Taking the Measure of Congress: Reply to Chiou and Rothenberg
Author(s): Sarah A. Binder
Source: *Political Analysis*, Vol. 16, No. 2 (Spring 2008), pp. 213-225
Published by: Oxford University Press on behalf of the Society for Political Methodology
Stable URL: http://www.jstor.org/stable/25791929
Accessed: 10/07/2014 22:13

Advance Access publication February 6, 2008

Political Analysis (2008) 16:213–225
doi:10.1093/pan/mpm033

Taking the Measure of Congress: Reply to Chiou and Rothenberg

Sarah A. Binder
The Brookings Institution and George Washington University,
1775 Massachusetts Avenue, NW, Washington, DC 20036
e-mail: binder@gwu.edu

Chiou and Rothenberg raise important questions about how to measure key concepts in the study of legislative stalemate in the U.S. Congress. In challenging my choice of measures to capture bicameral differences, Chiou and Rothenberg argue that my findings are the artifact of measurement error. In this reply, I review the hurdles involved in measuring policy views over time and across institutions and suggest that the preferred measure of Chiou and Rothenberg falls short for measuring bicameral differences. Second, I assess the extent to which measurement choices affect the robustness of my findings about the determinants of gridlock. Drawing on new measures and model specifications, I show that my results are robust to alternative specifications. I conclude with an assessment of the broader challenges posed by how we measure critical concepts in the study of congressional performance.

1 Introduction

Fang-Yi Chiou and Lawrence S. Rothenberg offer a valuable perspective on the difficulties of testing theoretical conjectures about the forces that drive patterns in American national politics. Our ability to explain the variation we observe in political arenas is strongly shaped by the validity and reliability of the measures we devise. In "Comparing Legislators and Legislatures: The Dynamics of Legislative Gridlock Reconsidered," Chiou and Rothenberg appropriately point out the high stakes of getting measurement right: Our understanding of the forces that drive congressional performance will be impaired if we choose patently inferior measures.

In this response, I take seriously the objections raised by Chiou and Rothenberg to measurement decisions in my initial work on legislative gridlock that appeared in the *American Political Science Review* in 1999. I do so in two ways. First, I review the challenges involved in selecting measures that capture preferences over time and across institutional venues. Chiou and Rothenberg's preferred Common Space measure, I suggest, avoids only some of the pitfalls involved in measuring preferences in these contexts. Second, I assess the extent to which measurement choices affect the robustness of my findings about the determinants of gridlock. Drawing on new measures and model specifications devised for a book-length treatment of stalemate (Binder 2003), I show that my initial results are not an artifact of measurement error and are in fact robust to alternative specifications. I conclude by reviewing the broader challenges posed by imperfect measurement of critical concepts in the study of congressional performance.

2 Contested Measures

Chiou and Rothenberg raise several challenges to my published work on legislative gridlock in which I show that electoral and institutional forces shape congressional performance: Divided government, the decline of the political center, and bicameral differences over public policy increase the likelihood of stalemate (Binder 1999, 2003). Their "principal quarrel" targets my use in 1999 of W-NOMINATE scores pioneered by Keith Poole and Howard Rosenthal (Chiou and Rothenberg 2008a). Applying such scores in the context I use them, Chiou and Rothenberg argue, introduces problems of measurement error. Instead, Chiou and Rothenberg advocate use of Poole's "Common Space" scores. Reanalyzing the determinants of legislative gridlock by substituting Common Space measures for W-NOMINATE measures, Chiou and Rothenberg conclude that the statistical results in Binder (1999) are an artifact of measurement error that stems primarily from the use of W-NOMINATE scores to measure bicameral differences in policy views. Chiou and Rothenberg suggest that much of what we think we know about congressional performance is not true. Contrary to a range of recently published work (see e.g., Coleman 1999; Tsebelis 2002; McCarty, Poole, and Rosenthal 2006), Chiou and Rothenberg suggest that party control and ideological alignments across parties and institutions have little bearing on congressional performance on major issues of the day.

In reestimating my gridlock models using Common Space scores to measure congressional preferences, Chiou and Rothenberg raise several additional concerns that flow from their analysis. First, finding no impact of institutional, electoral, or partisan forces, Chiou and Rothenberg question the dependent variable in Binder (1999, 2003), a measure that captures the frequency of gridlock as a function of both the number of stalemated issues (the numerator) and the agenda of potential enactments (the denominator). Second, Chiou and Rothenberg raise doubts about the alternative measure of bicameral differences that I develop in *Stalemate* to eliminate reliance on W-NOMINATE scores (Binder 2003). They doubt its usefulness as a method of capturing preferences across different institutional venues.

Before turning to the broader issue of how to measure preferences across institutions and over time, it is helpful to clear away some of the underbrush that lies between our two analyses. Chiou and Rothenberg compare my analysis in Binder (1999, Table 3) with their new analysis that appears in their Table 1 (Chiou and Rothenberg 2008a). In their analysis, Chiou and Rothenberg revise the independent variables by substituting Common Space scores for W-NOMINATE scores; as the authors point out, Common Space scores were just becoming available when Binder (1999) appeared in print. When Common Space scores are substituted, Chiou and Rothenberg find little support for my conjectures about the forces that drive variation in the frequency of deadlock.

In publishing *Stalemate* (Binder 2003), I too conducted a wholesale revision of the analysis that appeared in the *APSR* in 1999. I revamped the variables and estimation in several ways. First, in revising my statistical model, I eliminated the variable that tapped the impact of the ideological diversity of Congress.[1] As I noted at the time (see Binder 2003, 193, note 25), the variable suffered from a measurement problem since it relied on W-NOMINATE scores to measure preferences across chambers and the variable did not help to test my theoretical interests in the impact of party control, moderation, and bicameralism on legislative outcomes. For these two reasons, the variable was not included

[1]Chiou and Rothenberg (2008b) observe that the variable is missing from the results reported in Table 1.

Table 1 Estimating the dynamics of legislative gridlock with alternative measure of bicameral differences

	(1) Frequency of gridlock on all salient issues	*(2) Frequency of gridlock on only salient "first-dimension" issues*
Divided government	.421 (0.154)**	.442 (0.170)**
Partisan moderation	–.016 (0.006)**	–.013 (0.006)*
Bicameral differences (all conference reports)	7.501 (3.626)*	–
Bicameral differences (first-dimension conference reports)	–	8.017 (3.919)*
Time out of majority	–.050 (0.035)	–.039 (0.038)
Budgetary situation	.017 (0.011)	.020 (0.012)
Public mood (lagged)	–.004 (0.015)	–.008 (0.016)
Constant	–.160 (0.985)	–.093 (1.081)
N	24	24
F	3.64	2.78
Adjusted R^2	.408	.318

Note. Cell entries are weighted least-squares logit estimates for grouped data (standard errors in parentheses). $*p < .05$, $**p < .01$ (one-tailed t-tests). Estimates generated via Stata 8.0's *glogit* routine.

in the estimations reported in Binder (2003), and thus, it does not appear in the estimations reported in Table 1.

Second, in revising the statistical models in Binder (1999), I devised a new measure of bicameral differences in Binder (2003) that does not rely upon a general ideological score to capture differences in preferences in two different institutional contexts. Instead, as I explain in detail below, the measure uses chamber voting on conference reports to isolate and compare the two chambers' positions on final votes on identical proposals. This move eliminates the need to choose between variants of NOMINATE scores, neither of which are perfectly suited to capture bicameral differences. Third, in additional analysis (see Binder 2003, Table 4.4), I revised my measure of the filibuster threat by replacing W-NOMINATE scores with the DW variant of NOMINATE scores that allows for over-time comparison of senators' preferences.

Only one variable in Binder (2003) still uses W-NOMINATE scores to measure congressional preferences: the measure of partisan moderation. This variable measures the mean percentage of legislators who are located closer to the ideological midpoint between the two parties than to their own party median.[2] Chiou and Rothenberg acknowledge that in this context "W-NOMINATE only has to serve as an ordinal measure, lessening the amount of comparability-induced measurement error" (2008a). In other words, the way W-NOMINATE scores are used here mitigates the problems associated with a static preference measure.[3]

[2]More precisely, I divide the percentage of moderates by the distance between the party medians to account for movement in party medians over time. After calculating separate House and Senate moderation scores, I average the two series to produce a single partisan moderation score for each Congress.

[3]Chiou and Rothenberg (2008b) observe that the variable measuring partisan moderation in Binder (1999) was revised in Binder (2003). As I observed at the time, the revised measure of moderation does not appreciably affect the statistical results (see Binder 2003, 175, note 19, and 193, note 26). When I rerun the analysis that appears above in Table 1 (column 1) using my 1999 measure of partisan moderation, the only noticeable change is the statistical significance of the coefficient for bicameral differences, which falls from $p < .05$ to $p < .1$ (one-tailed test). The statistical significance of the coefficient for partisan moderation is unaffected. The results in Table 1 (column 2) are unaffected by using the 1999 measure. I revised the measure in response to suggestions raised by a reviewer of Binder (2003).

When I reestimate the model with the new measures, the primary results from my original analysis are sustained (see Table 1). My findings in the 1999 *APSR* piece, in other words, are not artifacts of measurement. The frequency of deadlock increases in the presence of divided government, with increasingly polarized Congresses, and with growing bicameral policy differences. Because of the improvements in the specification of the models that I published in 2003, the appropriate comparisons to the new analysis of Chiou and Rothenberg are the models in *Stalemate* (2003) rather than those that appeared in the *APSR* in 1999. The key difference between Binder (2003) and Chiou and Rothenberg (2008a) is thus how we measure bicameral differences over time: Chiou and Rothenberg agree with my assessment published in 2003 that the deficiencies of W-NOMINATE scores necessitate a new way of measuring bicameral differences. Thus, I concentrate the discussion here on a consideration of the challenge of capturing interchamber differences in policy views.

3 Congressional Preferences in Time and Space

Chiou and Rothenberg advocate the use of Common Space scores, particularly as a means of measuring bicameral differences. Noting that W-NOMINATE scores are "not intended for comparisons between chambers and across Congresses" (Chiou and Rothenberg 2008a), Chiou and Rothenberg identify Common Space scores as "a new measurement solution for the task of across-chamber and over-time comparison." Are Common Space scores the solution for this particular measurement challenge? To answer this question, a closer assessment of the measurement challenge is in order, as well as a consideration of the advantages and disadvantages entailed in the use of Common Space scores and alternative measures.[4]

The key challenge in capturing bicameral differences over an extended period of time is to devise a common policy space over which to observe legislators' policy views across chambers and over time. Absent such a common policy space, we cannot be sure of the comparability of what we measure to be legislators' preferences in two different chambers or at two different points in time. Bailey (2007) articulates the dilemma: "Simply put, no matter how well preferences are estimated within an institution, they are not comparable across institutions without clear points of reference" (434). Both Herron (2004) and Bailey (2007) observe the limitations of using NOMINATE-based scores to compare legislators' policy views over time. Even Common Space scores—designed to ensure comparability over time—are potentially limited in their applications unless we can be certain that there is "no fundamental change in the underlying voting space" (Poole 2005, 139). How vote choices at any given time map onto the underlying policy dimension is the critical issue. If we suspect that the agenda of issues that comes to a vote changes over time or differs across chambers, then it may be difficult to discern the meaning of differences in ideal points when comparing legislators' ideal points in different congresses or chambers.

[4]My concerns about Common Space scores apply to their use in measuring bicameral differences over decades of congressional history. My views echo those of Keith Poole (2004) when he notes the limitations of the assumptions underlying the calculation of Common Space scores. The scores may be useful for locating legislators within a chamber and within a narrow time frame. Thus, in Murphy, Binder, and Mann (2004), Common Space scores are used to spatially locate the 2004 Democratic and Republican presidential and vice presidential candidates who had served in the Senate or who had taken positions on Senate roll call votes (i.e., the incumbent president, George W. Bush). Granted, Vice President Richard Cheney is also plotted on the same scale, although the text emphasizes the ideological fit of Cheney to the Republican House colleagues with whom he had previously served.

Clearly, there are advantages to using Common Space scores to measure bicameral differences. First, there is no shortage of votes over which to calculate legislators' preferences. Because Common Space coordinates are generated from all but near-unanimous and unanimous votes, several hundred votes per Congress are typically available for scaling, whereas thousands of votes are available for longer serving legislators (whose entire voting history in Congress is incorporated into the calculation of a single career-long Common Space coordinate).

Second, Poole (1998) does attempt to create a common reference point around which to calibrate legislators' ideal points. To do so, he assumes that legislators have fixed preferences over time and, for those who have served in both the House and Senate, across chambers. Unlike the DW variant of NOMINATE that allows legislators' preferences to move in a linear fashion over time, Common Space scores rely on the assumption of fixed preferences over time and across institutions as a "glue" to tie the two chambers into a single ideological rank ordering. Common Space scores are the "industry-standard" for researchers seeking a readily available metric with extensive cross-sectional and over-time reach for comparing legislators' ideal points.

Are Common Space scores appropriate for every application? Recent analysis of Common Space scores suggests otherwise. As Bailey (2007) demonstrates, the mapping of policy issues onto the underlying second dimension (related to civil rights and liberties) of Poole and Rosenthal has clearly changed over time. If we assume that the mapping of policy has remained constant over time, then Common Space coordinates offer some interesting anomalies. Comparing the Common Space scores for pro-segregationist senators from mid-century (such as Democratic senators James Eastland and Theodore Bilbo of Mississippi) to those of modern southern Democrats (such as Ernest Hollings of South Carolina or Charlie Stenholm of Texas) reveals that their coordinates on both the first and second dimensions are markedly similar. But southern Democrats today, like Hollings, are clearly more moderate on racial issues than their southern segregationist forebears.

Consider, for example, Senator Hollings' second-dimension coordinate (.614). As Bailey (2007, 437) observes, Hollings' second-dimension Common Space coordinate is quite similar to that of Eastland (.736), who asserted on the Senate floor in 1945 that "the Negro race is an inferior race. . .." Hollings's coordinate is even closer to that of Harry Byrd (D-Virginia) (.680), who argued that we should "exclude the Negro population" from voting. In contrast, Hollings voted with liberals on several civil rights measures, including the Voting Rights Act of 1982 and the Civil Rights Restoration Act of 1988. What it means to be liberal or conservative on racial issues has clearly changed over the past half-century, meaning that floor votes on racial and related issues today do not map onto the second dimension as they did 40 years ago. The first dimension is also likely susceptible to this problem. Common Space scores cannot detect any changes in the underlying agenda or in the mapping of issues onto the left-right dimension. If the meanings of the liberal and conservative labels shift over time, then comparisons of Common Space scores over the postwar period may fail to reliably measure preference change.

Poole (2004) is quite clear about the assumptions underlying Common Space scores. In fact, he is careful to point out that the "assumption that a legislator serving in both Chambers of Congress has the same coordinate throughout his or her career means that these coordinates should be used with caution" (Poole 2004). Poole's caution about the assumption of fixed preferences across the two chambers bears further consideration. As I show elsewhere (Binder 2003, 144), we can evaluate the fit of the assumption of fixed ideologies by comparing the voting behavior of legislators who served in both chambers within a given Congress. By isolating such legislators within a single Congress, we can

attempt to hold constant the issues on which legislators cast votes. And given different constituencies for a House and Senate member from the same state, we might expect to see different voting patterns in each chamber. Different floor rules in the two chambers—which would affect the types of proposals that come to a vote—might also affect legislators' revealed preferences as they move from the House to the Senate.

Because the issues that map along the left-right dimension in the contemporary period tend to match partisan divisions over the agenda (Poole and Rosenthal 1997), we can use legislators' party loyalty scores in the Congress in which they served in both chambers to compare their behavior in the two chambers. These scores provide considerable evidence that one's revealed preferences are likely to change when moving from the House to the Senate. Robert Griffin (R-Michigan), for example, shows an 80% party loyalty score for his House service in the 89th Congress, compared to a 72% party loyalty score for his Senate service in that Congress. Small-state senators also exhibit considerable change in their party loyalty, such as Quentin Burdick of North Dakota, whose party loyalty went up 12 percentage points when he switched chambers in the 86th Congress. We can infer from changes in party loyalty within a single Congress that legislators' revealed preferences are unlikely fixed across the two chambers.

Chiou and Rothenberg dismiss such evidence of cross-chamber variability in preferences, noting that party loyalty scores may be a function of other forces such as institutional rules that alter the set of votes on which we observe legislators' loyalty. This is precisely the dilemma that arises in assuming fixed ideologies and treating the array of House and Senate votes as reflecting revealed preferences on a shared policy space. As Chiou and Rothenberg recognize in thinking about comparisons of party loyalty scores, we cannot assume identical choice situations across the sets of votes that go into the calculation of Common Space scores. We might instead imagine that the distribution of cut-points and status quo points associated with different votes in the House and Senate would vary across the two chambers, even as the ideological alignment of House and Senate members might not change. In other words, unless we can fix a reference point between the two chambers, we cannot hope to detect ideological differences based only on information about voting alignments (Bailey 2007). The differences in party loyalty scores in two different institutional contexts are suggestive that cut-points in the two chambers' votes may differ substantially, differences that cannot be detected with a general ideological score that assumes legislators' ideologies are the same in both chambers.

4 Alternative Measure of Bicameral Differences

Given both the advantages and disadvantages of using Common Space coordinates to measure bicameral differences in policy views, it seems reasonable to evaluate the pros and cons of the alternative measure I proposed in Binder (2003). Instead of measuring preference differences with a general ideological score, the alternative measure uses floor votes on conference reports to capture basic differences in the two chambers' propensity to support identical policy proposals with identical status quo points. I argue here that the benefits of this measure outstrip its weaknesses and strongly make up for the deficiencies of measures based on general ideological scores.

Rather than assuming that the agenda of issues considered by both the House and Senate of each Congress is identical, I isolate the set of conference reports considered by both the House and Senate in each Congress. As Chiou and Rothenberg point out, the disadvantage of this method is the small number of votes on conference reports in several of the earlier Congresses in the time period (1947–2000), even including voice votes. Still,

the advantages of the measure are several. Moreover, some of the limitations of the measure as suggested by Chiou and Rothenberg do not tarnish its validity or reliability.

First, by looking at voting only on conference reports, I fix the range of bill proposals over which legislators cast votes, freezing identical choice situations that pit identical proposals against identical status quos. With the underlying policy space fixed, I can calculate a straightforward measure of the percentage of legislators approving the conference report in each chamber and then calculate the difference.[5] The measure is thus directly interpretable as a proxy for the difference in the chambers' central tendencies on policy issues in any given congress. If increases in the differences in the average policy stances of the two chambers over time are consequential in explaining increases in the frequency of gridlock, we should see a statistically significant relationship between increases in bicameral differences and the frequency of deadlock. The strong suit of the conference report measure is its ability to identify identical choice situations in the two chambers, eliminating the need to assume fixed preferences of legislators across careers and in different chambers with different sets of constituencies.

Second, Chiou and Rothenberg contend that the measure "is almost certainly not measuring preference differences exclusively." The grounds on which they reach that conclusion do not hold up well under scrutiny. I consider each of their concerns in turn.

4.1 *Correlation with Other Measures*

Chiou and Rothenberg raise the concern that the conference report measure does not correlate neatly with measures of bicameral differences based on first-dimension W-NOMINATE scores or Common Space scores. Putting aside the measure based on W-NOMINATE scores, since both Chiou and Rothenberg and I dismiss it as a robust measure of bicameral differences, I find an interesting relationship between the measure of bicameral differences based on conference reports and the measure based on first-dimension Common Space scores. The conference report and Common Space measures are strongly correlated before the 1970s, moving in tandem with a Pearson's r correlation of .87 (see Fig. 1a). For the period after the early 1970s, the relationship falls apart, showing a −.43 correlation in Fig. 1b. Because our two measures are highly correlated before 1970, my finding that bicameral differences encourage stalemate is primarily driven by the observations in the period after 1970. Divergence in the two measures after 1970 thus becomes critical to explaining the different results that Chiou and Rothenberg and I reach.

Why is the correlation of the two measures so much weaker after 1970? Several potential explanations eventually need to be sorted out. One possibility is that the change in the relationship of the two variables is due to my inclusion of voice and recorded votes in determining chamber preferences on conference reports, as Chiou and Rothenberg (2008a, 2008b, note 17) suggest.[6] Interestingly, between 1947 and 1970, voice votes make up on average 94% of the conference report votes each congress; between 1971 and 2000, the average percentage drops to 71%. If my decision to treat voice votes as unanimous chamber agreement is unreasonable, the decline in the proportion of voice votes suggests

[5]Chiou and Rothenberg (2008b) argue that calculating the average of the absolute difference in House and Senate support for each conference report requires an "implausible assumption" that the conference report votes should be equally weighted. Devising a measure that accounts for variation in issue salience across votes in calculating preferences is an important next step in the use of roll call votes to locate legislators' preferences.

[6]There are fewer than seven recorded votes in both chambers on conference reports between the 80th and 88th Congresses, with the exception of the 11 recorded votes in both chambers in the 81st. Including voice votes and considering them as unanimous chamber agreement is a necessity given the paucity of recorded votes in both chambers in the early congresses.

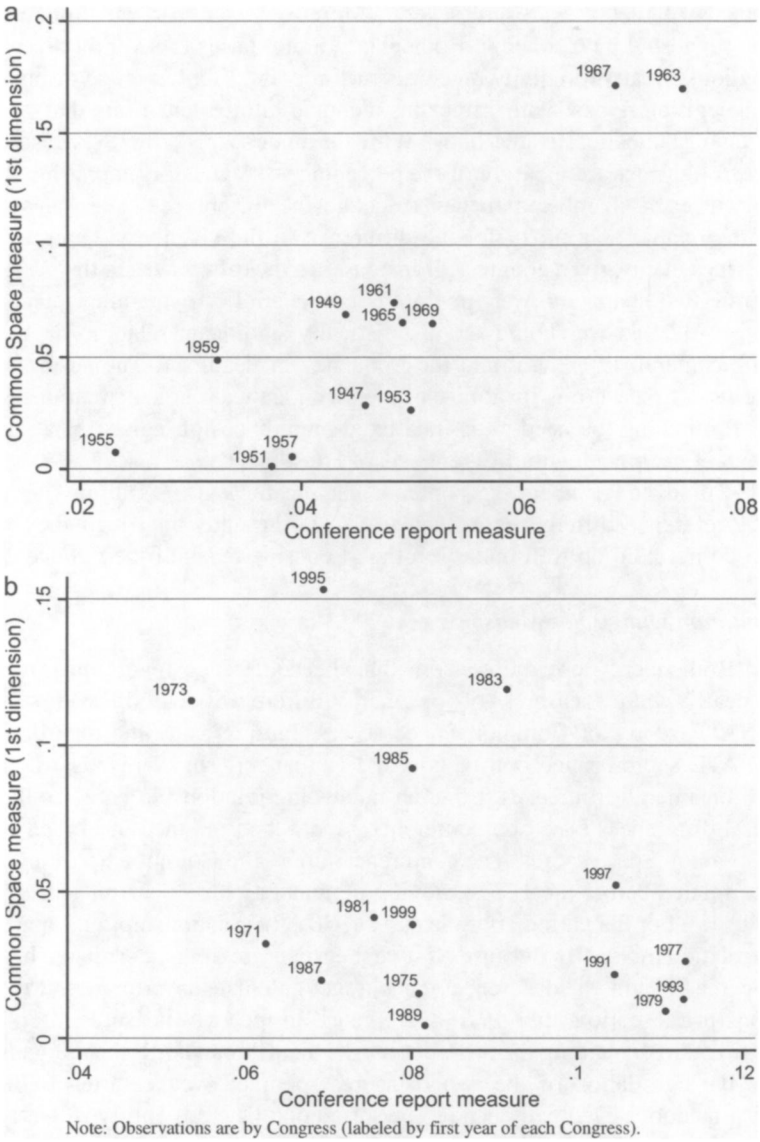

Fig. 1 (a) Comparing measures of bicameral differences, 1947–1970. (b) Comparing measures of bicameral differences, 1971–2000. For both (a) and (b) observations are by Congress (labeled by first year of each Congress).

that my measure of bicameral differences is better suited to the period after 1970 than before. Given that the two measures of bicameral differences are strongly correlated in the period before 1970, this suggests that neither would be particularly well suited to detecting bicameral differences before 1970.

Another possibility is that something happened in the early 1970s to affect the comparability of Common Space scores over the long period of history studied here. Three concerns are paramount. First, Poole (2004) urges caution with the use of Common Space scores when he notes that "The Common Space coordinates are adjusted so that the two

dimensions are *equally salient and lie within a unit circle* [emphasis in original]." Given the effective disappearance of the second (racial) dimension starting in the 1970s, comparing Common Space scores before and after that period may be problematic and may help account for the lack of correlation between the two measures of bicameral differences after 1970.

Second, as Roberts and Smith (2003) and Smith (1989) demonstrate, the composition of the roll call record was affected by the decisions in 1970 to allow recorded votes in the Committee of the Whole (COW) (the primary location for voting on amendments) and in 1973 to introduce electronic voting for COW and House votes. The absence of identical rule changes in the House and Senate and the addition of more partisan-recorded amendment votes in the House roll call record starting in the early 1970s (as well as agenda changes over the longer period) may affect the comparability of Common Space across chambers over time.

Third, the assumption of fixed legislator preferences over time may be problematic when using Common Space scores over the full time series for bicameral comparisons. The 1970s featured large shifts in public opinion, as well as high levels of turnover in Congress. The impact of turnover, however, was more dramatic for the House than for the Senate. Almost half of the House members serving in 1975 had not been in office at the start of the decade, whereas three-quarters of sitting senators that year had been a member of Congress at the start of the decade. Given House turnover, given the movement of House members to the Senate and not vice versa, and given that Common Space scores are based on entire voting careers, House Common Space scores may better capture preferences than do Senate scores at this point—particularly if legislators' preferences are less stable than assumed by Common Space.

Chiou and Rothenberg infer that the break in the correlations in Fig. 1 stems from the deficiencies of the conference report measure. Without a better sense of the cause of the breakdown, the inference may be premature. There are several potential explanations for the weak correlation of the two measures, suggesting that we lack a perfect measure to capture bicameral differences over a long stretch of political history.

4.2 *Strategic Voting Contamination*

Chiou and Rothenberg raise an additional charge about the inclusion of voice votes on conference reports, suggesting that the unanimity assumption might be wrong. Instead, they suggest that voice votes could represent strategic decisions by the minority party to avoid recorded votes on measures they oppose, but expect to pass. The U.S. Constitution and Senate rules, however, are biased in favor of the casting of roll call votes and against the ability of a minority to block them. As detailed in Article 1, Section 5 (the "Rule-making" clause), of the Constitution, the yeas and nays of the members of either chamber on any question can be demanded if one-fifth of those present agree.[7] So even if a minority in either chamber wished to avoid a floor vote, a small group of majority party members could spoil that plan. The minority coalition on adoption of a conference report is unlikely to be able to avoid a floor vote, increasing the likelihood that the unanimity assumption is correct.

Chiou and Rothenberg are implicitly concerned about the potential mix of strategic and sincere voting on adoption of conference reports. By isolating votes on conference reports, however, I have radically reduced the likelihood of incorporating strategic voting in

[7]Given that a majority of the Senate constitutes a quorum, a minimum of 11 senators would be required to secure a recorded floor vote. In the House, with a quorum of 218, the assent of 44 members is sufficient to secure a recorded vote.

calculating chamber positions on conference reports. Up or down votes on adoption of conference reports are votes on final passage. As Krehbiel, Meirowitz, and Woon (2005) make clear, "only once voting reaches its last stage does it simplify to a binary choice, when we can then be certain that sophisticated voting does not occur."[8] In contrast, NOMINATE-based measures incorporate all types of roll call votes, including amendments on which sophisticated voting might be expected. As Krehbiel, Meirowitz, and Woon (2005, 9) thus conclude, "... tests using Nominate (or other large-N) ratings must be treated with suspicion, too, because our usual interpretations of the recovered ideal points and cutpoints rest on an assumption that the inputs for the rating scheme are sincere votes." Given the increase in amendment votes that appear in the roll call record after the 1970s (Roberts and Smith 2003), the conference report measure (in comparison to the Common Space coordinates) provides greater certainty that the measure reflects legislators' sincere preferences over the five decades studied here.

4.3 *Dimensionality*

Chiou and Rothenberg raise the concern that the conference report measure of bicameral differences necessarily mixes first- and second-dimensional issues. This is potentially more of a concern for the data before the 1970s, given the essential disappearance of the second dimension after that period. Still, in order to reduce cross-contamination of the dimensions and related voting alignments, I recalculated the bicameral difference measure including only conference reports that addressed first-dimension issues. Only 21 of the more than 2900 conference reports considered between 1947 and 2000 were measures that primarily addressed issues of civil rights or civil liberties. Thus, the impact on the measure of bicameral differences is essentially imperceptible; not surprisingly, the correlation of the two measures (one including all conference reports and one including only first-dimension conference reports) is .9987. I also purged the dependent variable of all issues related to civil rights and civil liberties. When I rerun the analysis (see Table 1, column 2), the estimates are essentially the same.[9] Bicameral differences still strongly help account for variation in the frequency of deadlock, and they do so across all five levels of issue salience identified in Binder (2003). The mixing of first- and second-dimension issues does not undermine the validity or reliability of the conference report measure.

4.4 *Preferences and Cut-points*

Chiou and Rothenberg raise one final objection to the conference report measure, observing that very strong assumptions have to be made about the distribution of cut-points, proposal locations, and status quo points in order for the conference report measure to capture true preference differences. Chiou and Rothenberg are certainly correct that the conference report measure does not allow us to pinpoint voting alignments spatially, as does a measure that incorporates ideal points estimated from voting positions.

That said, it is not clear that comparisons of legislators' NOMINATE scores allow us to detect differences in the locations of cut-points in two different sets of votes. As noted

[8]Chiou and Rothenberg (2008b) suggest that voting on conference reports might also be strategic: Legislators' votes on conference reports could be shaped by anticipation of votes on future conference reports. The theoretical basis for the conjecture remains to be elaborated.

[9]The dependent variable includes all issues (except civil rights and civil liberties) on which the *New York Times* wrote four or more editorials in a single Congress. The bicameral differences variable is generated from only conference reports on first-dimension issues. Results from the additional gridlock models—calculated at varying levels of issue salience—are available on request from the author.

above, when Roberts and Smith (2003) compare NOMINATE scores based on votes cast in the COW (which are votes on amendments) to NOMINATE scores based on votes in the full House (which are primarily final passage votes), they find markedly little difference in legislators' NOMINATE scores.[10] Noticeably, however, party cohesion scores differ quite markedly over the two types of votes. Moreover, when Roberts and Smith (2003, 311–2) compare the location of cut-points in the two types of votes, they find a large difference in the proportion of cut-points located between the party medians for the two types of votes. The different pattern in cut-points appears to shape the frequency of partisan divisions in the two contexts, "even if members' underlying policy positions are the same in the two procedural settings" (Roberts and Smith 2003, 311).

Chiou and Rothenberg recognize this possibility in their Fig. 2. But it is not clear that measures of the alignment of legislators' ideal points can capture differences in the makeup of voting coalitions that underlie votes on conference reports. Indeed, when I recalculate chamber medians from Common Space scores based solely on final passage votes, these new chamber medians are very strongly correlated with the chamber medians based on the traditional Common Space scores calculated over nonunanimous votes.[11] Measures of the spatial alignment of legislators do not appear to be able to detect differences in the coalitions that emerge with different cut-points. Because the conference report–based measure ensures that we are comparing votes with identical cut-points and status quo points, it seems that a comparison of voting propensity in favor of conference reports may more accurately detect differences in the policy views of the House and Senate than do fixed Common Space coordinates that assume away institutional differences across chambers and ideological movement over time.

5 Implications and Conclusions

> Although political methodologists are well aware of measurement issues and the problems that can be created, such concerns are not always front and center when we are doing substantive analytic research (Chiou and Rothenberg 2008a).

Chiou and Rothenberg divide the discipline into methodologists and substantive scholars. I would instead suggest that substantive scholars of Congress and legislative politics do focus on getting measurement right. Our ability to explain variation in the political world is only as strong as our ability to design measures that closely capture key political concepts. There is, of course, rarely a perfect measure, and thus, the selection of measures often requires a tradeoff. What are the advantages of one measure over another? Do the drawbacks of a popular measure outweigh its advantages?

Given my need to capture cross-institutional and over-time variation in legislators' policy views, the difficulties of devising appropriate measures loom large. For the array of reasons explored here, it appears that a bicameral measure based on vote outcomes on conference reports outperforms a measure based on Common Space coordinates. Most important in comparing policy differences across institutions and over time is to find a common reference point to bridge the institutions (Herron 2004). Common Space scores do this by using the legislators who served in both chambers as the glue, making the assumption that legislators' ideologies are the same in both institutions and over time. Bailey's (2007) "XTI" approach uses position taking and voting on Supreme Court cases

[10] According to Roberts and Smith (2003, 311), the correlation between the two sets of DW-NOMINATE scores (those based on COW votes and those based on House votes) is above .97 for all Congresses since 1971.

[11] I appreciate Keith Poole's willingness to calculate the new set of Common Space scores based only on final passage votes in the two chambers.

by members of Congress and presidents as a method of relating civil rights and civil liberties policy views of the justices to those of presidents and legislators.[12]

My approach, in contrast, uses sincere voting on identical conference reports in the two chambers to freeze the policy space and thus to directly compare contemporaneous decisions of the House and Senate on identical proposals, with identical status quo points. If the greatest challenge in measuring House-Senate differences is to create a common policy space over which to observe bicameral differences, then the conference report measure has a considerable advantage over a measure based on Common Space coordinates. The NOMINATE-based scores have many superb applications. Their underlying assumptions, however, limit their value in capturing bicameral differences over long periods of time. This is especially so for periods in which the salience of the two dimensions and the meaning of the second dimension vary considerably; such concerns encourage in this context use of voting on conference reports rather than Common Space coordinates.[13]

To be sure, as Chiou and Rothenberg (2008a, 2008b) argue, it would be preferable if there were sufficient numbers of recorded roll call votes on conference reports over the time period studied so that voice and recorded votes would not have to be tallied together. On balance, however, the array of advantages and disadvantages of the two measures seems to weigh heavily in favor of the conference report measure. Moreover, the impact of bicameralism when captured with my alternative measure makes stronger theoretical sense than the results obtained with the Common Space measure. In the former, increases in bicameral differences increase gridlock; in the latter, the two are negatively correlated. Following Tsebelis (2002), we should expect gridlock to rise as the two chambers diverge. Thus, I suspect that the conference report measure is better suited for estimating the frequency of deadlock.

Given the strength of the conference report measure and the limitations of the Common Space alternative, Chiou and Rothenberg's broader conclusions about my findings on legislative gridlock lack foundation. First, my findings are not the artifact of measurement error, since the results in Table 1 are robust to alternative measures. Second, although Chiou and Rothenberg are unable to generate similar results when they use Common Space–based measures and alternative measures of legislative gridlock (Chiou and Rothenberg 2008a, Table 3), the culprit is not necessarily the inclusion of a denominator in my measure of gridlock.[14] Most likely, the Common Space measures are affecting the analysis, rather than the inclusion of a denominator. A simple regression of the frequency of gridlock on divided government, after all, produces different results than a simple regression of the number of landmark laws enacted on divided government.[15] Our intuitions

[12]Chiou and Rothenberg (2008b) use Bailey's (2007) preference scores to devise a measure of bicameral differences. Bailey's measure, however, is estimated from data relating to civil rights and liberties and crime-related issues. As the evidence in Mayhew (1991, 207–213) illustrates, however, only 3.9% of the landmark laws enacted between 1947 and 1990 addressed such issues; similarly, only 4.8% of the salient issues discussed in my gridlock measure based on *New York Times* editorials targeted such issues.

[13]In the Appendix to Binder (2003, 141), I acknowledge Chiou and Rothenberg's (2008a) important point that my results hold with measures based on W-NOMINATE scores and conference report voting, but not with a measure based on Common Space scores (again, scores that were widely available in 2003, but not in 1999 when my original work was published).

[14]Chiou and Rothenberg (2003) recognize that "Binder's (1999) measure ... is likely the closest to the kind of systematic measure of gridlock level, allowing demand to vary rather than implicitly assuming that it is the same across periods" (512).

[15]More specifically, in the former, the coefficient for divided government is statistically significant in a one-tailed test at $p < .05$ (in a grouped logit estimation); in the latter, the coefficient for divided government is insignificant (in an ordinary least-square estimation).

about the forces that drive legislative outcomes seem to vary as we move our focus between alternative ways of conceptualizing congressional performance.

In my previous work on legislative stalemate, I have argued that conflict within Congress—both across the two chambers and between the parties—is as important in explaining patterns in the frequency of deadlock as is conflict between the branches. How partisan and institutional forces may block congressional action remains an interesting question worth evaluating with carefully tuned measures. Chiou and Rothenberg (2008a) have highlighted one of the most pressing challenges for students of American legislatures: devising the most appropriate means for "comparing political decision-makers in multiple institutions over time." Considerable progress has been made on this front in recent years; much innovation lies ahead.

References

Bailey, Michael A. 2007. Comparable preference estimates across time and institutions for the Court, Congress and presidency. *American Journal of Political Science* 51:433–48.

Binder, Sarah A. 1999. The dynamics of legislative gridlock, 1947–1996. *American Political Science Review* 93:519–33.

Binder, Sarah A. 2003. *Stalemate: Causes and consequences of legislative gridlock.* Washington, DC: Brookings Institution Press.

Chiou, Fang-Yi, and Lawrence S. Rothenberg. 2003. When pivotal politics meets partisan politics. *American Journal of Political Science* 47:503–22.

Chiou, Fang-Yi, and Lawrence S. Rothenberg. 2008a. Comparing legislators and legislatures: The dynamics of legislative gridlock reconsidered. *Political Analysis* doi:10.1093/pan/mpm021.

Chiou, Fang-Yi, and Lawrence S. Rothenberg. 2008b. The search for comparability: Response to Binder. *Political Analysis* doi:10.1093/pan/mpm031.

Coleman, John J. 1999. Unified government, divided government, and party responsiveness. *American Political Science Review* 93:821–35.

Herron, Michael C. 2004. Studying dynamics in legislators ideal points: Scale matters. *Political Analysis* 12:182–90.

Krehbiel, Keith, Adam Meirowitz, and Jonathan Woon. 2005. Testing theories of lawmaking. In *Social choice and strategic decisions: Essays in honor of Jeffrey S. Banks*, ed. David Austen-Smith and John Duggan. Berlin, Germany: Springer Publishers.

Mayhew, David R. 1991. *Divided we govern.* New Haven, CT: Yale University Press.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America.* Princeton, NJ: MIT Press.

Murphy, Alan, Sarah A. Binder, and Thomas E. Mann. 2004. Presidential candidates: Where do they stand? *The New York Times*, July 26.

Poole, Keith T. 1998. Recovering a basic space from a set of issue scales. *American Journal of Political Science* 42:954–93.

Poole, Keith T. 2004. Description of NOMINATE data. http://voteview.com/page2a.htm, Dated July 13, 2004 (accessed June 19, 2007).

Poole, Keith T. 2005. *Spatial models of parliamentary voting.* Cambridge: Cambridge University Press.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political–economic history of roll call voting.* New York: Oxford University Press.

Roberts, Jason M., and Steven S. Smith. 2003. Procedural contexts, party strategy, and conditional party voting in the U.S. House of Representatives, 1971–2000. *American Journal of Political Science* 47:305–17.

Smith, Steven S. 1989. *Call to order.* Washington, DC: Brookings Institution.

Tsebelis, George. 2002. *Veto players: How political institutions work.* Princeton, NJ: Princeton University Press.