Article

Classifying Judicial Selection Institutions

Greg Goelzhauser[1]

Abstract
The judicial selection classification problem is widely recognized but poorly understood. In this note, I identify the classification problem's three interrelated sources: ambiguous theoretical arguments, varying decision rules for categorizing merit selection states, and not accounting for interim selections in mixed systems. To demonstrate threats to inference posed by the classification problem, I replicate a study on opinion writing productivity in state supreme courts. I also offer straightforward suggestions for resolving the classification problem. Eliminating the classification problem will help ensure that inferences are comparable across studies with respect to the consequences of institutional design choices concerning state judicial selection mechanisms.

Understanding the consequences of institutional design choices with respect to judicial selection is an important issue in the study of state politics and American politics more generally. Leveraging institutional variation in the states, scholars have examined the relationship between selection institutions and fundamental outcomes of interest such as democratic performance, representation, legitimacy, and judicial quality. In addition to attracting considerable interdisciplinary academic interest, empirical inquiries into the effects of judicial selection institutions help inform ongoing policy debates. Unfortunately, "there is no generally accepted taxonomy of state selection methods" (Fitzpatrick 2017, 1739). This problem is well known, with scholars pointing out that "the literature has not arrived at a consistent methodology" (Choi, Gulati,

---

[1]Utah State University, Logan, UT, USA

Corresponding Author:
Greg Goelzhauser, Department of Political Science, Utah State University, 0725 Old Main Hill, Logan, UT 84322, USA.
Email: greg.goelzhauser@usu.edu

and Posner 2010 [CGP], 297) for classifying judicial selection institutions, and that "scholars rarely define their selection categories" (Epstein, Knight, and Shvetsova 2002, 204).[1] As a result of the classification problem, we are comparing different bundles of states across studies while ostensibly drawing inferences about the same institutional mechanisms. This frustrates efforts to aggregate knowledge and evaluate institutional performance.

In this note, I unravel the judicial selection classification problem. It primarily stems from differences over how to classify "merit selection" states, which necessarily results in spillover effects for other systems. While theoretically informed variation in operationalization should be encouraged, the present coordination failure is mostly driven by unstated assumptions and unexamined reliance on popular sources that do not make decision rules explicit. Furthermore, while less prevalent than differences over how to classify merit selection, ambiguous theoretical arguments and not accounting for interim selections in mixed systems can also be problematic. In addition to being problematic in their own right, these issues can exacerbate the merit selection classification problem. After discussing the classification problem's sources in detail, I demonstrate how it can pose threats to inference by replicating an important study on opinion writing productivity by state supreme court justices. I conclude with recommendations for solving the classification problem.

Sources of the Classification Problem

Ambiguous Theoretical Arguments

As an initial matter, it is often important to clearly specify whether the theoretical argument concerns judicial selection or retention institutions. The words "selection" and "retention" are often used interchangeably. This is appropriate in certain contexts, such as when the exclusive focus is on contestable elections. In this literature, which addresses fundamental questions concerning issues such as turnout, vote choice, advertising, and campaign spending (e.g., Bonneau and Cann 2015; Bonneau and Hall 2009; Hall 2015; Kritzer 2015), contests often pit incumbents against challengers, simultaneously making them selection and retention contests. Moreover, the operative selection and retention institutions are sometimes identical (e.g., initial gubernatorial appointment and subsequent reappointment). In other contexts, however, using both terms can generate confusion or be indicative of an underspecified theoretical argument. For example, as discussed in more detail below, arguments about "merit selection" or the "merit plan" may refer to commission-aided judicial selection or the combination of commission-aided selection and subsequent retention elections.[2] Theory should dictate whether the selection/retention distinction is relevant, but one heuristic is to determine whether the applicable institution could differ for a particular observation.

To help fix ideas, I refer to "selection" institutions as the mechanisms responsible for seating judges to an initial term, while "retention" institutions keep judges on the bench.[3] In the literature on the consequences of institutional design choices concerning state judicial institutions, scholars often develop or test theories about the effects

of selection *or* retention institutions depending on the underlying mechanism of interest. For example, scholars examining the characteristics of judges who join the bench may be interested in whether certain selection institutions sort differentially with respect to qualifications or diversity (e.g., Bratton and Spill 2002; Goelzhauser 2016, n.d.; Hurwitz and Lanier 2003). In contrast, scholars interested in how judges behave on the bench may be interested in examining whether retention mechanisms that vary the trade-off between independence and accountability influence litigation trends or the extent to which judicial decisions are responsive to voter preferences (e.g., Brace and Boyea 2008; Caldarone, Canes-Wrone, and Clark 2009; Hall 1992).

Of course, scholars are not limited to developing theories about selection *or* retention institutions. Many important research questions may have implications for selection *and* retention institutions. There are at least two general classes of potential theoretical arguments in this vein. First, scholars may be *separately* interested in the effect of selection and retention institutions on some outcome of interest. With respect to judicial quality, for example, theory might suggest that selection institutions differentially sort on talent and intelligence (e.g., Posner 2005, 1268), while retention institutions differentially incentivize judges to work hard and effectively (e.g., Goelzhauser 2012). Exemplifying this approach, Owens et al. (2015) examine the effect of state selection and retention methods on the probability of the U.S. Supreme Court granting review and reversing in cases decided by state supreme courts. Second, scholars may be interested in particular *pairwise combinations* of selection and retention institutions. Exemplifying this approach, Savchak and Barghothi (2007) examine judicial decision making in states that use a combination of commission-aided appointment and retention elections on the theory that the relevant constituencies differ.

To summarize, it is often (but not always) important to distinguish between selection and retention institutions. Outside contexts where these terms are understandably used interchangeability, scholars may be interested in the effects of (1) selection institutions, (2) retention institutions, (3) selection and retention institutions separately, or (4) pairwise combinations of selection and retention institutions. Specifying the category of interest is essential for several reasons. First, the set of relevant institutional mechanisms differs across categories, and this differential bundling has important empirical consequences. Second, as discussed in more detail below, any theory emphasizing selection (i.e., 1, 3, or 4) must address the fact that it is often a judge-level rather than state-level phenomenon due to the prevalence of mixed systems for interim appointments.[4] Third, this typology clarifies that the merit selection classification problem discussed below applies when the theoretical interest emphasizes selection institutions (i.e., 1, 3, or 4). Ultimately, it is important to stress that theory should dictate whether the distinction between selection and retention is relevant, and if so whether the underlying interest in any way emphasizes variation in selection institutions.

Merit Selection Differences

Differing schemes for categorizing merit selection is the next—and most important—factor driving the classification problem. Although classifying retention institutions is

mostly straightforward (but see Nelson, Caufield, and Martin 2013), selection categorizations often differ across studies with no explanation (Epstein, Knight, and Shvetsova 2002, 204). When Epstein, Knight, and Shvetsova made this point, it is no surprise that their examples of classification differences involved merit selection. There is no generally agreed upon definition of merit selection, which sets the stage for coordination failure. Moreover, different sources categorize merit selection systems in different ways, and scholars rely on different sources. In some instances, sources are not listed at all.[5] A review of 11 recent political science articles modeling selection systems empirically reveals that five did not list a source, three listed the American Judicature Society (AJS), and two listed print publications.[6]

Most scholars who do not list a source for categorizing merit selection states probably rely on information formerly provided by the AJS, now hosted by the National Center for State Courts. The online information provided by the *Judicial Selection in the States* repository is conveniently organized and easily accessible. Unfortunately, the AJS did not specify its classification rule. Moreover, some have suggested that categorizations by particular groups may be driven by political considerations, with the AJS overclassifying and Federalist Society underclassifying merit selection states (see Fitzpatrick 2017, 1740). Due to the AJS providing numerous documents categorizing selection institutions, and perhaps as a result of undocumented changes over time, different lists citing the AJS may even include different classifications for the same states.

To highlight the scope of classification disagreement, Figure 1 plots state inclusion on 10 different merit selection lists from contemporary and prominent sources. Each source reports having obtained the information in a different way or does not explain how the information was obtained (thus, lists that cited the AJS, for example, were not included beyond the published AJS list).[7] Thirty states appeared on at least one list, and 13 states appeared on every list. The shortest list includes 16 states, while the longest includes 28 states. None of the 10 lists include the same group of states.[8]

What explains this disagreement? The use of inconsistent terminology may be partly responsible for the confusion. The phrase "merit plan" (or "Missouri plan") traditionally denotes the combination of commission-aided gubernatorial appointment at the selection stage with an up or down popular vote (i.e., retention elections) at the retention stage. However, the phrase "merit selection" is sometimes used interchangeably with "merit plan," and sometimes used to refer only to commission-aided appointment without specifying the accompanying retention institution (as I use the phrase here). Relating back to the discussion distinguishing judicial selection and retention, scholars developing or testing theories about selection (but not retention) institutions will be interested in commission-aided appointments but not the accompanying retention mechanism. The 13 states included on every list in Figure 1 are the 13 "merit plan" states, so classification convergence is reassuring but not surprising. Several of the states for which there is disagreement have merit selection but not retention elections (e.g., New York). It is important to emphasize, however, that no list included *only* merit plan states. Thus, while clarifying terminology going forward would be helpful, it is not the sole—and perhaps not even an important—source of the merit selection classification problem.
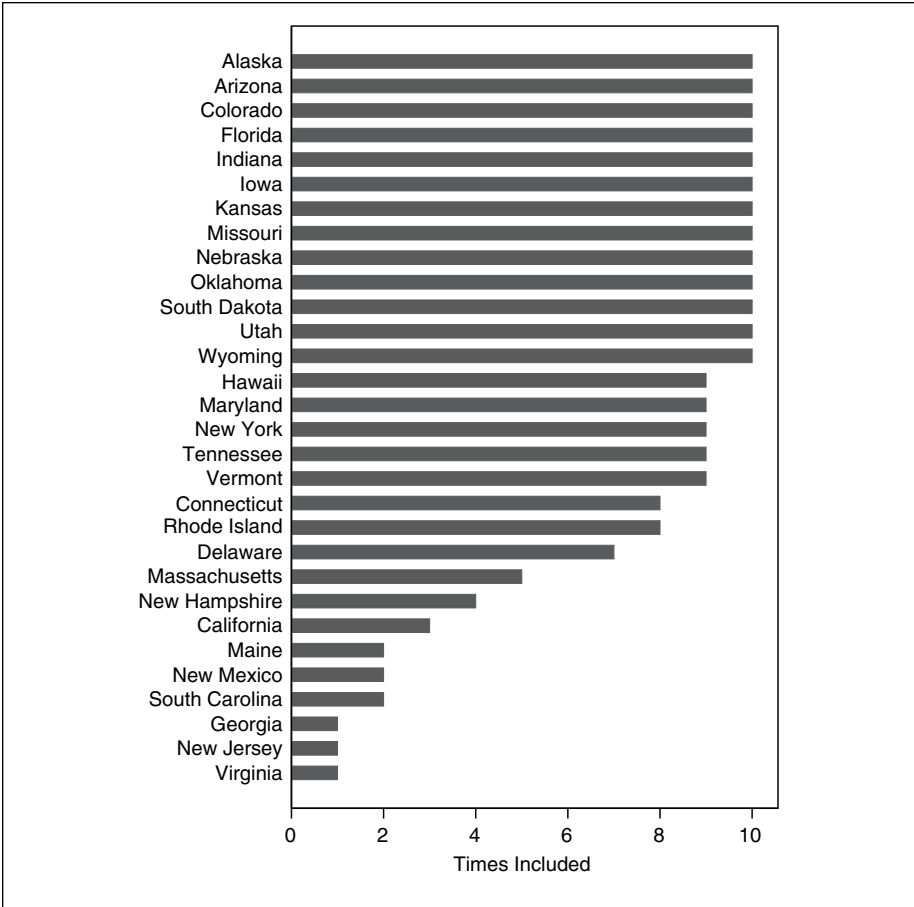
Figure 1. Inclusion on 10 merit selection or merit plan lists.

Categorizing states with "voluntary" merit selection poses a second difficulty. In contrast to states with merit selection implemented by constitution or statute, governors unilaterally devise voluntary systems through executive order (Lowe 1971; Vandenberg 1983). Under voluntary merit selection, governors often select all of the commissioners, who in turn serve at the governor's pleasure, and commission nominations are nonbinding. In general, voluntary systems are less firmly entrenched, and changes can be made from one governor to the next or even during a single administration. In North Carolina, for example, Governor Bev Perdue (D) issued a "temporary modification" (Perdue 2012) to a previously issued executive order implementing merit selection. This modification effectively suspended the voluntary use of merit selection for the purpose of filling a vacancy on the North Carolina Supreme Court before Governor-elect Pat McCrory (R) took office.[9] Governor McCrory rescinded the original executive

order after taking office. In addition to North Carolina, which is rarely included on merit selection lists regardless of time period, voluntary system implementation leads to differences over classifying states such as Georgia and Maryland.

There is no objectively correct solution to the merit selection classification problem. However, one potentially attractive coordinating option is to consider merit selection in place when there is a statutory or constitutional rule dictating that a commission winnow applicants prior to gubernatorial appointment.[10] There are theoretical and prudential reasons for coordinating around this definition, which I refer to here as "formal merit selection," as a default rule. Theoretically, merit selection is designed to be an institutionalized external and ex ante constraint on gubernatorial decision making. Arguments in favor of merit selection often stress that commissions are designed to check governors by prioritizing qualifications over politics, and this is ostensibly realized in part by giving other entities such as bar associations, judges, and legislative leaders some role in commissioner selection. Indeed, this is what differentiates merit selection from unilateral elite appointment. In contrast, voluntary merit selection systems are self-imposed through executive order and typically styled as "advisory," with nonbinding recommendations for office and governors retaining power over commissioner selections (see, for example, Perpich 1989).[11] As a prudential matter, this default rule is also transparent, easily replicable, and straightforward to implement. Furthermore, it avoids potential difficulties associated with outsourcing classification decisions to private entities utilizing nontransparent decision rules with an unclear commitment to updates.

While having a default rule in place would be beneficial for coordination purposes, overrides should be encouraged to promote theory building and hypothesis testing on the consequences of variation in selection institution design. Scholars can also present robustness checks using alternative classification strategies. For example, theory might suggest that merely having a commission in place—regardless of how it was constituted or the extent of its powers—may impact selection outcomes. Or one might examine whether governors are more or less able to make ideologically proximate appointments under different institutional arrangements. The concept of "merit selection" is not inherently one thing or another, and nothing presented here should be taken as a suggestion to impose a particular understanding to the exclusion of others. However, having a theoretically and prudentially sensible default rule in place would promote coordination, which in turn will help researchers and policy makers better understand the consequences of particular institutional arrangements. Overriding the default rule should be encouraged when theory dictates examining the effects of alternative arrangements, with new classifications explicitly justified to facilitate shared understanding and transparent decision making.

Mixed Selection Systems

Not accounting for mixed selection systems is the classification problem's third potential source. Mixed states employ different institutions for interim and full-term selections. In these states, selection institutions are a judge-level rather than court-level or

state-level phenomenon. Interim selections are common. From 1964 through 2004, 52% of state supreme court justices who joined the bench in election states were initially appointed (Holmes and Emrey 2006). In 2016, 45% of state supreme court justices in election states were initially appointed (Berry and Lisk 2017). It is important to emphasize that scholars are already sensitive to the importance of interim selections. In the judicial elections literature, for example, candidates who were initially appointed to the bench are routinely distinguished for various reasons, including expectations concerning relative incumbent performance and the likelihood of attracting challengers (e.g., Hall 2001; Hall and Bonneau 2006; Streb and Frederick 2009). In certain judicial selection contexts, however, interim selections are sometimes ignored in practice. In the review of articles on state judicial selection described previously, seven included theoretical expectations where classifying interim rather than full-term mechanisms could have been relevant, but zero did so explicitly and three noted that their data did not account for interim selections. Sidestepping interim selections can result in threats to inference. If one were examining the relationship between selection institutions and diversity with the Holmes and Emrey (2006) data, for example, not accounting for interim selections would yield a 52% misclassification rate assuming one observation per judge. Moreover, this issue compounds the merit selection classification problem because numerous states have had formal or voluntary merit selection systems only for interim picks.

There are several reasons why the mixed selection system problem persists. As an initial matter, data on interim selections may be unavailable or difficult to collect. Notwithstanding available data from scholars studying election dynamics as described above, gaps remain depending on the court and sample period under consideration. Furthermore, scholars may be interested in selection systems as a whole. The strength of this justification may depend on context. For scholars interested in the relationship between selection (as opposed to retention) institutions on legitimacy, it may be justifiable to examine full-term mechanisms on the understanding that the empirical test is not concerned with how any particular judge was initially seated. For scholars interested in diversity, however, it may be important to code how particular judges were seated to get a sense of comparative institutional performance, and this often requires accounting for interim selections. Fitzpatrick (2017, 1739 n. 30) offers another justification: "Over time, even judges selected differently on an interim basis become folded into the state's full-term method." The difficulty with this proposition is that it potentially conflates selection and retention mechanisms. Regardless of how judges are initially selected, they are certainly folded into the state's retention method over time, but the initial selection method remains fixed.[12]

An Empirical Application

In this section, I demonstrate how the classification problem poses threats to inference in empirical work by replicating an important study on judicial productivity by CGP.[13] CGP evaluate the relationship between institutional design and opinion productivity as a measure of judicial performance, discussing potential selection and retention effects.

Traditional accounts generate different expectations with respect to productivity as a measure of judicial performance. Some suggest that appointers are more likely than electors to emphasize legal talent, perhaps leading to more productivity by appointed judges (emphasizing selection), while less independence with respect to retention may incentivize judges to work harder and publicly signal their predilections, perhaps leading to more productivity by elected judges (emphasizing retention). CGP (297) note that "the literature has not arrived at a consistent methodology" for classifying states and that their "approach differs from others," though they did not specify the basis for their classification decisions.[14] It appears they organized states based on particular combinations of selection and retention institutions (CGP 303), grouping states into four categories: appointment, merit selection, nonpartisan election, and partisan election (CGP 298).

It is important to emphasize that building off CGP's work to illustrate the classification problem's empirical consequences is not meant to be critical—quite the contrary. CGP's innovative article is perhaps the canonical study on state judicial selection mechanisms and institutional performance.[15] Importantly, CGP emphasize the theoretical link between judicial performance and both selection and retention institutions, making it well suited for highlighting the classification problem's depth. Moreover, CGP make their classification decisions with respect to particular states explicit, which is relatively rare and to be applauded. To evaluate the productivity hypotheses, CGP utilize justice-level data from the State Supreme Court Data Project (Brace and Hall 2002). They do not account for interim appointments. The dependent variable is the natural log plus one of the number of opinions (majority and separate) written by a state supreme court justice in a given year. CGP employ ordinary least squares (OLS) to fit the models. In addition to the key explanatory variables, they include a variety of judge-level, state-level, and court-level controls. While CGP (309, n. 13) present numerous robustness checks, I focus on a model they deemed "correctly specified" for simplicity.[16]

To illustrate how the classification problem poses threats to inference, Figure 2 plots estimated OLS coefficients comparing CGP's results with alternative specifications using different judicial selection classification schemes. Abbreviated corresponding tables of results for all models discussed here are presented in the appendix, but coefficient plots allow for quick visual comparisons across all pairwise combinations of selection institutions, some of which would necessarily be obscured in a results table due to omitted baselines. While there are numerous lists of merit selection states in the literature, I selected ones for comparison that (1) clearly distinguished selection and retention institutions and (2) included dates to ensure temporal compatibility.[17] While CGP do not account for interim selections, as is common in the literature, I coded individual justices by their initial selection method for every other list. As a result, the comparison models account for both different classification decisions and interim selections.[18]

The estimated OLS coefficients plotted in Figure 2 depict the effect of switching the omitted baseline from the first-listed to second-listed selection category among pairs on the vertical axis.[19] For example, the first estimate replicates CGP's result for
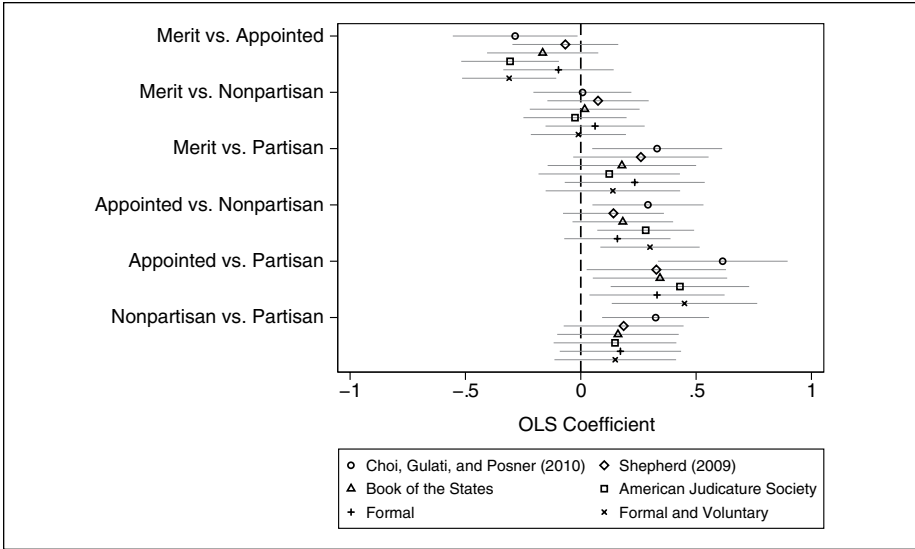
Figure 2. Selection system comparison.

appointment relative to a merit baseline and suggests that appointed judges are less productive on average than judges seated through merit selection. Figure 2 demonstrates that different classification choices and accounting for interim selections can affect substantive conclusions. If one were interested in drawing inferences solely about selection institutions, differences in the classification of merit selection states and accounting for interim selections alter inferences for four of the six pairwise selection system comparisons and otherwise impact effect size estimates.

All models yield results suggesting there are no productivity differences across the merit selection and nonpartisan election categories. Likewise, all models converge on the result that judges selected through partisan election are more productive than appointed judges. However, this similarity masks substantial effect size differences. As the outcome variable is log-transformed, I exponentiate estimated coefficients to provide a clearer sense of substantive impact. CGP's categorization generates the largest effect size, with an estimated productivity increase of 85% moving from appointment to partisan election. At the low end, Shepherd's list generates an estimated 39% increase in productivity.

The other four pairwise comparisons generate conflicting conclusions about productivity. With respect to the comparisons between (1) merit selection and partisan election and (2) nonpartisan election and partisan election, CGP's list is the only one that yields evidence of a significant difference—both in favor of partisan election, with productivity increases of 39% and 38%, respectively. For the comparison between merit selection and appointment, half of the lists produce evidence of a significant productivity difference disfavoring appointment with effect
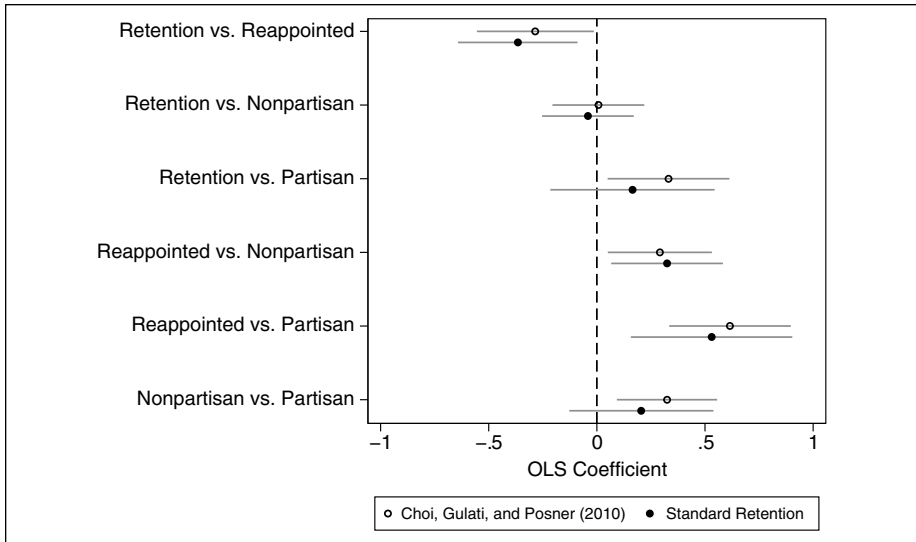
Figure 3. Retention system comparison.

sizes ranging from –27% to –25%, while the other half result in a failure to reject the null hypothesis of no difference. Similarly for the comparison between appointment and nonpartisan election, half of the lists produce evidence of a significant productivity difference favoring nonpartisan election with effect sizes ranging from 33% to 35%, while the other half result in a failure to reject the null hypothesis of no difference.

Figure 3 plots CGP's results with estimates from a model utilizing a standard coding of retention institutions.[20] Although CGP include states with life tenure in the appointment category (as is common in the literature), the standard retention model here includes a separate indicator for these states to maintain consistency with the emphasis on formal rules developed earlier (though the results are not included in Figure 3 for ease of comparison). While there is understandably more model convergence here than with respect to a focus on selection institutions, given that retention institutions are easier to categorize and interim selections are irrelevant, the results nonetheless emphasize the importance of theoretically clarifying which institutional arrangements are the subject of intended inference. Both classifications yield a failure to reject the null hypothesis of no productivity difference in four of six pairwise comparisons. But only CGP's classification yields evidence that partisan election outperforms retention and nonpartisan election, by 39% and 38%, respectively.

While CGP's results are often cited for propositions concerning selection and retention institutions, their classification groups states at least in part by pairwise combinations of selection and retention institutions. Figure 4 plots CGP's results with estimates from models combining the various selection schemes introduced
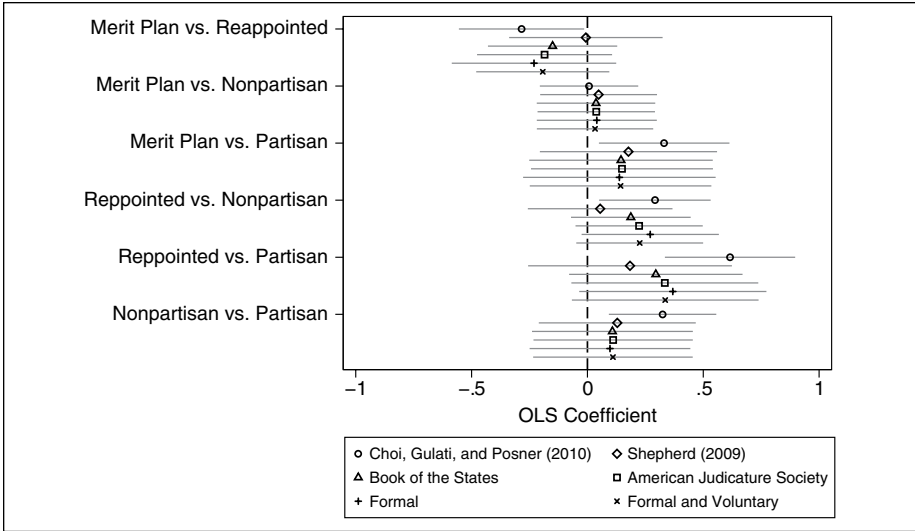
Figure 4. Comparison of combinations of selection and retention institutions.
Note. OLS = ordinary least squares.

earlier with the standard retention model just discussed. Recall that CGP group states into four categories. Their four-category grouping works well where the interest lies solely in selection institutions, and with the exception of a few life tenure states only slightly less well where the interest is solely in retention institutions. But due to the frequency of interim selections, which CGP do not account for, modeling particular combinations of selection and retention mechanisms is a more unruly endeavor. Examples of additional comparisons needed to exhaust the observed data utilized by CGP include merit selection and nonpartisan election (e.g., Idaho), appointment and retention election (e.g., California), appointment and nonpartisan election (e.g., Minnesota), and appointment and partisan election (e.g., Alabama). The complexity of studying pairwise combinations of selection and retention institutions induced by differences between full-term and interim selections has not been widely recognized in the literature. And note that the problem is compounded if there is disagreement about what constitutes merit selection with respect to interim appointments (e.g., Georgia with use of voluntary merit selection for interim vacancies) and if interest lies in merit selection as a retention institution (e.g., New York).

For ease of comparison with CGP's results, Figure 4 plots coefficients from models that include a single indicator for pairwise combinations of selection and retention mechanisms other than those of direct interest for CGP. Although the point estimates are similarly signed, the confidence intervals tend to be larger in the non-CGP models because there are fewer observations in the primary categories. As a

result, while CGP's model generates conclusions about meaningful differences with respect to five of six pairwise comparisons, with effect sizes ranging from –25% (appointment underperforming merit selection) to 85% (partisan election outperforming appointment), no other model generates a meaningful difference with respect to any pairwise comparison.

To summarize the potential substantive threats to inference, consider a seemingly simple question: how does merit selection compare to other systems with respect to judicial performance as measured by opinion productivity? The evidence presented here suggests that the answer depends on a number of factors even when holding the sample constant: does "merit selection" mean commission-aided gubernatorial appointment or the combination of commission-aided appointment and retention elections? Which states have merit selection under either definition—just those with constitutional or statutory implementing rules, those plus voluntary plans, or some combination? And is selection system coded as a state-level or judge-level variable? If the concept of interest with respect to "merit selection" is solely commission-aided gubernatorial appointment, it may underperform appointment by up to 27% or there may be no difference depending on the classification rule (with a significant difference in three of six models). Likewise, partisan election may outperform merit selection by up to 39% or there may be no difference depending on the classification rule (with a significant difference in three of six models).

If the concept of interest with respect to merit selection is the combination of commission-aided gubernatorial appointment and retention elections, and the comparison is with other pairwise combinations of selection and retention mechanisms, it may outperform appointment–reappointment by up to 25% or there may be no difference (with a significant difference in one of six models). Partisan election–partisan election, meanwhile, may outperform merit selection by up to 85% or there may be no difference between the two (with a significant difference in one of six models). In short, the normative and policy implications are unclear at best and a simple question about comparative institutional performance is surprisingly difficult to answer—even with a common sample—absent coordination around a reasonable default rule. Moreover, because decision rules for classifying merit selection states necessarily impact how other states are classified, the inferential problem can reach almost any selection institution or combination of selection and retention institutions.

Conclusion

The judicial selection classification problem is well known but poorly understood. As a result of the classification problem's persistence, studies of institutional performance often inadvertently invite inferences about institutional arrangements with a common name but different associated state groupings. Classification decisions that may be justifiable within the confines of a single study result in mismatches when compared with other studies relying on different assumptions or sources despite interest in the same underlying institutional mechanisms. Moreover, decisions regarding which of

several plausible classification rules to adopt can dictate resulting inferences about the impact of various selection institutions. Following a few best practices can help alleviate the classification problem:

1. Scholars should determine whether the distinction between selection and retention is relevant to their study, and if so clearly specify whether the theoretical argument emphasizes (1) selection institutions, (2) retention institutions, (3) selection and retention institutions separately, or (4) particular pairwise combinations of selection and retention institutions. If the distinction is not relevant (as in many studies concerning the dynamics of judicial elections) or the theoretical interest lies solely in (2), the classification problem probably does not apply.

2. If the theoretical interest lies at least in part with selection institutions (i.e., 1, 3, or 4 above), the next step is to explicitly state the decision rule for categorizing merit selection states. One attractive option is to adopt a formal definition of merit selection to include states with statutory or constitutional requirements to have a nominating commission winnow a pool of applicants prior to gubernatorial appointment. This potential default rule is supported by theoretical and prudential considerations. While coordinating around a default rule will help solve the classification problem, theoretically motivated experimentation should be encouraged as a way to better understand institutional performance.

3. If the theoretical interest lies at least in part with selection institutions (i.e., 1, 3, or 4 above), the next step is to account for interim selections if necessary. Accounting for interim selections is important in its own right given their prevalence, but failing to do so can also exacerbate the merit selection classification problem where such a system is used for interim selections. While it is straightforward to account for interim selections with individual-level data, creating a separate category for mixed systems is one way to manage the problem with aggregated data (e.g., court level or state level).

Increasing coordination will bring more transparency to the literature and broader judicial selection debate by ensuring that similar institutional arrangements are compared across studies. It is important to emphasize, however, that the ultimate goal is to make motivating assumptions explicit rather than to enforce uniformity for its own sake. There is no objectively correct way to deal with the classification problem, and theoretically informed experimentation should be encouraged. To the extent coordination can be improved, however, the literature and broader debate will benefit from coordination.

Appendix

This appendix provides table of results underlying the plots in Figures 2, 3, and 4.

Table A1. Selection Institutions.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Appointed	−0.284*	−0.067	−0.165	−0.307**	−0.097	−0.310**
	(0.161)	(0.137)	(0.144)	(0.126)	(0.143)	(0.122)
Nonpartisan	0.007	0.075	0.017	−0.025	0.062	−0.010
	(0.127)	(0.131)	(0.142)	(0.133)	(0.128)	(0.123)
Partisan	0.331*	0.261	0.178	0.123	0.234	0.139
	(0.168)	(0.175)	(0.192)	(0.183)	(0.181)	(0.174)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	998	998	998	998	998	998
AIC	1,844.229	1,863.708	1,857.927	1,839.951	1,862.756	1,839.321
BIC	2,030.647	2,050.126	2,044.346	2,026.369	2,049.174	2,025.739

Note. The dependent variable is the natural log plus one of the number of opinions written by a state supreme court justice in a given year. Standard errors clustered by state in parentheses. Model 1 replicates CGP (2010); Model 2 uses Shepherd's (2009) classification of merit selection; Model 3 uses the Council of State Governments (1996) classification; Model 4 uses the American Judicature Society's classification; Model 5 uses the formal classification (Goelzhauser 2016); and Model 6 uses formal plus known voluntary plans (Goelzhauser 2016). Merit selection is the excluded baseline.
*$p < .10$. **$p < .05$.

Table A2. Retention Institutions.

	Model 1	Model 2
Reappointed	−0.284*	−0.366**
	(0.161)	(0.165)
Nonpartisan	0.007	−0.041
	(0.127)	(0.126)
Partisan	0.331*	0.165
	(0.168)	(0.227)
No retention		−0.508
	-	(0.341)
Controls	Yes	Yes
Observations	998	998
AIC	1,844.229	1,863.429
BIC	2,030.647	2,054.754

Note. The dependent variable is the natural log plus one of the number of opinions written by a state supreme court justice in a given year. Standard errors clustered by state in parentheses. Model 1 replicates CGP (2010); Model 2 uses standard retention classifications including a no retention alternative not included by CGP. Retention elections is the excluded baseline.
*$p < .10$. **$p < .05$.

Table A3. Combinations of Selection and Retention Institutions.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Appointed–reappointed	−0.284*	−0.007	−0.150	−0.185	−0.231	−0.193
	(0.161)	(0.198)	(0.166)	(0.174)	(0.212)	(0.171)
Nonpartisan–nonpartisan	0.007	0.048	0.037	0.038	0.041	0.033
	(0.127)	(0.150)	(0.153)	(0.151)	(0.155)	(0.150)
Partisan–partisan	0.331*	0.177	0.145	0.149	0.138	0.143
	(0.168)	(0.228)	(0.236)	(0.234)	(0.248)	(0.234)
Other combinations	-	−0.017	−0.022	−0.024	−0.018	−0.033
		(0.128)	(0.129)	(0.128)	(0.151)	(0.124)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	998	998	998	998	998	998
AIC	1,844.229	1,882.671	1,880.154	1,879.305	1,877.215	1,879.156
BIC	2,030.647	2,073.995	2,071.478	2,070.629	2,068.539	2,070.480

Note. The dependent variable is the natural log plus one of the number of opinions written by a state supreme court justice in a given year. Standard errors clustered by state in parentheses. Model 1 replicates CGP (2010); Model 2 uses Shepherd's (2009) classification of merit selection; Model 3 uses the Council of State Governments (1996) classification; Model 4 uses the American Judicature Society's classification; Model 5 uses the formal classification (Goelzhauser 2016); and Model 6 uses formal plus known voluntary plans (Goelzhauser 2016). The merit plan is the excluded baseline.
*p < .10.

Notes

1. Similarly, Hurwitz and Lanier (2001, 87 n. 20) explain, "Categorizing the variants of selection systems necessarily entails some subjectivity; hence, our classifications may differ slightly from those of other studies."
2. Studies that incorporate commission-aided selection or retention elections demonstrate why it is sometimes important to distinguish selection and retention institutions. Not doing so can be confusing because, for example, the set of selection institutions would not include retention elections. For an example of using "selection" and "retention" interchangeably when the exclusive focus was on retention institutions, see Goelzhauser and Cann (2014, 130).

3. A similar point about distinguishing these terms has been made in the context of debates concerning institutional design and the trade-off between independence and accountability. This debate is more relevant for retention as opposed to selection institutions (e.g., Ware 2008, 406–408).

4. This may be true in certain instances for retention as well. In New Mexico, interim appointments must win a full term through partisan election after which subsequent terms are secured through retention election. Thus, at any given time for the New Mexico Supreme Court, for example, justices may be facing either partisan or retention elections.

5. In previous work, for example, I failed to provide any information about how the presence of a "nominating commission" was coded (Goelzhauser 2011, 768). I used information provided by the American Judicature Society (AJS).

6. I reviewed every article published in *State Politics & Policy Quarterly*, every article published over a 10-year span in the top-3 general interest journals (i.e., *American Political Science Review, American Journal of Political Science*, and *The Journal of Politics*), and articles from two law and courts journals: *The Journal of Law & Courts* (since first printing in 2003) and *Justice System Journal* (2007–16). This search yielded 73 articles on state courts, 11 of which dealt with judicial selection as that phrase is used here.

7. Many published lists simply cite the AJS as a source. To avoid merely duplicating the AJS list, I searched for lists generated by prominent organizations or published by scholars until reaching 10 total that either did not include a source or listed a source that differed from a previously collected list. The sources include lists by the AJS, now hosted by the National Center for State Courts (http://www.judicialselection.com/judicial_selection/methods/selection_of_judges.cfm?state=), Ballotpedia (https://ballotpedia.org/Judicial_selection_in_the_states), and the Federalist Society (http://www.statecourtsguide.com/). Lists from scholarly publications include Choi, Gulati, and Posner (2010); Garoupa and Ginsburg (2015); George and Yoon (n.d.); Goelzhauser (2016); Hanssen (2004); Newman and Isaacs (2004); and Shepherd (2009).

8. In the supplemental review of political science articles described in note 6, 5 of the 10 with an expressed interest in some variant of merit selection provided lists of states with counts ranging from 15 to 24. Of the five lists, four were distinct. Moreover, two lists citing the AJS as a source differed on how to classify three states.

9. The modifying order clarified that the commission was not an actual constraint, with Governor Perdue (2012) stating, "I will exercise my constitutional and statutory authority to fill the vacancies without first receiving nominations from the Commission." However, the order adds, "I plan to consult with individual members of the Commission as I deem appropriate to seek their input regarding the qualifications of specific candidates for appointment" (Perdue 2012). The modifying order closes by "urg[ing] future Governors to continue utilizing the Commission to assist them in filling judicial vacancies" (Perdue 2012). Whether merit selection was ever in place in North Carolina and whether it remained in place with the modifying order are difficult conceptual questions. Most scholars do not seem to consider merit selection to have ever been implemented in North Carolina. The *Judicial Selection in the States* resource lists the 2011 executive order as a formal change to interim selection methods but does not mention the order's subsequent modification or abolition. North Carolina's summary page does list merit selection as an interim selection method for superior court vacancies.

10. Some might argue that the AJS information is better known and more commonly used, thus making for a more natural default rule. While the AJS has ubiquity on its side, arguments against such an adoption include the fact that the decision rule utilized for classification decisions is unclear and thus difficult to evaluate. Moreover, this lack of a clear decision

rule complicates further application and extension absent updating by those who oversee the project. Utilizing the default rule proposed here circumvents these difficulties. If the AJS measure is preferred for some reason, however, coordination will nonetheless be improved if decisions are transparent.

11. Of course, the specifics of various formal and voluntary merit selection systems differ. However, no formal system has the combination of features that comprise the typical voluntary system. With respect to commissioner selection authority, Utah's governor makes each appointment but must select two individuals from a list provided by the state bar association.

12. As with merit selection groupings, it may be beneficial to experiment with how interim selections are treated in empirical studies. For example, governors may face distinct pressures when making interim versus full-term appointments, and interim appointments may be less salient. Given that these possibilities have not received much attention in the literature, I do not address them further here. I thank an anonymous reviewer for raising this issue.

13. CGP also evaluate independence and quality. Although the issues discussed here apply to each of these areas, I focus on productivity as an illustrative case because the goal of this note is to highlight and resolve the classification problem rather than replicate CGP for substantive purposes.

14. As indicated previously, I too have failed to specify coding rules or a source for classifying judicial selection institutions (Goelzhauser 2011). The information came from the AJS.

15. While there is a larger literature in political science on the link between retention mechanisms and performance (e.g., Brace and Hall 1997; Canes-Wrone, Clark, and Park 2012; Hall 2014), relatively little work has been done on the link between selection institutions and performance.

16. The model replicated here is CGP's Model 4 in Table 5 (CGP, 310).

17. The Shepherd (2009) classifications were for an empirical project that also used data from the State Supreme Court Data Project, though the actual empirical test focuses on retention institutions. Information from the *Book of the States* comes from its 1996 edition for compatibility with CGP's sample period. Lists of formal and voluntary merit selection states are from Goelzhauser (2016).

18. This approach means that we do not independently observe inferential differences due to selection systems holding the treatment of interim selections constant or interim selections holding classification schemes constant. Rather, the comparative results displayed here primarily demonstrate how different classification schemes can generate different results. That accounting for interim selections has important inferential implications is more obvious in light of their ubiquity. More generally, as noted previously, differences over how to categorize merit selection systems is the predominant and least well understood factor driving the classification problem.

19. The lines represent 90% confidence intervals to maintain consistency with CGP's discussion of significance at the $p < .10$ threshold.

20. In the standard model, Michigan and Ohio are coded as nonpartisan election states.

References

Berry, Kate, and Cathleen Lisk. 2017. *Appointed and Advantaged: How Interim Vacancies Shape State Courts*. https://www.brennancenter.org/sites/default/files/analysis/Appointed_and_Advantaged_How_Interim_Appointments_Shape_State_Courts.pdf (accessed March 6, 2016).

Bonneau, Chris W., and Damon M. Cann. 2015. *Voters' Verdicts: Citizens, Campaigns, and Institutions in State Supreme Court Elections*. Charlottesville: University of Virginia Press.

Bonneau, Chris W., and Melinda Gann Hall. 2009. *In Defense of Judicial Elections*. New York: Routledge.

Brace, Paul, and Brent D. Boyea. 2008. "State Public Opinion, the Death Penalty, and the Practice of Electing Judges." *American Journal of Political Science* 52 (2): 360–72.

Brace, Paul, and Melinda Gann Hall. 1997. "The Interplay of Preferences, Case Facts, Context, and Rules in the Politics of Judicial Choice." *Journal of Politics* 59 (4): 1206–31.

Brace, Paul, and Melinda Gann Hall. 2002. *State Supreme Court Data Project*. http://hdl.handle.net/1902.1/10712 (accessed March 6, 2016).

Bratton, Kathleen 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): 504–18.

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.

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.

Council of State Governments. 1996. *The Book of the States1996-97 Edition*. Lexington, VA: Council of State Governments.

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* 26:290–336.

Epstein, Lee, Jack Knight, and Olga Shvetsova. 2002. "Selecting Selection Systems." In *Judicial Independence at the Crossroads: An Interdisciplinary Approach*, eds. Stephen B. Burbank, and Barry Friedman. Thousand Oaks: Sage, 191–226.

Fitzpatrick, Brian T. 2017. "The Ideological Consequences of Selection: A Nationwide Study of the Methods of Selecting Judges." *Vanderbilt Law Review* 70 (6): 1729–54.

Garoupa, Nuno, and Tom Ginsburg. 2015. *Judicial Reputation: A Comparative Theory*. Chicago: Chicago University Press.

George, Tracey E., and Albert H. Yoon. n.d. "The Gavel Gap: Who Sits in Judgment on State Courts?" *American Constitution Society*. http://gavelgap.org/pdf/gavel-gap-report.pdf

Goelzhauser, Greg. 2011. "Diversifying State Supreme Courts." *Law & Society Review* 45 (3): 761–81.

Goelzhauser, Greg. 2012. "Accountability and Judicial Performance: Evidence from Case Dispositions." *Justice System Journal* 33 (3): 249–61.

Goelzhauser, Greg. 2016. *Choosing State Supreme Court Justices: Merit Selection and the Consequences of Institutional Reform*. Philadelphia: Temple University Press.

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

Goelzhauser, Greg, and Damon M. Cann. 2014. "Judicial Independence and Opinion Clarity on State Supreme Courts." *State Politics & Policy Quarterly* 14 (2): 123–41.

Hall, Melinda Gann. 1992. "Electoral Politics and Strategic Voting in State Supreme Courts." *Journal of Politics* 54 (2): 427–46.

Hall, Melinda Gann. 2001. "State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform." *American Political Science Review* 95 (2): 315–30.

Hall, Melinda Gann. 2014. "Representation in State Supreme Courts: Evidence from the Terminal Term." *Political Research Quarterly* 67 (2): 335–346.

Hall, Melinda Gann. 2015. *Attacking Judges: How Campaign Advertising Influences State Supreme Court Elections*. Stanford: Stanford University Press.

Hall, Melinda Gann, and Chris W. Bonneau. 2006. "Does Quality Matter? Challengers in State Supreme Court Elections." *American Journal of Political Science* 50 (1): 20–33.

Hanssen, F. Andrew. 2004. "Is There a Politically Optimal Level of Judicial Independence?" *American Economic Review* 94 (3): 712–29.

Holmes, Lisa M., and Jolly A. Emrey. 2006. "Court Diversification: Staffing the State Courts of Last Resort through Interim Appointments." *Justice System Journal* 27 (1): 1–13.

Hurwitz, Mark S., and Drew Noble Lanier. 2001. "Women and Minorities on State and Federal Appellate Benches." *Judicature* 85 (2): 84–92.

Hurwitz, Mark S., and Drew Noble Lanier. 2003. "Explaining Judicial Diversity: The Differential Ability of Women and Minorities to Attain Seats on State Supreme and Appellate Courts." *State Politics & Policy Quarterly* 3 (4): 329–52.

Kritzer, Herbert M. 2015. *Justices on the Ballot: Continuity and Change in State Supreme Court Elections*. New York: Cambridge University Press.

Lowe, Stanley R. 1971. "Voluntary Merit Selection." *Judicature* 55:161–68.

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.

Newman, Sandra Schultz, and Daniel Mark Isaacs. 2004. "Historical Overview of the Judicial Selection Process in the United States: Is the Electoral System in Pennsylvania Unjustified?" *Villanova Law Review* 49 (1): 1–54.

Owens, Ryan J., Alexander Tahk, Patrick C. Wohlfarth, and Amanda C. Bryan. 2015. "Nominating Commissions, Judicial Retention, Forward-Looking Behavior on State Supreme Courts: An Empirical Examination of Selection and Retention Methods." *State Politics & Policy Quarterly* 15 (2): 211–38.

Perdue, Beverly Eaves. 2012. "Temporarily Modifying the Judicial Selection Process Established in Executive Order No. 86.*" Executive Order No*. 137.

Perpich, Rudy. 1989. "Providing for Continuation of the Governor's Minnesota Judicial Merit Advisory Commission.*" Executive Order No*. 89-4.

Posner, Richard. 2005. "Judicial Behavior and Performance: An Economic Approach." *Florida State University Law Review* 32 (4): 1259–80.

Savchak, Elisha Carol, and A. J. Barghothi. 2007. "The Influence of Appointment and Retention Constituencies: Testing Strategies of Judicial Decisionmaking." *State Politics & Policy Quarterly* 7 (4): 394–415.

Shepherd, Joanna M. 2009. "The Influence of Retention Politics on Judges' Voting." *Journal of Legal Studies* 38 (1): 169–206.

Streb, Matthew J., and Brian Frederick. 2009. "Conditions for Competition in Low-Information Judicial Elections: The Case of Intermediate Appellate Court Elections." *Political Research Quarterly* 62 (3): 523–37.

Vandenberg, Donna. 1983. "Voluntary Merit Selection: Its History and Current Status." *Judicature* 66:265–78.

Ware, Stephen J. 2008. "Selection to the Kansas Supreme Court." *Kansas Journal of Law and Public Policy* 17 (3): 386–423.

Author Biography

Greg Goelzhauser is an Associate Professor in the Department of Political Science at Utah State University. He studies American political institutions with an emphasis on law and courts.