Measuring Supreme Court Case Complexity

Greg Goelzhauser
Utah State University

Benjamin J. Kassow*
University of North Dakota

Douglas Rice
University of Massachusetts

Case complexity is central to the study of judicial politics. The dominant measures of Supreme Court case complexity use information on legal issues and provisions observed postdecision. As a result, scholars using these measures to study merits stage outcomes such as bargaining, voting, separate opinion production, and opinion content introduce posttreatment bias and exacerbate endogeneity concerns. Furthermore, existing issue measures are not valid proxies for complexity. Leveraging information on issues and provisions extracted from merits briefs, we develop a new latent measure of Supreme Court case complexity. This measure maps with the prevailing understanding of the underlying concept while mitigating inferential threats that hamper empirical evaluations. Our brief-based measurement strategy is generalizable to other contexts where it is important to generate exogenous and pretreatment indicators for use in explaining merits decisions. (*JEL* K00, K40)

1. Introduction

Case complexity is a fundamental concept in empirical legal studies. Scholars are often interested in understanding how a complicated information environment shapes judicial behavior. Although measurement varies across contexts, scholars studying the US Supreme Court typically capture case complexity with the number of legal issues and provisions in a case—both of which are coded postdecision as part of the Supreme Court Database (SCDB). This approach is problematic. As an initial matter, the SCDB's issue variables are not valid complexity measures because

* Department of Political Science and Public Administration, University of North Dakota, Grand Forks, ND, USA benjamin.kassow@und.edu

Thanks to Justin Gross, Chuck Smith, Jeff Staton, and Nicole Vouvalis for helpful discussions. Thanks to Lisa Holmes, Chris Krewson, Kevin Quinn, Eric Waltenburg, and the anonymous reviewers for helpful feedback.

The Journal of Law, Economics, and Organization, Vol. 38, No. 1
doi:10.1093/jleo/ewaa027
Advance Access Publication January 12, 2021

the coding protocols generally impose a single-issue constraint (Shapiro 2009). Moreover, using postdecision measures to explain prior conduct is methodologically troublesome. As Clark et al. (2015: 38) note in an analogous context, concepts "measured only after many of the outcomes of interest flips the causal chronology, rendering disquieting many claims of causal inference." In particular, reversing the causal order introduces posttreatment bias and heightens endogeneity concerns.

We leverage merits briefs to devise a new latent measure of case complexity. In particular, we exploit a 1954 Supreme Court rule change requiring petitioners, and permitting respondents, to highlight legal issues and relevant provisions in merits briefs. Using issue and provision counts extracted from briefs, we develop a Bayesian measurement model to construct a latent measure of case complexity. Our measure is particularly useful for studying merits stage outcomes such as oral argument behavior, opinion assignment, bargaining, voting, separate opinion production, and opinion content. The measure can also be used to study postdecision outcomes such as hierarchical compliance, executive enforcement, and precedent treatment, though inferential tradeoffs and opinion-based alternatives should be considered in these contexts. Due to endogeneity and posttreatment bias concerns, our measure should not be used to explain premerits stage activity such as agenda setting or early merits stage activity such as brief construction or amicus participation.

2. Inferential Problems with the Standard Approach

Complex legal cases have complicated information environments.[1] Legal complexity encompasses characteristics such as "density, technicality, differentiation, and indeterminacy" (Schuck 1992: 3). Put simply, a complex legal environment "taxes cognition" (Katz and Bommarito 2014: 337). Law and courts scholars from across disciplines have long invoked this conceptualization of complexity when formulating theories about judicial decision making. A variety of measures has been used to capture case complexity empirically, including profit disgorgement (Choi and Pritchard 2017), litigation involving statistics (Baye and Wright 2011), number of defendants and expert witnesses (Heise 2004), cross-appeals (Hettinger et al. 2004), opinion clarity (Owens and Wedeking 2011), docket age (Eisenberg and Miller 2004), dockets listed on appeal (Samaha et al. 2020), expert evaluations (Johnson 1987), issue technicality (Vanberg 2001), and procedural background statements (Black and Owens 2009).

The standard approach to measuring Supreme Court case complexity emphasizes issues and provisions. Maltzman and Wahlbeck (1996) introduced this approach in a seminal article on vote switching. They argue

1. We thank an anonymous reviewer for suggesting the phrase "complicated information environment."

that case-level uncertainty increases the probability of postconference vote switching. Emphasizing case dimensionality as a source of uncertainty and complexity, Maltzman and Wahlbeck contend, "When a case is particularly complex and thus raises multiple legal issues, the likelihood that a justice will understand all aspects of the case at the time of the vote is low" (584). To measure complexity, they factor analyze three SCDB variables: the number of issues, provisions, and separate opinions in a case. In addition to becoming the standard for measuring Supreme Court case complexity, scholars have adopted similar measures—particularly issue counts—in federal circuit (e.g., Hettinger et al. 2004) and state supreme (e.g., Goelzhauser and Cann 2014) court contexts.

To motivate our measurement contribution, we begin by highlighting inferential problems with the standard approach, particularly when invoked to explain merits stage outcomes. As an initial matter, all three measures induce posttreatment bias when used as predictors at this stage. Developing an actor-based salience measure from oral argument activity and examining whether it taps into case complexity, Black et al. (2013b: 816 n.12) write, "We do not include the number of [separate] opinions [in our complexity measure] because that would be a post-hoc measure." Unfortunately, the same is true for the issue and provision variables. Correcting an analogous problem induced by using postdecision newspaper coverage to measure case salience, Clark et al. (2015: 42) note, "While claims of causation are always suspect in observational research, measuring a treatment by proxy after the observed outcome is a particularly troublesome form of posttreatment bias." This bias "can be in any direction [and] any size" (Montgomery et al. 2018; see also Acharya et al. 2016).

Variables observed postdecision may also be endogenous to merits stage outcomes. The general problem is well recognized. For example, Black et al. (2013: 816 n. 12) exclude separate opinions from a complexity measure used to explain case salience in part because the former "is likely a consequence of" the latter. Also with respect to salience, Clark et al. (2015: 42) note, "If decisions and related choices affect the [post-decision] salience measure. . .then even if salience itself affects decisions or choices, our estimate of that effect can be biased in an unknown direction." Making a similar point when substituting amicus participation for postdecision newspaper coverage as a salience measure, Carrubba and Zorn (2010: 819) write, "[B]riefs have the advantage of being filed prior to the decision itself, thus ameliorating concerns over endogeneity." The same concern applies to using the standard complexity variables to explain merits stage outcomes.

Questions have been raised about the reliability of the provision variables. The SCDB includes a broad "Legal Provisions Considered by the Court" variable and its granular components "Legal Provisions Supplement" and "Legal Provision Minor Supplement." Collectively, these variables capture constitutional clauses, statutes, and other

provisions listed in the summary section of the Lawyers' Edition of the Supreme Court Reports or a numbered holding in the reporter's syllabus. In a recoding project, Shapiro (2009) reports two types of systematic underinclusivity. The first derives from relevant provisions in cases not being listed in Lawyers' Edition summaries or numbered syllabi holdings. The second derives from provisions that are listed in these sections not being coded in the SCDB. In 2015, the SCDB codebook noted, "[W]e mistakenly allowed for only one type [o]f legal provision. . .. If a second or third legal provision warranted inclusion into the case record, entirely new and separate records needed to be created" (Spaeth et al. 2015: 58). The extent to which this change mitigated the underinclusivity reported by Shapiro is unclear.

Last, the issue variables are not valid measures of complexity. The SCDB includes a 14-category "Issue Area" variable and a more granular "Issue" variable. The codebook notes, "Although criteria for the identification of issues are hard to articulate, the focus here is on the subject matter of the controversy. . .rather than its legal basis" (Spaeth et al. 2019: 45). Issues are identified "on the basis of the Court's own statements as to what the case is about" (45). Coding is related to the provision variable such that "each legal provision should generally not have more than a single issue applied to it," except "when a preference for one rather than the other cannot readily be made" (45). Rather than highlighting legal issues, the "objective is to categorize the case from a public policy standpoint" (45). The issue variables accurately capture policy content (Rice 2017), but have been criticized for being underinclusive with respect to topics (Shapiro 2009) and endogenous to case outcomes (Harvey and Woodruff 2013).

Using the issue variables to measure complexity raises two concerns. First, the single-issue constraint artificially limits counts, thereby defeating the purpose of employing them to measure complexity. This is a fatal flaw with respect to using the SCDB's issue variables as proxies for complexity. As Rice (2017: 214) puts it, "To use the [issue variables] to analyze any subject involving the potential of multiple issues. . .is inappropriate at a very basic level." Second, although the SCDB is clear that the issue variables capture policy rather than legal content, scholars often invoke them as if they emphasize the latter. Shapiro (2009: 490) attributes this in part to terminology and explains, "Too often, scholars either appear to assume that by 'issue' [the SCDB] means 'legal issue,' or they fail to make a clear distinction between [the SCDB's] definition of issue and what the word 'issue' means to most lawyers and legal academics." Although slippage between concepts and measures is to be expected, there is little reason to think that the issue variables capture what scholars are interested in when measuring complexity, and in any event, the single-subject constraint renders the question moot from a causal inference perspective.

Not being able to use the SCDB's issue variables as complexity measures is unfortunate for two reasons. First, the argument that complexity increases with the number of legal issues has intuitive appeal. Chief

Justice Rehnquist (1987: 293), for example, once contrasted "a relatively simple [case], with only one real legal issue" from "complex [cases], with several interrelated issues." This logic is consistent with evidence of preference inseparability in multi-issue cases (Braman 2006). Second, scholars often use the issue variables to test theoretical arguments about complexity motivated in part by the presence of multiple legal issues. Maltzman and Wahlbeck (1996: 584) suggest that a "particularly complex [case]-raises multiple legal issues" and note Justice Douglas switching votes in *Dandridge v*. Williams (1970), from reverse on equal protection grounds to affirm on conflict preemption grounds. Similarly, Collins (2008: 863) contends, "[I]n a case with multiple issues, a justice might have ideologically incompatible preferences attached to each separate issue in the case." And Carrubba and Zorn (2010: 819) assert that "multiple legal issue[]. . .cases—which often decide jurisdictional or other threshold issues, as well as matters of substantive law—are necessarily of greater complexity than single-issue cases."

3. Measuring Complexity Predecision

We introduce a predecision complexity measure derived from litigant briefs. This section proceeds in three parts. First, we explain how we leverage a 1954 Supreme Court rule change regarding merits brief construction to capture legal issue and provision counts. This is the same information scholars previously sought from the SCDB, but our ex ante approach diminishes endogeneity and posttreatment bias concerns when studying merits stage outcomes. Our approach to measuring legal issues is also valid with respect to complexity. Second, we develop a latent variable model that uses extracted issue and provision counts to derive a latent measure of case complexity. Third, we demonstrate measurement validity through a series of case studies and comparisons with existing measures.

Before proceeding we pause to emphasize our aim. By using information about issues and provisions contained in merits briefs, our measure is best suited to explain merits stage outcomes such as oral argument behavior, opinion assignment, bargaining, voting, and separate opinion production. Due to endogeneity and posttreatment bias concerns, our measure is not suitable for studying premerits stage activity such as agenda setting or early merits stage activity such as brief construction or amicus participation. Where the interest is in ex post decision complexity, such as when studying compliance and implementation, our measure can be used while recognizing inferential tradeoffs. Due to issue avoidance, which is a rare but regular occurrence, our measure is overinclusive with respect to counting issues and provisions addressed in opinions. If issues and provisions are the mechanism of interest, however, the measure may still be worthwhile. Alternatively, there are a number of opinion-based complexity measures that may be more suitable at this stage absent a specific interest in issues and provisions. We discuss the postmerits use of our measure in more detail in the applications section.

4. Extracting Information from Briefs

Our predecision measurement strategy leverages merits briefs for information on issues and provisions. Since the Supreme Court's 1954 term, petitioners have been required to highlight relevant issues and provisions in merits briefs. Rule 24(1)(a) requires petitioners to list "the questions presented for review." Rule 24(1)(f) requires petitioners to list "the constitutional provisions, treaties, statutes, ordinances and regulations involved in the case." The "involved in the case" stipulation distinguishes these central provisions, which must be "set out verbatim," from others referenced in the brief that merely need to be cited in the table of authorities. Rule 24(2) exempts respondents from these requirements but permits them to include their own issue and provision sections if they are "dissatisfied with [the petitioner's] presentation."

We manually extracted issue and provision counts from 1954 through 2017 terms. We accessed merits briefs from a variety of sources, including the Making of Modern Law Database of US Supreme Court Records and Briefs (1954–78), Nexis Uni (1979–2006), and SCOTUSBlog (2007–17). We conducted supplemental searches in Westlaw and a microfiche repository for briefs missing from these collections.[2] To extract information on issues and provisions, we hand-coded counts from the relevant sections mandated by Rules 24(1)(a) and 24(1)(f).[3] As envisioned by Rule 24(2), and confirmed by existing empirical evidence (see McGuire and Palmer 1995: 701 n. 6), petitioners and respondents regularly disagree in their delineations, so we gather counts from both sets of briefs.[4] Overall, our sample period spans 60+ terms and 6000+ cases.

Figure 1 plots the distribution of issues and provisions for petitioners and respondents. All four variables exhibit positive skew, reflecting the fact that most cases fall within a relatively modest range of issues and provisions with some rare cases taking more extreme values. The distributions and measures of central tendency are similar across petitioners and respondents. For questions presented, the medians are 1, means about 2, standard deviations about 1.5, and ranges from 0 to 17 for petitioners and 0 to 19 for respondents. For provisions, the medians are 3, means about 4, standard deviations about 4, and ranges from 0 to 71 for petitioners and 0 to 63 for respondents.

2. We were unable to locate briefs for approximately 8% of the cases in our sample period despite checking numerous sources.

3. To examine intercoder reliability, we randomly selected the 1995 term and had two people code issue and provision counts from each brief. Respectively, the linear-weighted Krippendorff's α statistics (Krippendorff 2009; Antoine et al. 2014) for issues and provisions are 0.986 and 0.899, with 1 representing perfect agreement. A quadratic weighting yields respective issues and provisions values of 0.986 and 0.946.

4. Issues and provisions are, respectively, correlated at 0.57 and 0.73 across petitioners and respondents. Scholars interested in issue creation should consider the possibility that issues can be raised elsewhere and subsequently addressed by the Court without violating the norm against sua sponte consideration (Epstein et al. 1996).

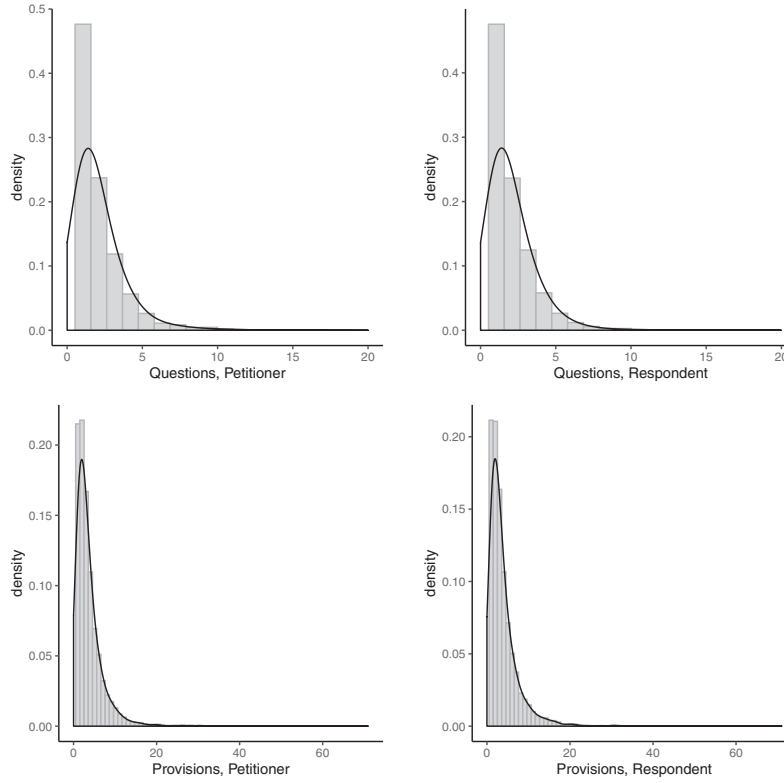

Figure 1. Distributions of Number of Questions and Number of Provisions Presented in Litigant Briefs. Black lines indicate density curves.

Figure 2 plots variable means across terms, as well as a smoothed fit via local polynomial regression. Two dynamics are notable. First, the long-term trend in each instance is a decrease in the average number of issues and provisions presented. These trends appear to have leveled off post-2000, with the exception of respondent provisions, which are trending upward. Second, and importantly, despite sourcing changes during the sample period, we find no evidence of any stark interruptions in the series. Rather, the trends are gradual over time.

The manifest variables plotted in Figure 2 indicate decreasing average case complexity over time. This does not mean that the law is less complex over time. Rather, cases *as presented to the Court* are becoming less complex in terms of issues and provisions. Although the mechanism behind these trends is unclear, we highlight two possibilities. First, the Supreme Court may be more regularly invoking its power to grant review limited to certain questions presented. In addition to directly limiting the number of issues, this may have a corresponding impact on the number of provisions. Second, the Supreme Court Bar has become more specialized

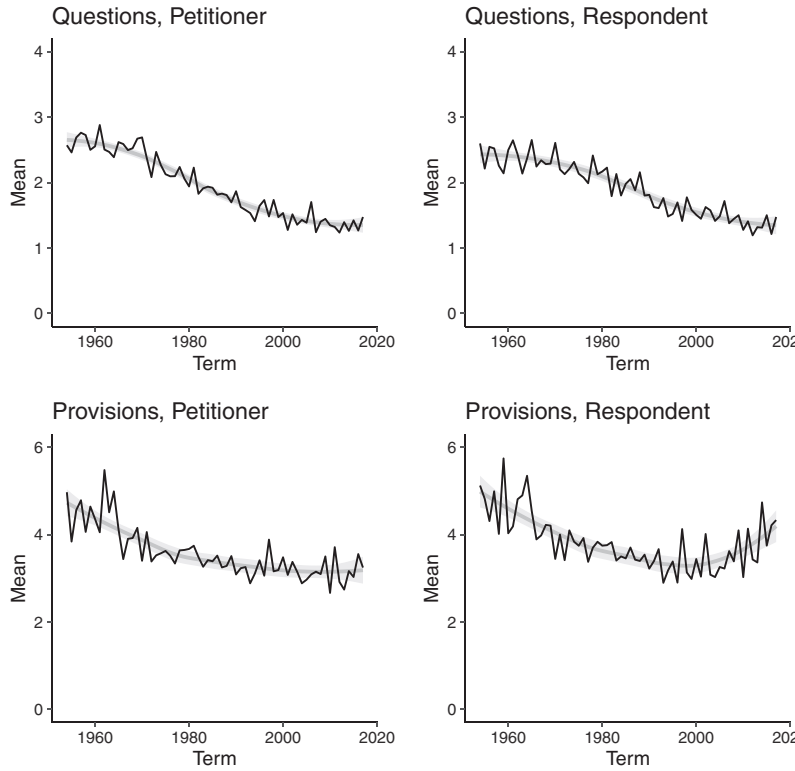

Figure 2. Number of Questions and Number of Provisions Presented in Litigant Briefs Over Time. Black lines indicate average values by term. Gray lines and shaded regions, respectively, indicate local polynomial regression fit and 95% confidence intervals.

(McGuire 1995; Goelzhauser and Vouvalis 2013; Biskupic et al. 2014), with a select group of repeat players regularly being brought on after intermediate appellate review for their expertise in securing Supreme Court review and winning on the merits. While exhausting potential issues makes sense during the early stages of litigation, better and more experienced attorneys may strategically waive preserved but nonessential issues after taking over a case on appeal to the Supreme Court. Strategic waiver may make sense as a way to signal the strength of any pressed issue, ensure that scarce brief space and oral argument time is reserved for a thorough presentation of the best arguments, and guard against repeat player reputation erosion that might result from decreasing average presentation quality by including substandard arguments.[5]

5. As a first cut at examining the possibility of a connection between attorney experience and strategic listings, we reviewed the second (1954) through eighth (2003) editions of Stern and Gressman's *Supreme Court Practice* treatise. There was little discussion concerning strategic behavior with respect to listing issues and provisions. Identifying the mechanism behind downward trends in case complexity is an important question for future research. Our

5. Modeling Latent Complexity

Using the information extracted from litigant briefs, we derive a latent measure of case complexity.[6] We assume that the count of issues and provisions identified by petitioners and respondents are Poisson random variables. Taking i to index the case and j to index the manifest variables, we therefore assume:

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij}) \tag{1}$$

where Y_{ij} is the count for case i of the number of provisions or the number of legal issues j. Our model is then:

$$\log(\lambda_{ij}) = \beta_j \theta_i - \alpha_j \tag{2}$$

In this formulation, β_j is the variable discrimination parameter on θ_i, and θ_i is the latent trait that otherwise explains variation in Y_{ij}, or case complexity. We include the term α_j as a means to scale the location of the manifest variable.

To resolve reflection invariance, we restrict the data such that *Communist Party v. Subversive Activities Control Board* (1961)—for which there were 28 provisions cited by both the petitioner and the respondent, 11 questions cited by the petitioner, and 13 questions cited by the respondent—take a higher value of the latent trait than *Wal-Mart v. Dukes* (2011)—for which the corresponding counts are all zero.[7]

We estimate the model using a Hamiltonian Monte Carlo algorithm through rstan (Carpenter et al., 2017). As priors, we set θ to normal(0,1), α to normal(0,5), and β to beta(0.5,0.5), or the Jeffreys' prior over a Bernoulli trial. We run four chains of 20,000 iterations each with a 15,000 iteration warm-up. Standard diagnostics provide strong evidence of convergence.[8]

Figure 3 plots latent complexity estimates and associated posterior intervals. Cases are ordered vertically by complexity, with actual values (gray dots) and 95% credible intervals (light gray bars) along the x-axis.

latent complexity measure can be used to explore a variety of important questions about how attorney expertise impacts presentation decisions.

6. Our latent measure is similar to the latent measure of case salience developed by Clark et al. (2015). We deviate from their specification by excluding term intercepts. With respect to salience, there are documented reasons (e.g., available newspaper space) to expect variation in media coverage over time on the basis of changes in the media's coverage of the Court rather than changes in case salience (see, e.g., Collins and Cooper 2012; Clark et al. 2015). With respect to complexity, we are not aware of similar exogenous temporal trends that would negatively impact estimation of the latent trait, though as discussed previously there may be identifiable reasons for the observed downward trends.

7. Zero counts are rare but sporadically occur perhaps due to rule noncompliance. Zero provision counts also occur in cases emphasizing prior precedent.

8. In Appendix, we plot the distribution of potential scale reduction statistic R-hats (Gelman and Rubin 1992). As that figure makes clear, all R-hats are well within standard benchmarks with values within a very small range of 0.999 and 1.001.

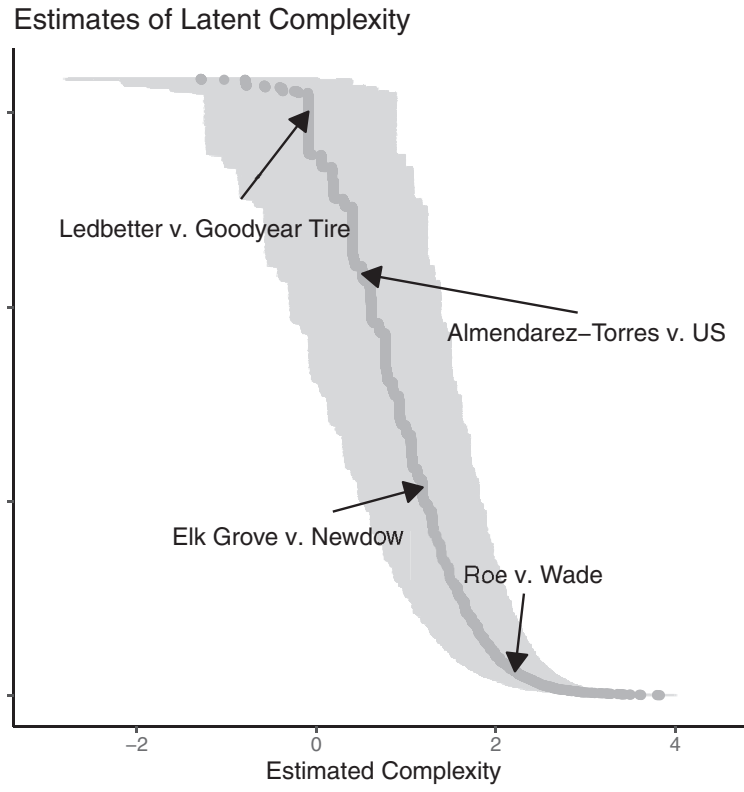

Figure 3. Estimates from Latent Variable Model of Case Complexity. Estimates of case complexity (dark gray dots) and associated 95% posterior intervals (light gray bars).

The distribution of cases is broad and reflects well on measurement validity, which we discuss in detail below. Observed sample values of latent complexity range from −1.3 to 3.8, with a mean of 0.9, and standard deviation of 0.7. The long right tail reflects the extremity observed above in the raw count distributions of the observed variables. Below, we elaborate on how the highlighted cases fit into this distribution.

Before doing so, we first examine variation in complexity over two important dimensions: issue area and time. In Figure 4, each panel represents a subset of the data defined by the issue area assigned by the SCDB. A few notable dynamics emerge. First, changes in average complexity over time are not consistent across issue areas. Second, there is some evidence of recent increases in average case complexity for federalism and federal taxation cases. Third, there has been a slow decline in the average complexity of civil rights and criminal procedure cases, two prominent topics on the Supreme Court's docket. These temporal differences may reflect the Court's shifting topic priorities over time, though this is a question we leave for future research.

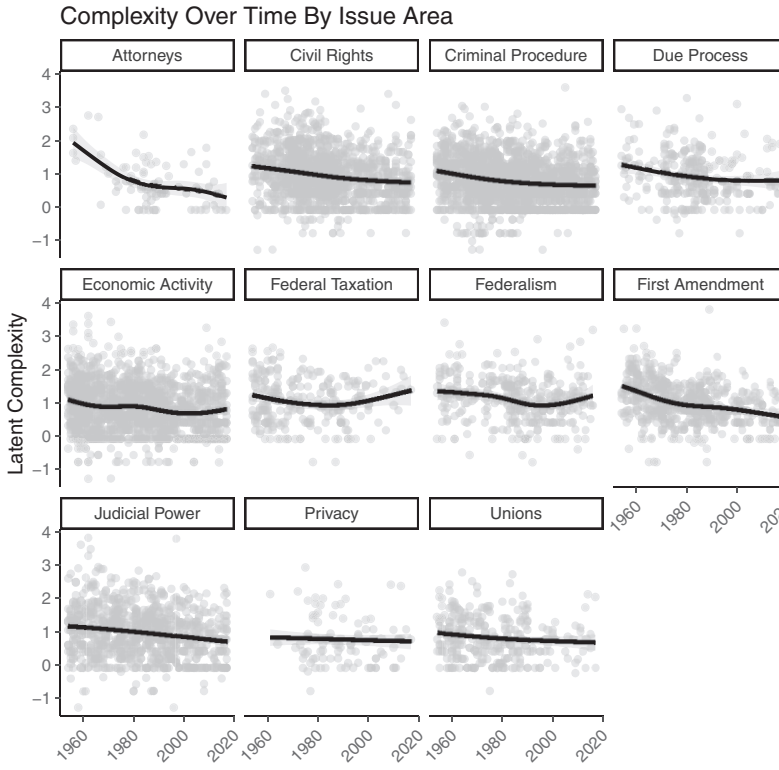

Figure 4. Latent Complexity by Issue Area over Time. Light gray dots indicate case-level estimate of latent complexity, and lines indicate smoothed fit of latent complexity (*y*-axis) by Supreme Court term (*x*-axis).

6. Model Evaluation

To evaluate how well our model fits the data, we rely on a predictive check approach from Wang and Blei (2019) that "compares the observed assignments with assignments drawn from the model's predictive distribution" (Wang and Blei 2019: 1580). If our model can generate predictions that closely match the observed values of our heldout manifest variables, then we can have more confidence in the model.

The approach operates as follows. First, we randomly holdout 20% of the manifest data, here y_{ij} or the observed counts. We pause to highlight that this is random across case (i.e., row) and manifest variable (i.e., column). That is, each manifest variable and observation remains, but a random 20% of the observed counts are replaced by missing values. Then, we fit our factor model to the remaining 80% of the data. From the fitted factor model, we generate 100 replicated datasets by probabilistically sampling the heldout counts from their predictive distribution. These counts are what we should observe according to the model. Finally, we compute

a test statistic of the expected heldout log-likelihood across both observed and replicated datasets, and calculate a predictive score as the proportion of instances where the test statistic for the replicated heldout data is lower than the test statistic on the observed heldout data.

Smaller values of the test statistic indicate the replicated data generated by the model are further from the observed data. Wang and Blei report models "with predictive scores larger than 0.1 to be satisfactory" (1581). For our model, the above approach yields a predictive score of 0.18, indicating our model can successfully assign predicted values in accordance with the actual, observed data. We turn, then, to evaluating our measurement approach across a number of different validity assessment criteria.

7. Measurement Validity

7.1 Face Validity

We consider face validity with several notable cases highlighted in Figure 3. In *Ledbetter v. Goodyear Tire & Rubber* (2007), the Court held 5-4 that a Title VII pay discrimination claim was time-barred because it fell outside the statute's 180-day filing window. The plaintiff argued that discriminatory evaluations from outside the filing window led to unlawful within-window practices. The majority concluded that only within-window actions could be used to prove discriminatory intent, while the dissent argued that the statute's broad remedial purpose and incremental nature of discrimination warranted a more flexible reading. After considerable debate during the 2008 election, Democrats led the effort to override the decision by passing the Lilly Ledbetter Fair Pay Act, which expanded the window for filing employment discrimination claims. While the underlying political and policy issues might be called complex, *Ledbetter* was relatively simple from an issues and provisions perspective. Both the SCDB and brief-based measures classify the case as involving one issue and one provision, but this similarity masks an important distributional difference. Generating a factor score of the SCDB's issue and provision variables yields an identical complexity estimate for cases, including *Ledbetter*, from about the 13th to 81st percentile. With our measure, *Ledbetter* is in the 5th percentile.

Highlighting the SCDB's underinclusion of relevant legal provisions, Shapiro (2009: 498–9) emphasizes *Almendarez-Torres v. United States* (1998). In *Almendarez-Torres*, the Court held 5-4 that a federal statute treated recidivism as a sentencing factor rather than a separate crime. It also held that failing to treat recidivism as an offense element does not violate the Constitution. Finding either reading of the statute plausible, the dissent would have treated recidivism as a separate crime in order to avoid the constitutional question. As Shapiro notes, the SCDB codes the case as involving a single statutory provision and does not pick up the constitutional dimension. Thus, the SCDB classifies *Ledbetter* and *Almendarez-Torres* as equally complex. Our measure, however, picks up the statutory

and constitutional elements of the case and correctly classifies *Almendarez-Torres* as more complex than *Ledbetter* from an issues and provisions perspective. Distributionally, the SCDB factor score lumps *Almendarez-Torres* with the cases grouped from about the 13th to 81st percentile, while our measure places it at about the 32nd percentile.

Next, we consider *Elk Grove v. Newdow* (2004). The Court granted cert to consider two questions: (a) whether a teacher-led recitation of the Pledge of Allegiance in a public school violated the Establishment Clause and (b) whether a noncustodial parent had standing to sue. By a 5-3 vote with one justice recused, the Court held that the noncustodial parent lacked standing and thus did not reach the constitutional question. The SCDB classifies *Newdow* as involving one issue and zero provisions—thereby making it less complex than *Ledbetter* and *Almendarez-Torres*. For scholars interested in how case complexity impacts merits outcomes, however, the endogeneity and posttreatment problems are clear: *Newdow*'s ex ante complexity may be the cause of its apparent ex post simplicity. Indeed, scholars often invoke the *Newdow* scenario of simultaneously confronting justiciability and merits questions as a quintessentially complex information environment (e.g., Carrubba and Zorn 2010: 819). Our measure picks up both dimensions and classifies *Newdow* as relatively complex. With one issue and zero provisions, *Newdow* receives the lowest possible SCDB complexity score along with about 13% of the distribution. Our measure places *Newdow* at about the 66th percentile.

Last, *Roe v.* Wade (1973) exemplifies how our measurement strategy captures complexity in a way that can help explain a variety of decision-making phenomena. In addition to what became the core constitutional holding, *Roe* presented questions concerning the Court's jurisdiction over grants of declatory relief, the scope of the lower court's injunction, standing, mootness, and abstention. The Court addressed questions concerning relief, justiciability, and abstention in Parts II and III of its opinion. The SCDB correctly classifies *Roe* as the most complex case we consider here. Although by design the SCDB only recognizes one issue, it identifies two provisions—the Fourteenth Amendment's Due Process Clause and Article III's cases or controversies requirement. Our measure captures these issues more comprehensively, but there is more to the story.

Another question presented in *Roe* asked whether the state statute was unconstitutionally vague. The Court ultimately concluded that its Fourteenth Amendment holding "makes it unnecessary. . .to consider vagueness" (164), but this passing reference belies the issue's ex ante importance. The lower court invalidated the statute on privacy and vagueness grounds. At the Court's first postargument conference, vagueness played a prominent role in the discussion with several justices supporting a disposition on that ground (Dickson 2001: 806–7). Justice Blackmun's first draft opinion invalidated the statute on vagueness grounds, and Justice White's first dissent emphasized vagueness. Blackmun's initial draft prompted return memos from Brennan and

Douglas expressing their understanding that a conference majority preferred to reach the privacy issue. Douglas added that "it [was] puzzling. . .-why [Chief Justice Burger] made the assignment at all except that he indicated he might affirm on vagueness," suggesting that Burger may have assigned the opinion strategically in order to secure a relatively narrow result on vagueness as opposed to privacy grounds.[9]

After a flurry of exchanges, Blackmun proposed rearguing *Roe* the following term. Holding the case over would allow further opportunity to forge a coalition on the disposition while also allowing replacements for Harlan (Rehnquist) and Black (Powell) to participate—votes that could have potentially impacted the outcome. The reargument suggestion prompted Douglas to write Burger, "If the vote of the Conference is to reargue, then I will file a statement telling what is happening to us and the tragedy it entails."[10] Ultimately, the Court voted to reargue and Blackmun's draft the following term addressed privacy rather than vagueness. Toward the end of the bargaining stage, Rehnquist congratulated Blackmun for "marshalling as well as I think could be done," while making one last effort to secure a narrower disposition—this time by invalidating the law only as applied to women seeking a first-trimester abortion.[11] Analogizing the request to earlier efforts to secure a narrower result on vagueness grounds, Blackmun refused, writing in response, "I still think [the statute] is vague and could not withstand careful analysis. . .[but] my vagueness approach. . .did not find favor."[12]

The story of how the vagueness question shifted from a central position in *Roe* to a largely forgotten single sentence in the opinion encapsulates the value of having an ex ante measure of case complexity. Arguably, *Roe*'s complexity impacted an array of merits stage outcomes of interest to judicial politics scholars, including strategic opinion assignment (e.g., Maltzman and Wahlbeck 2004), bargaining and accommodation (e.g., Maltzman et al. 2000), opinion content control (e.g., Carrubba et al. 2012), opinion writing scope (e.g., Fox and Vanberg 2014), separate opinion writing (e.g., Corley et al. 2013), and reargument (e.g., Hoekstra and Johnson 2003). Whereas any opinion-based measure would struggle to account for this ex ante complexity, our briefs-based measure incorporates the relevant information, which in turn helps to identify *Roe* as a relatively complex case. The SCDB factor score does lump *Roe* into a group of cases that fall between about the 82nd and 95th complexity percentile, but that owes more to its grouping of most cases below that threshold than a granular assessment of *Roe* as being particularly complex. Our measure places *Roe* at about the 96th percentile.

9. Letter from Douglas to Blackmun, May 19, 1972 (70-18). See also Letter from Brennan to Blackmun, May 18, 1972 (70-18). All letters referenced are available in the Supreme Court Opinion Writing Database (Wahlbeck et al. 2011).

10. Letter from Douglas to Burger, June 1, 1972 (70-18).

11. Letter from Rehnquist to Blackmun, November 24, 1972 (70-18).

12. Letter from Blackmun to Rehnquist, November, 27, 1972 (70-18).

7.2 Convergent and Discriminant Validity

Now we turn to convergent and discriminant validity. Convergent validity examines the extent to which our measure dovetails with other measures with which it should be theoretically related; discriminant validity examines the extent to which our measure diverges from those with which it should deviate (Quinn et al. 2010). Figure 5 plots comparisons of our latent complexity measure with four commonly employed case complexity proxies: the SCDB's number of issue areas coded for a case (upper left panel), the SCDB's number of issues coded for a case (upper right panel), the SCDB's number of legal provisions coded for a case (lower left panel), and the natural log of the number of amicus curiae briefs filed in a case (lower right panel).[13]

We first consider comparisons between our latent measure and the SCDB's issue and issue area variables. As discussed previously and elsewhere (e.g., Rice 2019), the SCDB's default coding rule is to assign a single issue or issue area to a case. Moreover, issue assignment is based on the primary policy dimension on which the Court decides a case. As Figure 5 makes clear, the overwhelming majority of cases are coded as involving one issue and issue area. In terms of convergent validity, however, we do see that as the number of assigned issues and issue areas increases, so too does the average and distribution of our latent complexity measure. Yet, perhaps more importantly, we find considerable variation within the SCDB's issue and issue area classifications. Specifically, we find that nearly the entire range of latent complexity is covered within each of the one or two issue or issue area categories.

These dynamics are further reflected in the coding of legal provisions. With a greater range in the SCDB's coding of provisions, we again find that our measure captures the distribution of complexity according to the opinion-based coding protocol, but also find considerable variation in latent complexity within each level. Taken together, our latent complexity measure relates to standard measures of case complexity, but also demonstrates the limitations of those measures for accurately capturing case complexity. By incorporating ex ante measures of issues and provisions instead of relying on opinion content, our latent predecision measure more accurately reflects the complexity distribution of cases as they reach the merits stage.

In the lower right panel of Figure 5, we plot latent complexity (x-axis) against the natural log of the number of amicus curiae briefs filed in a case (y-axis), with an overlaid smoothed conditional means fit. We find only tentative evidence that increased filings are associated with latent

13. Figures show estimate comparisons of latent complexity θ. The size of the boxplots is proportional to the number of cases within that category (i.e., the number of cases for which the SCDB coded the relevant number of issue areas, issues, or legal provisions). The lower right panel features a scatterplot with a smoothed conditional mean (blue line) and associated 95% interval (gray shaded region).

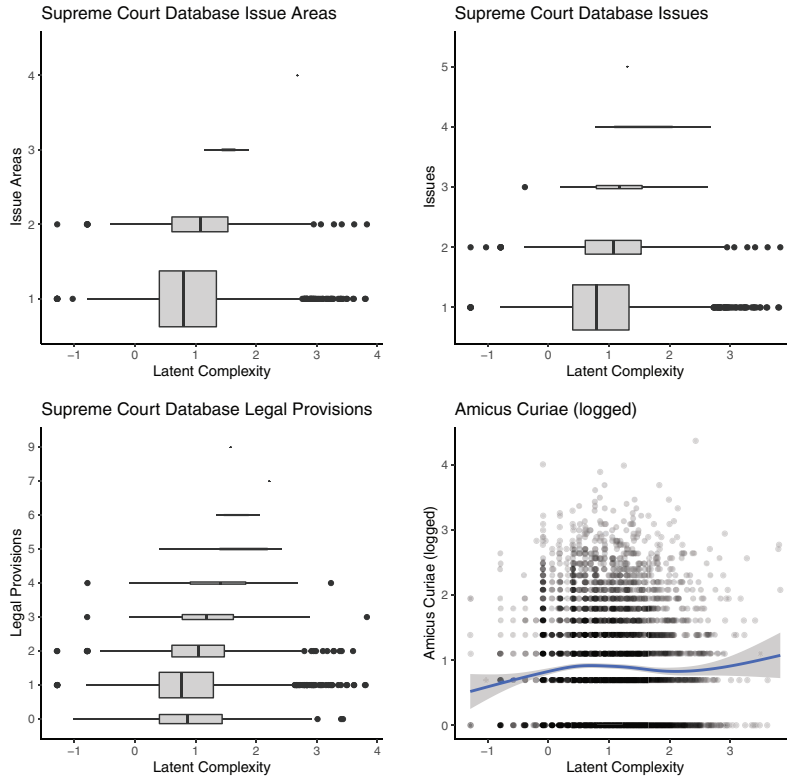

Figure 5. Latent Complexity Comparisons to Other Complexity Constructs.

complexity. The relationship is most evident at the extremes (i.e., very low and very high values of complexity) but nonexistent among the bulk of cases. Although amici participation is sometimes invoked as a proxy for case complexity, the lack of relationship evident here is not surprising given that amici participation more directly captures concepts such as case salience, interest group attention, and the scope of political conflict.

Last, we turn to a direct comparison between our measure and the original measure factor analyzing three SCDB variables: the number of issues, provisions, and separate opinions. In the left panel of Figure 6, we present a scatterplot of our latent measure (x-axis) against the factor scores (y-axis). As the plot makes clear, the two measures are weakly correlated; indeed, the Pearson's correlation coefficient is 0.13 between the factor score measure and our latent complexity measure, providing substantial evidence that the approaches are capturing similar yet substantively different dynamics. In the right panel, we explore whether the relationship between measures changes over time; as the plot makes clear, the correlation between the complexity measures by term indicates no temporal patterns.

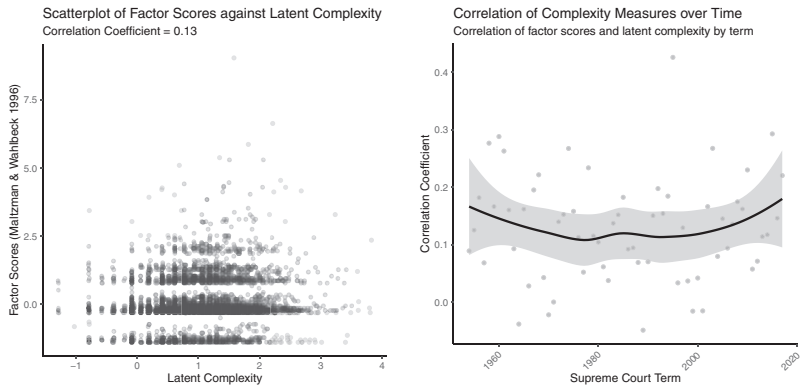

Figure 6. Latent Complexity Comparison to Factor Score.

7.3 Applications

In this section, we illustrate the usefulness of our latent complexity measure by replicating two existing studies—one concerning oral argument impact on merits voting and the other lower court compliance with Supreme Court precedent. The replications demonstrate that our measure yields theoretically intuitive results while enhancing inference credibility. Furthermore, we consider applications from distinct stages of the judicial process: merits and postdecision. Although our measure is best suited for explaining merits stage outcomes (e.g., voting), it can also be useful for studying postdecision outcomes (e.g., compliance). We discuss the tradeoffs involved with using our measure to study postdecision outcomes below.

7.4 Merits Stage

Oral arguments are central to the Supreme Court's decision-making process. Johnson et al. (2006: 104) examine the conditions under which oral argument quality is associated with merits voting. Specifically, they develop a theory emphasizing two conditional arguments. First, they argue that the effect of oral argument quality on a justice's vote will diminish as litigant-justice ideological divergence increases. Second, they argue that the effect of oral argument quality on a justice's vote depends on case complexity:

> [C]ases that come to the Supreme Court often focus on several issue dimensions, as well as on more than one constitutional or statutory question. It is in these cases where lawyers, who often spend years researching the case, have more information than the justices, even after briefs have been submitted. Because of the need for information in complex cases, the quality of arguments should have a greater effect than in less complex cases.

The sample includes 539 cases from 1970 through 1994. The outcome variable is an indicator for whether a justice voted to reverse (1 = yes,

0 = no). Oral argument quality is measured using Justice Blackmun's attorney performance grades. Specifically, Johnson, Wahlbeck, and Spriggs standardize Blackmun's different grading systems over time by creating *z*-scores, then subtract the respondent's grade from the petitioner's such that increasing values indicate better petitioner performance relative to the respondent. Following standard practice, complexity is measured with a factor analysis of the number of issues and provisions in the case as coded by the SCDB. As noted previously, however, this measure introduces posttreatment bias since it is only observed after oral argument quality.

Although Johnson, Wahlbeck, and Spriggs find support for the conditional ideology hypothesis, they do not reject the null hypothesis with respect to complexity. Table A1 presents full results. The noncomplexity results are similar across specifications. Using the SCDB measure, the estimated coefficient for the complexity interaction is negative and insignificantly different from zero. In contrast, using our predecision latent measure yields a more precisely estimated result indicating the conditional effect of oral argument quality on votes decreases on average as complexity increases. A subsequently proposed counter theory explains the result. Whereas Johnson et al. (2006: 107) advance the "hypothesis that the justices will give more weight to oral arguments in complex cases," Ringsmuth et al. (2013: 433) argue that justices are "less likely to be persuaded by oral arguments in more complex cases" because they "are less likely to be persuaded. . .in situations [with] more policy options" (433).

Figure 7 plots conditional marginal effects, revealing that the results are more similar than they first appear while nonetheless highlighting our measure's advantage. Both panels reveal statistically significant conditional marginal effects, but estimate imprecision in the SCDB measure model indicates an insignificant effect across about 12% of the sample at the highest complexity values, whereas the effect in the latent predecision model is insignificant across only about 1% of the sample (since there is little overlap across measures the high ends of the distributions do not necessarily contain the same subset of cases). In this instance, increasing measurement validity and excising posttreatment bias by substituting our predecision measure helps detect significant effects over a larger sample range, increases estimate precision, and adjudicates between competing theories about the influence of complexity on merits voting.

7.5 Postdecision

While our measure is best suited for the merits stage, it can also be useful for studying postdecision outcomes. There are two considerations when determining whether to use our measure in these circumstances. First, it is important to determine whether issues and provisions are part of the theoretical mechanism by which a complicated information environment impacts the outcome of interest. If not, postdecision measures such as issue technicality (Vanberg 2001), opinion clarity (Owens and Wedeking 2011),

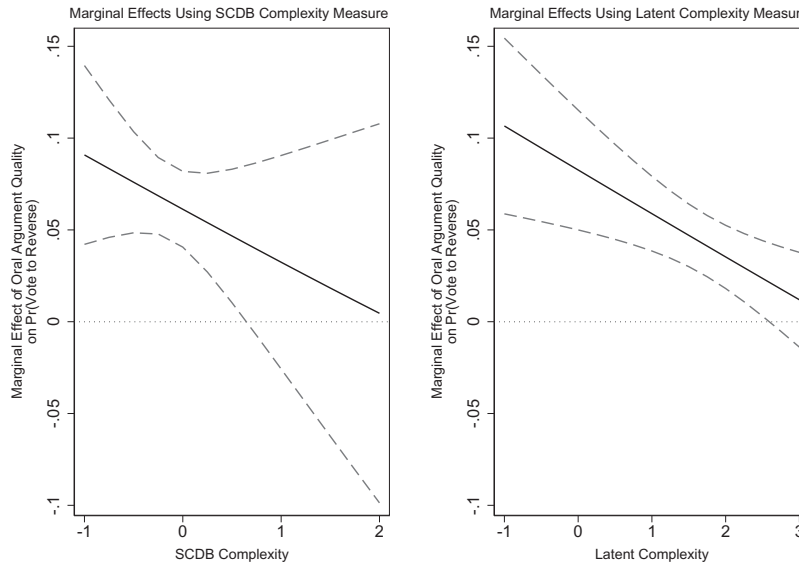

Figure 7. Conditional Marginal Effect of Oral Argument Quality on Voting Behavior across Values of SCDB Complexity (Left Panel) and Latent Complexity (Right Panel).

separate opinions (Westerland et al. 2010), and topic attention (Rice 2019) may be more suitable. Second, if issues and provisions are theoretically relevant for explaining postdecision outcomes as is often the case (e.g., Spriggs and Hansford 2001; Benesh and Reddick 2002; Corley 2009), it is important to understand that our measure is overinclusive with respect to issues and provisions in some instances due to avoidance. Although our measure accurately captures *Newdow*'s merits stage complexity, for example, it overstates it if the interest is understanding postdecision outcomes. Nonetheless, issue avoidance is a regular but rare occurrence (Solimine and Gely 2010; Goelzhauser 2011) and our measure is otherwise well suited for capturing the number of issues and provisions at issue in a case.[14]

To illustrate the usefulness of our measure when examining postdecision outcomes, we turn to Westerland et al.'s (2010) study of lower court compliance with Supreme Court precedent. The vertical relationship between higher and lower courts necessitates understanding the conditions under which agents comply with dictates established by their principals. To this end, Westerland et al. develop a theory explaining when we should expect to observe federal circuit courts comply with Supreme Court

14. Although empirical evidence suggests that avoidance through the use of justiciability doctrines (Goelzhauser 2011) and dismissing cases as improvidently granted (Solimine and Gely 2010) is rare, the extent to which the Court addresses issues and provisions addressed in merits briefs is unclear and an important question for future research.

precedent. Consistent with agency theory, they expect noncompliance to increase with preference divergence. The outcome variable classifies lower court treatments of a given Supreme Court precedent, moving from less to more compliant (1 = deviate, 2 = neutral, 3 = comply).

One of Westerland et al.'s theoretical arguments is that precedent clarity should be positively associated with lower court compliance. They employ two strategies for measuring precedent clarity. First, they include the number of concurring and dissenting opinions. Second, in addition to this postdecision measure, they include a measure of case complexity scoring the total number of issues and provisions as coded by the SCDB. Since case complexity is measured prior to observing lower court compliance, posttreatment bias is not an issue. However, the construct validity concerns discussed previously remain. Moreover, the use of issues and provisions in addition to a postdecision measure like the number of separate opinions illustrates that scholars value issues and provisions as complexity mechanisms that can help explain downstream behavior. Where that is true, our measure is a useful replacement for the SCDB's measures.

Westerland et al.'s sample includes 500 Supreme Court decisions from 1953 to 1990 that yield 10,000+ lower court citations. They find support for a number of their hypotheses, including the core prediction that noncompliance increases as the ideological divergence between the precedent-enacting and contemporary courts increases. Introducing our latent measure does not change the primary results.[15] Table A2 presents full results. With respect to complexity, though, their model does not reject the null hypothesis. When substituting our latent measure of complexity, however, the estimated coefficient is negative and statistically distinguishable from zero—evidence supporting Westerland et al.'s hypothesis that complex cases induce less compliance. Figure 8 plots marginal effects, with the left panel's confidence intervals illustrating how the SCDB's counts are clustered at low levels. Explaining the unexpected clarity finding, Westerland et al. note that their "measures of clarity may capture this concept only poorly" (902). By introducing a measure with more construct validity, we find evidence that supports their original insight that more complex cases generate more noncompliance.

8. Conclusion

Case complexity is central to the study of judicial politics. For decades, the number of issues and provisions in case as coded by the SCDB has been the dominant measures of case complexity. But these measures suffer

15. The only other changed result concerns the number of positive intervening Supreme Court treatments. The authors predict no relationship between positive treatments and compliance but counterintuitively find a positive relationship ($p = 0.047$). Estimating their model with the sample of cases for which we have latent complexity scores yields $p = 0.123$ and substituting our latent measure yields $p = 0.314$. These changes are not surprising given the slight power loss and theoretical expectation of no relationship.

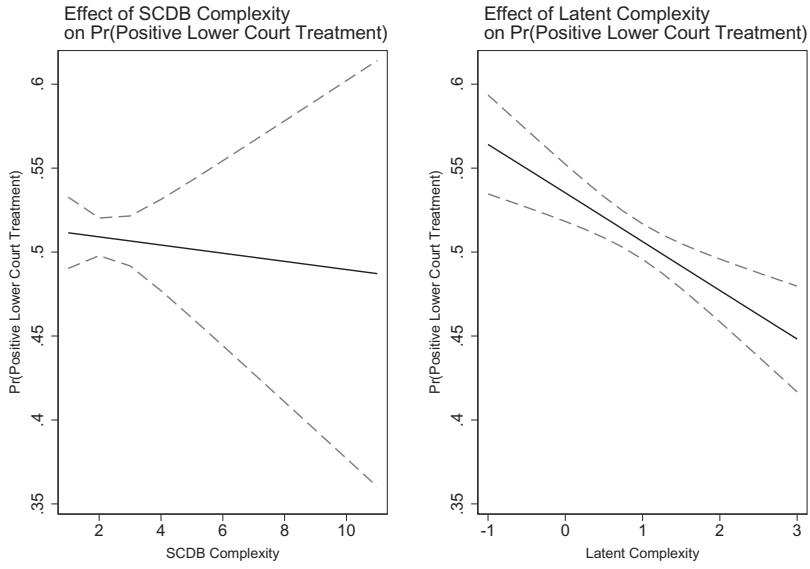

Figure 8. Effect of Complexity on Lower Court Compliance Using SCDB Complexity (Left Panel) and Latent Complexity (Right Panel).

from several limitations: the SCDB's issue variables are not valid measures of complexity, the reliability of the provision variables has been questioned, and including both sets of variables to explain merits stage outcomes heightens endogeneity and posttreatment bias concerns. We overcome these problems by extracting information on issues and provisions from merits briefs. Furthermore, we use the raw issue and provision counts from petitioner and respondent briefs to derive a latent measure of case complexity. In addition to mitigating several threats to inference, our measure is replicable, easy to update, and transportable to other contexts.

Moving forward, we offer general and specific guidance concerning case complexity. First, we encourage scholars to continue developing theories about case complexity and otherwise considering how case complexity might confound other relationships of interest. Second, we encourage scholars to revisit previous theoretical arguments concerning Supreme Court case complexity, where existing findings may be an artifact of the measurement problems detailed here. Third, we encourage scholars who invoke case complexity to precisely specify the underlying concept of interest independent of how available measures are constructed. Fourth, we encourage scholars to pay careful attention to the causal ordering of the theoretical argument to avoid conditioning on a posttreatment variable. With a carefully defined concept and attention to causal ordering, scholars will be well-positioned to examine the fit between existing measures and the underlying research question. To this

end, we consider the use of various measures at three stages of the judicial process:

(1) Premerits (e.g., agenda setting, amicus brief filings): Our measure should not be used to capture case complexity at the premerits stage due to endogeneity and posttreatment bias concerns. With respect to agenda setting, for example, the Court can shape the number of issues and provisions observed in merits briefs by limiting review to certain questions presented and adding others the parties did not propose. One alternative measure at this stage is the clever indicator of procedural complexity advanced by Black and Owens (2009); while they leverage cert pool memos, which are not widely available, the logic can be extended to lower court opinions. Another alternative would be to implement our measure using briefs requesting Supreme Court review. Most of the available alternative measures of case complexity are not well suited for this stage.

(2) Merits (e.g., oral argument, bargaining, voting, separate opinion production): Our measure is best suited to capture case complexity at the merits stage. By leveraging information from merits briefs, our measure avoids posttreatment bias in research settings that plausibly consider behavior that occurs after these briefs are filed. In addition, most of the available alternative measures of case complexity are not well suited for this stage because they are observed only postdecision. There are two merits-stage use exceptions. First, our measure should not be used to explain features of merits briefs themselves since doing so would be circular. Second, our measure should not be used to explain amicus curiae activity; it is common, for example, for the respondent's merits brief to be filed after amici file merits briefs, which means our measure may be endogenous to amicus brief content.

(3) Postmerits (e.g., congressional responses, executive implementation, lower court compliance): Our measure can be used at the postmerits stage while recognizing its limitations and considering alternative measures. Our measure's key limitation at this stage is potential overinclusivity due to avoidance. Existing evidence shows that avoidance is a regular but rare occurrence. For any given research setting, it is important to consider the extent to which underinclusiveness due to avoidance may be problematic before using our measure of case complexity. If issues and provisions are the mechanisms of interest, our measure can be used, but we encourage robustness checks given the number of available decision-based measures at this stage. Alternative measures include the number of separate opinions, opinion clarity, the SCDB's provision variables, and Rice's (2019) estimates of topic attention.

There is much to consider in future work. As an initial matter, the literature would benefit from new measures of case complexity—particularly at the premerits and merits stages where there are currently few alternatives. Using briefs is a promising approach that can be extended to capture premerits issues and provisions from the agenda-setting stage. Furthermore, additional measures of complexity can be extracted using other sources of information in briefs, including arguments and precedent citations (Goelzhauser and Vouvalis 2014). Given the wealth of information in briefs more generally (see Hazelton et al. 2019), they are a promising source of predecision information that is only systematically available now through the SCDB's opinion-based coding protocols, including whether a case raises a constitutional question and whether there is an underlying circuit split. Moreover, our approach can be adapted to other contexts, including lower federal courts, state courts, and international courts.

Data Availability Statement

The measure of complexity and all materials to replicate the findings reported in this study will be posted on JLEO's website.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects," 110 American Political Science Review 512–29.

Almendarez-Torres v. United States. 1998. 523 U.S. 224.

Antoine, Jean-Yves, Jeanne Villaneau, and Anais Lefeuvre. 2014. "Weighted Krippendorff's Alpha is a More Reliable Metrics for Multi-Coders Ordinal Annotations: Experimental Studies on Emotion, Opinion and Coreference Annotation." Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 550–9.

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," 54 *Journal of Law and Economics* 1–24.

Benesh, Sara C., and Malia Reddick. 2002. "Overruled: An Event History Analysis of Lower Court Reaction to Supreme Court Alteration of Precedent," 64 *Journal of Politics* 534–50.

Biskupic Joan, Janet Roberts, and John Shiffman. 2014. "The Echo Chamber: A small group of lawyers and its outsized influence at the U.S. Supreme Court." *Reuters*, December 8. https://reut.rs/2Une21m.

Black, Ryan C., and Ryan J. Owens. 2009. "Agenda Setting in the Supreme Court: The Collision of Policy and Jurisprudence," 71 *Journal of Politics* 1062–75.

Black, Ryan C., Rachel A. Schutte, and Timothy R. Johnson. 2013a. "Trying to Get What You Want: Heresthetical Maneuvering and U.S. Supreme Court Decision Making," 66 *Political Research Quarterly* 819–30.

Black, Ryan C., Maron W. Sorenson, and Timothy R. Johnson. 2013b. "Toward an Actor-Based Measure of Supreme Court Case Salience: Information-Seeking and Engagement during Oral Arguments," 66 *Political Research Quarterly* 804–18.

Braman, Eileen. 2006. "Reasoning on the Threshold: Testing the Separability of Preferences in Legal Decision Making," 68 *Journal of Politics* 308–21.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language," 76 *Journal of Statistical Software*.

Carrubba, Cliff, Barry Friedman, Andrew D. Martin, and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?," 56 *American Journal of Political Science* 400–12.

Carrubba, Clifford J., and Christopher Zorn. 2010. "Executive Discretion, Judicial Decision Making, and Separation of Powers in the United States," 72 *Journal of Politics* 812–24.

Choi, Stephen J., and A.C. Pritchard. 2017. "The SEC's Shift to Administrative Proceedings: An Empirical Assessment," 34 *Yale Journal on Regulation* 1–32.

Clark, Tom S., Jeffrey R. Lax, and Douglas Rice. 2015. "Measuring the Political Salience of Supreme Court Cases," 3 *Political Research Quarterly* 37–65.

Collins, Paul M. Jr. 2008. "The Consistency of Judicial Choice," 70 *Journal of Politics* 861–73.

Collins, Todd A., and Christopher A. Cooper. 2012. "Case Salience and Media Coverage of Supreme Court Decisions: Toward a New Measure," 65 *Political Research Quarterly* 396–407.

Communist Party v. Subversive Activities Control Board. 1961. 367 U.S. 1.

Corley, Pamela C. 2009. "Uncertain Precedent: Circuit Court Responses to Supreme Court Plurality Opinions," 37 *American Politics Research* 30–49.

Corley, Pamela C., Amy Steigerwalt, and Artemus Ward. 2013. *The Puzzle of Unanimity: Consensus on the United States Supreme Court*. Stanford, CA: Stanford University Press.

Williams Dandridge v.. 1970. 397 U.S. 471.

Dickson, Del, ed. 2001. *The Supreme Court in Conference: 1940-1985*. New York: Oxford University Press.

Eisenberg, Theodore, and Geoffrey P. Miller. 2004. "Attorney Fees in Class Action Settlements: An Empirical Study," 1 *Journal of Empirical Legal Studies* 27–48.

Elk Grove v. Newdow. 2004. 542 U.S. 1.

Epstein, Lee, Jeffrey A. Segal, and Timothy Johnson. 1996. "The Claim of Issue Creation on the U.S. Supreme Court," 90 *American Political Science Review* 845–52.

Fox, Justin, and Georg Vanberg. 2014. "Narrow versus Broad Judicial Decisions," 26 *Journal of Theoretical Politics* 355–83.

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences," 7 *Statistical Science* 457–72.

Goelzhauser, Greg. 2011. "Avoiding Constitutional Cases," 39 *American Politics Research* 483–511.

Goelzhauser, Greg, and Vouvalis Nicole. 2014. "Measuring Case Complexity." Paper presented at the 72nd Annual Conference of the Midwest Political Science Association, Chicago, April 3–6.

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

Goelzhauser, Greg, and Nicole Vouvalis. 2013. "State Coordinating Institutions and Agenda Setting on the U.S. Supreme Court," 41 *American Politics Research* 819–38.

Hazelton, Morgan L.W., Rachael K. Hinkle, and James F. Spriggs II. 2019. "The Influence of Unique Information in Briefs on Supreme Court Opinion Content," 40 *Justice System Journal* 126–57.

Harvey, Anna, and Michael J. Woodruff. 2013. "Confirmation Bias in the United States Supreme Court Judicial Database," 29 *Journal of Law, Economics, & Organization* 414–60.

Heise, Michael. 2004. "Criminal Case Complexity: An Empirical Perspective," 1 *Journal of Empirical Legal Studies* 331–69.

Hettinger, Virginia A., Stefanie A. Lindquist, and Wendy L. Martinek. 2004. "Comparing Attitudinal and Strategic Accounts of Dissenting Behavior on the U.S. Courts of Appeals," 48 *American Journal of Political Science* 123–37.

Hoekstra, Valerie, and Timothy Johnson. 2003. "Delaying Justice: The Supreme Court's Decision to Hear Rearguments," 56 *Political Research Quarterly* 351–60.

Johnson, Charles A. 1987. "Law, Politics, and Judicial Decision Making: Lower Federal Court Uses of Supreme Court Decisions," 21 *Law & Society Review* 325–40.

Johnson, Timothy R., Paul J. Wahlbeck, and James F. Spriggs II. 2006. "The Influence of Oral Arguments on the U.S. Supreme Court," 100 *American Political Science Review* 99–113.

Katz, Daniel Martin, and M.J. Bommarito II. 2014. "Measuring the Complexity of the Law: The United States Code," 22 *Artificial Intelligence and Law* 337–74.

Krippendorff, Klaus. 2009. "Testing the Reliability of Content Analysis Data: What is Inolved and Why," in K. Krippendorff and M.A. Bock, eds., *The Content Analysis Reader*, 350–357. Thousand Oaks, CA: Sage Publications.

Ledbetter v. Goodyear Tire & Rubber. 2007. 550 U.S. 618.

Maltzman, Forrest, and James F. Spriggs II, and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. New York: Cambridge University Press.

Maltzman, Forrest, and Paul J. Wahlbeck. 1996. "Strategic Policy Considerations and Voting Fluidity on the Burger Court," 90 *American Political Science Review* 581–92.

———. 2004. "A Conditional Model of Opinion Assignment on the Supreme Court," 57 *Political Research Quarterly* 551–63.

McGuire, Kevin T. 1995. "Repeat Players in the Supreme Court: The Role of Experienced Lawyers in Litigation Success," 57 *Journal of Politics* 187–96.

McGuire, Kevin T., and Barbara Palmer. 1995. "Issue Fluidity on the U.S. Supreme Court," 89 *American Political Science Review* 691–702.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It," 62 *American Journal of Political Science* 760–75.

Owens, Ryan J., and Justin P. Wedeking. 2011. "Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions," 45 *Law & Society Review* 1027–61.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs," 54 *American Journal of Political Science* 209–28.

Rice, Douglas R. 2019. "Measuring the Issue Content of Supreme Court Opinions," 7 *Journal of Law and Courts* 107–27.

——— 2017. "Issue Divisions and US Supreme Court Decision Making," 79 *Journal of Politics* 210–22.

Wade Roe v.. 1973. 410 U.S. 113.

Rehnquist, William H. 1987. *The Supreme Court: How It Was, How It Is*. New York: William Morrow & Co.

Ringsmuth, Eve M., Amanda C. Bryan, and Timothy R. Johnson. 2013. "Voting Fluidity and Oral Argument on the U.S. Supreme Court," 66 *Political Research Quarterly* 429–40.

Samaha, Adam M., Michael Heise, and Gregory C. Sisk. 2020. "Inputs and Outputs on Appeal: An Empirical Study of Briefs, Big Law, and Case Complexity," 17 *Journal of Empirical Legal Studies* 519–55.

Schuck, Peter H. 1992. "Legal Complexity: Some Causes, Consequences, and Cures," 42 *Duke Law Journal* 1–52.

Shapiro, Carolyn. 2009. "Coding Complexity: Brining Law to the Empirical Analysis of the Supreme Court," 60 *Hastings Law Journal* 477–539.

Solimine, Michael E., and Rafael Gely. 2010. "The Supreme Court and the Sophisticated Use of DIGs," 18 *Supreme Court Economic Review* 155–76.

Spaeth, Harold, Lee Epstein, Ted Ruger, Sarah C. Benesh, Jeffrey Segal, and Andrew D. Martin. 2015. *The Supreme Court Database Code Book* (Release 2) (released March 14, 2016).

———. 2019. *The Supreme Court Database Code Book* (Release 1) (released September 13, 2019).

Spriggs, James F. II, and Thomas G. Hansford. 2001. "Explaining the Overruling of U.S. Supreme Court Precedent," 63 *Journal of Politics* 1091–111.
Vanberg, Georg. 2001. "Legislative-Judicial Relations: A Game-Theoretic Approach to Constitutional Review," 45 *American Journal of Political Science* 346–61.
Wahlbeck, Paul J., James F. Spriggs II, and Forrest Maltzman. 2011. *The Burger Court Opinion-Writing Database*. http://supremecourtopinions.wustl.edu.
Wal-Mart v. Dukes. 2011. 564 U.S. 338.
Wang, Yixin, and David Blei. 2019. "The Blessings of Multiple Causes," 114 *Journal of the American Statistical Association* 1574–96.
Westerland, Chad, Jeffrey A. Segal, Lee Epstein, Charles Cameron, and Scott Comparato. 2010. "Strategic Defiance and Compliance in the U.S. Court of Appeals," 54 *American Journal of Political Science* 891–905.

Appendix A
Model Diagnostics

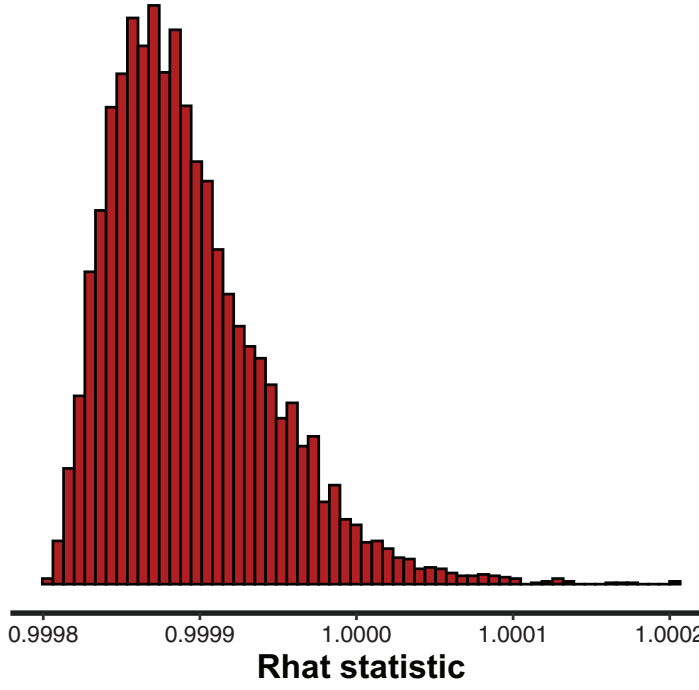

Figure A1. Distribution of Potential Scale Reduction Statistic R-hat. R-hat is a measure of the variance of all chains compared to within-chain variances. Generally, the chains have converged if we do not observe any values above 1.1.

Full Replication Results

Table A1. Replicating Johnson et al. (2006)

	Model 1	Model 2	Model 3
Oral argument grade	0.203*	0.254*	0.341*
	(0.045)	(0.045)	(0.069)
Ideological compatibility with appellant	0.315*	0.306*	0.307*
	(0.038)	(0.035)	(0.034)
SCDB complexity	0.072	0.076	
	(0.074)	(0.064)	
Latent complexity			0.015
			(0.040)
Ideological compatibility × Oral argument grade	0.024*	0.014	0.016
	(0.011)	(0.013)	(0.011)
Oral argument grade × SCDB complexity	−0.088	−0.117	
	(0.095)	(0.105)	
Oral argument grade × Latent complexity			−0.097*
			(0.033)
US appellant	0.474*	0.421*	0.401*
	(0.098)	(0.095)	(0.105)
US appellee	−0.788*	−0.900*	−0.856*
	(0.097)	(0.116)	(0.126)
S.G. appellant	0.324*	0.443*	0.490*
	(0.104)	(0.105)	(0.095)
S.G. appellee	−0.210	−0.096	−0.122
	(0.133)	(0.111)	(0.120)
D.C. appellant	0.407*	0.404*	0.414*
	(0.101)	(0.110)	(0.108)
D.C. appellee	0.068	0.128	0.100
	(0.143)	(0.126)	(0.118)
Law professor appellant	−0.761*	−1.008*	−1.041*
	(0.169)	(0.197)	(0.197)
Law professor appellee	−1.555*	−1.565*	−1.581*
	(0.203)	(0.204)	(0.207)
Clerk appellant	−0.248*	−0.269*	−0.281*
	(0.095)	(0.103)	(0.104)
Clerk appellee	−0.165	−0.113	−0.153
	(0.192)	(0.204)	(0.205)
Elite law school appellant	0.026	0.019	0.019
	(0.111)	(0.104)	(0.106)
Elite law school appellee	−0.126	−0.160*	−0.159*
	(0.083)	(0.071)	(0.072)
Difference in litigating experience	−0.128*	−0.130*	−0.138*
	(0.014)	(0.013)	(0.013)
Constant	0.282*	0.289*	0.273*
	(0.053)	(0.046)	(0.075)
Observations	3331	3053	3053

Note: The dependent variable is an indicator scored 1 if a justice votes to reverse and 0 otherwise. Logistic regression is used to fit the models. Robust standard errors are in parentheses.

*$p < 0.05$. Model 1 replicates the results presented by Johnson et al. (2006: 109, Table 3, Model 2); Model 2 replicates those results on the sample of cases for which we have latent complexity scores; and Model 3 replaces the SCDB measure with our predecision latent measure.

Table A2. Replicating Westerland et al. (2010)

	Model 1	Model 2	Model 3
Contemporary SC-enacting SC distance	−0.635*	−0.696*	−0.622*
	(0.141)	(0.149)	(0.150)
Panel-enacting SC distance	−0.069	−0.034	−0.031
	(0.099)	(0.103)	(0.103)
Panel-contemporary SC distance	0.013	0.074	0.066
	(0.125)	(0.131)	(0.131)
Age of SC precedent	−0.011*	−0.011*	−0.009*
	(0.003)	(0.003)	(0.003)
SCDB complexity	−0.008	−0.009	
	(0.028)	(0.030)	
Latent complexity			−0.116*
			(0.030)
Dissenting opinions	−0.001	0.023	0.026
	(0.031)	(0.033)	(0.033)
Concurring opinions	0.065*	0.077*	0.086*
	(0.032)	(0.034)	(0.034)
Positive SC treatments	0.014*	0.011	0.007
	(0.007)	(0.007)	(0.007)
Negative SC treatments	−0.006	−0.002	−0.006
	(0.011)	(0.011)	(0.011)
Positive LC treatments	0.088*	0.087*	0.088*
	(0.006)	(0.006)	(0.006)
Negative LC treatments	−0.051*	−0.052*	−0.051*
	(0.006)	(0.006)	(0.006)
Tau1	−1.013*	−0.986*	−1.044*
	(0.079)	(0.083)	(0.063)
Tau2	−0.079	−0.061	−0.118
	(0.079)	(0.082)	(0.063)
Observations	10,198	9259	9259

Note: The dependent variable is ordered and scored 1 when a lower court deviates from a precedent, 2 when it treats a precedent neutrally, and 3 when it treats the precedent positively. Ordered logistic regression is used to fit the models. Robust standard errors are in parentheses.

*$p < 0.05$. Model 1 replicates the results presented by Westerland et al. (2010: 898, Table 3); Model 2 replicates those results on the sample of cases for which we have latent complexity scores; and Model 3 replaces the SCDB complexity measure with our predecision latent measure.