

The Structure of Political Choices: Distinguishing Between Constraint and Multidimensionality*

William Marble[†] Matthew Tyler[‡]

July 16, 2020

Abstract

In the literatures on public opinion and legislative behavior, there are debates over (1) how constrained preferences are and (2) whether they are captured by a single left-right spectrum or require multiple dimensions. But insufficient formalization has led scholars to equate a lack of constraint with multidimensional preferences. In this paper, we refine the concepts of constraint and dimensionality in a formal framework and describe how they translate into separate observable implications for political preferences. We use this discussion to motivate a cross-validation estimator that measures constraint and dimensionality in the context of canonical ideal point models. Using data from the public and politicians, we find that American political preferences are one-dimensional, but there is more constraint among politicians than among the mass public. Furthermore, we show that differences between politicians and the public are not explained by differences in agendas or the incentives faced by the actors.

*For helpful discussions and comments, we thank Adam Bonica, David Broockman, Justin Grimmer, Andy Hall, and Jonathan Rodden. We also thank Adam Bonica for sharing NPAT data. We are grateful to participants at the Political Economy Breakfast and the Workshop for Empirical American Politics at Stanford. An R package for implementing the MultiScale algorithm is available at <https://github.com/matthewtyler/MultiScale>.

[†]Ph.D. Candidate, Department of Political Science, Stanford University. wpmarble@stanford.edu

[‡]Ph.D. Candidate, Department of Political Science, Stanford University. mdtyler@stanford.edu

1 Introduction

Ideological scaling methods have long been a mainstay in legislative studies, and scholars are increasingly applying these methods to disparate sources of data beyond roll-call votes. The goal is to extract a simple, low-dimensional summary measure of ideology from votes, survey responses, or other types of political data.¹ Typically, researchers seek to align political actors on a simple left-right political spectrum, which can be used to characterize public opinion and to study representation. Despite the prevalence of ideological scaling methods, there remain unresolved debates about how to interpret the resulting estimates – especially when applied to non-institutional actors such as survey respondents.

First, there is debate over the dimensionality of political conflict. In the study of American politics, the default setting is to estimate a one-dimensional left-right model, in line with conventional wisdom about the dimensionality of contemporary Congress (cf. Poole and Rosenthal, 1997). Yet, some researchers suggest as many as eight dimensions are needed to explain Congressional voting patterns (Heckman and Snyder, 1997), and recent empirical work on American institutions has found value in accounting for multidimensional preference structures (Jeong et al., 2014; Crespin and Rohde, 2010). Other scholars argue that one dimension is not enough because voters *think* about politicians in multiple dimensions rather than just one (Ahler and Broockman, 2018). This is perhaps because the political preferences of voters are best captured by two or more dimensions (Treier and Hillygus, 2009).

Second, there is debate over how constrained attitudes are in the public relative to politicians, which determines the interpretability of any ideal point estimates. Some authors claim that most citizens do not have well-formed political opinions, let alone opinions that can be meaningfully placed on a left-right spectrum (e.g., Converse, 1964; Kinder, 2003). In an extreme view, policy attitudes are unstable and entirely idiosyncratic, meaning that scaling methods have little hope

¹Foundational work in scaling roll-call votes includes Poole and Rosenthal (1997), Clinton, Jackman and Rivers (2004), and Heckman and Snyder (1997). More recent studies applying similar methods to survey responses include Shor and McCarty (2011) and Jesse (2009). Bonica (2013) extends these methods to campaign finance data, and Bond and Messing (2015) apply them to social media data.

of recovering a useful estimate of ideology. A slightly weaker formulation is that citizen preferences are somewhat constrained, but that they are not amenable to a low-dimensional summary (Broockman, 2016; Lauderdale, Hanretty and Vivyan, 2017). In this case, only a small portion of variance in survey responses can be explained by a single dimension.

A common sentiment is that public opinion is multidimensional, while political conflict among the parties is one-dimensional. A natural implication is that a higher-dimensional model should better describe public opinion data. Under this view, low constraint and multidimensionality are synonyms (Broockman, 2016). A population exhibiting high constraint must also have one-dimensional political preferences, and a population exhibiting low constraint has multidimensional preferences.

In this paper, we seek to distinguish between these two notions, dimensionality and constraint, in the context of ideal point models. We point out that dimensionality refers to the effective number of separate issues that are commonly understood and acted upon by all voters. In the language of Heckman and Snyder (1997), the dimensionality is the number of “attributes” of policy choices that are needed to rationalize votes. Constraint, in contrast, refers to how much political actors rely on these attributes (e.g., left-right ideology) in forming opinions on particular policies rather than idiosyncratic reasons. An example of an idiosyncrasy in American politics would be an otherwise fiscally conservative voter favoring generous unemployment benefits (perhaps because they were once unemployed themselves). In a highly constrained population of actors, knowing an actors’ opinions on one set of issues should enable accurate prediction of further opinions, relative to an appropriately chosen null model. In an unconstrained population, most policy attitudes are idiosyncratic and thus unrelated to each other.

From this new perspective, constraint and dimensionality are orthogonal concepts. Theoretically, any population can exhibit any level of constraint with any level of dimensionality. Instead of unidimensionality and high constraint going hand-in-hand, we could in fact observe low constraint and unidimensionality together or high constraint with multidimensionality. Thus, even if we are convinced a population like the mass public exhibit low constraint on average, it remains

an empirical question whether public opinion should be characterized by one or more dimensions. Similarly, a high level of constraint in Congress does not guarantee that members of Congress vote in a manner consistent with one-dimensional political preferences.

Drawing on these ideas, we propose an out-of-sample model validation procedure for ideal point models that enables us to estimate the dimensionality of political preferences and the associated level of constraint for a given population. In contrast, extant model validation efforts in the literature have focused on in-sample fit or ad-hoc measures of out-of-sample fit. We document evidence of significant overfitting in ideal point models, illustrating the importance of a theoretically motivated out-of-sample validation strategy.

We apply the validation procedure to the workhorse quadratic-utility ideal point model commonly used to estimate ideal points. With an array of data sets that encompass both politicians and the public, we draw three main empirical conclusions.

First, we find no evidence that multidimensional models of ideal points explain preferences better than one-dimensional models – in fact, due to overfitting, higher-dimensional models can perform worse than a model that does not estimate ideal points at all. Second, we find that ideal point models are considerably less predictive when applied to the public. In contrast with politicians, voter responses are dominated by idiosyncratic, rather than ideological, preferences. This suggests the public has low constraint, at least relative to politicians. We also document some heterogeneity in the public, suggesting that ideal point models are more informative about individual issue attitudes for some groups than for others. Third, we decompose this difference in model performance between the public and politicians. We find that nearly all of the divergence can be attributed to differences in the constraint of the actors, rather than different measurement tools or disparate incentives faced by the actors. When applied to high-quality survey data of politicians, scaling methods perform nearly as well as when applied to roll-call votes. We also take advantage of paired data sources of politicians and the public to show that this conclusion is not driven by differences in the agenda or survey design.

These results suggest caution in applying ideal point estimation methods to surveys in the

mass public. The resulting estimates do indeed explain some of the variation in stated preferences. However, the variance in preferences for particular policies that is explained by ideal points is considerably lower than for politicians. Idiosyncratic preferences — rather than spatial preferences — tend to dominate voter attitudes. These results suggest that scholars should not limit themselves to ideal point estimates when studying political attitudes in the mass public (Ahler and Broockman, 2018).

The rest of the paper is structured as follows. First, we discuss the differences between dimensionality and constraint. Then, we propose out-of-sample validation procedures to measure them. Next, we address the debate about the dimensionality of political conflict. Finally, we examine differences in constraint across populations and contexts, while addressing possible explanations for the divergence between elites and the mass public.

2 Constraint and Multidimensionality

At least as far back as Converse (1964), scholars of public opinion have been aware of the fact that American voters do not fit as cleanly inside ideological lines as, say, members of Congress or state legislators. A common sentiment in this literature is that voters' policy attitudes are not derived from a coherent ideological framework. Instead, attitudes are idiosyncratic or at least not structured in the same way as politicians'. This perspective emphasizes a notion of ideology as *constraint*. Constraint here refers to the degree to which policy attitudes on some issues are predictive of policy attitudes on other issues. For example, if there is high degree of ideological constraint in the population, then knowing a voter's preferences for welfare spending should allow one to infer their preferred tax rate. If there is a low degree of constraint, then knowing the voter's preference about the welfare spending tells us little about their preferred tax rate.

The primary evidence for the lack of constraint comes from the low inter-item correlations between survey responses (Converse, 1964) and lack of knowledge about which issues "go together" (Freder, Lenz and Turney, 2020). Summarizing one view, Kinder (2003, 16) writes that

“Converse’s original claim of ideological naïveté stands up quite well, both to detailed reanalysis and to political change.” This view implies that a one-dimensional spatial model is simply not useful for understanding public opinion.

In contrast, some scholars have attempted to salvage the idea of constrained voters by arguing that a multidimensional model provides a more reasonable picture of how voters perceive politics. For example, Treier and Hillygus (2009) write, “Our analysis documents the multidimensional nature of policy preferences in the American electorate... [F]ailing to account for the multidimensional nature of ideological preferences can produce inaccurate predictions of voting behavior.” In some of these arguments, the additional dimensions are considered to be just as important as the first. For instance, Lauderdale, Hanretty and Vivyan (2017) claim that including a second dimension nearly doubles how much variance in stated preferences is attributable to ideology.² An implied sentiment is that observed levels of constraint increase with a more flexible notion of ideology that encompasses multiple dimensions.

The goal of this paper is to distinguish between these two notions, constraint and multidimensionality, and to provide rigorous measures of them. The dimensionality of policy attitudes refers to the number of distinct underlying issues that are common to all people responding to the survey (or voting on roll-call votes). For example, we may think of policies as occupying a space with both “economic” and “moral” issue dimensions that are understood in the same way by all political actors. Multidimensionality simply refers to the presence of multiple such issues. The level of constraint, in contrast, refers to how much knowledge of someone’s policy attitudes on some issues helps us predict their policy attitudes on other issues *through* the common policy space. There may be many idiosyncratic factors affecting individuals’ policy attitudes that have nothing to do with the common policy space. The level of constraint refers to how much variance the common policy space explains relative to the idiosyncratic components.

There are no trade-offs between constraint and multidimensionality: either can appear with

²The stated predictive power of multiple dimensions varies greatly from study to study. For instance, Ansolabehere, Rodden and Snyder (2006) use measures of both economic and a moral preferences, but discount the importance of the moral dimension for explaining vote choice.

or without the other. Whether voters have multidimensional preferences has little to do (logically) with whether they have constrained preferences. A natural implication of this point is that there is a limit to how well we can predict political attitudes from other attitudes, since modeling additional dimensions of ideology has diminishing returns.

2.1 Formalizing Constraint and Multidimensionality

We now formalize constraint and multidimensionality in the context of canonical ideal point models. This will result in some observable implications that we can use to empirically identify the level of constraint and dimensionality in a population given a sample of its political choices.

We couch our discussion in terms of the familiar quadratic-utility spatial voting model used in Clinton, Jackman and Rivers (2004, hereafter CJR).³ Specifically, we will suppose the political actors we are studying — whether they be members of Congress, survey respondents, and so on — have an ideal point γ_i located in some common D -dimensional Euclidean space. When considering a choice between two policy proposals, a “yea” policy located at point ζ_j and a “nay” policy located at ψ_j , voters’ utility is a function of the distance between their ideal point γ_i and the policy positions, plus a mean-zero idiosyncratic preference shock. Under the assumption of normally distributed shocks, it is simple to obtain an expression for the probability actor i chooses the yea option on choice j , denoted $y_{ij} = 1$. Averaging over preference shocks, it is given by

$$P(y_{ij} = 1 \mid \alpha_j, \beta_j, \gamma_i) = \Phi(\alpha_j + \beta_j^\top \gamma_i), \quad (1)$$

where α_j and β_j are functions of the policy locations and the distribution of shocks.⁴

³In this article, we are agnostic as to whether or not the assumptions behind the CJR model are “correct” or can be interpreted as structural. For our purposes, it is enough that the models predict or describe behavior. So, for instance, we are interested in these models even if the ideal points being measured are actually some reduced form combination of true ideological, partisan, and constituent incentives. Our theoretical discussion also does not rely on particular functional form assumptions.

⁴In particular, $\alpha_j = (||\psi_j||^2 - ||\zeta_j||^2)/\sigma$, and $\beta_j = 2(\zeta_j - \psi_j)/\sigma$ where σ^2 is the variance of idiosyncratic preference shock (Clinton, Jackman and Rivers, 2004). If we allow the idiosyncratic preferences to have non-zero mean, then we merely have to shift α_j accordingly.

We use this model to characterize the distinction between multidimensionality and constraint. Implicit in the model is the dimensionality of the common policy space. The parameters γ_i and β_j lie in some D -dimensional Euclidean space \mathbb{R}^D representing the space of possible policies. For instance, Treier and Hillygus (2009) consider a 2-dimensional policy space to reflect economic and social issues. We might label the positive end of this space in both directions to refer to “conservative” policies, so an actor with $\gamma_i = (-1.5, 3.2)$ would prefer “liberal” economic policies and “conservative” social policies.

To further unpack this, focus on the linear predictor of actor i and choice j :

$$\alpha_j + \beta_j^\top \gamma_i = \alpha_j + \sum_{d=1}^D \beta_{jd} \gamma_{id}. \quad (2)$$

Equation 2 suggests that choice j is analogous to a (generalized) linear regression. The “intercept” α_j and “coefficients” β_j change from choice to choice depending on the alternatives on offer, but the “covariates” γ_i stay the same for actor i across choices. For instance, the mapping from economic and moral preferences to a tax policy question will differ from how those same preferences map onto an immigration policy question (different intercepts and slopes), but the underlying economic and moral preferences (covariates) stay the same.

Thus, we can think of the dimensionality D as the number of underlying preferences needed to explain expected surpluses. When D is small, there are only a few key attributes that meaningfully distinguish between policies in expectation — all of the other variables that determine utilities are too idiosyncratic to be organized into a common policy space. However, when D is large, there is a greater variety of systematic political conflict. The residual incentives for voting yea or nay are still idiosyncratic, but the systematic component of surpluses are much more complex and involve a higher number of trade-offs between issue dimensions. With high-dimensionality, actors might be balancing preferences along, say, tax policy, morality policy, immigration policy, foreign policy, etc., provided these preferences are sufficiently uncorrelated.⁵

⁵The dimensionality is the number of linearly independent issues. If tax, morality, immigration, and foreign policy preferences are highly correlated, then there might only be one or two (linearly independent) dimensions. This

If multidimensionality is the correct number of covariates needed to model expected choice in Equation 2, then constraint is the amount of variation those covariates can explain. As we alluded to previously, constraint captures the idea that knowing an actor's ideal policy, γ_i , improves our ability to predict their choices. In the linear regression analogy, constraint is similar to the R^2 statistic: how much better we do in prediction after conditioning on the covariates γ_i .

To see precisely how knowledge of γ_i might improve our ability to predict choices, recall that the probability of a yea vote conditional on γ_i is given by Equation 1, $P(y_{ij} = 1 \mid \alpha_j, \beta_j, \gamma_i) = \Phi(\alpha_j + \beta_j^\top \gamma_i)$. How does the probability of a yea vote change if we don't have knowledge of the ideal point γ_i ? To compute this, we must imagine drawing an actor at random, which means we must draw a value of their ideal point from the population distribution. For illustration, suppose we draw the ideal points from a standard multivariate normal: $\gamma_i \sim N(0, I_D)$. Then the population average probability of a yea vote on choice j is given by

$$P(y_{ij} = 1 \mid \alpha_j, \beta_j) = \Phi\left(\frac{\alpha_j}{\sqrt{1 + \|\beta_j\|^2}}\right), \quad (3)$$

for all actors i . This is essentially an intercept-only probit model $\Phi(\delta_j)$ with intercept $\delta_j = \alpha_j / \sqrt{1 + \|\beta_j\|^2}$ that varies for each item j .⁶

Equation 3 is the appropriate null model for understanding to what extent ideal policy preferences explain choices. When equations 1 and 3 are different, then knowledge of an actor's ideal policy helps explain their choices. Before knowing someone's ideal policy, the prediction for their choice should be the same for everyone, Equation 3. If choices are unconstrained, then idiosyncratic components determine choices completely and thus ideal points γ_i would be uninformative. In that case, there would be no difference between predictions made with ideal points and predictions made without them. However, once we know an actor's ideal policy, and there is a high degree of constraint, our prediction for their choice should alter dramatically as we go

is analogous to multicollinearity in linear regression.

⁶The population average in equation 3 is found by observing that $\epsilon_{ij} - \beta_j^\top \gamma_i$ follows a $N(0, 1 + \|\beta_j\|^2)$ distribution, where $\epsilon_{ij} \sim N(0, 1)$ is the utility error corresponding to equation (1). Our notion of constraint does not depend on our choice of distribution for γ_i . If γ_i has density h , then $P(y_{ij} = 1 \mid \alpha_j, \beta_j) = \int \Phi(\alpha_j + \beta_j^\top \gamma) h(\gamma) d\gamma$, which is always just an intercept-only model for each choice j .

from $P(y_{ij} = 1 \mid \alpha_j, \beta_j)$ to $P(y_{ij} = 1 \mid \alpha_j, \beta_j, \gamma_i)$.

This discussion of constraint was specific to the individual actor i . To get a population measure of constraint, we can average over the population how much our predictions improve. This gives us the expected predictive power of ideal points relative to the null model.

To summarize our discussion, multidimensionality is a property of the agenda and how preferences are organized among the population as a whole. It is analogous to the (correct) number of covariates in a linear regression model. In contrast, constraint is how much individuals use these organized preferences to select the choices on offer. If the organized preferences matter, and we know actors' ideal policies, we will make much different predictions than we would without that structure. Consequently, under high constraint, ideal points explain a large amount of variation in choices, while under low constraint ideal points are only mildly predictive of choices. Critically, multidimensionality and constraint are not mutually exclusive: any population/agenda combination can have high or low dimensionality and high or low constraint. There are no (logical) trade-offs between constraint and multidimensionality, and whether voters are multidimensional has nothing to do with whether they are constrained.

2.2 Empirical Implications

Having clarified the distinction between multidimensionality and constraint, we turn towards measuring these concepts in common data on political choices, such as surveys and roll-call votes.

First, we treated the dimensionality D as a fixed number in our theoretical discussion. Indeed, when fitting the ideal point model laid out above, we must make a choice about the dimensionality of the model we wish to fit. However, we do not know the true dimensionality D *a priori*, so it makes sense to think about D as a parameter we can infer. Measuring the dimensionality is therefore tantamount to learning the value of $D \in \{1, 2, \dots\}$ that leads to the best approximation of the true data-generating process.

After learning the best-fitting dimensionality, we can assess the level of constraint by comparing the fraction of choices predicted by our fitted model to the fraction of choices predicted

by the fitted null model that does not include ideal points. In a highly constrained population, nearly all of the variation that cannot be explained by the null model will be explained by the best-fitting ideal point model. However, in an unconstrained population, the best-fitting ideal point model will explain only slightly more variation than the null model. This perspective highlights that constraint is not binary, but is a matter of degree. Idiosyncratic preferences surely exist; the question is how important those preferences are in comparison to the systematic components determining political choices. In the next section we describe our method for estimating this degree of constraint in any particular population-agenda combination.

3 Out-of-Sample Validation for Ideal Point Models

Our plan to measure dimensionality and constraint relies on estimating the predictive performance of fitted ideal point models. The goal of this section is to (1) explain why performance should be measured out-of-sample, rather than in-sample, for both substantive and methodological reasons and (2) explain how we estimate out-of-sample performance. By in-sample we mean fitting and evaluating a model with the same data. By out-of-sample we mean first fitting an ideal point model and then evaluating how well it predicts choices not used in fitting the model.

3.1 Arguments for Out-of-Sample Validation

First, out-of-sample prediction is more directly aligned with the original definitions of constraint. For instance, Converse (1964) clearly had out-of-sample prediction in mind:

Constraint may be taken to mean the success we would have in predicting, given initial knowledge that an individual holds a specified attitude, that he holds certain further ideas and attitudes (Converse, 1964).

In other words, given some choices A, constraint is our ability to predict other choices B. It would not make sense to be given choices A and measure constraint as our ability to predict A with itself.

An out-of-sample measure of constraint is more consistent with the notion that belief systems and ideologies are bundles of ideas, structured together through the common policy space.

Second, on methodological grounds, in-sample estimates of fit are biased towards measuring higher constraint and higher dimensionality. More complex models tend to overfit to training data (Hastie, Tibshirani and Friedman, 2009). Even if political preferences are low-dimensional, higher-dimensional ideal point models will tend to overfit and thus overstate the true dimensionality of preferences. The threat of threat of overfitting is quite real in our data, and we present ample evidence of this in Section 5.

3.2 Out-of-Sample Validation for Ideal Point Models

To measure out-of-sample predictive power, we must have some data that is not in the sample used to fit the ideal point model parameters. A natural first reaction would be to simply drop some actors or some votes from the analysis, and then predict choices for those actors or votes. However, every actor and every vote in the ideal point model has a parameter that must be modeled — namely, α_j, β_j for votes and γ_i for actors — so we cannot exclude whole actors (rows) or whole votes (columns) from the model fitting process.

Our proposed validation scheme gets around this obstacle by randomly selecting actor-vote pairs, corresponding to individual cells the in data matrix Y , and hiding them from estimation. If we only randomly remove a few cells — perhaps 10% of the matrix Y — then almost all of any actor's choices will still be available for learning γ_i . Similarly, we will still have roughly 90% of the choices for vote j , so the vote parameters α_j, β_j can still be estimated with ease. With estimates of these parameters in hand, we can go back and evaluate how well the ideal point model can explain the held-out choices (cells). Figure 1 illustrates the hold-out strategy.

Below, we implement two versions of this cell hold-out strategy to perform out-of-sample validation of ideal point models. The first approach is the simple train-test split we described above: one training sample is used to fit the model with 90% of the cells observed and a performance is evaluated on a test sample with the 10% of cells that were randomly held out. A second strategy is

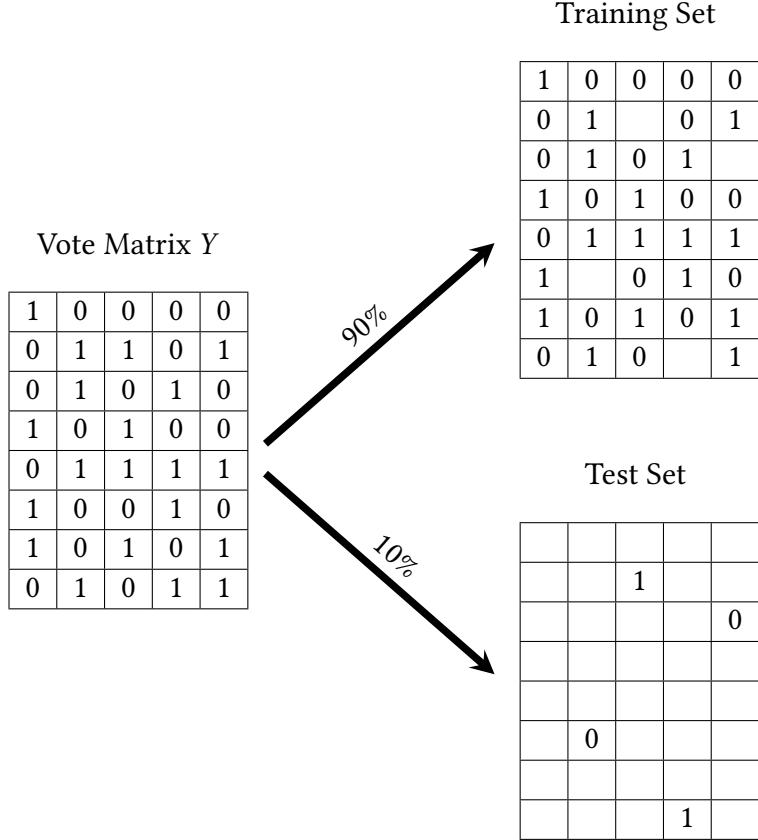


Figure 1: An illustration of our out-of-sample validation strategy with $N = 8$ actors and $J = 5$ votes. Cells are randomly sampled to be in the test set.

a cross-validation approach where we randomly divide all the cells into 10 groups and treat each of the 10 groups as a test sample in 10 different model fits. For each of these 10 model fits, we use the other 9 non-test groups as the training sample.⁷ This second approach uses all of the data to estimate out-of-sample performance, since each cell appears exactly once in a test sample.⁸

We show in a small simulation study, reported in Figure 4 in Online Appendix C, that the proposed strategy can accurately recover the true dimensionality of the data-generating process with data sets similar in size to those used in this paper.

In terms of identifying the most appropriate number of dimensions, our approach is an appli-

⁷Chapter 7 of Hastie, Tibshirani and Friedman (2009) describes cross-validation in greater depth.

⁸We note that both strategies are only estimating ideal point parameters with 90% of the data available, as opposed to using 100%. Under the assumption that more data leads to better predictions, this necessarily biases us towards finding weaker out-of-sample performance across all model specifications. However, based on simulations and experiments where we apply the method to even smaller training samples, we do not find the loss of the first 10% of cells to matter in any meaningful way for our ultimate conclusions.

cation of model selection from statistics and machine learning. We treat each possible dimension D as a separate model and identify the best model based on optimizing a statistic (Hastie, Tibshirani and Friedman, 2009). In our case, that statistic is hold-out accuracy, but we could have also made a case for using other well-established model-selection criteria such as AIC (Akaike, 1973), BIC (Schwarz, 1978), etc. We have chosen hold-out accuracy because it is both substantively more interpretable (how *much* constraint?) and closer to the original notion of constraint put forward in Converse (1964).

Political scientists have previously sought to validate ideal point models in a variety of ways. The most common strategy has been to focus entirely on in-sample measures of fit (i.e., without a hold-out sample). For example, in their work on congressional roll-call voting, Poole and Rosenthal (1997) report that in-sample accuracy does not increase beyond two dimensions. Jessee (2009) presents data supporting the same conclusion for survey respondents.

3.3 Estimating Ideal Point Models

Our proposed cross-validation approach requires re-fitting a multidimensional ideal point model on each data set 10 separate times. To do this quickly while allowing for large amounts of missing data (due to item nonresponse), we wrote software based largely on the formulas of Imai, Lo and Olmsted (2016).⁹ It is described in detail in the Online Appendix A and is available online.^{10,11} Also in the Online Appendix B, we show that our one-dimensional ideal point estimates are highly correlated with one-dimensional DW-NOMINATE scores ($r = 0.97$) for members of Congress (Poole and Rosenthal, 1997) and Shor-McCarty scores ($r = 0.86$) for state legislators (Shor and McCarty, 2011).

⁹We did not use their package because, at the time of writing, it would only fit one-dimensional models and had trouble recovering good estimates with a large degree of missing data.

¹⁰We use the same software default priors as the emIRT package of Imai, Lo and Olmsted (2016). In particular, the priors are $\alpha_j \stackrel{\text{ind.}}{\sim} N(0, 5^2)$, $\beta_j \stackrel{\text{ind.}}{\sim} N(0, 5^2 I_D)$, $\gamma_i \stackrel{\text{ind.}}{\sim} N(0, I_D)$. Note that while the priors assume the parameters are independent, they are not necessarily independent in the posterior. Thus, the ideal points are possibly correlated in the posterior and in fact we find they often are.

¹¹The software is available at <https://github.com/matthewtyler/MultiScale>.

4 Data Sources

We use several sources of data to evaluate the performance of ideal point models. These data sets are drawn from typical uses in the literature on scaling and cover both politicians and the mass public. For survey questions with more than two ordered response options, we binarize the answers by classifying whether the answer is greater than or equal to the mean. Online Appendix D further describes the variables used in the analysis.

Senate Voting Data. As a benchmark, we use roll-call votes from the 109th Senate (2005–2007). These data are included in the R package `pscl`, and contain 102 actors voting on 645 roll calls. About 4 percent of the roll call matrix is missing.

NPAT. As a source of survey data among elites, we use data from Project Votesmart’s National Political Courage Test, formerly known as the National Political Awareness Test (NPAT). The NPAT is a survey that candidates take. The goal of the survey is to have candidates publicly commit to positions before they are elected. For political scientists, the data are useful because they provide survey responses to similar questions across institutions. As such, one prominent use of the NPAT data is to place legislators from different states on a common ideological scale (Shor and McCarty, 2011).

While there are some standardized questions, question wordings often change over time and across states, requiring researchers to merge together similar questions. The full matrix we observe has 12,794 rows and 225 columns. Unlike the roll-call data, however, there is a high degree of missingness: 79 percent of the response matrix is missing.

State Legislator Survey. We additionally use Broockman’s (2016) survey of sitting state legislators. This survey contains responses from 225 state legislators on 31 policy questions. Question topics include Medicare, immigration, gun control, tax policy, gay marriage, and medical marijuana, among others. Only about 5 percent of the response matrix is missing. Additionally, we use a paired survey of the public that is also reported in Broockman (2016). A subset of the questions are identical to those asked of state legislators. We only use the first wave of the survey, in which there are 997 respondents and no missing data.

2012 ANES. We use questions from the 2012 American National Election Studies Time Series File, drawn from the replication material of Hill and Tausanovitch (2015). There are 2,054 respondents and 28 questions. These questions cover a broad swath of politically salient topics, including health insurance, affirmative action, defense spending, immigration, welfare, and LGBT rights, as well as more generic questions about the role of government. About 10 percent of the response matrix is missing. We also use various subsets of the ANES to investigate heterogeneity in the public.

2012 CCES. Finally, we use data from the 2012 Cooperative Congressional Election Survey. In particular, we focus on the “roll call” questions, where respondents are asked how they would vote on a series of bills that Congress also voted on. These data have been used to jointly scale Congress and the public (Bafumi and Herron, 2010). There are 54,068 respondents, answering 10 such questions on the 2012 CCES. The questions cover bills such as repealing the Affordable Care Act, ending Don’t Ask Don’t Tell, and authorization of the Keystone XL pipeline. Only about 3 percent of the response matrix is missing.

We also match these survey responses to the corresponding roll-call votes in the Senate. These votes took place in the 111th, 112th, and 113th Congresses. We are able to match 9 questions to roll-call votes.¹² A full list of the votes used for scaling is available in Online Appendix D.

5 Evidence on Dimensionality

This section has two goals: first, to demonstrate the existence of overfitting in higher-dimensional ideal point models; and second, to test for which dimensionality best explains variation in political choices. We focus on two sets of data: the 109th Senate (2005 - 2007) as a sample of political elites and the 2012 American National Election Study (ANES) as a sample of the mass public. We estimate the ideal point models for these data sets separately for $D \in \{0, 1, \dots, 25\}$ dimensions. A $D = 0$ dimensional model only includes an intercept term for each vote. As a measure of model

¹²We could not match the Bowles-Simpson budget question to a roll call vote, because it never got a floor vote in Congress.

fit, we focus on the estimated accuracy — i.e., the proportion of responses for which the observed choices is most likely according to the model fit. We use the simple 90/10 training/test split we described above. The difference between the accuracy in the training and test sets conveys the degree to which overfitting occurs at that dimensionality.

Figure 2 displays the results. For both samples, overfitting occurs within a few dimensions, suggesting that there is actually quite a bit of harm in attempting to model idiosyncrasy with additional dimensions. In the public, statistically significant overfitting occurs even after using just one dimension. In an effort to accommodate a low-dimensional structure model of policy preferences, the fitted model starts to find connections between idiosyncratic preferences that do not generalize beyond the training data. When these connections are applied to the data in the test set, the model is overconfident in its ability to explain idiosyncratic responses that are actually impossible to predict — resulting in a decrease in accuracy relative to the training set. It is clear from this figures that it would be a mistake to compute accuracy only using the training set, since doing so over-estimates the generalizability of the model.

The conclusion from these figures is that the best-fitting model is very low dimensional for both samples. In the public, model fit decreases beyond a single dimension. In the Senate, a one dimensional model is statistically no worse than any other model. In contrast to the conjectures offered in the literature on political attitudes, the overfitting problem is much more severe in the ANES than in the Senate. For both samples, we can confidently conclude that a one-dimensional model is the most reasonable model.¹³

It might seem possible that these results are a property of the statistical model being applied, rather than a feature of the political actors in these data sets.¹⁴ However, when these same ideal

¹³Figure 2 is somewhat suggestive that a two-dimensional or three-dimensional model might be a better fit in the Senate. But these supposed improvements are small and not statistically significant.

¹⁴A cautious reader might worry that we are only finding one-dimensionality because we do not include polynomial or interaction terms for the γ_{id} components. Fortunately for our purposes, purely linear models with additional orthogonal dimensions approximate lower-dimensional models that have polynomials and interactions (e.g., Maraun and Rossi, 2001). The idea behind this is that many orthogonal dimensions can approximate a nonlinear function of a few dimensions (or another factor structure, see Paisley and Carin, 2009). Thus, if we were omitting important nonlinear γ_{id} terms, then we would find that higher-dimensional models fit the data well. Since we find that one-dimensional models are no worse than any other model, it must be that the nonlinear functions of γ_i are not that important for explaining the response variables.

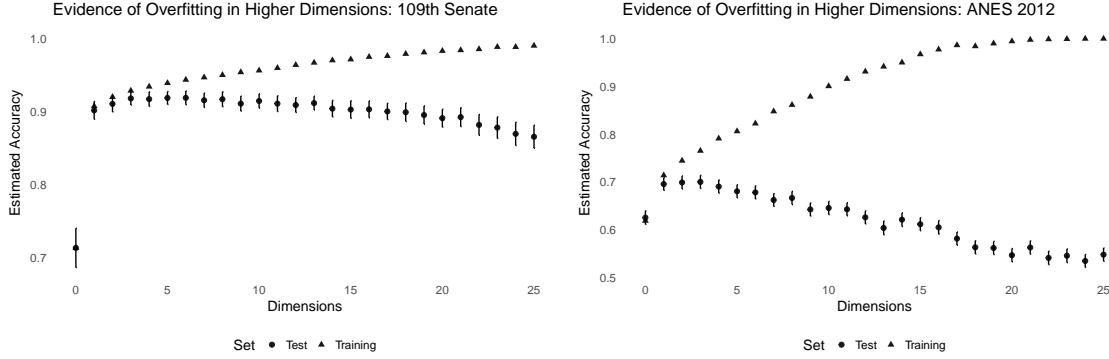


Figure 2: Estimated accuracy within the training and test sets for models fit with latent dimensions $D \in \{0, 1, \dots, 25\}$. Model fit deteriorates after three dimensions for the Senate data and after one dimension for the ANES data. Error bars show 95% confidence intervals, clustered at the respondent level.

point models are applied to, say, ratings of movies on Netflix written by movie-watchers, the negative consequences of overfitting do not present themselves until after 30 or 60 dimensions of movie preferences are assumed (Salakhutdinov and Mnih, 2008).¹⁵ It is hard to interpret overfitting results in political data as artifacts of the statistical model being used when those same models can uncover higher dimensions of latent structure in other choice settings. As a whole, the public appears to discriminate over many more attributes when making entertainment choices than when answering questions about politics and policy. This fact should not be too surprising, given that the parties neatly organize policies into two competing bundles. The entertainment market is much more fragmented.

6 Evidence on Constraint

Using the same framework, we now turn to a systematic investigation of ideological constraint. As noted above, we conceptualize constraint as how much better we can predict one set of policy opinions if we know another set of policy opinions, compared to an appropriately chosen null model. In the context of the item-response theory model, this corresponds to a comparison of

¹⁵The Netflix studies do not use binary data, but they use analogous ideal point techniques to uncover latent structure in movie choice data.

the performance of an ideal point model, which allows responses to vary depending on an actor’s ideal point, to an intercept-only model, in which predicted responses do not depend on the actor’s ideal point. In the extreme case of no constraint, the predictive performance will be identical, and the difference between a one-dimensional model and a intercept-only model will be negligible. At the other extreme of perfect constraint, a model that includes ideal points will dramatically improve upon the intercept-only model.

The literature suggests that we should expect higher levels of constraint among politicians than among the mass public. We use a number of data sources from both of these populations, which enables a direct test of this hypothesis and allows us to estimate just *how much more* constrained politicians are than citizens. Our analysis is a “difference-in-differences” approach that compares the improvement in predictive performance among politicians to the improvement among the public. We are therefore interested in higher-precision estimates of predictive performance, so we turn to 10-fold cross validation as outlined in Section 3.

For each data set, we estimate models with $D \in \{0, \dots, 5\}$ dimensions, 10 times each, holding out a 10% sample each time to be used as a test set. For each holdout response, we calculate the likelihood, given the estimated model parameters, of the observed response, and classify its accuracy based on whether the likelihood is greater than 0.5. We then calculate the average accuracy across all holdout responses for each model. Given the results in the previous section indicating that responses are best described as one-dimensional, our key measure of constraint is the increase in accuracy moving from the $D = 0$ intercept-only model to the $D = 1$ one-dimensional model.¹⁶

6.1 Constraint Among Elites and the Public

The main results are shown in Figure 3. The left-hand panel plots the average cross-validation accuracy for each data set. The right-hand panel plots the increase in accuracy for $D \in \{1, \dots, 5\}$

¹⁶We use accuracy as the measure of model fit in this section to keep in line with the existing literature and for ease of interpretation. However, the substantive conclusions drawn in this section are not sensitive to this choice. Online Appendix E shows the same results using the likelihood of the observed hold-out responses as the measure of model fit.

compared to the intercept-only model.

Beginning at the top of the figures, the solid squares show the cross-validation accuracy of ideal point models for roll-call votes in the 109th Senate. The left-hand panel shows that nearly 90 percent of votes are correctly classified by a one-dimensional ideal point model. The right-hand panel shows that this is about a 20 percentage point increase over the intercept-only model. As discussed in the previous section, there is little additional gain in accuracy for the Senate data moving beyond a single dimension, though the out of sample performance does not degrade either.

Next, the hollow square shows the results when applied to data from Broockman's (2016) survey of state legislators. A one-dimensional model can accurately classify roughly 83 percent of responses. Again, this is an increase of over 20 percentage points compared to the intercept-only model. Here, the performance of the model begins to decay once we estimate more than a single dimension. This result again underscores the one-dimensionality of political constraint among political elites.

The one-dimensional model performs less well when applied to the NPAT data, as illustrated by the solid diamonds. A one-dimensional model correctly classifies only about 75 percent of responses — an increase of less than 10 percentage points over the intercept-only model. This increase is less than half of the gain achieved with the other two sources of elite data. There is a mild increase in accuracy associated with a second dimension, though the increase is only about 2 percentage points.¹⁷

Notwithstanding the NPAT results, the picture that emerges from this exercise confirms the conventional wisdom that politicians — at both the national and state level — are highly constrained in their preferences. Two-thirds of inexplicable votes under the null model are now predictable due to the inclusion of a one-dimensional ideal point.

¹⁷Our suspicion for why the NPAT results differ is data quality. The NPAT data are highly non-standardized, with question wordings varying across time and space. These peculiarities require researchers to combine similar questions. However, this data cleaning may undermine the assumption of a commonly understood policy space. There is also a high degree of missingness in this data set, which may violate the ignorability assumptions necessary for estimation with missing data.

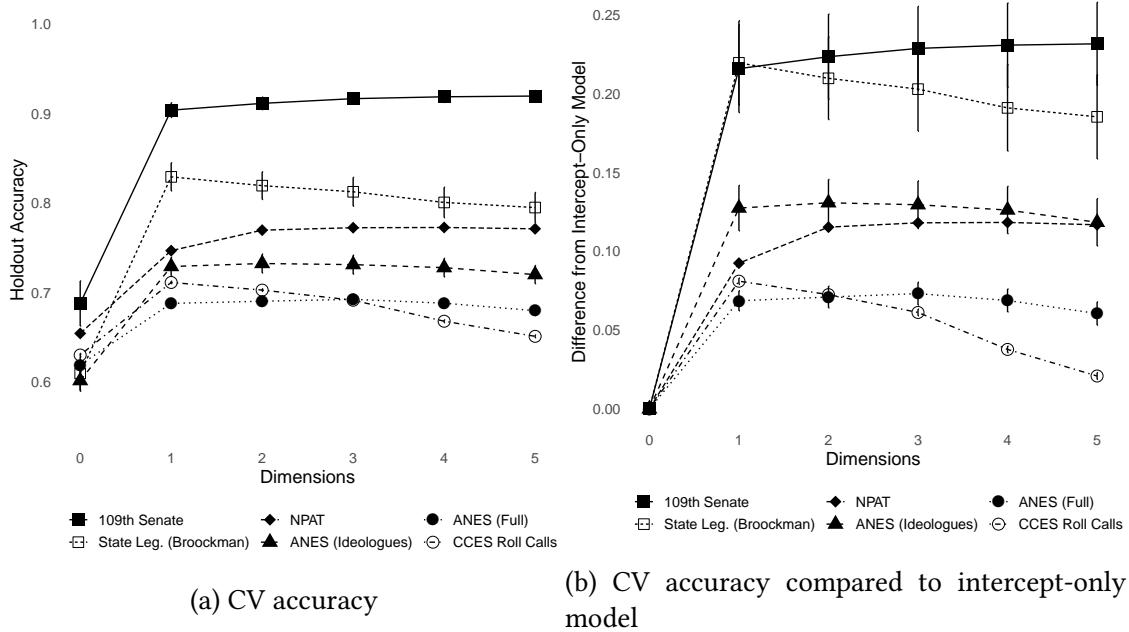


Figure 3: (Left) Cross-validation accuracy for models up to five dimensions. (Right) Increase in percent of accurately classified votes compared to an intercept-only model. Error bars show 95% confidence intervals, clustered at the respondent level.

Such a dramatic increase does not hold for the public. As noted above, we use data from the 2012 ANES and CCES. The nature of the questions included differs between these sources. For the ANES, the questions are typical of public opinion research. The CCES questions, however, ask respondents how they would vote on particular roll-call votes that were actually voted on in Congress. Despite the differences in question format, our substantive conclusions are identical for both data sets.

Consider the solid circles in Figure 3, which correspond to the full sample of ANES respondents. The left-hand side shows that an intercept-only model correctly classifies about 62 percent of responses. Adding a one-dimensional ideal point increases this classification accuracy to about 69 percent. Similarly, an intercept-only model correctly classifies about 63 percent of CCES roll call responses, compared to 71 percent accuracy for the one-dimensional model. In both cases, additional dimensions do not increase the performance of the models, and, in the case of the CCES, degrades the performance.

These results suggest that classification accuracy increases by about 12 percent in a one-

dimensional model compared to an intercept-only model.¹⁸ Despite using a different methodology, Lauderdale, Hanretty and Vivyan (2017) come to a similar conclusion; they report that about 1/7th of the variation in survey responses can be explained by a one-dimensional ideal point, while the rest they attribute to idiosyncratic or higher-dimensional preferences.

Overall, we take these results to mean that there is some constraint in the public, but the relationship between the estimated ideal points and the survey responses is much noisier among the public than it is among elites. For the ANES and CCES, a one-dimensional ideal point model only increases classification accuracy by about 7 or 8 percentage points compared to an appropriate null model. Only about 20% of unpredictable votes under the null model are now predictable using ideal point models. This number pales in comparison to the 67% of inexplicable-turned-predictable votes for the national and state political elites.

6.2 Investigating Heterogeneity in the Public

Of course, there is heterogeneity in the level of constraint in the public, and the overall results may mask constraint among a meaningful subset of the public. As a preliminary test of this possibility, we re-run the ANES analysis after subsetting to people who say self-identify as ideological (i.e., answer 1, 2, 6, or 7 on a 7-point ideology scale). This group of people is likely to have better-formed opinions about political issues and to perceive a common policy space, implying that we may observe more constraint in this population. These results are shown in the solid triangles in Figure 3. As expected, ideal point models have higher accuracy among this subset than among the ANES respondents as a whole. A one-dimensional model can correctly classify 73 percent of responses among this subset, compared to only 60 percent that are correctly classified by the intercept-only model. The right-hand panel also shows that the absolute increase in accuracy is actually larger than the increase for the NPAT. Still, compared to the Senate roll-call votes or the state legislator survey, the increase in accuracy is relatively small, again highlighting the higher level of constraint among elites than the public.

¹⁸Relative to the baseline, there is an 11% increase in accuracy for the ANES ($[69 - 62]/62 = .11$) and a 12% increase for the CCES ($[71 - 63]/63 = .12$).

A fuller picture is presented in Figure 4, which shows the average increase in classification accuracy for each respondent from a one-dimensional model, relative to the intercept-only model, across a number of political and sociodemographic variables. The black points show the results when the cross-validation procedure is run on the full dataset, while the gray points show results when it is run on each subset separately.¹⁹ The political variables include whether the respondent was contacted by or donated to a campaign, the respondent’s self-report ideology, party identification, and whether the respondent correctly places the Democratic Party to the left of the Republican Party on an ideology scale (Freder, Lenz and Turney, 2020). The socioeconomic variables include age, education, gender, family income, and race.

The results suggest that there is indeed heterogeneity in the public. Broadly, we find significantly more constraint among conservatives than liberals, Republicans than Democrats, and those who are more knowledgeable and engaged with politics. Similarly, we find greater constraint among people who are older, more highly educated, and richer.

Nonetheless, only in a few subgroups – namely, conservatives and Republicans – does the increase in out-of-sample predictive accuracy come close to the increase seen among politicians.²⁰ For the vast majority of subgroups we study, a one-dimensional ideal point model does help predict responses better than an intercept-only model, but the gains are very modest in comparison to politicians.

6.3 Explaining the Public-Politician Divide

Our results thus far suggest that ideal point estimates extract relatively little information from survey responses of the mass public – at least when compared to politicians. Our preferred

¹⁹These two approaches have different trade-offs. The former allows item parameter estimates to be influenced by all groups – even those with different levels of constraint – potentially obscuring the structure of public opinion in particular subgroups. The latter allows item parameter estimates to vary across subgroups, solving this concern. However, if a subgroup is highly homogeneous, then an intercept-only model may perform very well – leaving little potential for a one-dimensional model to improve predictions. In practice, the two approaches tell similar stories.

²⁰Interestingly, while our two approaches typically provide similar results, they diverge for conservatives and Republicans, with the combined approach showing larger gains than the split-sample approach. This suggests that conservatives and Republicans are highly distinct from the rest of the population, but there is not that much heterogeneity in survey responses among those who identify as conservatives or Republicans.

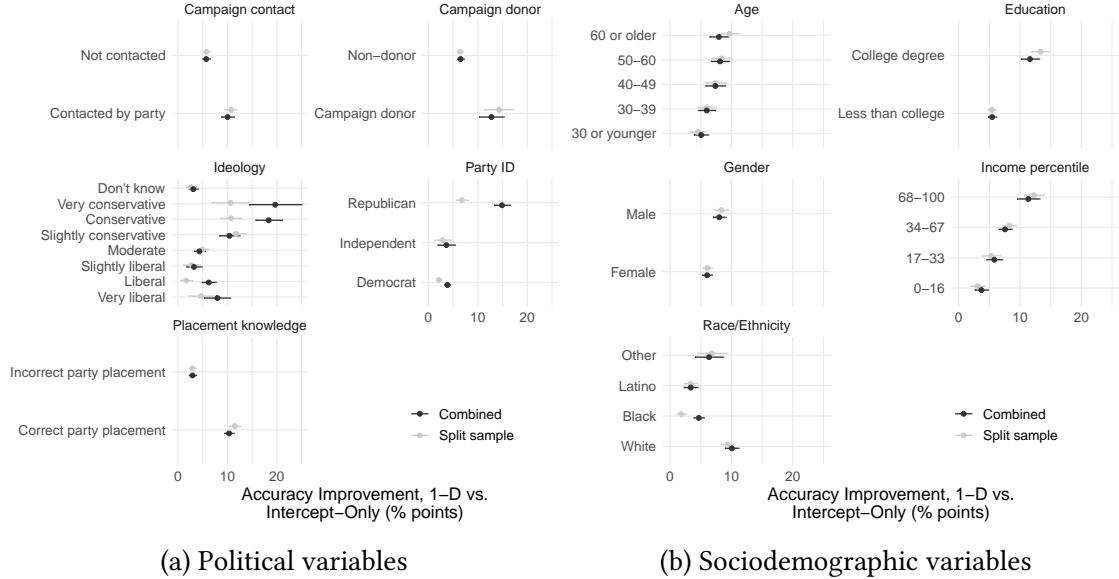


Figure 4: Constraint among subsets of the public. The points show the average respondent-level increase in classification accuracy obtained by a one-dimensional ideal point model relative to the intercept-only model. Bars show 95% confidence intervals. The black points show results when all respondents are used in the cross-validation procedure, while the gray points show results when cross-validation is performed on each subset separately. “Placement knowledge” refers to whether respondents place the Democratic Party to the left of the Republican Party on a 7-point ideology scale (Freder, Lenz and Turney, 2020).

interpretation is that the public simply has a lower degree of ideological constraint than political elites.

But there are at least three alternative explanations that we probe in this section. The first is that surveys and roll-call votes are very different environments. Survey respondents face few incentives to thoughtfully consider their responses before answering, which may lead to an increased amount of noise in their responses. In contrast, roll-call votes in Congress are “real-world” actions that provide obvious incentives to vote in particular ways.

Second, even if one grants that survey respondents reveal genuine preferences, one might object on the grounds that the set of topics covered in the data sets are not the same. If roll-call votes in Congress are simply better tools for discriminating ideology than survey questions, we would over-estimate the degree of constraint in Congress relative to the public.

Third, given the results of the previous section showing some evidence of heterogeneity in

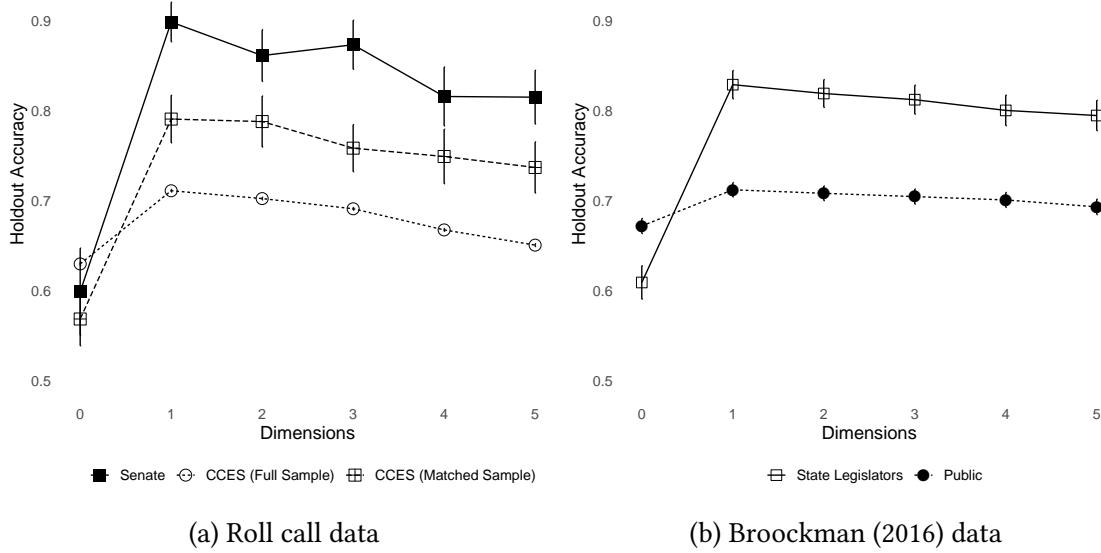


Figure 5: Comparison of cross-validation accuracy in the public and among elites, holding the survey items constant. Error bars show 95% confidence intervals, clustered at the respondent level. Even with common agendas and incentives, politicians exhibit a higher degree of constraint than the public.

the public, it could be that the increased constraint among politicians is simply explained by the unrepresentative demographics of politicians.

To assess the plausibility of these hypotheses, we take advantage of two paired datasets, in which politicians and the public respond to identical questions. Recall that the CCES questions correspond to roll-call votes that were recently held in Congress. This feature allows us to compare the performance of scaling methods using the exact set of issues in Congress and on the CCES. If differing agendas are the cause of the divergent results above, then we should see the divergence in constraint between the public and political elites shrink when restricting ourselves to a common agenda. Additionally, we can run the analysis on both the full set of CCES respondents as well as a subset that is selected to mirror — to the extent possible — the demographics of the Senate.²¹ If either the agenda or demographics explain our results, then this sample should show a similar degree of constraint as Senate roll-call voting.

The left-hand panel of Figure 5 shows the roll-call measures for the Senate and the CCES, for

²¹Details are in Online Appendix F.

both the full sample and the sample matched to Senate demographics. Despite including only 8 roll-call votes in the Senate, a one-dimensional model accurately classifies nearly 90 percent of votes — compared to less than 60 percent in the intercept-only model. In contrast, the accuracy among the public on the same questions goes from 63 percent to 71 percent in the full sample. The CCES sample that is demographically matched is somewhere in between, with accuracy increasing from 63 percent in the intercept-only model to 79 percent in the one-dimensional model. This finding provides confirmation of the prior result that there is heterogeneity in the level of constraint across the public. Still, even among this group of respondents who are demographically similar to Senators, their survey responses are still less predictable than Senators' roll call votes. Some, but not all, of the increased constraint seen among politicians can be explained by their demographics.

Finally, returning now to the first objection, if there are different incentives created by the roll-call context, we might still observe differences. To probe this question, we take advantage of a parallel survey of the mass public that Broockman (2016) conducted along with the state legislator survey. Here, legislators' responses were anonymous, so the incentive structure inherent to survey-taking is the same for both the mass public and elites.

The results are shown in the right-hand panel of Figure 5. They tell an even more stark story when the survey context is held fixed: politicians' responses are quite predictable, while the public's are not. This result suggests that it is not the use of surveys *per se* that is driving the difference between politicians and the public.

These results suggest that the driving force behind the divergent performance of ideal point models in the public relative to elites is primarily the differing levels of constraint of politicians as politicians — not a different agenda, different incentives faced by actors operating in public and private, nor the demographic characteristics of politicians, although this last element does play a part.

7 Conclusion

The main contributions of this paper are twofold. First, we formalize the notions of multidimensionality and constraint in a theory of political choices based on the canonical spatial voting model. From this discussion, we highlight observable implications that relate to the dimensionality of political conflict and the level of constraint in a given population. Second, we propose an out-of-sample validation strategy to evaluate empirically the structure of political choices in a broad range of data sources. We focus on out-of-sample validation both for substantive reasons — “constraint” naturally refers to how well a person’s opinion on one set of issues predicts her opinion on others — and for methodological reasons — in-sample measures of model fit are biased towards finding more dimensions and more constraint than are actually present.

The importance of out-of-sample validation is apparent from our empirical results: we find that ideal point models that contain more than a single dimension are counterproductive. Political choices in the United States, whether by survey respondents or Senators, are best approximated as one-dimensional.

In the Senate, this result is unsurprising. However, conventional wisdom holds that political opinions among the mass public may be more nuanced than a single left-right scale, implying that a higher-dimensional structure may exist. We find no evidence of such a higher-dimensional structure. If anything, moving beyond a single dimension tends to produce worse inferences in the public than among politicians due to overfitting.

Next, we turn to the issue of constraint. We operationalize constraint as the increase in predictive performance that can be achieved by a model that explicitly incorporates an actor’s ideal point, relative to a model that doesn’t. Using cross-validation, we show that political elites are highly constrained, while members of the mass public are relatively unconstrained. About 2/3 of the choices among politicians that cannot be predicted by an intercept-only model can be predicted when we estimate a model with a one-dimensional ideal point. In contrast, only about 20% of survey responses in the mass public that are unpredictable in an intercept-only model become predictable when the model includes an ideal point.

Using a series of paired data sets, we show that this difference in predictive performance cannot be attributed to the survey instrument, nor to differences in the agenda, nor to differing incentives faced by politicians and regular citizens. The most likely explanation, in our view, is the most simple: politicians organize politics in a more systematic way than most citizens. There is also evidence of heterogeneity in the public. Ideal point models fare better at predicting individual issue attitudes among people who look demographically more similar to politicians.

Substantively, these results suggest caution when applying ideal point models to survey responses from the mass public. While the public is best approximated as having one-dimensional ideal points, this ideal point does not predict attitudes on any given issue particularly well. In the public, it appears that idiosyncratic, rather than ideological, preferences explain the majority of voter attitudes. Our work suggests that scholars of public opinion should pay heed to both ideological and idiosyncratic portions of policy attitudes.

References

- Ahler, Douglas J. and David E. Broockman. 2018. “The Delegate Paradox: Why Polarized Politicians Can Represent Citizens Best.” *Journal of Politics* 80(4):1117–1133.
- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*. Budapest: pp. 267–281.
- Anscombe, Stephen, Jonathan Rodden and James M Snyder. 2006. “Purple America.” *Journal of Economic Perspectives* 20(2):97–118.
- Bafumi, Joseph and Michael C. Herron. 2010. “Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress.” *American Political Science Review* 104(03):519–542.
- Bond, Robert and Solomon Messing. 2015. “Quantifying Social Media’s Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook.” *American Political Science Review* 109(1):62–78.

- Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57(2):294–311.
- Broockman, David E. 2016. "Approaches to Studying Policy Representation." *Legislative Studies Quarterly* 41(1):181–215.
- Clinton, Joshua D., Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2):355–370.
- Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. David Apter. New York: The Free Press pp. 206–261.
- Crespin, Michael H. and David W. Rohde. 2010. "Dimensions, Issues, and Bills: Appropriations Voting on the House Floor." *Journal of Politics* 72(4):976–989.
- Freder, Sean, Gabriel S. Lenz and Shad Turney. 2020. "The Importance of Knowing 'What Goes With What': Reinterpreting the Evidence on Policy Attitude Stability." *Journal of Politics*.
- Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2 ed. New York: Springer.
- Heckman, James J. and James M. Snyder. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *The RAND Journal of Economics* 28(0):S142–S189.
- Hill, Seth J. and Chris Tausanovitch. 2015. "A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization." *Journal of Politics* 77(4):1058–1075.
- Imai, Kosuke, James Lo and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110(4):1–20.
- Jeong, Gyung-Ho, William R. Lowry, Gary J. Miller and Itai Sened. 2014. "How Preferences Change Institutions: The 1978 Energy Act Author(s):" *Journal of Politics* 76(2):430–445.
- Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(01):59.
- Kinder, Donald R. 2003. Belief Systems After Converse. In *Electoral Democracy*, ed. Michael MacKuen and George Rabinowitz. Ann Arbor: University of Michigan Press.

- Lauderdale, Benjamin E., Chris Hanretty and Nick Vivyan. 2017. "Decomposing Public Opinion Variation into Ideology, Idiosyncrasy and Instability." *Journal of Politics* pp. 1–9.
- Maraun, Michael D. and Natasha T. Rossi. 2001. "The Extra-Factor Phenomenon Revisited: Unidimensional Unfolding as Quadratic Factor Analysis." *Applied Psychological Measurement* 25(1):77–87.
- Paisley, John and Lawrence Carin. 2009. Nonparametric Factor Analysis with Beta Process Priors. In *Proceedings of the 26th International Conference on Machine Learning*. pp. 777–784.
- Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press.
- Salakhutdinov, Ruslan and Andriy Mnih. 2008. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the 25th international conference on Machine learning*. ACM pp. 880–887.
- Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6(2):461–464.
- Shor, Boris and Nolan M. McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3):530–551.
- Treier, S. and D. S. Hillygus. 2009. "The Nature of Political Ideology in the Contemporary Electorate." *Public Opinion Quarterly* 73(4):679–703.

Online Appendix for “The Structure of Political Choices: Distinguishing Between Constraint and Multidimensionality”

William Marble* Matthew Tyler†

July 16, 2020

*Ph.D. Candidate, Department of Political Science, Stanford University. wpmarble@stanford.edu

†Ph.D. Candidate, Department of Political Science, Stanford University. mdtyler@stanford.edu

Contents

A Technical Appendix: MultiScale Algorithm	2
B External Validation for the MultiScale Algorithm	4
C Simulation Study of Cross-Validation Estimator	6
D Data Appendix	8
E Alternative Measure of Fit	12
F Constructing the Matched CCES-Senate Sample	13
G Predicting Dyadic Agreement	14

A Technical Appendix: MultiScale Algorithm

The Parametric Model

N voters and J binary questions to vote on. The vote matrix is $Y \in \{0, 1\}^{N \times J}$. For some $D \in \mathbb{N}$, Let $\alpha_j \in \mathbb{R}$, $\beta_j \in \mathbb{R}^D$ and $\gamma_i \in \mathbb{R}^D$ for each $j = 1, \dots, J$ and $i = 1, \dots, N$. We assume the following latent variable model generates the binary vote matrix Y .

$$y_{ij} = I(s_{ij} > 0)$$

$$s_{ij} = \alpha_j + \beta_j^\top \gamma_i + \epsilon_{ij}, \quad \epsilon_{ij} \stackrel{\text{ind.}}{\sim} N(0, 1),$$

where we have assumed $\sigma = 1$, since it is not identified. Note that for $\theta = (\{\alpha_j\}_{j=1}^J, \{\beta_j\}_{j=1}^J, \{\gamma_i\}_{i=1}^N)$, this implies the reduced form likelihood

$$p(Y | \theta) = \prod_{i=1}^N \prod_{j=1}^J [\Phi(\alpha_j + \beta_j^\top \gamma_i)]^{y_{ij}} [1 - \Phi(\alpha_j + \beta_j^\top \gamma_i)]^{1-y_{ij}}.$$

Let $R \in \{0, 1\}^{N \times J}$ denote the matrix of observation statuses. That is $r_{ij} = 1$, if the (i, j) th cell of Y is observed and $r_{ij} = 0$ if it is missing. We assume the data are missing at random, $P(R | Y_{\text{obs}}, Y_{\text{mis}}, \theta, \omega) = P(R | Y_{\text{obs}}, \theta, \omega)$, and that the parameters ω that determine R are distinct from the structural voting parameters θ , meaning that we can ignore the likelihood of R (Section 6.2, Little and Rubin, 2014). The resulting (ignorable) likelihood is

$$p(Y_{\text{obs}} | \theta) = \prod_{i=1}^N \prod_{j=1}^J \left\{ [\Phi(\alpha_j + \beta_j^\top \gamma_i)]^{y_{ij}} [1 - \Phi(\alpha_j + \beta_j^\top \gamma_i)]^{1-y_{ij}} \right\}^{r_{ij}}.$$

We assume standard priors on θ ; specifically,

$$\xi(\theta) = \prod_{j=1}^J N \left(\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix}; \mu_{ab}, \Sigma_{ab} \right) \prod_{i=1}^N N(\gamma_i; \mu_Y, \Sigma_Y),$$

where $\mu_{ab} \in \mathbb{R}^{D+1}$, $\mu_Y \in \mathbb{R}^d$ and $\Sigma_{ab} \in \mathbb{R}^{(D+1) \times (D+1)}$, $\Sigma_Y \in \mathbb{R}^{D \times D}$ are positive definite matrices.

The Algorithm

We consider just the log likelihood to illustrate how we extend the algorithm of Imai, Lo and Olmsted (2016). Let $m_{ij} = \alpha_j + \beta_j^\top \gamma_i$ and S_{obs} be the values of S that correspond to the Y observed values Y_{obs} . The complete-data (complete here is with respect to S_{obs} , not the values of Y for which $r_{ij} = 0$) log likelihood is given by

$$\begin{aligned} & \log p(Y_{\text{obs}}, S_{\text{obs}} \mid \theta) \\ &= \log \prod_{i=1}^N \prod_{j=1}^J \left[N(s_{ij} \mid m_{ij}, 1) \right]^{r_{ij} \left(I(y_{ij}=1)I(s_{ij} \geq 0) + I(y_{ij}=0)I(s_{ij} < 0) \right)} \\ &= \sum_{i=1}^N \sum_{j=1}^J r_{ij} \left(I(y_{ij} = 1)I(s_{ij} \geq 0) + I(y_{ij} = 0)I(s_{ij} < 0) \right) \log N(s_{ij}; m_{ij}, 1). \end{aligned}$$

But this is the same complete-data log likelihood found in Imai, Lo and Olmsted (2016, Appendix A), except for the insistence on only using the observed data as observations. Therefore we can take their update equations and restrict ourselves to only using observed data. Specifically, iterate between

$$s_{ij} \leftarrow m_{ij} + (2y_{ij} - 1) \frac{\phi(m_{ij})}{\Phi((2y_{ij} - 1)m_{ij})}$$

$$\gamma_i \leftarrow \left(\Sigma_Y^{-1} + \sum_{j=1}^J r_{ij} \beta_j \beta_j^\top \right)^{-1} \left(\Sigma_Y^{-1} \mu_Y + \sum_{j=1}^J r_{ij} \beta_j (s_{ij} - \alpha_j) \right)$$

$$\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \leftarrow \left(\Sigma_{ab}^{-1} + \sum_{i=1}^N r_{ij} \begin{bmatrix} 1 \\ \gamma_i \end{bmatrix} \begin{bmatrix} 1 \\ \gamma_i \end{bmatrix}' \right)^{-1} \left(\Sigma_{ab}^{-1} \mu_{ab} + \sum_{i=1}^N r_{ij} s_{ij} \begin{bmatrix} 1 \\ \gamma_i \end{bmatrix} \right).$$

B External Validation for the MultiScale Algorithm

In the following figures, we plot several validation measure for the MultiScale algorithm. They show that for several data sources, the MultiScale estimates correlate highly with other measures of ideology.

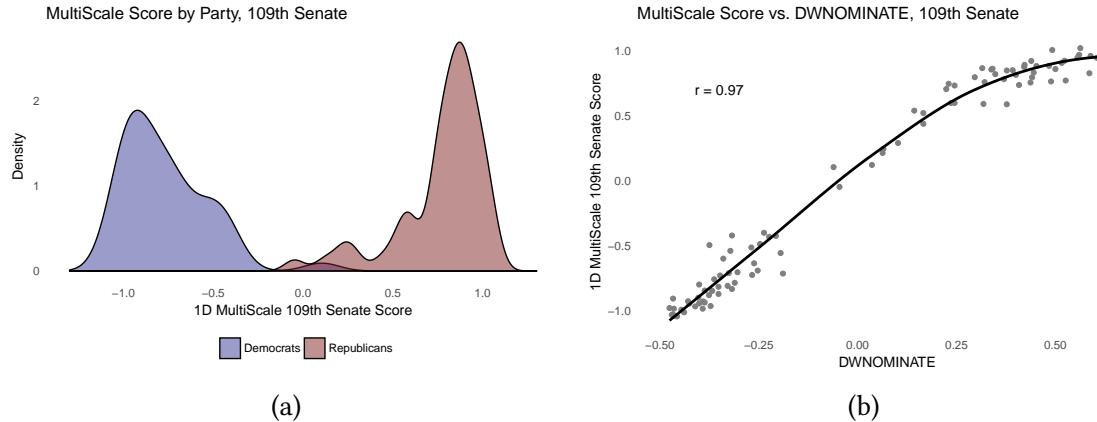


Figure 1: Comparison of MultiScale ideal points among Senators in the 109th Congress. The left-hand side shows that Democrats are almost universally to the left of Republicans. The right-hand side shows that MultiScale scores are highly correlated with DW-NOMINATE scores.

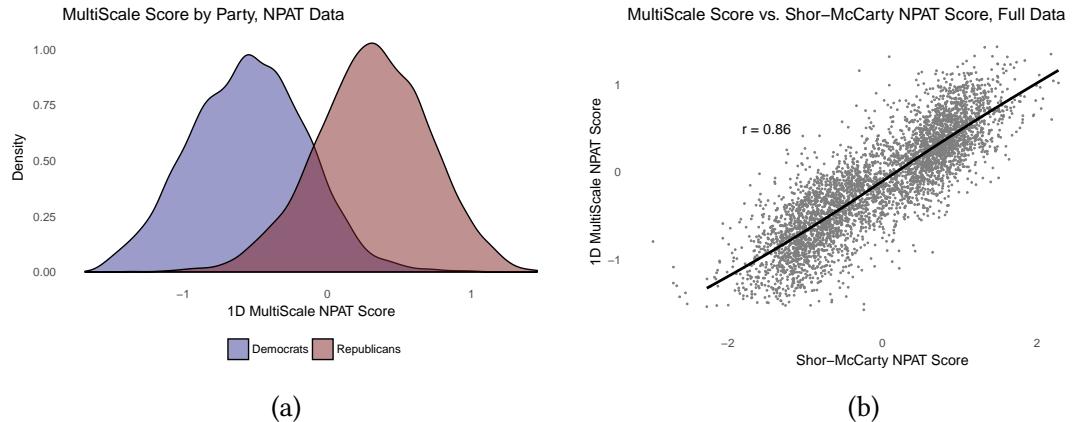


Figure 2: Comparison of MultiScale ideal points among politicians using NPAT data. The left-hand side shows that Democrats are consistently to the left of Republicans. The right-hand panel shows the correlation between MultiScale scores estimated with the NPAT data to the Shor-McCarty NPAT scores.

MultiScale Score by Party, ANES 2012

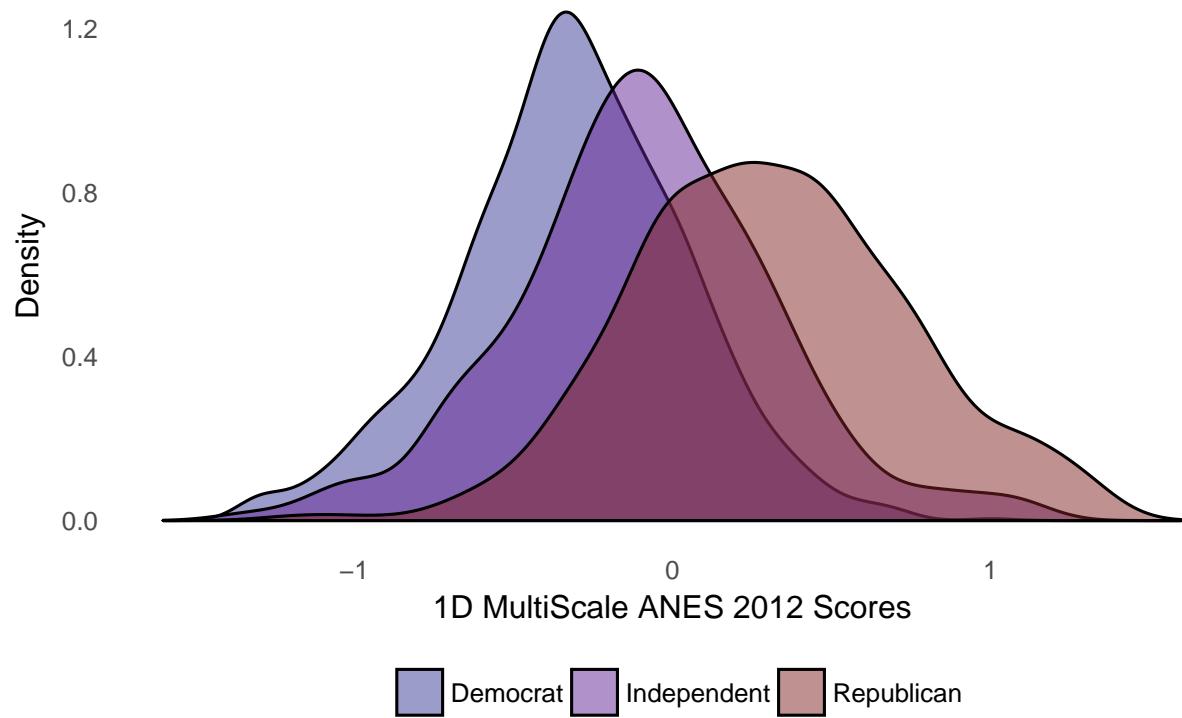


Figure 3: Comparison of MultiScale ideal points from the ANES by party.

C Simulation Study of Cross-Validation Estimator

To illustrate that our proposed method of out-of-sample validation can accurately recover the latent dimensionality of political choices, we conduct a small simulation study.

We simulate data sets according to the spatial voting model laid out in Section 2 in the main text. First, we fix the number of actors and choices (N and J , in the notation of the paper). Then, we simulate a series of choice matrices Y^D generated according to a $D \in \{0, \dots, 5\}$ dimensional ideal point model. In particular, the voter surplus for voter i on choice j is modeled as

$$s_{ij} = \alpha_j + \beta_j^{D^\top} \gamma_i^D + \epsilon_{ij} \quad (1)$$

$$\epsilon_{ij} \stackrel{\text{ind.}}{\sim} N(0, 1) \quad (2)$$

$$Y_{ij}^D = I(s_{ij} > 0), \quad (3)$$

where we explicitly denote the dimensionality with superscripts. We draw α_j independent standard normal (separately for each dimension) and γ_i^D from a multivariate standard normal. To ensure that all D dimensions are in fact relevant to the choice, we restrict each element of β_j^D to be either -1 or 1 , chosen randomly.

For each simulated data set, we run the cross-validation procedure outlined in the text. We estimate the predictive error associated with estimating models that assume 0 to 5 dimensions. If the cross-validation procedure can correctly measure the dimensionality of the data, the accuracy should be maximized when we estimate a model that assumes the same dimensionality that the data were generated with.

We repeat this exercise twice. First, we simulate data sets that are approximately the same size as the Senate data used in the main text, with $N = 102$ and $J = 645$. Second, we simulate data sets approximately the same size as the ANES, with $N = 1,000$ and $J = 20$.

The results are shown in Figure 4. The left-hand panel shows the results for the Senate-sized data and the right-hand panel shows the results for the ANES-sized data. In all cases, the estimated dimensionality is the same as the true dimensionality, providing evidence that the

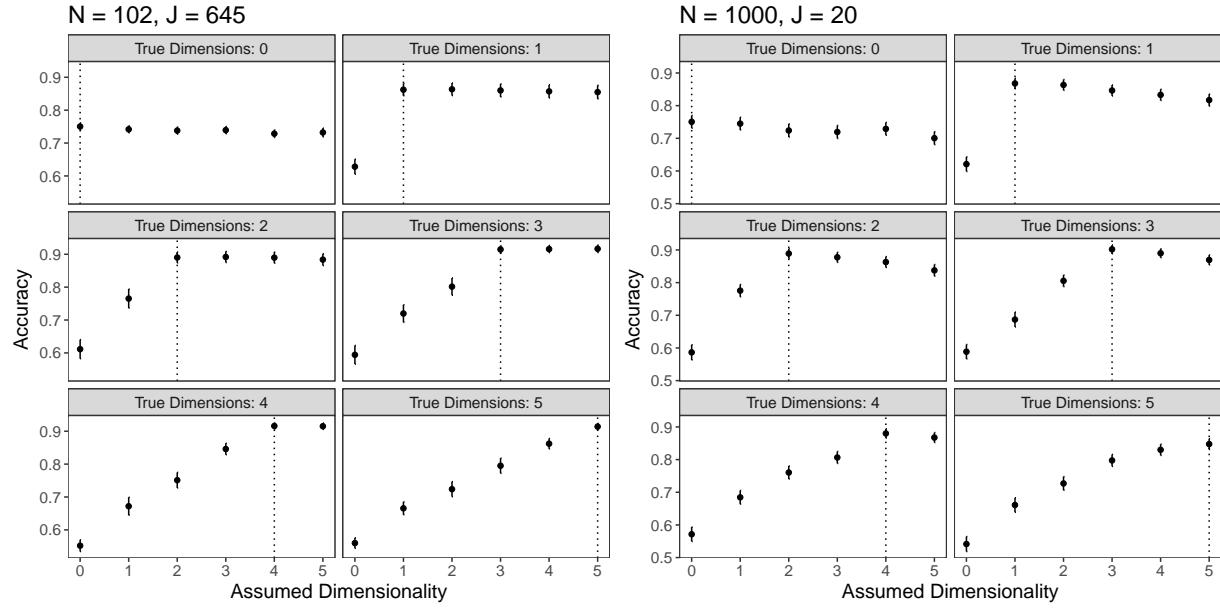


Figure 4: Simulation results. Left-hand panel shows Senate-sized data; right-hand panel shows ANES-sized data. Dotted vertical lines indicate the dimensionality of the true data-generating process.

validation strategy proposed in the paper can accurately recover the dimensionality of the data-generating process.

D Data Appendix

In this Appendix, we present the items used from each of the data sources referenced in the main text. For items that had more than two responses categories, we binarize them by splitting them at the mean (treating ordinal variables as cardinal). We could split them at the median with substantively similar results. Table 1 shows the number of rows and columns in each data source, along with patterns of missingness and number of questions for which we observe $K = 1, 2, \dots, 7$ responses. The subsequent tables list and describe the variables used from each data source.

Data source	Rows	Columns	% Missing	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$	$K = 7$
109th Senate	102	645	4.5	101	544	0	0	0	0	0
NPAT	12,794	225	79.1	0	225	0	0	0	0	0
Legislator Survey	225	31	5.5	0	19	0	0	0	1	11
2012 ANES	2,054	28	10.2	0	6	9	1	7	0	5
2012 CCES Roll Calls	54,068	10	3.2	0	10	0	0	0	0	0

Table 1: Description of data sources used in the main analyses. The columns labeled $K = 1, 2, \dots, 7$ show the number of questions with K observed responses. For example, $K = 1$ means the responses to a given item were unanimous, $K = 2$ means there were two observed responses for a given item, and so on.

Broockman (2016) State Legislator Variables

Variable	Description
iq_vouchers	The government should provide parents with vouchers to send their children to any school they choose, be it private, public, or religious. (Binary)
iq_medicalpot	Allow doctors to prescribe marijuana to patients. (Binary)
iq_taxesover250k	Increase taxes for those making over \$250,000 per year. (Binary)
iq_overturnroe	Overturn Roe v. Wade. (Binary)
iq_privitsocialsec	Allow workers to invest a portion of their payroll tax in private accounts that they can manage themselves. (Binary)
iq_gaymarriage	Same-sex couples should be allowed to marry. (Binary)
iq_unihealth	Implement a universal health care program to guarantee coverage to all Americans, regardless of income. (Binary)

iq_medlawsuits	Limit the amount of punitive damages that can be awarded in medical malpractice lawsuits. (Binary)
iq_guncontrol	There should be strong restrictions on the purchase and possession of guns. (Binary)
iq_illegalim	Illegal immigrants should not be allowed to enroll in government food stamp programs. (Binary)
iq_enda	Include sexual orientation in federal anti-discrimination laws. (Binary)
iq_affaction	Prohibit the use of affirmative action by state colleges and universities. (Binary)
iq_unfunding	The US should contribute more funding and troops to UN peacekeeping missions. (Binary)
iq_fundarts	The government should not provide any funding to the arts. (Binary)
iq_dealhpenalty	I support the death penalty in my state. (Binary)
iq_repealcapgainstax	Repeal taxes on interest, dividends, and capital gains. (Binary)
iq_epaprohibit	Prohibit the EPA from regulating greenhouse gas emissions. (Binary)
iq_birthcontrolmandate	Health insurance plans should be required to fully cover the cost of birth control. (Binary)
iq_subsidizeloans	The federal government should subsidize student loans for low income students. (Binary)
eq_guns	Which statement comes closest to describing your views on gun control? (1-7 scale)
eq_health	Which statement comes closest to describing your views on the issue of health care? (1-7 scale)
eq_immigration	Which statement comes closest to describing your views on immigration? (1-7 scale)
eq_taxes	Which statement comes closest to describing your views on taxes? (1-7 scale)
eq_abortion	Which statement comes closest to describing your views on abortion? (1-7 scale)
eq_environment	Which statement comes closest to describing your views on pollution and the environment? (1-7 scale)
eq_medicare	Which statement comes closest to describing your views on Medicare, the government's program for covering the elderly's health care costs? (1-7 scale)
eq_gays	Which statement comes closest to describing your views on rights for gays and lesbians? (1-7 scale)
eq_affirmativeaction	Which statement comes closest to describing your views on affirmative action in higher education? (1-7 scale)
eq_unions	Which statement comes closest to describing your views on unions? (1-7 scale)

eq_education	Which statement comes closest to describing your views on public funding for private school education? (1-7 scale)
eq_contraception_version2	Which statement comes closest to describing your views on birth control? (1-7 scale)

2012 ANES Variables, from Hill and Tausanovitch (2015)

Variable	Description
VCF0806	R Placement: Government Health Insurance Scale
VCF0809	R Placement: Guaranteed Jobs and Income Scale
VCF0823	R Opinion: Better off if U.S. Unconcerned with Rest of World
VCF0830	R Placement: Aid to Blacks Scale
VCF0838	R Opinion: By Law, When Should Abortion Be Allowed
VCF0839	R Placement: Government Services/Spending Scale
VCF0843	R Placement: Defense Spending Scale
VCF0867a	R Opinion: Affirmative Action in Hiring/Promotion [2 of 2]
VCF0876a	R Opinion Strength: Law Against Homosexual Discrimination
VCF0877a	R Opinion Strength: Favor/Oppose Gays in Military
VCF0878	R Opinion: Should Gays/Lesbians Be Able to Adopt Children
VCF0879a	R Opinion: U.S. Immigrants Should Increase/Decrease [2 of 2]
VCF0886	R Opinion: Federal Spending- Poor/Poor People
VCF0887	R Opinion: Federal Spending- Child Care
VCF0888	R Opinion: Federal Spending- Dealing with Crime
VCF0889	R Opinion: Federal Spending- Aids Research/Fight Aids
VCF0894	R Opinion: Federal Spending- Welfare Programs
VCF9013	R Opinion: Society Ensure Equal Opportunity to Succeed
VCF9014	R Opinion: We Have Gone Too Far Pushing Equal Rights
VCF9015	R Opinion: Big Problem that Not Everyone Has Equal Chance
VCF9037	R Opinion: Government Ensure Fair Jobs for Blacks
VCF9040	Blacks Should Not Have Special Favors to Succeed
VCF9047	R Opinion: Federal Spending- Improve/Protect Environment
VCF9048	R Opinion: Federal Spending- Space/Science/Technology
VCF9049	R Opinion: Federal Spending- Social Security
VCF9131	R Opinion: Less Government Better OR Government Do More
VCF9132	R Opinion: Govt Handle Economy OR Free Market Can Handle
VCF9133	R Opinion: Govt Too Involved in Things OR Problems Require

2012 CCES Variables

Variable	Description
CC332A	roll-call votes - Ryan Budget Bill
CC332B	roll-call votes - Simpson-Bowles Budget Plan
CC332C	roll-call votes - Middle Class Tax Cut Act
CC332D	roll-call votes - Tax Hike Prevention Act
CC332E	roll-call votes - Birth Control Exemption
CC332F	roll-call votes - U.S.-Korea Free Trade Agreement
CC332G	roll-call votes - Repeal Affordable Care Act
CC332H	roll-call votes - Keystone Pipeline
CC332I	roll-call votes - Affordable Care Act of 2010
CC332J	roll-call votes - End Don't Ask, Don't Tell

Matched roll-call votes, Senate

We extracted data on roll-call votes corresponding to the CCES questions from voteview.com.

Roll-call votes are on final passage, where applicable. In the case of issues that were voted on multiple times, we take the vote closest to the 2012 election.

Issue	Congress	Vote Number
Affordable Care Act	111th	396
Repeal Don't Ask, Don't Tell	111th	281
Tax Hike Prevention Act	111th	276
Ryan budget	112th	277
Middle Class Tax Cut Act	112th	184
US-Korea Free Trade Agreement	112th	161
Affordable Care Act Repeal	112th	9
Birth Control Exemption	112th	24
Keystone Pipeline	113th	280

E Alternative Measure of Fit

In this appendix we replicate the plots from the main text, except instead of accuracy we use the average likelihood of the observed responses. Working with the likelihood is slightly less interpretable, but has the advantage of being able to distinguish between correct classifications that are “just barely” correct (e.g., 51% likelihood of observed response) and correct classifications that have a higher degree of confidence (e.g., 95% likelihood of observed response).¹ The substantive conclusions remain unchanged.

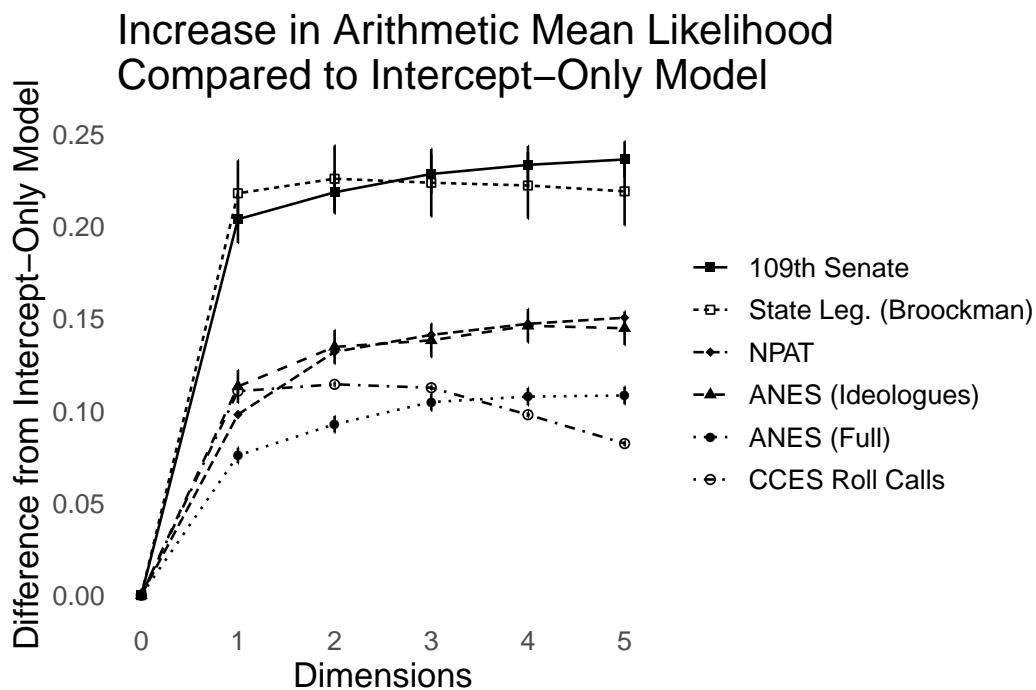


Figure 5: Increase in average cross-validation likelihood of observed response over an intercept-only model. Error bars show 95% confidence intervals clustered by respondent.

¹Technically, the likelihood is a proper scoring rule while accuracy is not, meaning that the likelihood is maximized by the true model. Given that the substantive conclusions drawn are not sensitive to the use of accuracy and it is more easily interpretable, we focus on that in the main text.

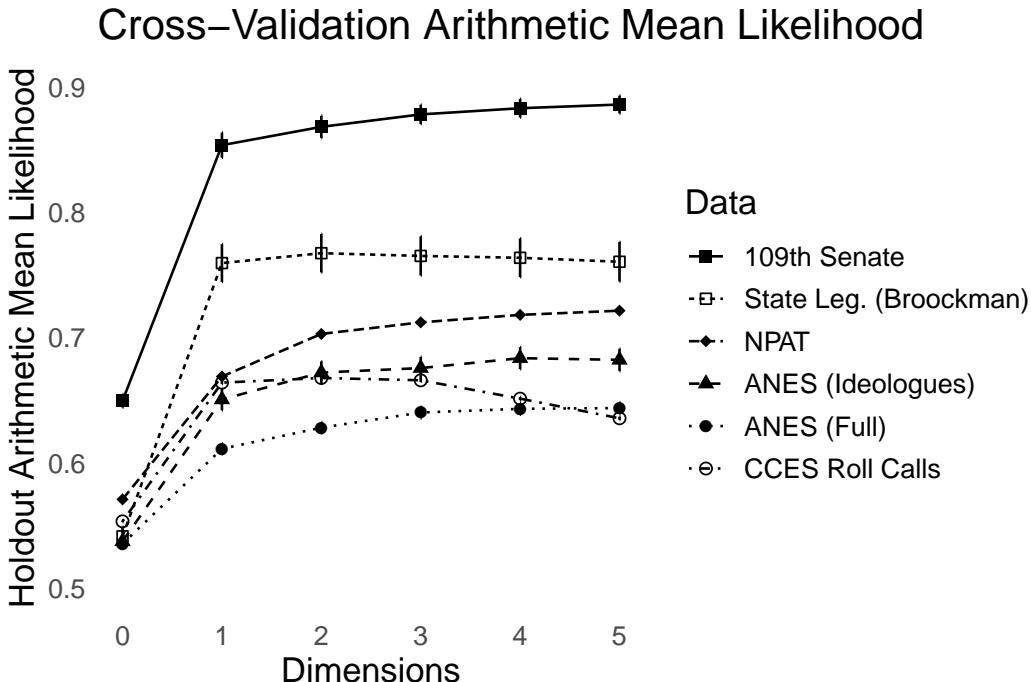


Figure 6: Average cross-validation likelihood of observed response. Error bars show 95% confidence intervals clustered by respondent.

F Constructing the Matched CCES-Senate Sample

In Section 6.3, we compare Senate roll call votes to the CCES responses of a subset of respondents who are demographically similar to Senators. Here, we describe the matching process.

We begin with data on the demographics Senators collected by Carnes (2013) and the Congressional Research Service (Petersen, 2012). Compared to the general public, Senators skew much richer, highly educated, older, white, and male. Even with the large sample size afforded by the CCES, it is difficult to find a sample that matches the demographics of Senators exactly.

Still, we create a more similar sample, we first subset the CCES to respondents whose family incomes exceed \$150,000, who identify as either Democrats or Republicans, who are at least 35 years old, and who have at least a college degree. This leaves us with a sample of around 1,200. Then, within this sample, we use a raking procedure to construct weights that target the demographics of Senators on the following variables: age (in 10-year bins), sex, education (college or graduate school), and party. We use the R package `anesrake` to construct the weights (Pasek,

2018). We then sample 500 respondents in proportion to their weights to create a matched sample. Table 3 compares the distribution of age, party, education, and sex in the Senate to the matched sample.

Variable	Senate (%)	CCES (%)	Difference
Age			
30-39	0.3	0.2	-0.1
40-49	11.3	15.0	3.7
50-59	32.0	33.0	1.0
60-69	33.3	33.2	-0.1
70+	22.7	18.6	-4.1
Party			
Democrat	47.7	51.2	3.5
Republican	52.3	48.8	-3.5
Education			
Post-graduate	47.0	56.0	9.0
No post-graduate	53.0	44.0	-9.0
Sex			
Male	85.3	80.2	-5.1
Female	14.7	19.8	5.1

Table 3: Balance table comparing demographics of Senators in 108th-110th Congresses to the matched CCES sample. All CCES respondents reported a family income of at least \$150,000, are at least 35 years old, and have at least a college degree. Senate demographics drawn from Carnes (2013) and Petersen (2012).

G Predicting Dyadic Agreement

Another way of demonstrating the limited gains from fitting a higher-dimensional model is by examining agreement scores between pairs of respondents and relating them to estimated distance between their ideal points in D dimensions. Intuitively, respondents who tend to respond to the same questions should be ideologically similar, and thus the distance between their ideal points should be small. If a higher-dimensional model provides a better description of survey responses, we should see a meaningfully different relationship between agreement scores and estimated ideal points across different choices of D .

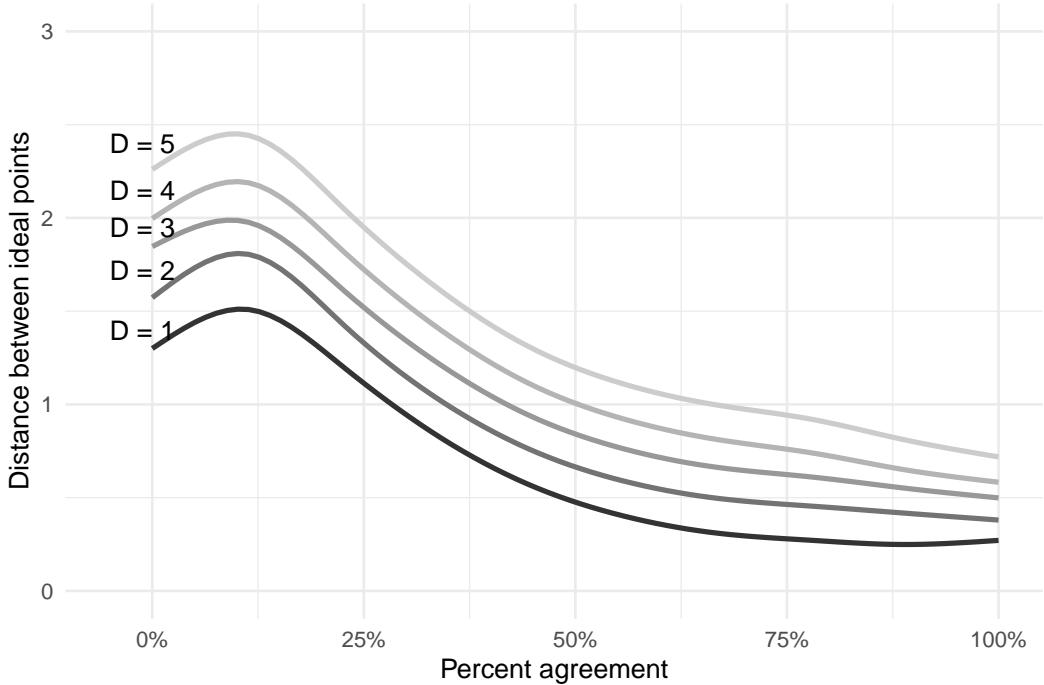


Figure 7: Relationship between ideal point distance and agreement score for each pair of 2012 ANES respondents. The reason for the non-monotonicity for low agreement scores is small sample size: respondents who answer 0 questions in the same way tend to have answered very few common questions, so those questions have relatively little influence on the ideal point estimates.

Formally, denote respondents $i = 1, \dots, N$ and questions $j = 1, \dots, J$. For the (undirected) respondent pair (i, k) , denote the set of questions to which they both provide a response S_{ik} . If the respondents provide the same response to question $j \in S_{ik}$, then we say $x_{ik}^j = 1$, and 0 otherwise. The agreement score for (i, k) is then $y_{ik} = \frac{1}{|S_{ik}|} \sum_{j \in S_{ik}} x_{ij}^j$, which is simply the proportion of times that i and k answer in the same way.

We compute this quantity for each respondent pair in the 2012 ANES (for a total of 2 million pairs), and also compute the Euclidean distance between their estimated ideal points in $D = 1, \dots, 5$ dimensions. Ideal points are estimated using the full data set, without any holdout data.

Figure 7 plots the distance between ideal points against the agreement score for each respondent pair. While there is an intercept shift, with respondents tending to be farther apart from each other in higher-dimensional space, the basic relationship between agreement scores and distance is the same across dimensions. This result further reinforces the idea that there is little additional

insight gained in the public beyond one dimension.

References

- Broockman, David E. 2016. "Approaches to Studying Policy Representation." *Legislative Studies Quarterly* 41(1):181–215.
- Carnes, Nicholas. 2013. *White-Collar Government: The Hidden Role of Class in Economic Policy Making*. Chicago: University of Chicago Press.
- Hill, Seth J. and Chris Tausanovitch. 2015. "A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization." *Journal of Politics* 77(4):1058–1075.
- Imai, Kosuke, James Lo and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110(4):1–20.
- Little, Roderick J A and Donald B Rubin. 2014. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Pasek, Josh. 2018. *anesrake: ANES Raking Implementation*. R package version 0.80.
URL: <https://CRAN.R-project.org/package=anesrake>
- Petersen, R. Eric. 2012. "Representatives and Senators: Trends in Member Characteristics Since 1945."
URL: <https://fas.org/sgp/crs/misc/R42365.pdf>