *American Politics Research* 31 (2): 191–211.
- Wohlfarth, Patrick C. 2009. "The Tenth Justice? Consequences of Politicization in the Solicitor General's Office." *Journal of Politics* 70 (1): 224–37.

Author Biographies

Ryan J. Owens is a Lyons Family Faculty Scholar and associate professor of political science at the University of Wisconsin-Madison. He is also an Honorary Fellow in the Institute for Legal Studies.

Alexander Tahk is an Assistant Professor of political science at the University of Wisconsin-Madison.

Patrick C. Wohlfarth is an Assistant Professor of Government in the Department of Government and Politics at the University of Maryland, College Park.

Amanda C. Bryan is an Assistant Professor in the Department of Political Science at Loyola University Chicago.