In this article, we challenge the conclusion that the preferences of members of Congress are best represented as existing in a low-dimensional space. We conduct Monte Carlo simulations altering assumptions regarding the dimensionality and distribution of member preferences and scale the resulting roll-call matrices. Our simulations show that party polarization generates misleading evidence in favor of low-dimensionality. This suggests that the increasing levels of party polarization in recent Congresses may have produced false evidence in favor of a low-dimensional policy space. However, we show that focusing more narrowly on each party caucus in isolation can help researchers discern the true dimensionality of the policy space in the context of significant party polarization. We re-examine the historical roll-call record and find evidence suggesting that the low-dimensionality of the contemporary Congress may reflect party polarization rather than changes in the dimensionality of policy conflict.
1. INTRODUCTION

For nearly a generation, congressional research has advanced empirically using estimates of member “ideologies” generated from scaling analyses. Two questions that many who use scaled roll-call estimates would like to answer are: “How many dimensions are there?” and “What do the dimensions actually mean?” The standard answer is that there are between one and two dimensions (and today only one), and that the first dimension is a left-right ideological dimension that structures most of congressional politics (or at least roll-call voting).

In this article, we challenge this low-dimensionality conjecture and with it the further claim that multidimensional preferences map down to just one liberal-conservative dimension. We show that moderate to high levels of bimodality in the distribution of legislators’ preferences necessarily leads scaling procedures to suggest a single dimension – regardless of the true dimensionality of the policy space. That is, party polarization on the order of that found in the contemporary US Congress obscures the true dimensionality of the policy space.

To demonstrate how party polarization downwardly biases estimates of dimensionality, we conduct Monte Carlo experiments varying first, the true dimensionality of the policy space and, second, the distribution of legislators’ preferences. We scale these simulated roll-call matrices and show that, if the two parties polarize sufficiently along even a few policy dimensions, scaling procedures will estimate just one or two dimensions, whether there truly are one, twenty, or anything in between. Thus, finding that a low number of dimensions can explain a large proportion of the variation in the roll-call record does not imply that legislators are making decisions based on preferences arrayed in a low-dimensional space. At the macro level, these results suggest that existing analyses of the roll-call record provide ambiguous evidence in favor of low dimensionality. At the micro level, they raise questions as to how scaling estimates should be interpreted and utilized.

We then show through simulation that analyzing the dimensionality of roll calls within each of the two party caucuses provides information about the true dimensionality of legislators’ preferences even in presence of significant party polarization. (Indeed, it is precisely in the presence of significant polarization that focusing on each caucus separately is most informative.) This suggests a plan of attack in which the scaling of inter-party roll-call votes can be combined with scaling intra-party roll-call votes. The first yields insight into the extent of polarization between the parties on partisan issues, and the second illustrates the true complexity of the policy space.

We therefore re-examine the dimensionality of the US Senate, scaling only the intra-party roll-call record and show there is less evidence in favor of the low-dimensional conjecture than

---

1Throughout, we use the term “preference” to indicate the point in policy space each member acts to achieve through roll-call voting. We are agnostic as to whether these ideal points derive from personal beliefs, constituency and electoral pressures, or both. This might also be termed each members’ “induced ideal point.” See Section 3.1 for additional discussion of this issue.
is commonly supposed. Indeed, examining each caucus separately, the evidence in favor of the low-dimensionality conjecture is relatively weak.

*Implications for theories of political conflict:* While these results may seem narrowly methodological, we believe that they are nonetheless of significant theoretical and practical importance for two main reasons. First, at the broadest level, the low-dimensional conjecture is a critical assumption that informs many important theories, formal or otherwise, in the literatures on elections, legislative institutions, and inter-branch relations. Our results speak to the need for expanded attention to theoretical models of politics robust to assumptions about the number of dimensions. For instance, there is a dramatic difference in what spatial models say about politics if the space is or is not *exactly* one-dimensional. In one dimension there is a median voter. If, however, the space is perturbed even infinitesimally away from a pure single dimension, there is no median, and a great many results evaporate (Kramer 1973). Yet, models that are exceptionally fragile to dimensionality assumptions continue to proliferate in the literatures on elections, Congress, and inter-branch relations. In many cases, these assumptions are justified either implicitly or explicitly via references to the scaled roll-call analyses discussed below. Brady and Volden (2006), for instance, state that:

In addition to the above reasons to focus on the main policy dimension despite the possibility of multiple dimensions, there is strong empirical support for the existence of a main policy dimension for a number of issues. Poole and Rosenthal (1997) address the history of roll-call voting in the Congress and find that preferences along a single dimension can account for about three-fourths of the votes of members of Congress on a wide range of issues (Brady and Volden 2006, p. 9).

*Implications for the interpretation of scaled estimates:* Second, our results challenge common substantive interpretations of ideal points estimates produced for each member of Congress. While combining all roll-call behavior into a single unidimensional score is a useful data reduction technique, the resulting estimates may not correspond well with the substantive “ideological” meaning with which they are sometimes ascribed. We show that increased inter-party polarization on even

---

2While not strictly requiring a single dimension, nearly all applications of Romer-Rosenthal agenda setting are also based on exacting unidimensionality assumptions for the simple reason that they nearly always require a median voter to exist (Romer and Rosenthal 1978). Pivot point models are in the same category (e.g., Krehbiel 1998). Moreover, many derivations of Duvergerian-style results (Palfrey 1989), prominent models of elections and government formation under proportional representation (Austen-Smith and Banks 1988), informational models of Congress (Gilligan and Krehbiel 1989; Krehbiel 1992), and others (e.g., Iversen and Soskice 2001; Persson and Tabellini 2000) require a very exacting form of unidimensionality. Many, if not all, of their derivations simply collapse if the unidimensionality assumption fails to the slightest possible degree (Kramer 1973). It is even the case that many of the results used to study $n$-dimensional policy spaces are built on repeated application of median voter logic (Shepsle and Weingast 1987; Laver and Shepsle 1990).
a few policy dimensions can lead to a dramatic distortion of the recovered space. Basically, scaled analyses of the entire record emphasize issues that divide the parties while underemphasizing those issues that divide one or both parties internally. This means that when scaling a legislative body that is deeply divided on just one or two issues, preferential distinctions on less polarized dimensions are swamped by the larger inter-party conflict. This leads to a conflation of dimensions such that legislator preferences in multiple distinct issue areas appear to map onto a single dimension – even when preferences on each dimension are distinct and uncorrelated.

Thus, relying on single-dimensional scaled roll-call scores may obscure the true nature of member preferences and political conflict itself. The estimated first dimension will represent the issues that divide the parties at any given moment with no necessary “ideological” meaning. What political science (and the media) call liberal and conservative may be whatever divides the parties and nothing more. Single-dimensional scores, therefore, will represent the positions of legislators on those few polarizing issues while preferences on less partisan dimensions will be obscured.

Substantively, this seems unsatisfactory. It leads to a confusing interpretation of ideology in Congress. No matter whether parties separate on a single policy dimension, a few, many, or all dimensions, it will all be lumped haphazardly in the term “ideology.” Ideology therefore means something different if the partisan cleavage happens to involve only economic and welfare policies, or includes civil rights, or includes abortion and family values, or even incidental policies or pork-barrel measures designed by leaders to serve as the basis for running for re-election.

Discussion: Our interpretation of roll-call estimates stands in contrast with the usual view that the first dimension is a liberal-conservative ideology, and that it is those preferences that are the most important causes of vote choices. From this standpoint, the very clear pattern of decreased dimensionality in recent Congresses is interpreted as ideology becoming more central to all decision-making, leading to high levels of partisanship and polarization. While that set of causal claims is consistent with the observed patterns in the scaled roll-call voting record, in this article we provide an alternative. Party polarization, even on a subset of underlying issue dimensions, results in scaled member preferences that appear increasingly unidimensional regardless of the true number of underlying dimensions. Thus, as parties have become stronger and more unified on a subset of issues, this has led to the illusion of reduced dimensionality.

Indeed, interpreting unidimensionality in the roll-call record as a result of increasingly powerful parties is consistent with other findings in the literature. For instance, many roll calls with no apparent ideological content but that divide the parties (e.g., distributive votes) map neatly onto the

---

3Thus, as we show below, it is not possible to distinguish between (1) a data-generating process where many dimensions are mapped down to just one due to constraint (Enelow and Hinich 1984; Hinich and Munger 1994) and (2) a data-generating process where many dimensions only appear to map onto one as a result of partisan teamsmanship on a subset of roll calls (c.f., Lee 2009).
single left-right dimension (Lee 2009). Moreover, this distortion of the policy space may explain why issues that were historically unrelated to the main left-right dimension map onto it completely, but only once those issues divide Democrats from Republicans (Karol 2009; Lee 2009). At one time, Civil Rights did not map onto the first dimension of conflict, but now it does (Carmines and Stimson 1989; Poole and Rosenthal 2007). The first dimension previously did not include abortion, but now it does (Adams 1997; Karol 2009). Our analysis suggests that these changes may simply be a result of the changing position of the parties rather than any more fundamental alteration in the relationship between these policies in the minds of members, the public, or anyone else. As Lee (2009) notes, “Any issue on which members of the two parties take opposing stands, whether or not it has any ideological content, will map on the first dimension...” (p. 52). Thus, significant changes in party position from conflict replacement (Schattschneider 1960) or conflict extension (Layman and Carsey 2002) may be masked as distinct policy dimensions, are subsumed into the broader “liberal-conservative” dimension once the parties divide sufficiently.

Thus, our results speak to the need for empirical scholars to increase attention to more aspects of the roll-call record than one-dimensional scaling scores. In the presence of the high levels of polarization that characterize the contemporary Congress, analyses of the first dimension will reveal factors that cause Democrats to differ from Republicans – and little else. A one-dimensional-dominant result may reflect party “teamsanship,” pure left-right ideology, or anything in between. Using standard empirical techniques, we can tell only that parties are divided from one another, but not if they are divided on one issue, many issues, or even, in the sense of Lee, none at all.

Theories that seek to uncover more fine-grained differences in legislative behavior must move beyond party and beyond explaining variation along the main Republican-Democratic axis. Yet many empirical analyses of legislative behavior and inter-branch negotiations, our own included, rely largely or exclusively on unidimensional scaling estimates (c.f. Krehbiel 1998; Cox and McCubbins 2005; Cameron 2000). This point applies equally to studies seeking to place voters and legislators on a single ideological dimension to explore linkages between voter attitudes and members roll-call votes (e.g., Bafumi and Herron 2010; Jessee 2009, 2010).

Nonetheless, we wish to emphasize at the outset that our aim is to introduce more caution and skepticism into the theoretical and empirical analysis of legislative bodies, not to attack any method of roll-call scaling in particular or roll-call scaling generally. Scholars have implemented various multidimensional scaling techniques, with Keith Poole and Howard Rosenthal’s work being the most well known (Poole and Rosenthal 1997, 2007). Their procedure represents a significant advance that fully deserves its popularity and widespread adoption. Led by Poole and Rosenthal’s own seminal research (e.g., Poole and Rosenthal 1991, 1997, 2007), the analysis of scaled roll-call estimates has spawned hundreds of path-breaking articles and books that have significantly

---

4Citations redacted for blind review.
advanced our understanding of Congress, inter-branch relations, elections, and democracy itself.\(^5\)

Yet, W-NOMINANTE and its numerous cousins\(^6\) are no more appropriate for answering every question in legislative research than are Likert scales appropriate for answering every question in political behavior. Like all measurement techniques, W-NOMINATE and its kin are based on specific assumptions that condition the scope of their applicability (c.f. Herron 2004; Clinton 2012). Our argument here is not that scaling procedures are “wrong,” only that they are ill equipped to uncover the true dimensionality and nature of legislators’ preferences in the context of polarization.

This rest of this article proceeds as follows. We begin by discussing past research on the dimensionality of Congress and approaches to estimating dimensionality in general. In Section 3, we present the details of our Monte Carlo simulations and then analyze the results. In Section 5, we analyze the post-war US Senate, focusing on the intra-party roll-call record.

2. DERIVING DIMENSIONALITY FROM ROLL-CALL SCALING

Our claim in this article is that the task of identifying the “true” number of dimensions in the Congress lies outside the scope of widely employed scaling procedures – at least as they have been applied to date. Or, to be more precise, the observation of a small number of dimensions resulting from the application of such scaling procedure to the entire roll-call record is insufficient to support the inference that the true number of dimensions is actually small. As we show, a surprisingly strong bias towards low dimensionality is present even when the stringent behavioral assumptions of the measurement models are met. When the preferences of two subpopulations are polarized on even a few dimensions, almost any number of true dimensions will appear to map onto just one dimension. Further, our simulations suggest that one approach to checking the true dimensionality of the space in the presence of polarized subpopulations is to examine roll calls within the polarized groups (i.e., within party caucuses). Before turning to our simulations and analysis, however, it is worthwhile to step back and examine common approaches to conceptualizing and evaluating dimensionality in Congress.

\(^5\)Poole and Rosenthal themselves have been careful to emphasize the limitations of a simple unidimensional model. While their empirical analyses illustrate the dominance of the first dimension in explaining roll calls in the contemporary era, their broader treatment of the historical record has rested more firmly on the two (or “one-and-a-half”) dimensional model (c.f., Poole and Rosenthal 2007).

\(^6\)See also Heckman and Snyder (1997); Clinton and Meirowitz (2001); Martin and Quinn (2002); Clinton, Jackman and Rivers (2004); Bafumi et al. (2005); and Poole (2005).
2.1. Does the basic space of Congress have one dimension or many?

Our challenge to the low-dimensionality claim may seem quixotic to some scholars of Congress because, at first blush, the evidence in favor of a simple political space appears so compelling. Consider, for example, the data displayed in Figure 1, which analyzes the empirical roll-call record in the US Senate from 1945 to 2010. The left panel shows the aggregate proportional reduction in error (APRE) associated with the first three dimensions of W-NOMINATE. The right panel shows the difference in APRE for each dimension, reflecting the marginal gains in explanatory power as we add dimensions to the W-NOMINATE model. In essence, these plots show that one or two dimensions alone are sufficient to explain most of the variation in the empirical roll-call record, especially for the contemporary Congress. The APRE for a one- or two-dimensional model is very high, approaching the maximum value of unity. Moreover, the marginal improvement in APRE for each additional dimension, shown in the right panel, is fairly modest throughout and shrinks towards zero. Indeed, as shown in the right panel, even adding the second dimension does not improve APRE in the contemporary Senate.

[Figure 1 about here.]

Recent challenges to the low-dimensionality conjecture: Despite this strong evidence, however, recent scholarship has raised serious doubts about the low-dimensionality of the roll-call record. Crespin and Rohde (2010) and Roberts, Smith and Haptonstahl (2009), for instance, analyze roll calls in specific issue areas and uncover substantial evidence in favor of a larger number of dimensions. Norton (1999) shows the same when focusing on roll calls related only to gender issues. This work parallels a growing body of research that investigates how violations of W-NOMINATE’s stringent assumptions result in systematic patterns of errors and misclassifications of roll-call votes. These findings suggest that NOMINATE scores and related measures are not detecting member preferences so much as providing a summary of each member’s observed behavior.

---

7APRE is a common metric for evaluating model fit in roll-call analyses. It ranges from zero to one, with larger numbers indicating superior fit. More precisely, the APRE for $n$ roll calls is,

$$APRE = \frac{\sum_{i=1}^{n} (\text{Minority vote size}_i - \text{NOMINATE classification errors}_i)}{\sum_{i=1}^{n} \text{Minority vote size}_i}$$

See Poole and Rosenthal (2007, p. 36-37) for additional details.


9In addition to those cited above, an incomplete list of recent empirical work in this vein would include Snyder and Groseclose (2000); Ansolabehere, Snyder and Stewart (2001); Cox and Poole (2002); Roberts and Smith (2003); Roberts (2007); Smith (2007); Masket (2007); and Patty (2008).
which is “endogenous to the legislative context” (Shepsle and Weingast 1994). The origins of that behavior are manifold, and might include members’ personal beliefs, but also voting procedures, pressure from party leaders, and the influence of outside actors. In total, this stream of scholarship does not cast doubt on the value of scaling procedures or their “correctness,” but rather suggests that caution is needed in interpretation, since roll-call behavior is not an unmediated reflection of a legislator’s ideology (c.f., Snyder 1992a; Clinton 2012).

A particularly important strain of research has focused on the role that party institutions and party leaders may play in altering the data-generating process to obscure the true dimensionality of the policy space (Dougherty, Lynch and Modonna 2012).

Hurwitz, Moiles and Rohde (2001) show evidence of multidimensionality in voting on agricultural appropriations, where party pressure is less likely to play a role. Jenkins (1999; 2000) uses data from the Confederate Congress to show that, in the absence of strong political parties, the structure, stability, and low-dimensionality of the roll-call record evaporates. Similar findings exist for state legislatures without strong two-party systems (Welch and Carlson 1973; Wright and Schaffner 2002). In a comprehensive review of the content of roll calls in the US Senate, Lee (2009) argues that much of the structure of scaling estimates – including the evidence supporting low dimensionality – is a result of partisan “teamsmanship” and may have little to do with members’ ideology.

Preference distributions and dimensionality: In this article, we make a different and somewhat broader point. Our aim is not to show that party institutions, committee gatekeeping, or agenda control influence the estimated dimensionality of legislators’ preferences. Rather, we want to show that scaling procedures will underestimate the true dimensionality of Congress in the presence of even moderately strong party polarization even when all other conditions are calibrated to maximize the probability of recovering the true dimensionality.

First, we show that the patterns in Figure 1, the patterns most commonly used to justify the low-dimensionality conjecture, are themselves consistent with the true number of latent policy dimensions being either small or large. The number of dimensions will be severely underestimated in the presence of the significant party polarization that characterizes the contemporary era, even if the parties polarize on only a small subset of dimensions. More pervasive and significant polarization will almost necessarily suggest a single dimension.

Second, we re-examine the empirical roll-call record stretching back nearly 70 years and focus on the record created within the two party caucuses. Our analysis shows there is little evidence in favor of a unidimensional Congress and suggests that it may be polarization rather than any fundamental change in the dimensionality of the political conflict that is responsible for the statistical patterns commonly interpreted as supportive of the low-dimensionality conjecture.

\textsuperscript{10}For a similar argument focusing on committee gatekeeping, see Snyder (1992a,b).
2.2. Establishing dimensionality

We begin by briefly considering some of the difficulties inherent in establishing the dimensionality of any data matrix. One possible method for examining the dimensionality of the roll-call record would be to draw on theory to specify conditions under which a certain number of dimensions affect voting. To the best of our knowledge, however, none of the widely applied scaling procedures takes the number of dimensions as a parameter derived from theory. Instead, it is only a maintained assumption that preferences are defined over a space with a finite number of dimensions.

Poole and Rosenthal, for example, embed their scaling in the spatial model of legislative behavior that has been the workhorse of positive theory since Black (1948), Downs (1957), and Enelow and Hinich (1984). They follow a standard set of assumptions ranging from the substantive (e.g., sincere voting) to the technical (e.g., preferences are defined over a policy space measured via a Euclidean metric). However, Poole and Rosenthal assume nothing about the dimensionality of the space other than that there are \( p \) dimensions, where \( p \) is some positive integer. To the best of our knowledge, the same is true for all other scaling procedures prominent in the literature. Our point here is not only to remind the reader of these well known considerations, but to remind the reader that the theory upon which most applications of scaling rely is not a source for addressing these questions. The theory is one of choice given preferences, not one of the nature of preferences.

Given that dimensionality is not something that flows from the spatial models behind most scaling procedures, how do we – how does anyone – know how many dimensions are appropriate? There are two main methods, both of which suffer from a common problem: the number of dimensions is a subjective judgment by the researcher.

**Comparative model fit:** One approach, and the method we will rely on below, is to scale the data under a number of different dimensionality assumptions and then compare the results. An inference, to the extent that this can be said to be an inference, is made by comparing the estimates from a model with a maintained assumption that there is a single dimension to one that assumes there are exactly two dimensions, and those, in turn, to the model that assumes there are exactly three dimensions, and so on. Thus, our inferences rely on statistics similar to those reported in Figure 1. We fit multiple models, examine the results, and determine which model seems “adequate.”

However, it is important to realize that this decision is a judgment call. No effort is made to fit models of various dimensionalities and then conduct formal likelihood-ratio tests. Indeed, for many procedures (c.f., W-NOMINATE) there are not even any formal tests (but see Poole, Sowell and Spear 1992). Poole and Rosenthal, for instance, primarily rely on comparative fit indices, such as the APRE statistics shown in Figure 1. For other methods (c.f., item response models), it may be possible to fit nested models and calculate Bayes factors, although we are aware of
no studies that have done so. Instead, researchers rely on comparative fit indices such as the Bayesian information criterion (BIC) or the Akaike information criterion (AIC), which penalize for model complexity (and hence added dimensions). However, while these metrics are in some sense “standard,” the degree to which they penalize for model complexity is nonetheless arbitrary. In the end, the adequacy of any model is determined by the number of observations that we are comfortable describing as external to the model or as “random.” Yet, it is always possible to account for more of the data within the spatial model by adding dimensions.\textsuperscript{11} It is in this sense that the question of how many dimensions best describe the data is a judgment.

This point is illustrated in Figure 2. The left-hand panel shows that, for this hypothetical roll call (given previously derived member preferences), if the voting data are placed onto a single dimension, the model yields a significant number of errors. The right panel, on the other hand, shows that by expanding the dimensionality, it is possible to draw a single dividing line that yields no errors and does not require altering member positions on the first dimension.

![Figure 2 about here.]

Additional heuristics: A second approach is to use one of several heuristics in the literature to identify the appropriate number of dimensions. The most widely used are the Kaiser (1960) eigenvalue-greater-than-one rule, the “elbow-test” proposed by Cattell (1966), and the parallel analysis test (Horn 1965).\textsuperscript{12} Each of these heuristics is designed to help scholars make a judgment regarding whether “enough” of the data’s structure is explained by a specific number of dimensions. The remaining errors are again attributed to noise. But eigenvalue and elbow rules are not tests in any strict sense; they merely provide guidance for researchers as to when adding dimensions will reduce the number of errors “sufficiently.” Moreover, these methods extract information from the roll-call matrices themselves rather than on any output of the scaling methods. Thus, in our analyses below, we rely on the first approach, with a specific focus on the APRE statistics that have been the dominant focus in the literature (c.f., Jackman 2001; Poole and Rosenthal 2007).

3. SIMULATING ROLL-CALL RECORDS

With this discussion in mind, we now turn to the Monte Carlo simulations. In each, we generate ideal points (i.e., preferences) of members from a known distribution with a known number of dimensions. We simulate observations by having members vote according to known rules to create a roll-call record. In Section 4, we then scale this simulated data.

\textsuperscript{11}Indeed, we can perfectly account for all of the observed data in the roll-call record for $M$ legislators if we use $q$ dimensions, where $q \in [1, M-1]$ is the rank of the roll-call matrix.

\textsuperscript{12}See Brown (2006, Chapter 2) for additional discussion of these tests.
While it would be possible to include additional complications, such as majority-party agenda control (Cox and McCubbins 2005) or bill events (Clinton and Meirowitz 2001), the simple simulations below are designed to make our point as cleanly as possible. We are not attempting to faithfully replicate the “true” data-generating process of Congress (whatever that may be), and we readily acknowledge that there may be other causes of artificially low-dimensional estimates left unexplored. Rather, we aim to show that inter-party polarization in the distribution of legislator preferences can by itself downwardly bias estimates of dimensionality.

The code used to generate our main results is included in Appendix A. The results from an alternative set of simulations, where the vote-margin of the simulations exactly matches the empirical record, are shown in Appendix B. Finally, Appendix C shifts slightly from our main focus on dimensionality itself to examine the conditions under which analyzing each party in isolation improves the accuracy of ideal-point estimates.

3.1. Simulation details

Distribution of member ideal points: Member ideal points, denoted $x_i$, are drawn from a multivariate normal distribution $x_i \sim N(0_p, I_{p \times p})$, where $p$ denotes the assumed number of dimensions. These points represent policy outcomes that members vote to achieve. We are agnostic as to whether these decisions are guided by personal beliefs, constituency pressure, or both. Each of the $p$ issue dimensions represents a distinct area of public policy. Conceptually, we can think of these policy domains as being latent traits (e.g., support for intervention in foreign wars) that guide decisions on multiple specific policies or acts of government (e.g., support for the 1993 US intervention in Somalia). In our simulations, we consider low-dimensional worlds where these policy dimensions are broad (e.g., liberalism-conservatism), high dimensional worlds where we imagine as many as 20 more specific latent dimensions (e.g., civil liberties, national security, crime, gay rights, etc.), and multiple states in between. The fewer dimensions, the broader and more encompassing the latent policy traits and policy dimensions are assumed to be.

We hold the number of legislators constant at $M = 101$ members. To approximate party polarization, we assume that members come from two subpopulations whose mean locations are a distance, $D$, apart in the policy space along each separating dimension. As $D$ increases from zero, the two clusters of legislators become increasingly distinct. We might imagine that as $D$ increases, the legislature consists of increasingly polarized Democrats and Republicans. Note, however, that nothing here distinguishes Democrats from Republicans other than their policy preferences.

More formally, we assume that $\frac{M+1}{2}$ members are distributed $x_i \sim N(\mu_p, I_{p \times p})$ while the re-

---

13We set $M = 101$ to be identical to the US Senate. This parameter does not affect the results shown below.
remainder are distributed $x_i \sim N(-\mu_p, I_{p \times p})$.\footnote{In previous work, we found no significant effect for majority size (Aldrich, Montgomery and Sparks 2010).} We allow that the populations may not be polarized on every dimension simultaneously. Thus, we add a parameter, $p_D \in [0, 1, \ldots, p]$, that indicates the number of dimensions on which the distributions are separated. Thus, if $p_D = 2$ and $p = 4$, then $\mu_p \equiv (\frac{0}{2}, \frac{0}{2}, 0, 0)$. Note that conceptually we are defining the degree of polarization on a given issue dimension to be equivalent to the distance between party means on that dimension.

**Voting behavior:** It is here that we mould our analysis around W-NOMINATE in particular, so we can be as faithful as is feasible to the exact assumptions of this procedure.\footnote{We do this to ensure our results do not rest on irrelevant quirk in our simulations.} Thus, the assumed behavior of members in our simulations is designed to be consistent with assumptions behind the W-NOMINATE procedure. In each simulation, we generate $N$ observations (i.e., votes) for each of the $M$ members. That is, we ask members to cast a vote comparing a single status quo, $a_j$, and a single proposal, $b_j$. Members have Euclidean preferences, and $d_{ijp}^{(a)} \equiv \|x_{ip} - a_{jp}\|$. We define the probability of member $i$ voting for the proposal on roll call $j$ as\footnote{This is the only major difference between our simulations and the assumptions of W-NOMINATE, which assumes that}

$$P_{ij} = \Phi \left[ \beta \left\{ \sum_{p=1}^{P} w_p(d_{ijp}^{(b)})^2 - \sum_{p=1}^{P} w_p(d_{ijp}^{(a)})^2 \right\} \right], \quad (1)$$

where $w_p$ indicates the weight of dimension $p$ in member preference and $\Phi$ is the cumulative function of the standard normal distribution.\footnote{For simplicity, we assume in our simulations that $w_p = \frac{1}{p} \forall p$.}

**Generating random roll calls:** To offer a fair test, it is necessary to generate status quo points that result in cut-points (or separating hyperplanes) spread throughout the space occupied by members. Therefore, rather than attempting to mimic the distribution of cut-points and majority sizes observed in the empirical record (e.g., Poole and Rosenthal 1991), our simulations generate roll calls randomly distributed through the policy space, an approach that is now common in the literature (c.f., Stiglitz and Weingast 2010; Hirsch 2011). This represents a more agnostic approach, and reduces the likelihood that our results are a product of how “nonrandom selection of roll calls may affect the ability to estimate ideal points that accurately reflect the preferences responsible for generating the observed votes” (Clinton 2012, p79). (We present results from an alternative approach to generating random roll calls in Appendix B.)
We use the following procedure to generate randomly distributed separating hyperplanes while still enabling us to easily calculate the voting probabilities shown in Equation 1. First, we randomly select (with replacement) one member to be the proposer, whom we assume proposes her own ideal point in the multidimensional space. Second, we randomly draw a cut-point, $c_j$, from the distribution $c_j \sim N(0, \frac{1}{T}I_{p \times p})$. The parameter $T$ serves to ensure that the cut-points do not result in an overly large number of unanimous or nearly unanimous votes as the dimensionality of the space expands. We then project across this cut-point to specify a status quo. That is, the status quo position on dimension $p$ is chosen as $a_{jp} = c_{jp} - (|c_{jp} - b_{jp}|)I[b_{jp} > c_{jp}] + (|c_{jp} - b_{jp}|)I[b_{jp} < c_{jp}]$ where $I(\cdot)$ is the usual indicator function.

**Overview of simulations:** For each parameter setting, we generate a hypothetical congress and ask all $M$ members to vote on $N$ different roll calls according to the procedures described above. By holding constant several nuisance parameters (c.f., $M$, $w_p$), the latent parameter’s space is not overly large. We vary only the five parameters as shown in Table 1. In total, we ran 3,672 simulations. Each simulation results in a roll-call matrix. We analyzed these matrices using the W-NOMINATE package in R (Poole et al. 2011).

**Discussion:** Before moving on to our results, it is worth emphasizing that our focus on W-NOMINATE reflects the prominence of this scaling procedure in the literature rather than any flaw or fault inherent to this method. All of the findings here have been replicated using simple principal component analysis (Aldrich, Montgomery and Sparks 2010). Moreover, using a reduced parameter sweep, the results have been reproduced using both Bayesian item response models (Clinton, Jackman and Rivers 2004) and optimal classification (Poole 2000, 2005) (results available upon request). Our use of W-NOMINATE should not be interpreted as either a critique of the method or an attempt to attribute the widely held belief in the unidimensionality of Congress to these authors. In addition, the results below are not sensitive to our reliance on APRE as a metric of model fit. Although we believe our focus on APRE reflects its widespread use for evaluating dimensionality, we have also analyzed these data using the eigenvalue-greater-than-one test and reached identical conclusions (Aldrich, Montgomery and Sparks 2010).

---

18 We found that $T = 1 + \sqrt{p}$ served to ensure adequate variance in the roll-call matrices.

19 We did not run simulations where the number of separating dimensions ($p_D$) would be larger than the number of dimensions ($p$). For the intra-party analysis below, we did not include simulations with party separation above five, as the number of valid roll calls became quite small. In addition, no intra-party W-NOMINATE estimates were created for four simulations with a small number of roll calls ($N = 1000$) and a high degree of polarization ($D = 5$) where the W-NOMINATE procedure repeatedly failed.

20 See Footnote 5. In any case, these different estimation techniques provide largely identical estimates and model-fit statistics (Clinton and Jackman 2009; Carroll et al. 2009).
Finally, we emphasize that our method of generating random roll calls is not intended to reflect the “true” data-generating process in Congress or any other legislative body. We could, for instance, allow only some party “leader” to propose changes to the status quo, which would reflect the gatekeeping power of the majority leadership (Cox and McCubbins 2005). However, relying on such variants to simulate a roll-call record introduces selection biases that have been shown to alter estimates of ideal points (e.g., Snyder 1992b; Clinton 2007, 2012) and is likely to alter estimates of dimensionality. The scheme we implement here is a multidimensional equivalent to the random generation of roll-call cut points for one-dimensional simulations (c.f., Stiglitz and Weingast 2010; Hirsch 2011; Clinton 2012). We believe that our random proposer approach represents an agnostic assumption that is least likely to artificially impose a low-dimensional solution on W-NOMINATE. Further discussion of how the cut-points generated here compare with the empirical record in the US Senate is provided in Appendix B.

4. SIMULATION RESULTS FOR A PARTISAN LEGISLATURE

In this section, we present results from simulations described in Section 3 to support our two major claims: (1) When scaling the entire roll-call record, it is not possible to determine whether low-dimensional scaling results reflect true low dimensionality in policy preferences or high levels of party polarization; (2) It is possible to distinguish between these two sources of low-dimensionality by analyzing the within-party record. In Section 5, we turn to testing the empirical implications of these simulations by analyzing the intra-party roll-call record of the US Senate.

4.1. Low-dimensionality in the context of party polarization

To introduce our method for presenting results, we begin by analyzing the results from a subset of the simulations. We then more systematically support our major theoretical claims. Throughout, we use the empirical analysis of the actual US Senate (shown in Figure 1) as a benchmark.

A very basic example of the kind of analysis we will be conducting is displayed in Figure 3.21 Figure 3 shows results from simulations in which the true dimensionality is low ($1 \leq p \leq 10$) and there is no partisan polarization ($D = 0$). The horizontal axes show the true number of dimensions ($p$) used in our simulations. The vertical axis on the left panel shows APRE. The vertical axis on the right panels shows the difference in APRE as each new dimension is added.

As a reference, the shaded region shows the range of observed values of APRE1 and APRE3-APRE2 in the empirical roll-call matrices in the US Senate from 1945 to 2010 (see Figure 1). Thus,  

---

21In analyzing the simulated roll-call matrices, we exclude any roll calls where more than 95% of the members voted the same way. Throughout, we excluded any W-NOMINATE analysis using fewer than 100 roll calls.
the shaded rectangles indicate the boundaries of APRE statistics that we have actually observed in
the real world, which are commonly interpreted as supporting a one- or two-dimensional policy
space. (For the sake of clarity, we do not produce the estimates from every Congress separately.)

In interpreting this figure (and those below), we focus on two specific questions. First, in
looking at the left panel, we ask: For what setting of \( p \) (the true number of dimensions) are the
APRE results as high or higher than the empirical roll call record? Recall that high APRE scores
indicate that just a few dimensions are able to explain most of the variation in the roll-call record.
In the case of Figure 3, we see that only when the true number of dimensions are actually very low
(between one and three) do the APRE statistics match the results we obtain from the analysis of
the empirical record. That is, APRE1 quickly diminishes as \( p \) increases.

In looking at the right panel, we ask: For what settings of \( p \) are the difference in APRE scores
as low or lower than the empirical record? Recall that small difference scores indicate that adding
additional dimensions fails to improve the explanatory power of the model. In this case, APRE2-
APRE1, APRE3-APRE2, and APR4-APRE3 begin quite small, but quickly increase to areas well
outside the shaded region. Comparing these results with Figure 1, we can see that only when the
true dimensionality is low are the APRE differences comparable to the empirical record.\(^{22}\) This
again suggests that, in the absence of party polarization, W-NOMINATE will only suggest a low-
dimensional space when the underlying space is actually low-dimensional.

\[ \text{[Figure 3 about here.]} \]

**Conclusion 1:** W-NOMINATE yields a close approximation to the true number of
dimensions that actually generated the data when (i) the true number of dimensions is
low, and (ii) there is no [party] polarization.

However, this positive finding – that scaling procedures only indicate a small number of di-
mensions in a truly low dimensional world – does not hold when we allow member preferences to
be distributed according to a mixture of normal distributions. We do this to mimic the real-world
polarization of Democrats and Republicans. Figure 4 shows examples (in one dimension) of these
distributions for several parameter settings of \( D \). The final panels of Figure 4 also show the dis-
tribution of the members of the 86\(^{th}\) Senate and 109\(^{th}\) House as estimated by the first dimension
of W-NOMINATE. These plots demonstrate that the level of polarization we consider in our sim-
ulations is no greater (and generally less) than what we observe in the actual roll-call record. For
example, the 86\(^{th}\) Senate (1959–1961) is perhaps the least polarized Senate in the post-war era,
and it reflects a partisan separation similar to \( D = 2 \) or \( D = 3 \) in our simulations.

\(^{22}\)Broadly speaking, we are looking to see for what values of \( p \) the APRE1 statistic (red) appears in the grey box
for the left panel. For the right panel, we are interested in the conditions where the APRE3-APRE2 statistic (blue) is
in the grey box.
Figure 5 shows how APRE statistics vary as a function of polarization for simulated Congresses of different dimensionalities. Within each panel, the evidence in favor of low dimensionality becomes stronger as party polarization increases. That is, the amount of the roll-call record explained by APRE1 goes up. In addition, the relative increase in APRE resulting from adding each dimension falls towards zero, indicating that adding additional dimensions does not improve the explanatory power of the model.

This pattern is consistent whether the true dimensionality in the simulations is one (the left panel) or 20 (the right panel). Notably, for even quite modest levels of polarization (e.g., D=2), these APRE results are in the region (shown in the grey shaded rectangles) that correspond to the analyses of the actual US Senate and are displayed in Figure 1. In addition, the results shown in Figure 5 come with $p_d \in (1, 2, 3)$, meaning that the parties are differentiated along only a few dimensions. These patterns are even more stark when parties separate on more dimensions.

**Conclusion 2:** As polarization increases, the number of estimated dimensions goes to one, regardless of the true number of dimensions.

Before moving onto the next section, it is worth emphasizing the degree and severity of the underestimation of dimensionality that result from minor changes to the assumed distribution of member preferences. Scaling procedures are, after all, designed to summarize large amounts of data using a reduced number of parameters. A small amount of downward bias in the number of suggested dimensions would not be shocking and indeed is to be expected. However, these results indicate that if the parties are polarized on only a single dimension – and even if this limited polarization is less than what we observe in the contemporary Congress – scaled roll-calls will suggest low-dimensionality whether there are actually one, five, or twenty dimensions.

### 4.2. Intra-party analysis

The results presented thus far are relatively straightforward. When the preferences of legislators are bimodal – reflecting party polarization on a few issues – standard scaling procedures will suggest a low number of policy dimensions regardless of the “true” complexity of the policy space.

---

23 Although not central to our main argument, a corollary of Conclusion 1 and Conclusion 2 is:

**Corollary 1:** When the estimated number of dimensions is high, this is an *indicator* of low levels of party polarization.
Given the widespread acceptance and use of the unidimensionality assumption for analysis of congressional action, this is by itself an important result. Yet, this finding is limited in that it provides no guidance as to how to better estimate the dimensionality of the space in the context of significant polarization. Instead, we have shown that apparent low-dimensionality may result either from voting in a low-dimensional setting or from bimodality. In this Section, we go further by proposing one approach to grappling with dimensionality in the presence of polarization. Intuitively, the results above show that dramatic inter-party differences along a subset of issue dimensions obscure distinctions in member preferences along less partisan dimensions. This suggests that we can gain additional leverage by scaling each of the two party caucuses separately.

Figure 6, therefore, shows intra-party APRE statistics for simulated roll-call records analyzed above, but now analyzing each party caucus separately. Specifically, we take all of the simulated roll calls and scale the responses from members of each party separately. In practice, due to the requirements of the W-NOMINATE procedure, we must drop roll calls where less than 5% of the members vote against the majority in the caucus for each intra-party analysis.

Unlike Figure 5, little in Figure 6 changes as the degree of party polarization increases. The flatness of each line shows that increased inter-party polarization has relatively little influence on APRE statistics as measured within each party. This confirms our notion that analyses of each party in isolation are less biased by party polarization.

Conclusion 3: Scaling intra-party observations approximates the true number of dimensions even in the context of significant [party] polarization.

[Figure 6 about here.]

5. EMPIRICAL INTRA-PARTY SCALING ANALYSIS

In the previous section, we showed through Monte Carlo simulations that party polarization can interfere with correctly recovering the dimensionality of the “true” world from an observed roll-call record. Further, one plausible approach to determining whether low-dimensional statistical estimates reflect true low-dimensionality in the context of party polarization is to focus on the intra-party roll-call records. With these findings in mind, we now turn to analyzing the claim that preferences in Congress have become more unidimensional over time.

We re-analyze the roll-call record in the US Senate from 1945 to 2010, as we did in Figure 1. However, we now calculate W-NOMINATE scores using only the intra-party roll-call record.24

As noted above, we exclude any W-NOMINATE analysis involving fewer than 100 roll calls. However, the APRE estimates do become somewhat more noisy as the degree of polarization increases and the number of roll calls included in the analyses goes down as the parties divide internally less often.

The major consequence of this procedure is the loss of many roll calls that divide Democrats and Republicans on
This means that we generate two APRE estimates for each Congress – one for Democratic senators and one for Republicans. The results are presented in Figure 7, which shows clearly that there is very little evidence of a secular trend towards low-dimensionality over the course of the 20th Century when we focus on the within-party roll-call record.

[Figure 7 about here.]

If anything, the APRE statistics suggest that there might be a slightly larger number of dimensions in recent years. Certainly, there is nothing like the dramatic evidence in support of low-dimensionality that is revealed by analyzing both parties at once (shown in Figure 1).

[Figure 8 about here.]

To make our point a bit more explicit, Figure 8 places the empirical record against simulations with either a low or a high number of dimensions. In each, the solid lines show APRE statistics for the entire record (either simulated or empirical), while the dashed lines show the APRE statistics as measured only within the two parties. Although this does not constitute a formal statistical test, Figure 8 suggests that the actual record in the Senate is more in line with a high dimensional world with increasing levels of polarization than with a mostly, and increasingly, unidimensional world. That is, the APRE statistics for the entire roll-call record have changed substantially since the 1960s, but there has been little change in the within-party record. This mirrors our simulation results, and suggests that the dimensionality of Congress has not decreased over the last four decades, but rather the parties have polarized.

near-party-line votes. This results from the fact that the NOMINATE procedure cannot be utilized for roll-calls that are nearly unanimous. Since we are analyzing each caucus in isolation, roll calls in which the caucus is united are automatically dropped by the scaling software. However, beyond this, we have done nothing to select specific types of “divisive” roll calls.

One important implication of this is that our estimates are generated with fewer roll calls than in Figure 1. However, we restrict our analysis to instances where there are more than 100 roll calls with sufficient variation within a given caucus. One possible criticism, however, is that our analysis may now be inaccurate due to the loss of important variance in the record. However, as we showed above, there is little useful information about the true dimensionality of preferences in party line votes in a polarized Congress.

To be abundantly clear, this analysis differs significantly from past work that focused on distinguishing between estimates recovered when using only party-line and lopsided votes (c.f., Snyder and Groseclose 2000; Cox and Poole 2002). (For a recent discussion, see: http://voteview.com/blog/?p=234.) Here, we are not subsetting the dataset by traits of the roll calls (e.g., whether or not they are lopsided), but by traits of the members. We scale members of each party entirely separately. This analysis makes no direct contribution to the debate on the usefulness of lopsided votes, and the differences between the results in Figure 7 and those reported in previous studies reflect our entirely different empirical strategy.

The simulations show that it is by examining the party caucuses in isolation that we may better distinguish between low and high-dimensional spaces in the context of high levels of inter-party polarization. Of course, this prevents us from using the ideal points generated from these intra-party analyses in conjunction as they are not on the same scale. However, our focus throughout this article is the dimensionality of the policy space rather than on ideal point estimation per se. A brief discussion of the “accuracy” of the W-NOMINATE estimates when each party is analyzed in isolation is presented in Appendix C.
6. CONCLUSION: TAKING MULTIPLE DIMENSIONS SERIOUSLY

One of the accepted truisms of American politics research is that we can accurately describe preferences over public policy with only one or two dimensions. Scholars and pundits conceptualize policies and political figures as fitting onto a single underlying liberal-conservative continuum. Only on occasion is this supplemented with a second dimension (e.g., “social issues”). Indeed, the bulk of contemporary research on American politics implicitly or explicitly accepts Poole and Rosenthal’s famous conclusion that “one-and-a-half” issue dimensions adequately encapsulates every era of the nation’s political history, and that the contemporary Congress is virtually or actually unidimensional (Poole and Rosenthal 1991, 1997, 2007).

In this article, we have shown that finding that a low number of estimated dimensions can explain a large proportion of the variation in roll-call voting does not necessarily imply that the process that generates the roll calls is itself low-dimensional. If parties are polarized – even along a few dimensions – we will observe low-dimensional estimates across the entire range (1 – 20) of true dimensional states. Thus, analyses of the complete roll-call record will provide at best ambiguous evidence in favor of low dimensionality. To better distinguish among possible sources of low-dimensional statistical results, we have shown that examining intra-party roll-call records can provide improved leverage. Moreover, applying this approach to the empirical roll-call record for the US Senate from 1945–2010 shows that the evidence in favor of the low-dimensional conjecture is weaker than is commonly appreciated.

Before concluding, it is important to note the limitations of the above findings. To begin with, we emphasize that our goal here is to examine the degree to which it is possible to clearly establish the dimensionality of the roll-call record in the context of strong party polarization. We argue above that examining comparative fit indices such as APRE for the entire roll-call matrix will obscure the “true” dimensionality of the policy space. Yet, this does not change the fact that, as the empirical APRE results in Figure 1 show, a great deal of legislative behavior can be predicted using only one or two dimensions. The genius of W-NOMINATE is that it is so well suited to generating “ideal point” estimates that can explain the largest number of roll calls using the fewest possible number of dimensions. However, as our simulations show, effectively summarizing a data matrix and recovering the “true” dimensionality of the data-generating process can be distinct tasks calling for different approaches to the data.

Second, the simulations above are far from exhaustive. Different assumed data-generating processes will certainly yield different results. This may be a particularly concern for our intra-party analysis. We assume that votes that divide one or both of the parties appear randomly on the agenda. However, many theories of Congress specify particular conditions under which such roll calls may be expected to come to the floor (Cox and McCubbins 2005; Dougherty, Lynch
While the results above are sufficient for supporting our main argument – that it is possible to recover better dimensionality estimates in the context of polarization by looking within each party – further simulations may be necessary to explore the degree to which focusing exclusively on the intra-party roll-call record is appropriate under the various alternative assumptions in the formal literature.

Finally, the results above suggest several areas for further research for improving our understanding of the policy space. For one, these results emphasize the need for exploring ways to directly compare models with different dimensionalities in a statistically informed manner. One plausible approach is suggested by Ghosh and Dunson (2009), who provide methods for computing Bayes factors for models of differing dimensionalities with uninformative priors. Another path forward may be to implement scaling procedures that allow multiple dimensions without assuming they affect all roll calls, similar to the constraints typical in confirmatory factor analysis (Brown 2006; Erosheva and Curtis 2011). These more subtle approaches to summarizing the roll call record promise to provide estimates that still reduce dimensionality of the data without confining all political disagreement to a single dimension.

References


### A. CODE APPENDIX

```r
# Load packages
library(mvtnorm); library(akima); library(pscl);library(wnominate);
library(arm); library(snow); library(oc); library(MASS); library(plyr)
library(fields); library(foreach); library(multicore); library(doMC)

# Generate voter preferences / ideal points:
preferenceGenerator <- function(mPSize, nObs, pSep, nDim, nSDim, normV){
  voterParty <- rep(x = c(1, -1), times = c(mPSize, nObs - mPSize))
  alpha1 <- rep(x = c(pSep/2, 0) * 1, times = c(nSDim, nDim - nSDim))
  alpha2 <- rep(x = c(pSep/2, 0) * -1, times = c(nSDim, nDim - nSDim))
  majorityPartyPreferences <- rmvnorm(n = mPSize, mean = alpha1,
                                     sigma = diag(normV, nDim))
  minorityPartyPreferences <- rmvnorm(n = nObs - mPSize, mean = alpha2,
                                      sigma = diag(normV, nDim))
  allVoterPreferences <- data.frame(rbind(majorityPartyPreferences,
                                          minorityPartyPreferences))
  allVoterPreferences$Party <- voterParty
  return(allVoterPreferences)
}
```
# Generate cut points.
cutPointGenerator <- function(nDim, nVotes, normM, normV) {
cpMeans <- rep(normM, nDim)
cpCov <- diag(normV, nDim)
cutPoints <- rmvnorm(n = nVotes, mean = cpMeans, sigma = cpCov/(1+sqrt(nDim)))
return(cutPoints)
}

# Takes ideal points and cut points, draws proposers and proposals, # infers status quo points
generateProposalsAndStatusQuos <- function(iPoints, cPoints) {
  nObs <- nrow(iPoints)
  nDim <- ncol(cPoints)
  nVotes <- nrow(cPoints)
  whoProposes <- sample(x = c(1:nObs), size = nVotes, replace = T)
  proposalPoints <- as.matrix(iPoints[whoProposes, 1:nDim])
  proposerParty <- iPoints$Party[whoProposes]
  idealPointArray <- as.matrix(iPoints[, 1:nDim])
  statusQuoPoints <-
    (cPoints - abs(cPoints - proposalPoints)*(proposalPoints > cPoints) +
    abs(cPoints - proposalPoints)*(proposalPoints < cPoints))
  statusQuoPoints <- as.matrix(statusQuoPoints)
  output <- list(proposer = whoProposes, proposalPoints = proposalPoints,
              statusQuoPoints = statusQuoPoints)
  return(output)
}

# Takes ideal points and an object that contains proposal and status quo points # Returns a roll-call matrix
generateRollCallsStandard <- function(.iPoints, pASQ, beta=300) {
  nDim <- ncol(pASQ$proposalPoints)
  distanceFromProposal <- rdist(x1 = .iPoints[, 1:nDim],
                               x2 = pASQ$proposalPoints)
  distanceFromStatusQuo <- rdist(x1 = .iPoints[, 1:nDim],
                               x2 = pASQ$statusQuoPoints)
  util.diff <- beta*((distanceFromStatusQuo^2)-(distanceFromProposal^2))
  probVote <- matrix(rbinom(n=length(util.diff), prob=pnorm(util.diff), size=1),
                      nrow=nrow(util.diff))
  return(probVote)
}

## Take roll-call matrix and estimate WNOMINATE.
dwNominateFunction <- function(rcMat, nDim) {
  nDimensionsToUse <- nDim
  nDimensionsToUse[nDimensionsToUse > 10] <- 10
  nDimensionsToUse[nDimensionsToUse < 4] <- 4
  rollCall <- rollcall(data = rcMat, yea = 1, nay = 0)
  nominateObject <- wnominate(rcObject = rollCall, dims = nDimensionsToUse,
                               minvotes = 20, lop = .05, trials = 1,
                               polarity = rep(1, nDimensionsToUse),
                               verbose = FALSE)
  return(nominateObject)
### Take all objects created from one parameter set, and select things to save from them.

```r
summarizeSimulation <- function(wNom, whichRow, label = "genericLabel"){
  rowParameters <- unlist(parametersToSweep[whichRow, ])
  nDim <- rowParameters["nDimensions"]
  effectiveNumberOfEigens <- sum(wNom$eigenvalues) ^ 2 / sum(wNom$eigenvalues ^ 2)
  names(effectiveNumberOfEigens) <- "eNEigenvalues"
  first10Eigens <- wNom$eigenvalues[1:10]
  names(first10Eigens) <- paste0("Eigen", 1:10)
  nDimensionsToUse <- nDim
  nDimensionsToUse[nDimensionsToUse > 10] <- 10 # Artificial capping at 10
  nDimensionsToUse[nDimensionsToUse < 4] <- 4 # Artificial flooring at 4
  correctClass <- rep(NA, 10)
  names(correctClass) <- paste("correctClassD", c(1:10), sep="")
  correctClass[1:nDimensionsToUse] <- wNom$fits[1:nDimensionsToUse]
  apre <- rep(NA, 10)
  names(apre) <- paste("apreD", c(1:10), sep="")
  apre[1:nDimensionsToUse] <- wNom$fits[(nDimensionsToUse+1):(nDimensionsToUse*2)]
  gmp <- rep(NA, 10)
  names(gmp) <- paste("gmp", c(1:10), sep="")
  gmp[1:nDimensionsToUse] <- wNom$fits[(nDimensionsToUse*2+1):(nDimensionsToUse*3)]
  nLegislatorsScaled <- sum(!is.na(wNom$legislators$coord1D))
  nRollCallsScaled <- sum(!is.na(wNom$rollcalls$midpoint1D))
  output <- data.frame(parameterRow = whichRow, t(rowParameters),
                        effectiveNumberOfEigens, t(first10Eigens),
                        t(correctClass), t(apre), t(gmp),
                        pUnity$meanPartyUnityScore,
                        pUnity$nPartyUnityVotes,
                        nLegislatorsScaled,
                        nRollCallsScaled, label)
  return(output)
}
```

## Function to generate party-only subsets.

```r
partySubsetScaler <- function(iPoints, rcMat, nDim, pUnity, whichRow){
  leftRollCalls <- rcMat[iPoints$Party == -1, ]
  rightRollCalls <- rcMat[iPoints$Party == 1, ]
  leftNominateObject <- dwNominateFunction(rcMat = leftRollCalls,
                                           nDim = nDim)
  rightNominateObject <- dwNominateFunction(rcMat = rightRollCalls,
                                           nDim = nDim)
  leftSimulationSummary <- summarizeSimulation(wNom = leftNominateObject,
                                                pUnity = pUnity,
                                                whichRow = whichRow,
                                                label = "leftPartyOnly")
  rightSimulationSummary <- summarizeSimulation(wNom = rightNominateObject,
                                                pUnity = pUnity,
                                                whichRow = whichRow,
                                                label = "rightPartyOnly")
  output <- data.frame(rbind(leftSimulationSummary, rightSimulationSummary))
  return(output)
}
```
B. PARTY-LINE VOTING AND VOTE MARGINS

One possible concern is that our results are driven by high levels of party-line voting in simulations with significant amounts of polarization – a phenomenon not observed in the empirical record. However, the degree of party-line voting in most of the simulations shown above is not substantially greater, and frequently less, than is observed in the empirical record. Figure B.1 shows the distribution for the percentage of party-unity votes\textsuperscript{27} in the US Senate from 1945–2009 plotted against the same distribution from our simulations (each observation is a Congress). These results are for all simulations with party polarization of 2, 2.5, and 3. Note that the frequency of party unity votes in the simulated Congress is not dramatically outside of the range observed in the actual Senate. Indeed, not until the set of simulations with polarization of $D = 4$ does the median level of party unity votes in the simulations exceed that observed for the 111\textsuperscript{th} Senate (75.1\%).

As a further robustness check, we conducted a second set of simulations such that the vote margins\textsuperscript{28} in our simulated matrix exactly match the empirical record. These supplementary simulations differ from those in the main text in the following ways:

1. We created a roll-call matrix with 20,000 votes.

2. We “reverse code” half of them, implicitly switching the proposal and status quo points (although not changing the cut-points themselves).

3. We look at 526 quantiles of roll-call margins of all votes in the Senate from 1960–2009. For each of these quantiles, we randomly choose (with replacement) from among the simulated roll calls most similar in terms roll-call margins. This results in a roll-call matrix with a Senate-typical 525 votes, where the distribution of vote margins matches the empirical record exactly.\textsuperscript{29}

\textsuperscript{27}Party unity votes are defined as roll calls where the majority of one party votes in opposition to the majority of the other party.

\textsuperscript{28}The vote margin for any given roll call is simply the number of yea votes divided by the total number of votes cast.

\textsuperscript{29}Due to the sampling in this procedure, a small number of roll calls appear more than once in the simulated record. The median number for a given simulation was seven total duplicates and the maximum was 28. However, few roll
4. To ensure that we observe roll calls with the correct number of yea and nay votes, regardless of the dimensionality of the space or the distribution of member preferences, we drew cut-points from a larger area of the implied policy space.\textsuperscript{30}

Figure B.1: Distribution of % party unity votes by Congress

Distributions of party unity votes observed in simulated (D=2, D=2.5, and D=3) and empirical (1945-2009) Senates. Each observation represents the percentage of votes that are party unity votes for a single Congress (simulated or empirical). Note that the number of party unity votes observed for moderately high levels of polarization is not dramatically different from what has been observed in the empirical record.

Figure B.2 shows the density of vote margins for one simulated roll-call matrix generated using the above procedure. Figures B.3 and B.4 replicate Figures 5 and 6 in the main text with this alternative procedure. As can be seen, although the patterns here are somewhat less dramatic, our overall findings are robust to matching the vote margins to the empirical record.

C. COMPARING IDEAL POINT ESTIMATES FOR FULL-CHAMBER AND SINGLE-PARTY NOMINATE ESTIMATES

In the main text, we focus exclusively on how well NOMINATE aids researchers in recovering the true dimensionality of the policy space. We find that focusing on each party in isolation fa-

\textsuperscript{30}Specifically, of the 20,000 total simulations, one-third were drawn from a normal distribution with \( N_p(-\mu, \frac{1}{1}I) \), \( N_p(\mu, \frac{1}{1}I) \), and \( N_p(0, \frac{1}{1}I) \), where \( T = (1 + p^{1/3}) \). The \( T \) parameter is necessary to prevent the creation of a roll-call matrix with (nearly) all unanimous votes in high dimensional simulations.
Figure B.2: Exemplar density of roll-call margins for the empirical US Senate and a matched simulated record.

The top panel shows the distribution of roll call margins for all roll calls cast in the 60th–111th US Senates. The bottom panel shows the distribution from an exemplar simulated record where $p = 2$, $p_d = 2$, $D = 4$, and $\beta = 0.5$. 
Figure B.3: APRE for different values of party separation in the empirically matched roll-call record ($p_d \in (2, 3)$)

The left panel shows results when the true dimensionality is $p = 1$. The center and right panels show APRE statistics for increasing numbers of dimensions. In the context of significant polarization, W-NOMINATE suggests a small number of dimensions regardless of the true value of $p$. 
Figure B.4: Intra-party APRE for different values of party separation in the empirically matched roll-call record ($p_d \in (2, 3)$)

The left panel shows results when the true dimensionality is $p = 1$. The center and right panels show APRE statistics for increasing numbers of dimensions. In the context of significant polarization, W-NOMINATE analysis of the intra-party record only suggests a small number of dimensions when $p$ is actually small.
cilitates identifying the dimensionality of the space in the context of polarization. However, some researchers may be interested in how well this procedure does at recovering the true positions of legislators in the context of polarization.

In this appendix, therefore, we are interested in ascertaining whether fitting party-specific W-NOMINATE subspaces more accurately recovers “true” ideal points. To do this, we ran 25 iterations each of a reduced parameter sweep, varying only the $\beta$ and Party Separation ($D$) parameters. We generated ideal points in a three-dimensional space ($p = p_d = 3$) and fit three-dimensional W-NOMINATE models. We did this first using all legislators simultaneously (the common practice), and then for each party separately.

For each set of results, we employed a procrustes rotation (Oksanen et al. 2013) to fit the space as closely as possible back to the original “true” ideal point space, and saved the resulting minimized mean squared error from this rotation. Intuitively, this statistic captures how closely NOMINATE ideal point estimates align with the initial configuration of legislator ideal points in three dimensions.

Figure C.1 displays these fit statistics for NOMINATE scores based on the full chamber (x-axis) contrasted against to those scores based on each party scaled separately (y-axis). The diagonal line shows the point at which each model predicts equally well. Points falling to the left of this line suggest that the full-chamber results are superior. The figure shows that RMSE is lower (meaning model fit is better) for the full-chamber estimates when party separation is low, whereas the intra-party NOMINATE estimates are better when party separation is extremely high.
Figure C.1: Comparing the fidelity of NOMINATE configurations to “true” ideal points.

The figure shows the RMSE for 250 simulations estimated at four different levels of party polarization. The ideal points are estimated using the full full chamber and each party in isolation. A procrustes transformation is used to recover the original space and evaluate the “accuracy” of the estimates. The diagonal line shows the point at which each model performs equally. The typical full-chamber approach does better on average, but its relative performance decreases as a function of polarization.
Figure 1: Empirical APRE Results for the US Senate (1945–2010)

The left panel shows the APRE statistics for the first three dimensions of a W-NOMINATE analysis of the US Senate from 1945–2010. The right panel shows the marginal improvement in APRE as each dimension is added. The points show the estimates from each Senate, and the lines are loess curves. This evidence strongly suggests an increasingly low-dimensional Senate.
Figure 2: Misclassifications as error or additional dimensions?

The panels show the location of a hypothetical status-quo position (SQ) and a proposed alternative (P). The left-hand panel shows the one-dimensional preferences of members (estimated from previous roll calls) and their votes on some new roll call. The Y’s represent yea votes and N’s represent nay votes. The right panel shows these same members plotted on two dimensions. By adding a second dimension, it is possible to draw a line that perfectly divides yea and nay votes while keeping their ordering on the first dimension constant. In general, it is always possible to add dimensions to better predict the data.
Figure 3: APRE results with no party polarization

The left panel shows the raw APRE scores from simulated roll calls, while the right panel shows the difference in APRE values as each dimension is added. The bold points and lines show the mean result at each parameter setting. The lighter points show the results from each simulation. The grey shaded region shows the range of observed values of APRE1 and APRE3-APRE2 in the empirical roll-call matrices for the US Senate from 1945 to 2010 (see Figure 1). When polarization is low, analysts will identify a small number of dimensions only if there are actually few dimensions.
Figure 4: Visualizing simulated and empirical mixture distributions

The first seven panels show a random draw of 20,000 observations from the mixture distribution used in our simulations. The final two panels show the distribution of 1st dimensional W-NOMINATE scores for the 86th Senate (1959–1960) and 109th House (2005–2006), as examples of low and high levels of empirically observed polarization. The level of polarization in our simulations is no greater than what we observe in the empirical roll-call record.
Figure 5: APRE for different values of party separation (D) \((p_d \in 1, 2, 3)\)

The left panel shows results when the true dimensionality is one \((p = 1)\). The center and right panels show APRE statistics for increasing numbers of dimensions. The bold points and lines show the mean result at each parameter setting. The lighter points show the results from each simulation. In the context of significant polarization, W-NOMINATE suggests a small number of dimensions regardless of the true value of \(p\).
Figure 6: Intra-party APRE for different values of party separation ($p_d \in (1, 2, 3)$)

The left panel shows results when the true dimensionality is $p = 1$. The center and right panels show APRE statistics for increasing numbers of dimensions. The bold points and lines show the mean result at each parameter setting. The lighter points show the results from each simulation. In the context of significant polarization, W-NOMINATE analysis of the intra-party record only suggests a small number of dimensions when $p$ is actually small.
Figure 7: Empirical APRE results for intra-party and inter-party analysis of the US Senate (1945–2010)

The left panels show the absolute proportional reduction in error (APRE) for the first three dimensions of a W-NOMINATE analysis of the US Senate from 1945-2010. The right panel shows the marginal improvement in APRE as each dimension is added. The results in the top panel are calculated for each party caucus separately. The the bottom panels show estimates when all members are included together in the same measurement model. There is little evidence of decreasing dimensionality in the intra-party record.
Figure 8: Comparing inter- and intra-party APRE statistics of simulated roll-call matrices with empirical APRE statistics of the US Senate (1960–2010)

In the left and center panels, the points show the average APRE scores for analyses of simulated roll-call records. The solid lines show results when the entire record is used, while the dashed lines show results when examining the intra-party record. The right panel shows the same APRE results for the US Senate from 1960–2010, a period of increasing party polarization.
Table 1: Parameter values for Monte Carlo simulations

<table>
<thead>
<tr>
<th>Parameter symbol</th>
<th>Interpretation</th>
<th>Simulated values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td># of dimensions</td>
<td>1, 2, \ldots, 10, 15, 20</td>
</tr>
<tr>
<td>$N$</td>
<td># of roll calls</td>
<td>1000, 2000, 5000</td>
</tr>
<tr>
<td>$D$</td>
<td>distance between subpopulations on each separating dimension</td>
<td>0, 0.5, 1, \ldots, 6</td>
</tr>
<tr>
<td>$p_d$</td>
<td># of dimensions on which subpopulations are separated</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>$\beta$</td>
<td>scaling parameter for probabilistic voting</td>
<td>0.5, 1, 2</td>
</tr>
</tbody>
</table>