

(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?

Stephen Jessee University of Texas at Austin

Abstract: *Estimating the ideological positions of political elites on the same scale as those of ordinary citizens has great potential to increase our understanding of voting behavior, representation, and other political phenomena. There has been limited attention, however, to the fundamental issues, both practical and conceptual, involved in conducting these joint scalings, or to the sensitivity of these estimates to modeling assumptions and data choices. I show that the standard strategy of estimating ideal point models using preference data on citizens and elites can suffer from potentially problematic pathologies. This article explores these issues and presents a technique that can be used to investigate the effects of modeling assumptions on resulting estimates and also to impose restrictions on the ideological dimension being estimated in a straightforward way.*

Replication Materials: The data, code, and any additional materials required to replicate all analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <http://dx.doi.org/10.7910/DVN/IIYGGX>.

Many theories and hypotheses in political science deal with the ideological positions of citizens in relation to those of candidates, elected representatives, or other political elites. In recent years, scholars have used new combinations of survey and statistical techniques to estimate the ideology of survey respondents and political elites on the same scale using their positions on specific policy proposals. While these new measures have shown potential to provide new insights in areas such as voting behavior and representation, little attention has been paid thus far to the properties of these estimates and to potential issues that can arise in these joint scaling exercises.

In their classic article, Aldrich and McKelvey (1977) introduce a method that estimates the ideological positions of political actors (e.g., parties, candidates, legislators) on the same scale as citizens based on survey respondents' placements of these actors on ideological perception scales (see also Palfrey and Poole 1987, or

for a Bayesian implementation of this model that allows for its use in the presence of missing data and produces uncertainty estimates for stimuli positions, Hare et al. 2015).¹ Other works have estimated ideological positions by scaling citizen preferences for or ratings of candidates (Hinich, Cahoon, and Ordeshook 1978; Weisberg and Rusk 1970). More recently, a sizable literature has developed estimating policy-based ideal points for citizens and political elites on the same scale. This work has leveraged survey questions answered by ordinary citizens that can, in one way or another, be matched to positions taken by legislators or candidates.

For example, Jessee (2009) uses such ideology estimates for citizens and presidential candidates to test predictions related to spatial voting theory (see also Jessee (2010a, 2010b, 2012)), whereas Bafumi and Herron (2010) estimate the ideological positions of members of Congress and their constituencies in order to assess the characteristics of representation. Furthermore, Jessee

Stephen Jessee is Associate Professor of Government, University of Texas at Austin, 1 University Station A1800, Austin, TX 78712 (sjessee@utexas.edu).

The author would like to thank Devin Caughey, Benjamin Lauderdale, Jeff Lewis, Neil Malhotra, Scott Moser, Michael Peress, Kevin Quinn, James Scott, Boris Shor, Chris Tausanavitch, Chris Warshaw, Chris Tausanavitch, Teppei Yamamoto, and especially Alexander Tahk for helpful comments. Earlier versions of this research were presented at the Ideology: Structures, Causes and Consequences Mini-Conference at the Department of Government, Georgetown University; Washington University Political Economy Speaker Series; and Conference on Ideal Point Models at MIT, WPSA, MPSA, and APSA. Part of this research was supported by a grant from the Office of the Dean of the College of Liberal Arts at the University of Texas at Austin.

¹Many other works have used some variant of estimation based on survey respondents' ideological placements (e.g., Adams, Merrill, and Grofman 2005; Alvarez and Nagler 1995; Brady and Sniderman 1985).

American Journal of Political Science, Vol. 60, No. 4, October 2016, Pp. 1108–1124

© 2016, Midwest Political Science Association

DOI: 10.1111/ajps.12250

and Malhotra (2013) and Malhotra and Jessee (2014) analyze survey respondents' stated positions on specific Supreme Court cases to estimate citizens' ideologies alongside those of individual justices and the Court as a whole. Variants of this strategy have also been applied to lower levels of government, with Shor and McCarty (2011) estimating the ideological positions of state legislators across states, and Tausanovich and Warshaw (2014) estimating the ideologies of American cities as well as their government outputs.

Techniques related to joint scaling have also estimated ideological positions for other types of actors based on many different types of data. These include Groseclose and Milyo (2005), who estimate the positions of media outlets alongside members of Congress; Bailey (2007), who estimates a single ideological scale for courts, Congress, and the president across time; and Bonica (2014), who estimates the positions of candidates and donors from campaign contribution data.

These studies have provided important insights, but the joint scaling methods used rely on several strong assumptions. In particular, it is usually assumed that there is a single ideological dimension that structures both citizen and legislator views across different policies in the same way. The literature to date has paid little attention to these concerns. In fact, the standard approach to joint scaling involves applying some sort of scaling procedure to a data set that includes bridge items, assuming (usually tacitly) that the model will estimate the correct dimension—that is, the dimension relevant for the theory or hypothesis under study.

This article addresses these issues, beginning by asking whether joint scalings of citizens and legislators are robust to seemingly innocuous factors such as the number of respondents included in the data. I analyze two data sets in which the policy positions of ordinary citizens are measured on the same issues as those of elected officials, identifying potential problems with the standard approach to joint scaling. I consider what should be done in the face of discrepancies between the structure of ideology in different groups—does this render the entire enterprise of joint scaling futile, or does there remain a useful way forward? I introduce an approach for estimating the ideology of members of multiple groups on the same scale under the constraint that the ideological dimension is structured based on the data from one particular group. This approach can be used to assess the similarity between the ideological dimensions underlying the policy views of citizens and legislators as well as to impose desired structure on estimates. I conclude by arguing that while it is centrally important for researchers to explore the validity of joint scaling assumptions and

results, the question of how to structure these estimations should ultimately be driven by substantive and theoretical concerns more than specific thresholds or statistical tests.

Jointly Scaling Groups with “Bridging” Ideal Point Analyses

The basic idea underlying ideal point modeling is that an ideological space, typically consisting of one or a small number of dimensions, underlies the revealed preferences of political actors. These data are seen as indicators, which are generated stochastically based on each actor's underlying ideal point and the characteristics of the policies being voted on. Ideal point models thus provide a way to uncover a latent space that structures the preferences of the actors under study, reducing a large number of variables into a single-dimensional or low-dimensional representation of preferences.

Once a set of indicators has been chosen that is thought to tap the latent trait of interest, researchers must choose a model and method for estimating these underlying values. Many recent works scaling respondents and legislators together have used the ideal point model from Clinton, Jackman, and Rivers (CJR; 2004). Other alternatives include NOMINATE (Poole and Rosenthal 1985) and factor-analytic techniques (e.g., Heckman and Snyder 1997). In practice, the specific form of the ideal point model tends to have only a minor impact on the resulting estimates. Because it has been most commonly used in recent joint scalings studies, I focus on the CJR model here.

The CJR ideal point model assumes that each actor, indexed by i , casts votes on a series of proposals, indexed by j based on quadratic utility functions over alternatives subject to independent normal disturbances, which can be shown to yield the following probit-link ideal point model for policy positions:

$$P(y_{ij} = 1) = \Phi [\beta_j (x_i - \gamma_j)], \quad (1)$$

where β_j is policy j 's discrimination parameter, indicating the strength and direction of the relationship between an actor's ideal point x_i and his or her likelihood of supporting the policy, and γ_j represents the cutpoint for vote j , which lies halfway between the “yea” and “nay” alternatives (or “support” and “oppose” positions).²

When estimating the ideological positions of actors of multiple types (e.g., legislators and ordinary citizens) on the same scale, it is typically necessary to observe

²Note that the specification here differs from CJR in transforming the difficulty parameter α_j by dividing by β_j to produce a more interpretable cutpoint parameter.

common items between the two groups, often called “bridge” items. Under this setup, β_j and γ_j can be assumed to be the same for the two groups on each of the bridging items. This allows researchers to pool members of the two groups together and estimate their ideology on the same scale. A key assumption here is that ideological space underlying the preferences of the two groups is structured in the same way.

Following Equation (2), we can consider two (possibly identical) ideological spaces, with each defined by the relationship for actors from a given group between ideological position x_i and the likelihood of supporting each policy proposal.³ In other words, these spaces can be defined by β_j and γ_j , which are now allowed to vary not just across policy items j , but also across groups so that we obtain

$$P(y_{ij} = 1) = \Phi [\beta_{g(i),j} (x_i - \gamma_{g(i),j})], \quad (2)$$

where $g(i)$ is the group of actor i .

If the item parameters are different for the two groups, it is not clear how we can compare the ideal points between the two groups. There are multiple reasons why we may worry about this. For example, if a given application uses survey questions about specific votes in Congress, we may worry that the questions do not correspond perfectly with the policies being voted on. This could be because of the wording of the questions or the context in which the decisions are being made by survey respondents as opposed to legislators. Alternatively, it could be that even though the actual items are identical (or nearly so) between the two groups, the structure of the ideological dimension is simply different. For example, support for a certain policy could be strongly related to ideological position for members of Congress, but not for ordinary citizens.

One way to think of the assumptions underlying these bridging estimations is very rigidly—either the item parameters are all exactly equal between the two groups or they are not. But this sharp approach takes very literally a model that is intended as an approximation of the process by which people take positions on various policies. For example, in Congress, it is clear that different types of members (e.g., Tea Party Republicans or Blue Dog Democrats) have different structures underlying their preferences. The political science literature on latent traits estimation of political ideology includes many examples of choosing parsimony over complexity, even when a literal interpretation of the model relying on formalized hypothesis tests might suggest a different strategy. For example, ideal point

³It is also possible to have more than two classes of actors, each with their own ideological space.

models of congressional voting typically assume only one or two dimensions (Clinton, Jackman, and Rivers, 2004; Poole and Rosenthal 1985), citing this as a useful balance of explanatory power and parsimony.⁴ The question here, following the classic quote, is not whether bridging ideal point models are wrong, but whether they are useful.

Data: Senate Representation Survey

The Senate Representation Survey, previously analyzed in Jessee (2009) and Jessee (2012), presents a particularly good test case for bridging ideal point analyses. Fielded between December 2005 and January 2006, the survey includes policy questions written to correspond to specific Senate roll-call votes from 2004 and 2005.⁵ The survey was administered online to 5,871 respondents from the Polimetrix (now YouGov/Polimetrix) online panel. The sample was not constructed to be representative at the national level. In particular, because one of the aims of the study was to analyze respondent perceptions of their senators, at least 100 respondents from each state were included in the sample. The sample also includes higher levels of political information on average than nationally representative surveys, such as the 2004 American National Election Studies survey, and includes fewer weak partisans and minorities. A list of the 27 policy questions analyzed here is shown in Table 1.

The Senate Representation Survey is particularly well suited for examining the assumptions of bridging ideal point analyses for several reasons. First, it contains a large number of questions on a wide range of policies that were voted on in the Senate. The policies include more mainstream issues such as gun control and raising the minimum wage as well as more obscure policies such as bankruptcy reform and overtime regulations, on which fewer respondents may have well-thought-out views. In this way, the Senate Representation Survey might be thought to represent a “hard test” for joint scaling because it was designed in part to assess whether ordinary citizens had meaningful opinions on policies that typically receive lower levels of public attention. The variety of different policy types included in the survey is also helpful for testing which type(s) of items may

⁴But see Tahk (2005) for a different approach to assessing dimensionality.

⁵In order to focus only on policy items, two questions on Supreme Court nominees were dropped. One policy item restricting ammunition sales was also dropped because it was a Republican substitute to a stronger Democratic measure and therefore was likely to be perceived very differently by respondents and legislators.

TABLE 1 Senate Votes Used in the Senate Representation Survey

Bill Number	Title	Senators	Respondents
		Yea-Nay Votes	Y-N-DK %
– HR 4250	Jumpstart Our Business Strength Act	78-15	44-32-23
S. Amdt. 1085 to HR 2419	Remove Funding for “Bunker Buster” Nuclear Warhead	43-53	52-41-8
– S 1307	Central American Free Trade Agreement	61-34	45-39-15
– S 256	Bankruptcy Abuse Prevention and Consumer Protection Act	74-25	54-30-16
S. Amdt. 367 to HR 1268	Remove Funding for Guantanamo Bay Detention Center	27-71	46-45-9
+ HR 1308	Working Families Tax Relief Act	92-3	79-10-12
S. Amdt. 2937 to HR 4	Child Care Funding for Welfare Recipients	78-20	50-38-13
S. Amdt. 1026 to HR 2161	Prohibiting Roads in Tongass National Forest	39-59	56-31-13
S. Amdt. 1626 to S 397	Child Safety Locks Amendment	70-30	75-21-4
– S. Amdt. 3584 to HR 4567	Stopping Privatization of Federal Jobs	49-47	50-35-16
– S. Amdt. 3158 to S 2400	Military Base Closure Delays	47-49	48-36-16
+ S. Amdt. 44 to S. 256	Minimum Wage Increase	46-49	67-29-4
– S 397	Protection of Lawful Commerce in Arms Act	65-31	74-19-6
S. Amdt. 2799 to S. Con. Res. 95	Cigarette Tax Increase	32-64	59-37-4
S. J. Res. 20	Disapproval of Mercury Emissions Rule	47-51	71-12-17
S. Amdt. 278 to S. 600	Family Planning Aid Policy (Mexico City Policy)	52-46	50-44-6
+ S. Amdt. 2807 to S. 600	Raise Tax Rate on Income over One Million Dollars	40-57	62-32-6
+ S. Amdt. 3379 to S. 2400	Raise Tax Rate on Highest Income Bracket	44-53	49-44-6
+ HR 1997	Unborn Victims of Violence Act	90-9	68-24-9
+ S. Amdt. 3183 to S. 2400	Federal Hate Crimes Amendment	65-33	49-42-9
S. Amdt. 902 to HR 6	Fuel Economy Standards	28-67	70-22-8
S. Amdt. 826 to HR 6	Greenhouse Gas Reduction and Credit Trading System	38-60	48-36-16
+ S. Amdt. 1977 to HR 2863	Banning Torture by U.S. Military Interrogators	90-9	57-38-5
S. Amdt. 1615 to S. 397	Broaden Definition of Armor Piercing Ammunition	31-64	70-22-8
+ S. Amdt. 168 to S. Con. Res. 18	Prohibit Drilling in Arctic National Wildlife Refuge	49-51	48-48-4
– S. Amdt. 3107 to S. 1637	Overtime Pay Regulations	52-47	44-44-12
– S. 5	Class Action Fairness Act	72-26	53-22-24

Note: Table shows Senate vote totals and percentages of 2004 survey respondents supporting, opposing, and saying “don’t know” to each surveyed policy. Leftmost column shows coding of “easy” and “hard” issues, represented by + and –, respectively. Full question wordings for Senate Representation Survey are listed in supporting information section 1.

be the most appropriate and which may be the most problematic in bridging applications.

Assessing the Performance of Joint Scaling

One way to assess the performance of joint scaling is to estimate the ideal point model in Equation (1) separately for respondents and for senators, and then compare the

estimated ideal points (x) from these separate scalings to those from a full joint scaling of these two groups together. This exercise produces extremely high correlations between separate and joint ideal point estimates from the Senate Representation Survey data: .98 for the estimated ideal points of senators and well over .99 for respondents.⁶ At first glance, these high correlations

⁶All estimates based on the standard CJR model are produced using the ideal function in the `pscl` library in R (Jackman 2009). Each set of estimates here was based on 250,000 iterations of the sampler, discarding the first 50,000 as burn-in and recording every

might be thought to indicate that these bridging estimates are well behaved. But these correlations do not tell us how close to being equal these sets of estimates actually are for two reasons (Achen 1977). First, correlation is only a measure of linear association, not equality. Second, and more fundamentally for our purposes, the estimates from the joint and separate estimations are not directly comparable.⁷ Therefore, we need other techniques in order to assess the viability of this joint scaling exercise.

Another way to think about the estimated dimension in joint scaling applications is to view it as a compromise, loosely speaking, between the dimensions structured by each of the groups being analyzed here. The degree of compromise in a given joint scaling application—how close the jointly estimated dimension is to the separately estimated dimensions for each group—is dictated by the model's fit to the data under different parameter values. When the dimensions underlying the views of the two different groups differ, the fit of the pooled model can be dramatically affected by factors that are not central to the phenomena under study, but are instead external or arbitrary.

A useful thought experiment is to consider what would happen to the estimated ideological dimension if the ratio of the number of respondents to senators in our data set were different. Because this ratio is dictated mostly by factors apart from the underlying political dynamics we seek to study, such as the size of one's research account, we should hope that it is not strongly impacted by it.

Although we cannot create new respondents for the already fielded survey, we can drop respondents to create a smaller data set consisting of a different balance of respondents to senators. The lower three panes of Figure 1 show the results of this exercise, comparing the estimates from the full joint model using all 5,871 respondents to those using only 1,000, 500, and 111 respondents, respectively, the last of these being equal to the number of senators used in the scaling. For each simulation, a given number of respondents is randomly sampled without replacement from the full survey sample. The ideal point model is then estimated using these respondents along with all 111 senators, pooling them together and assuming a single common ideological dimension. For each of these sample sizes, 100 such simulations are run, each using a new random sample of respondents of the specified

50th iteration thereafter. Evidence of convergence was strong after several thousand iterations.

⁷The scales would be comparable if, for example, we fixed some item parameters to the same specific values across these three scalings. But doing this would assume that the ideology scales were the same for the individual groups, at least on those items, which is undesirable given that this is what is being tested.

size to ensure that the results are not driven by the particular subset of respondents chosen for a given sample size.⁸

Looking at Figure 1, it is clear that the overall character of the estimated ideal points changes systematically with the number of respondents. As the number of respondents gets smaller, the estimated ideal points of respondents appear more moderate relative to those of senators. The logic behind this is that the estimated dimension in these models is a compromise between the senator and respondent dimensions. When respondents constitute the overwhelming majority of the data, as in the top pane of Figure 1, the item parameters are estimated mostly based on respondent choices. When the numbers of respondents and senators in the data become more equal, as in the lower panes of Figure 1, the item parameters are based on a more equal, compromise between the structure of the two groups' ideological dimensions.

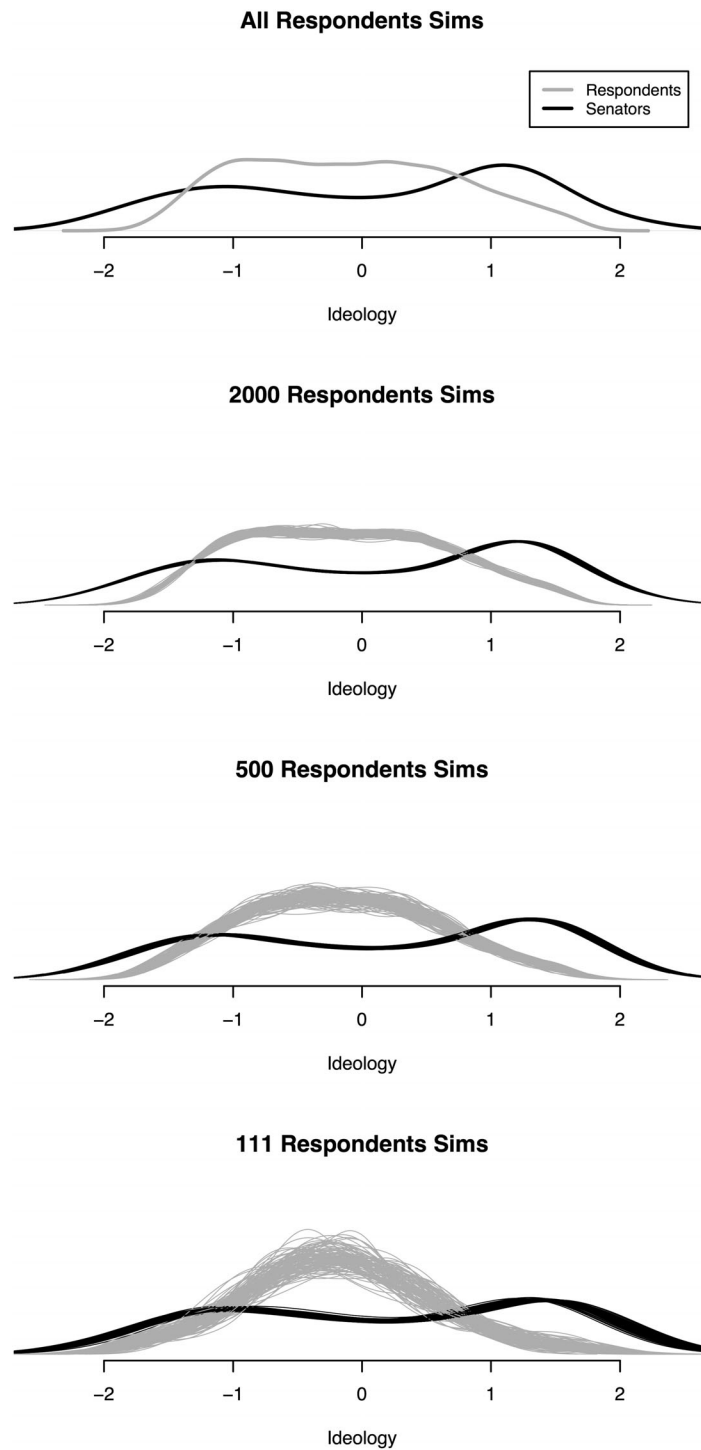
The systematic variation in the overall character of the estimates shown in Figure 1 is obviously not desirable. A researcher who runs a survey with a large number of respondents would learn something different about, for example, the relative polarization of respondents and senators than someone who ran a smaller survey. This is not just due to the extra uncertainty that comes from a smaller data set, but relates to the character of the estimates under these two setups.

A Method for Group-Based Ideology Estimation

This section describes a technique for estimating the ideal points of a set of actors from multiple groups while restricting the estimated ideological dimension to be structured only by the positions of a specific subgroup of actors. Researchers may want such a technique to impose such structure on the ideological dimension in a given ideal point estimation, rather than simply pooling all of the data together and using whichever dimension is estimated by the model. This motivation is all the more important given that ideal point estimation bridging two groups can suffer from the pathologies illustrated above under the standard approach. It may also be useful to compare the estimated ideal points and item parameters structured by each group separately as a diagnostic exercise, seeking

⁸Because of the large number of estimations, the sampler for each is run for 25,000 iterations, with the first 5,000 discarded as burn-in and every iteration thereafter recorded. The sampler appears to converge rapidly, and the recorded samples appear to provide a reasonable amount of information, particularly since only posterior means are examined.

FIGURE 1 Characteristics of Respondent and Senator Ideology Estimates Differ Sharply Based on Respondent Sample Size Used



Note: Panes show densities of estimated ideal points for senators and respondents from joint ideal point models estimated for random samples of given size from respondents along with all 111 senators in the data set. Top pane shows estimates using all 5,871 respondents, whereas lower panes show densities for 100 different estimates, each using a different random sample of respondents of the specified size.

to understand whether the dimensions underlying the preferences of each group differ meaningfully.

The CJR ideal point model discussed above is estimated using a Gibbs sampler that produces draws from the posterior distribution over the model's unknown parameters. This is accomplished by cycling through samples from the conditional posterior distributions for each set of parameters while fixing all other parameters at their most recently sampled values. In order to conduct a restricted ideal point estimation where the ideological dimension is structured based only on the preferences of one group, this process can be modified so that the item parameters β and γ are sampled at each iteration from the conditional posterior given the ideal points x and latent utility differences y^* of one particular group. In other words, the sampling procedure is identical to the one used in CJR except that inferences about the item parameters, which structure the underlying ideological dimension, are affected only by the policy positions of the chosen subgroup of actors. This procedure is equivalent to running the standard model on only the data from the group structuring the ideological dimension and then mapping the ideological positions of the out-group members into this ideological space by sampling from the conditional posterior of their ideal points given the item parameter values at each iteration of the sampler.⁹ Section 3 in the supporting information describes this process in more detail. This group-based scaling procedure will be available in future versions of the *pscl* R package.

Using this technique, it is possible to compare the ideological spaces, including ideal point and item parameter estimates, that structure the preferences of different groups while still estimating the ideology of all actors in the data and allowing for direct comparisons between the two sets of group-based estimates. Figure 2 shows the results of this exercise for the Senate Representation Survey, comparing estimates for senator and respondent ideal points that result from restricting the sampler to let senator or respondent preferences, respectively, structure the underlying ideological dimension.¹⁰ In contrast to the separate estimation strategy discussed above, this group-based scaling produces estimates that can be meaningfully compared on the same scale. This

is achieved by imposing the same identifying restriction on the ideal points across the two scalings: At each iteration, the estimates, including ideal points (x_i) and item parameters (β_j and γ_j), are rescaled such that the average of the mean respondent ideal point and the mean legislator ideal point is 0 and the average of the variance of respondent ideal points and the variance of senator ideal points is 1, and the space is oriented such that higher ideal point values represent more conservative ideological positions.¹¹ This means that we can assess how close these two sets of estimates are to being equal, not just how strong the relationship is between them.

Because correlation does not speak directly to how close to equal two variables are, this would suggest using a statistic such as the mean squared difference (MSD) between the two sets of estimates. This measure, however, does not have an easily interpretable scale. Here, I standardize the measure by dividing by the standard deviations of each variable and rescale the measure to be bounded between 0 and 1, calling the resulting statistic the standardized mean squared difference (sMSD), defined as

$$\text{sMSD} = \frac{1}{1 + \frac{\frac{1}{n} \sum_{i=1}^n (x_{i,(1)} - x_{i,(2)})^2}{\sigma_{(1)}\sigma_{(2)}}}, \quad (3)$$

where $x_{i,(1)}$ and $x_{i,(2)}$ are the estimated ideal points for actor i from scalings based on Groups 1 and 2, respectively, and $\sigma_{(1)}$ and $\sigma_{(2)}$ are the standard deviations of the two sets of estimates. This measure has several useful properties. First, it approaches 0 as the two sets of estimates become farther apart and equals 1 when $x_{i,(1)}$ and $x_{i,(2)}$ are identical for all actors i . The measure is also invariant to linear transformations applied to the two scales together.

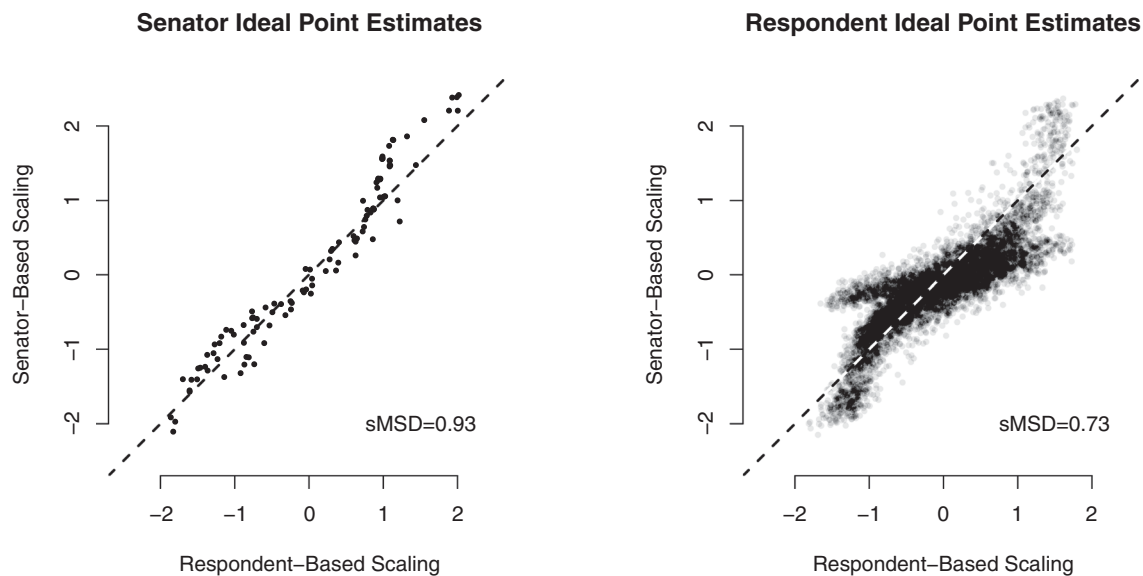
Looking at Figure 2, it is obvious that while the estimated ideal points for senators are nearly identical whether the dimension is structured based on the preferences of senators or respondents, the estimates for respondents are less similar. The sMSD for senators is .93, but for respondents it is .73. There are many respondents whose ideal points are quite different depending on which group structures the estimates, being much more moderate in the senator-based scaling, but more ideologically extreme under the respondent-based scaling. These results are similar when using all Senate

⁹One could also imagine an analogous procedure for maximum-likelihood-based estimators such as NOMINATE, where the conditional maximization of the item parameters is based only on the ideal points and positions of the chosen group.

¹⁰The group-based ideal point model is estimated using a modified version of the ideal function in the *pscl* library in R. All estimates are based on runs of 250,000 iterations with the first 50,000 iterations discarded as burn-in, recording every 50th iteration thereafter. Estimations appeared to converge rapidly.

¹¹This identifying restriction allows respondents as a whole and senators as a whole to have the same influence, loosely speaking, on identifying the ideal point space. Although different identifying restrictions do change many of the values calculated below (including, notably, sMSDs) the overall pattern of findings remains the same.

FIGURE 2 Respondent-Based and Senator-Based Ideal Point Estimates from Senate Representation Survey Show Small Differences for Senators, Large Differences for Respondents



Note: Plot compares ideal point estimates (posterior means) from respondent- and senator-based scalings separately for senators (right pane) and respondents (left pane). Respondent estimates are plotted with transparency to better show overlapping points.

votes from the 108th and 109th sessions instead of only the bridge items included in the survey.¹²

The densities of the estimated ideal points under the two group-based scalings, shown in Figure 3, also show significant differences in their overall characteristics. In particular, the senator-based estimates show a much more moderate distribution of respondent ideologies relative to those of senators, whereas the respondent-based estimates show only a slightly higher variance for the senator ideal point estimates as compared to respondents.

The densities of respondent-based ideal points, plotted in Figure 3, look similar to the pooled joint scaling in the top pane of Figure 1, whereas the senator-based densities look more similar to those based on all senators and only a subsample of 111 respondents seen in the bottom pane of Figure 1. This makes sense given that the higher the proportion of respondents in the data, the more the estimated dimension will be similar to that for respondents. The key advantages of the group-based procedure, however, are that the dimension being estimated can be chosen directly, rather than loosely affected by dropping some number of actors from one

group, and also that ideal points are estimated for all actors, whether or not they are members of the group chosen to structure the estimated dimension.

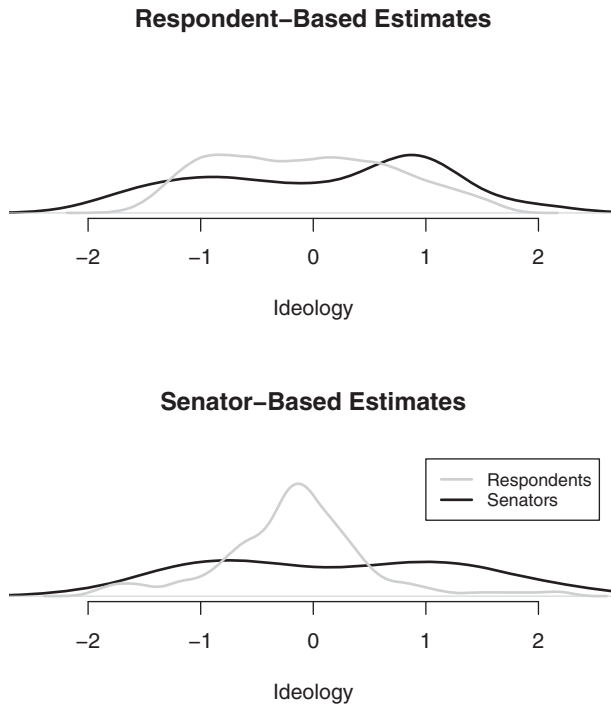
One way to understand why these two sets of estimates differ is to examine the estimated item parameters under the two setups. If the ideological dimensions for senators and respondents are structured similarly, we should observe similar estimates of the item parameters for each policy, whether from senator- or respondent-based scalings. Figure 4 shows the estimated discrimination parameters and cutpoints (β_j s and γ_j s from Equation 2) for these two scalings. For the discrimination parameters, the posterior means are plotted, whereas for the cutpoints, the posterior medians are used.¹³

Figure 4 shows that the signs of the discrimination parameter estimates are the same for 25 out of the 27 items, but the estimates exhibit little if any association beyond this, suggesting that while policies seen as liberal (conservative) by senators also tend to be seen as liberal (conservative) by respondents, the degree of ideological distance perceived between supporting and opposing the policies does not seem to be similar for the two groups. To put it differently, the degree of ideological divisiveness

¹²Note that the full-roll call matrix is only used for the Senate-based scaling, as it is inappropriate to estimate a respondent-based scaling for all Senate votes when respondents do not take positions on the vast majority of roll calls.

¹³This is because the cutpoint parameters $\gamma_j = \frac{\alpha_j}{\beta_j}$ approach $\pm\infty$ as $\beta_j \rightarrow 0$, making posterior means unreliable. See Section 4 in the supporting information for more information.

FIGURE 3 Densities of Ideal Point Estimates from Senate Representation Survey Show Large Differences under Respondent-Based and Senator-Based Scalings



Note: Densities of respondent and senator ideal points are plotted from respondent-based and senator-based scalings of all respondents and senators from the Senate Representation Survey.

for each of the items is not strongly associated between the two scalings.

The two policies for which the discrimination parameters are estimated to have opposite signs for respondents and senators are themselves quite different. S. Amdt. 3158, which proposed that a planned round of military base closures should be restricted solely to bases outside of the United States, had an estimated discrimination parameter of $-.20$ in the senator-based scaling and $.03$ in the respondent-based scaling, with the 95% highest posterior density regions (HPDs) for both estimates overlapping zero.^{14,15} This suggests that although the signs of the estimates differ between the two groups, the policy does not seem to be very ideologically divisive

¹⁴HPDs are defined as the smallest region of the parameter space that contains the specified posterior probability (in this case, 95%) for a given parameter. HPDs can loosely be thought of as a Bayesian analogue of confidence intervals.

¹⁵The posterior probability that the senator-based and respondent-based discrimination parameters for S. Amdt. 3158 have the same sign is $.19$.

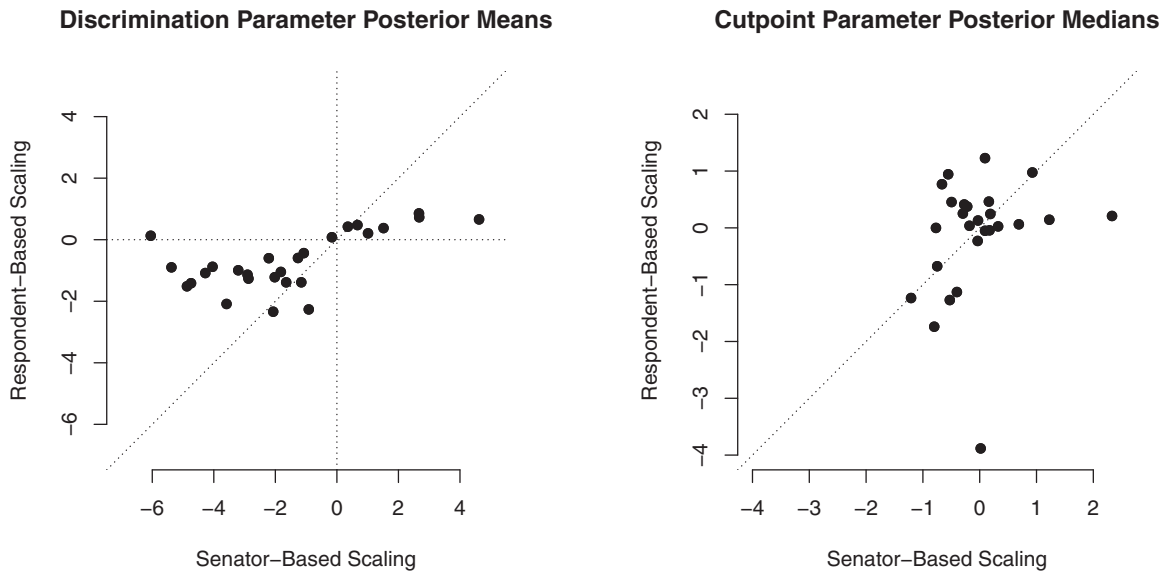
for either group. Therefore, this might not be thought to be a severe violation of the assumption that the item parameters are the same for the two groups. By contrast, the senator-based and respondent-based discrimination parameters for S. Amdt. 3107, which would have altered overtime pay regulations, show much larger differences. The estimated value for senators of -6.09 indicated that the measure was highly ideological, with support for the amendment being the more liberal position. For respondents, the discrimination parameter is estimated to be $.08$, which implies that support for the amendment was a conservative position, albeit a very mildly divisive one.¹⁶

The estimated cutpoints (γ_{js}) also show considerable variation between senators and respondents. The biggest outlier among the cutpoint estimates is clearly HR 1308, whose posterior medians are $-.01$ and -3.91 for senators and respondents, respectively. There is, however, a large amount of uncertainty in these estimates, particularly for the respondent-based scaling. This is due to the fact that the discrimination parameter for respondents is estimated to be quite close to zero. Therefore, it is difficult to tell whether the large discrepancy between the two estimates is due to sampling error or whether it reflects a true difference between the relationship between ideology and positions on this policy between senators and respondents. The second largest outlier (albeit a much milder outlier) among the cutpoint parameter estimates is S. Amdt. 1977, which proposed to prohibit torture of detainees in U.S. military custody, limiting interrogation techniques to those authorized in the U.S. Army Field Manual on Intelligence Interrogation. Even after accounting for uncertainty in the cutpoint estimates, it seems clear that the cutpoint for this policy is much closer to zero for respondents than for senators, indicating that moderate respondents are more likely to be indifferent or close to indifferent on this measure, whereas moderate or even slightly conservative senators were far more likely to support than oppose the measure.

One approach to dealing with the differential item functioning indicated by the outlying points in both panes of Figure 4 is to drop the worst of such offenders. In the present analysis, this might suggest omitting S. Amdt. 3107 and HR 1308, which had the largest differences in estimated discrimination and difficulty parameters, respectively, between the two group-based scalings. When reestimating the two group-based scalings dropping the item with the most outlying discrimination parameter (S. Amdt. 3107), the sMSD between the

¹⁶The 95% HPDs for the senator-based and respondent-based discrimination parameter for S. Amdt. 3107 both did not overlap zero.

FIGURE 4 Item Parameters from Senate Representation Survey Show Large Differences between Respondent-Based and Senator-Based Scalings



Note: Figure plots estimates of discrimination (left pane) and difficulty (right pane) parameters from respondent- and senator-based scalings of the Senate Representation Survey.

two sets of ideal point estimates remains roughly the same (.94) for senators and rises to .80 for respondents. Dropping HR 1308, which has by far the most discrepant cutpoint estimates, results in virtually no change in the correspondence between the estimates (sMSD of .94 and .73 for senator and respondent ideal points, respectively). Finally, dropping both of these items simultaneously produced roughly the same degree of correspondence between senator and respondent ideal point estimates from the two scalings (sMSDs of .93 and .80, respectively) as dropping S. Amdt. 3107 alone. Although this selective item deletion approach shows some promise to increase the correspondence between respondent- and senator-based estimates, it is not clear where to stop given that several remaining policies have a similar level of discrepancy, suggesting that we should either stop after dropping one or two, or proceed to drop many more, neither of which is a particularly satisfying option.

Overall, the parameters from these two sets of full scalings show similarities but are far from identical. The relationship between the senator-based and respondent-based discrimination parameters appears roughly linear on average, but with considerable individual variation. The magnitude of respondent-based β s is clearly smaller than that from the senator-based scaling. This suggests that while respondents and senators tend to agree on which policy proposals are liberal or conservative, senators discriminate much more sharply based on

ideology in their position taking. That the magnitude of the discrimination parameters for respondents might be a fraction of those for senators is equivalent to the error variance in the utility-based voting model being higher for respondents than senators (in the standard CJR model and the group-based model here, the error variance is fixed to 1 for all actors on all items). This could be seen as unsurprising given that legislators are essentially professional position takers who might be expected to do so with relatively low amounts of error. The cutpoint estimates, while not wildly divergent in most cases, also did not show strong correspondence between the two groups. Perhaps most importantly, the estimated ideal points, which are typically the focus of interest in political science scaling applications, were somewhat similar but far from identical under the two group-based estimations.

One way to think about how to interpret the sMSD values in this application is to ask how similar group-based scalings would be according to this metric if the ideological dimension were actually identical for the two groups. With the aim of answering this question, I conducted a set of simulations in which data are sampled from the predictive distribution, setting ideal points and item parameters equal to their posterior means from the full joint scaling (detailed results are presented in Section 5 in the supporting information). The distribution of sMSD values from these simulations ranged roughly from .97 to .99 for senators and from .95 to .99 for

respondents. Both of the observed values (.93 and .73 for senators and respondents, respectively) fall well outside of these ranges. Although the aim of these group-based scaling evaluations is not to provide sharp hypothesis tests of identical dimensions between the two groups, the discrepancy observed here suggests that there are significant differences between the dimensions structuring policy views for senators and survey respondents.

It bears keeping in mind that the Senate Representation Survey poses a “hard test” for the assumption that ordinary citizens and political elites have their ideologies structured in the same way. Many of the policy items included in this survey could be considered obscure or complex. Even with this hard test, however, the item parameters are found to have large positive correlations, albeit with a slope clearly less than 1, and the estimated ideal points, which are typically the main parameters of interest in political science applications, are quite similar, particularly after dropping the most problematic of the items. These mixed results beg the question of how common-scale ideal point estimation between citizens and political elites might fare when applied to different types of data sets, such as those including the types of issues ordinary citizens are more likely to have thought about and formed meaningful opinions on.

Data: 2008 Cooperative Congressional Election Study

The 2008 Cooperative Congressional Election Study (CCES) was fielded to an online sample of 32,800 respondents from the Polimetrix/YouGov online panel during October and November 2008 (see Ansolabehere 2011). Various versions of the CCES have been fielded since 2006, but the 2008 version was chosen because it contains the largest number of policy questions pertaining to specific House and Senate roll-call votes. In total, the CCES included eight items that directly corresponded to votes taken during the 110th House and Senate. Table 2 lists the policies as well as the House and Senate vote margins and the percentage of respondents supporting, opposing, and saying “don’t know” to each.¹⁷ Although the number of bridging items in the CCES is smaller than in the Senate Representation Survey, it has two attractive features. First, the CCES sample contains more than five times as many respondents as the Senate Representation Survey. Second, the CCES contains items for which we know the

¹⁷The CCES also included an item about a constitutional amendment to define marriage as between one man and one woman, but this was not included in the analysis because it was not directly voted on by the House and Senate during the 110th Congress.

positions of respondents, senators, and House members. We can thus compare not only how the structure of respondent and legislator ideology may differ, but also how the structure of House and Senate ideology may differ.

The CCES might be expected to be an “easier test” for the bridging assumptions implied by common-space scaling since its questions tend to pertain to more straightforward policies than many of those from the Senate Representation Survey. Another way to think of this is that the policy items asked in the Senate Representation Survey were closer to the type of proposals routinely voted on by legislators, whereas the CCES mostly asked about items that ordinary citizens might encounter and think about more routinely. It is instructive, then, to ask how the ideological dimensions underlying the preferences of legislators and ordinary citizens differ in the CCES data and how the overall character of these results compares to those for the Senate Representation Survey.

In order to answer these questions, three versions of the group-based ideal point model are estimated on the CCES data, letting the positions taken by House members, senators, and respondents, respectively, structure the estimated dimension. The scales from these three separate estimations are identified by post-processing each iteration of the sampler so that the mean legislator (House members and senators together) ideal point and the mean respondent ideal point sum to 0, the average of the legislator ideal point and respondent ideal point variances is 1, and higher ideal point values indicate more conservative positions.

Figure 5 plots the relationship between the estimated ideal points from these three scalings for House members, senators, and respondents separately. The most obvious feature across all of these plots is the very high degree of similarity between the estimates across all scalings and all types of actors. The sMSDs for these nine comparisons range from .94 to .99, which is quite high, particularly since these ideal points are estimated based on only eight items and therefore contain a considerable amount of measurement error. Comfortingly, these estimates all correlate fairly highly with other measures, such as, ideological self-placement and Bayesian Aldrich-McKelvey scores (Hare et al. 2015; see supporting information section 6).

The estimated item parameters from these three scalings are plotted in Figure 6. In all three cases, there is a positive relationship between the discrimination parameter estimates. The respondent-based discrimination parameters are very similar to those based on House member and senator preferences, with a strong linear relationship between the estimates and a relatively small amount of error. It is clear, however, that the slope of this linear relationship is less than 1, meaning that the

TABLE 2 Policy Items from 2008 CCES

Policy	Representatives Yea-Nay Votes	Senators Yea-Nay Votes	Respondents Y-N-DK %
Withdrawing troops from Iraq within 180 days	171-256	28-71	47-41-13
Increasing minimum wage to \$7.25	315-117	95-3	72-21-7
Allow federal funding of stem cell research	247-177	63-35	53-30-16
Allow U.S. spy agencies to eavesdrop on overseas terrorist suspects without first getting a court order	294-129	70-28	59-27-14
Fund a \$20 billion program to provide health insurance for children in families earning less than \$43,000	265-160	68-32	58-26-16
Federal assistance for homeowners facing foreclosure and large lending institutions at risk of failing	241-173	84-13	39-39-22
Extend the North American Free Trade Agreement (NAFTA) to include Peru and Columbia	286-132	78-18	31-34-35
U.S. Government's \$700 Billion Bank Bailout Plan	264-171	75-25	20-54-26

Note: Table shows House and Senate vote totals and percentages of 2008 CCES survey respondents supporting, opposing, and saying “don’t know” to each surveyed policy. Full question wordings for 2008 CCES are listed in supporting information section 2.

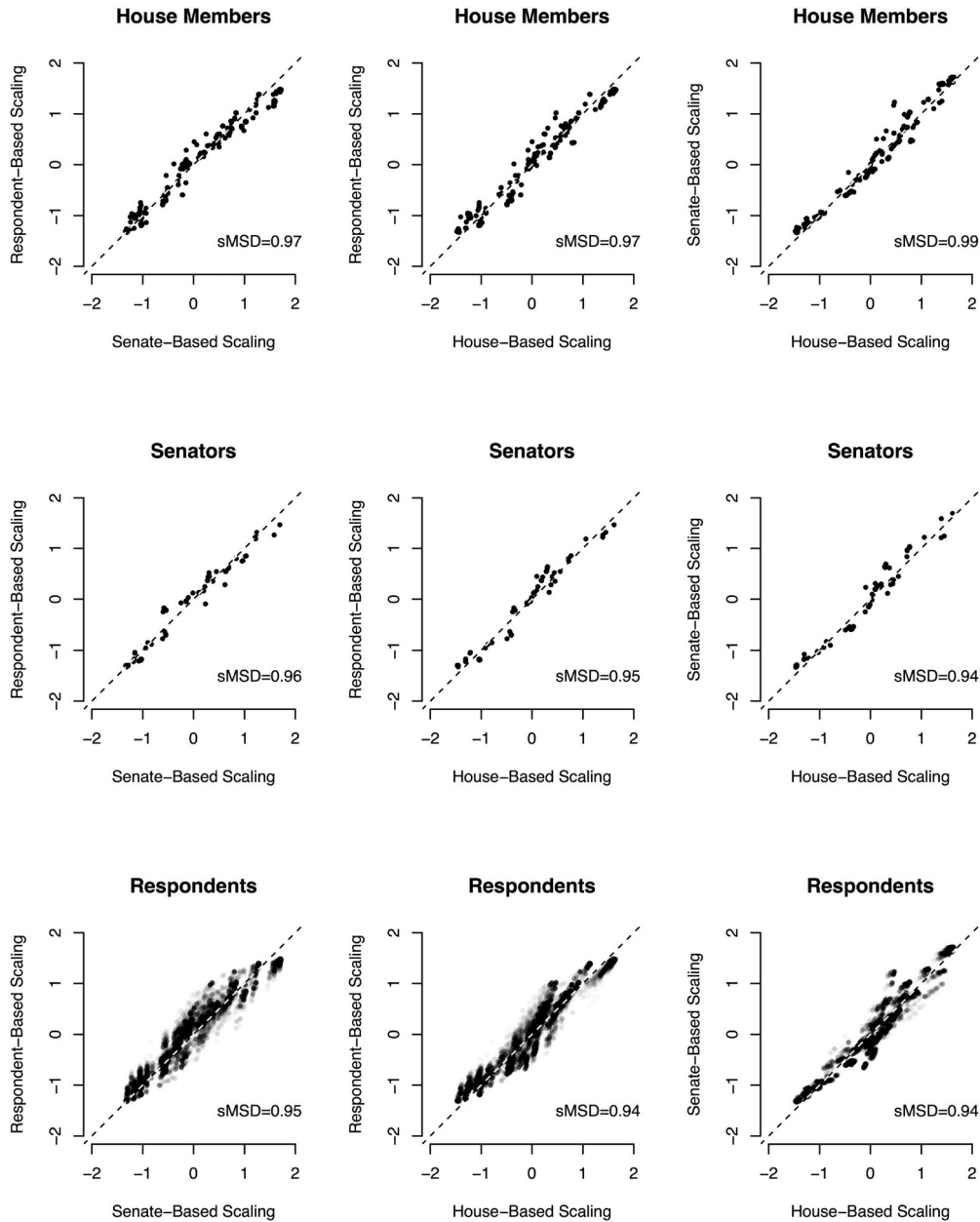
discrimination parameters used by respondents are smaller in magnitude (closer to 0) than those used by either House members or senators. This pattern is similar to the one in Figure 4 from the Senate Representation Survey, but the relationship is much stronger in the CCES data. As above, this can be interpreted to mean that respondents are “noisier” position takers, but that the basic pattern of how relatively liberal or conservative each issue is tends to be quite similar for all three of the groups considered here.

Although there is a fairly strong correspondence between the cutpoint estimates for senators and House members, the relationships for respondents and both sets of legislators are much weaker. As in the Senate Representation Survey, however, the estimated cutpoints tend to be clustered near the middle of the ideological scale. This means that even though there is not a high correlation between the cutpoints estimated for legislators and respondents, the actual distance between the estimates tends to be small, with one or two notable exceptions. The biggest outlier between respondent and both House or Senate estimates is HR 1424, the Emergency Economic Stabilization Act of 2008 (commonly known as the federal “bailout”). Although only 27% of respondents who took a position supported this policy, majorities voted for it in both the House and Senate. A more mildly outlying cutpoint between legislators and respondents is HR 3688, extending the NAFTA to include Columbia and Peru,

which was supported by large majorities in the House and Senate, but a minority of respondents. Overall, the cutpoints for the House- and Senate-based estimates are more similar to each other than the respondent-based estimates are to either of the legislator-based cutpoints.

As was done in the previous section for the Senate Representation Survey, we can compare the observed level of similarity between House-based, Senate-based, and respondent-based ideal point estimates for the CCES to what would be expected if the assumptions of the full joint ideal point model were true. Although the observed sMSD values for the Senate Representation Survey fall far below what would be expected under the joint model, the values observed for the CCES are actually only slightly lower on average than those from scaling data simulated assuming the joint model to be true. In fact, many of the simulated sMSD values are actually lower than the observed values (see Section 5 in the supporting information). Although ideal point models, particularly those used for joint scaling, are obviously not a literally true representation of how respondents and legislators generate policy positions, these results suggest that for the CCES, the group-based estimates are nearly as similar as we would expect them to be if the model’s assumptions were exactly true. By this standard, joint scaling appears to work well for this data set.

FIGURE 5 2008 CCES Ideal Point Estimates are Similar under House-, Senate-, or Respondent-Based Scalings



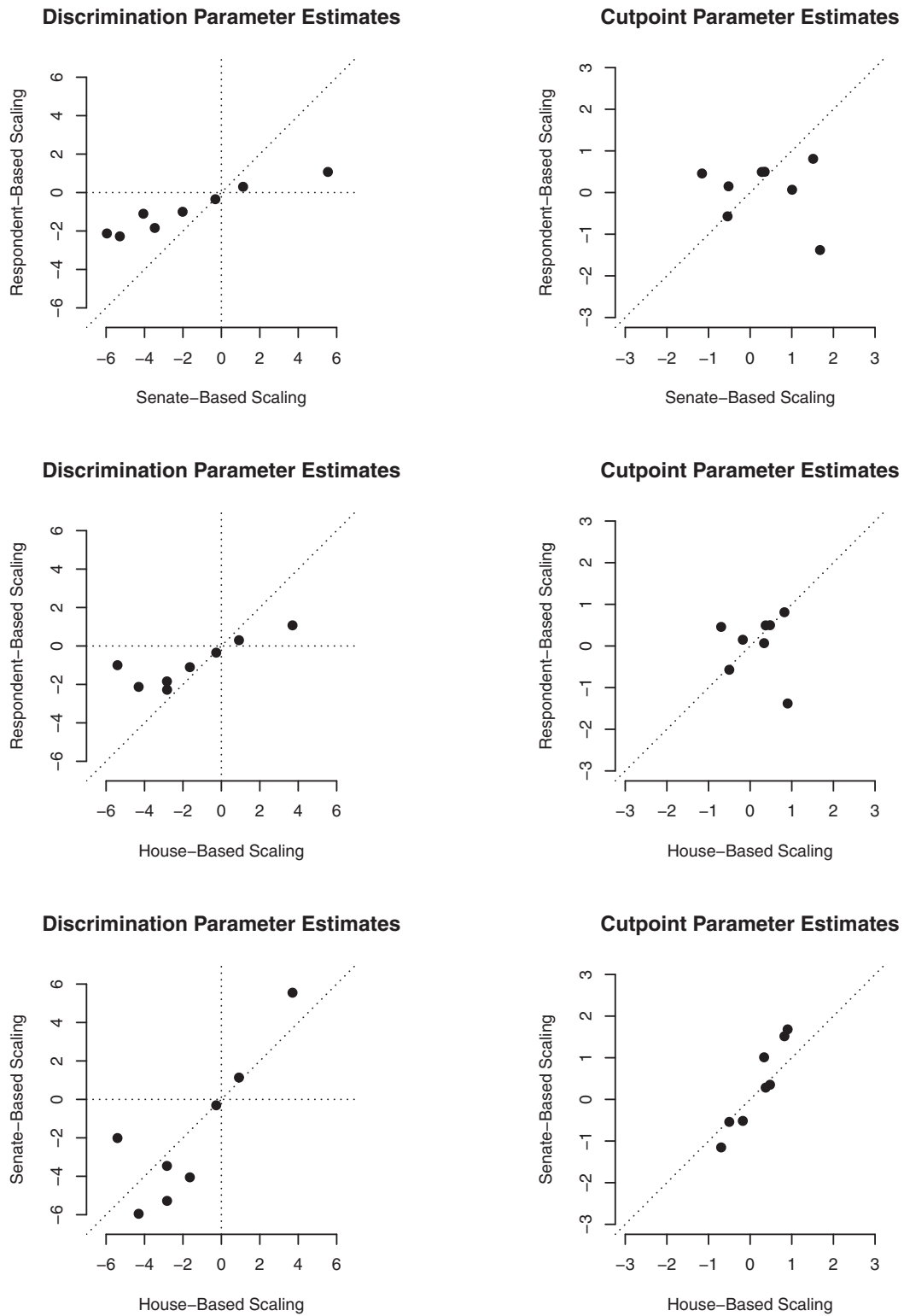
Note: Plots show estimates (posterior means) for ideal points of House members, senators, and respondents under House-, Senate-, and respondent-based scalings. Respondent estimates are plotted with transparency to show overlapping points.

“Hard” versus “Easy” Issues and the Performance of Joint Scaling

One way to assess whether the differing results for the Senate Representation Survey and the CCES were driven primarily by the different types of policy items used in the two surveys is to separately look at the “hard”

and “easy” policy items on the Senate Representation Survey. Because this survey included a much larger number of policy items than the CCES, we can create two different “subsveys,” each having eight policy questions (the same number as the CCES). To do this, I selected the eight items that are the least complex, most straightforwardly ideological and that respondents are most likely to have thought about, calling them “easy”

FIGURE 6 2008 CCES Item Parameter Estimates Show Some Differences under House- Senate- or Respondent-Based Scalings



Note: Plots compare estimates for item parameters under House-, Senate-, and respondent-based scalings of 2008 CCES data. Discrimination parameter estimates are posterior means, whereas cutpoint parameter estimates are posterior medians due to extreme draws from the posterior (see Section 4 in the supporting information for more information).

issues. I also selected the eight issues that are the most complex and that respondents are the least likely to have thought about and formed meaningful opinions on, calling these “hard” questions. The leftmost column of Table 1 shows these classifications. While somewhat subjective, these categorizations were made without reference to the estimated item parameters (e.g., the “easy” questions were not simply chosen based on which items’ discrimination parameters were largest in magnitude or most similar between the two groups). Respondents who answered at least five of a given issue type were included in this analysis, yielding 1,091 respondents for the “easy” issue scaling and 749 for the “hard” issue scaling.

The correspondence between respondent-based and senator-based ideal point estimates is indeed much higher when using the “easy” issue items than when using the “hard” ones. The sMSD for the “easy” issue scaling was .96 for senators and .86 for respondents. By contrast, these values were .75 and .51 for the “hard” issue scaling. Although these figures are likely to be affected somewhat by the smaller number of respondents, it seems clear that certain types of issues yield more comparable scales for the two types of actors. This finding should not be surprising given that legislators are essentially professional position takers, whereas most citizens are unlikely to have thought about many of the policies that are decided on by Congress. Furthermore, even the “easy” issue questions from the Senate Representation Survey might be considered more complex in both their topic and their presentation to respondents than most of the issues in the CCES. Therefore, while it is not clear that this is the only factor accounting for the strong differences between the two sets of results, it appears as though more straightforward questions on issues that citizens are more likely to have thought about produce more similar ideological dimensions between citizens and legislators.

Discussion

Ideal point models are a valuable tool for political scientists. But these models are not magic. There is no guarantee that an ideal point model will find the dimension of interest to researchers when fed a set of data. The burden is on researchers to provide clear thought about both the indicators chosen and what underlying dimension is relevant for a given application. The selection of a measurement approach is as much a substantive or theoretical issue as a statistical one.

A key question here is how literally ideal point models should be taken. On one hand, a key advantage

of these models is that they are often built up from clear microfoundations based on spatial utility. On the other hand, it is obvious that these models are abstractions, rather than literally true representations of how specific policy positions are generated. The most useful way to evaluate them is not to verify that the models are strictly true, but rather to assess whether they represent useful, rather than misleading, simplifications.

Along these lines, the findings above might be taken to imply that the variances of the utility disturbances for citizens and legislators are potentially different since smaller utility error standard deviations are equivalent to discrimination parameters farther from zero.¹⁸ In fact, similar heteroskedastic models have been estimated by Jessee (2009) and Lauderdale (2010). But this setup, while coherent within the ideal point framework, fundamentally changes the meaning of the estimated dimension. For example, if respondents have larger error variances than legislators, it would be possible for a legislator to have a more conservative ideal point than a respondent, but for the respondent to be more likely to support a given conservative proposal. More seriously, there is evidence that the relaxation of the homoskedasticity assumption actually increases the sensitivity of estimates to factors such as survey sample size (see Section 7 in the supporting information). This suggests that while thinking seriously about the formal assumptions of ideal point models remains important, the main practical issues in joint scaling are unlikely to be solved solely by more complex statistical models.

So how should researchers respond to the findings presented above? One response is to admit defeat. If the primary ideological dimensions underlying citizens’ policy views and legislators’ roll-call voting are not the same, one could argue that it is meaningless to talk about comparing the positions of citizens and legislators. But such criticisms could apply to virtually any estimates of ideology, including those within a single chamber of Congress containing different types of legislators (e.g., Democrats and Republicans, northerners and southerners).

Another response is to recognize that the results of joint scalings are typically used in subsequent analyses to test hypotheses of interest. In these situations, researchers could attempt to validate their hypotheses using both the citizen-based and legislator-based ideology estimates as described above. If the findings are similar using both

¹⁸In order to identify the ideal point model, it is assumed that the variance of utility errors for all actors is 1. If for one group an error variance were $\sigma_g^2 \neq 1$, this would be equivalent to letting their discrimination parameters for all questions equal β_j/σ_g but leaving the variance of their utility errors at 1 in this transformed model.

measures, this may provide some reassurance that the findings hold under both conceptualizations of ideology.

For example, group-based ideal point estimates can be used to test spatial voting in the 2004 and 2008 presidential elections, following Jessee (2012). Section 8 in the supporting information shows that these findings are very similar regardless of which estimates are used. This demonstrates that even in cases where the group-based ideal point estimates differ significantly, such as the Senate Representation Survey, the results of second-level hypothesis tests based on those group-based ideal points may yield the same substantive conclusions. But these findings are no guarantee that all such second-level results will be unaffected by these measurement choices. For example, Figure 3 above demonstrates that conclusions about the relative polarization of citizens and legislators may be quite different depending on whether researchers consider respondent-based, legislator-based, or full joint ideology estimates.

The most productive response to discrepancies between group-based ideal point estimates, I would argue, is to recognize that while many latent dimensions may exist in joint scaling applications, researchers should seek to estimate the dimension that is relevant for whatever theory or hypothesis they are studying. This dimension is unlikely to be “whatever the model estimates when fed this data set,” which, loosely speaking, is the dimension that explains the most variation in the full data. To determine which dimension is appropriate, we must look to the theory motivating our analyses. For example, in tests of spatial voting theory (e.g., Downs 1957; Enelow and Hinich 1984; Hotelling 1929), which posit that voters are more likely to cast their ballot for candidates whose ideological positions are closest to their own, a citizen-based ideological dimension might be most relevant given that the theory describes voters calculating ideological distances on which to base their decisions. Conversely, if one were interested in where candidates for office would fall on the ideological spectrum of sitting legislators if they were elected to Congress, a legislator-based scaling in which the candidates’ policy positions were matched to specific roll-call votes may be appropriate. The group-based scaling technique presented here provides a way to impose such restrictions on the dimension being estimated in joint scaling applications. More broadly, the choice of when researchers should seek a policy-based ideal point measure rather than an alternative such as the perceptual-based estimates from Aldrich and McKelvey (1977) should be driven by what one seeks to measure—the positions implied by the actual policy views of individuals or the location that they perceive for themselves relative to other political stimuli.

The group-based scaling method presented here also provides a way of examining which types of policy questions are most compatible with the assumptions underlying joint scaling, suggesting that simpler policies that citizens are more likely to have thought about tend to be treated more similarly by respondents and legislators. In fact, for the CCES, which included mostly these types of questions, respondent-based estimates showed roughly the same level of similarity with House- or Senate-based scalings as the House- and Senate-based estimates showed with each other. A question for further exploration is how the inclusion of nonbridge items for nonpolicy items (e.g., Bafumi and Herron 2010) affects estimates relative to scalings in which only bridge items are used (e.g., Jessee 2010a).

This article has examined some of the most important questions about the validity of bridging ideal point models, but the set of issues considered here is by no means complete. Although analyses estimating ideology across multiple groups have shown great promise in recent years, researchers must continue to work to assess their applicability, robustness, and validity. In the end, the central concern for researchers using this approach must be that estimates represent as closely as possible the substantive dimension(s) they seek to study. Although statistical tests can speak to this issue, they are far from a panacea. Careful thought about the data used for joint scaling as well as the assumptions of the model and structure imposed on the estimation remain the most important considerations in joint scaling applications.

References

- Achen, Christopher H. 1977. “Measuring Representation: Perils of the Correlation Coefficient.” *American Journal of Political Science* 21(4): 805–15.
- Adams, James F., Samuel Merrill III, and Bernard Grofman. 2005. *A Unified Theory of Party Competition*. Cambridge: Cambridge University Press.
- Aldrich, John H., and Richard D. McKelvey. 1977. “A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections.” *American Political Science Review* 71(1): 111–30.
- Alvarez, Michael R., and Jonathan Nagler. 1995. “Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election.” *American Journal of Political Science* 39(3): 714–44.
- Ansolabehere, Stephen. 2011. Cooperative Congressional Election Study, 2008: Common Content, July 2011. <http://hdl.handle.net/1902.1/14003>.
- 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(3): 519–42.

- Bailey, Michael A. 2007. "Comparable Preference Estimates across Time and Institutions for the Courts, Congress, and Presidency." *American Journal of Political Science* 51(3): 433–48.
- Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2): 367–86.
- Brady, Henry E., and Paul M. Sniderman. 1985. "Attitude Attribution: A Group Basis for Political Reasoning." *American Political Science Review* 79(4): 1061–78.
- Hinich, Melvin J., Lawrence Cahoon, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting." In *Graphical Representations of Multivariate Data*, ed. Peter C. C. Wang. New York: Academic Press, 243–78.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2): 355–70.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: HarperCollins.
- Enelow, James, and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge: Cambridge University Press.
- Groseclose, Tim, and Jefferey Milyo. 2005. "A Measure of Media Bias." *Quarterly Journal of Economics* 120(4): 1191–1237.
- Hare, Christopher, David A. Armstrong, Ryan Bakker, Royce Carroll, and Keith T. Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59(3): 759–74.
- 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." *RAND Journal of Economics* 28: 142–89.
- Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal* 39(153): 41–57.
- Jackman, Simon. 2009. pscl: Political Science Computational Laboratory (R library).
- Jessee, Stephen, and Neil Malhotra. 2013. "Public (Mis) Perceptions of Supreme Court Ideology: A Method for Directly Comparing Citizens and Justices." *Public Opinion Quarterly* 77(2): 619–34.
- Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(1): 59–81.
- Jessee, Stephen A. 2010a. "Partisan Bias, Political Information and Spatial Voting in the 2008 Presidential Election." *Journal of Politics* 72(2): 327–40.
- Jessee, Stephen A. 2010b. "Voter Ideology and Candidate Positioning in the 2008 Presidential Election." *American Politics Research* 38(2): 195–210.
- Jessee, Stephen A. 2012. *Ideology and Spatial Voting in American Elections*. New York: Cambridge University Press.
- Lauderdale, Benjamin E. 2010. "Unpredictable Voters in Ideal Point Estimation." *Political Analysis* 18(2): 151–71.
- Malhotra, Neil, and Stephen A. Jessee. 2014. "Ideological Proximity and Support for the Supreme Court." *Political Behavior* 36(4): 817–46.
- Palfrey, Thomas R., and Keith T. Poole. 1987. "The Relationship between Information, Ideology, and Voting Behavior." *American Journal of Political Science* 31(3): 511–30.
- Poole, Keith T., and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29(2): 357–84.
- Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3): 530–51.
- Tahk, Alexander M. 2005. The Signals in the Noise: A Cross-Validation Approach to Estimating the Dimensionality of Roll Call Voting. Unpublished manuscript. <http://www.tahk.us/papers/signals-noise.pdf>.
- Tausanovitch, Chris, and Christopher Warshaw. 2014. "Representation in Municipal Government." *American Political Science Review* 108(3): 605–41.
- Weisberg, Herbert F., and Jerrod G. Rusk. 1970. "Dimensions of Candidate Evaluation." *American Political Science Review* 64(4): 1167–85.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

1. Senate Representation Question Wording
2. 2008 CCES Question Wording
3. Description of Group-Based Scaling Gibbs Sampler
4. Gibbs Sampler Performance
5. sMSD Monte Carlo Simulations
6. Comparison to Bayesian Aldrich-McKelvey Scaling
7. Heteroskedastic Ideal Point Model Performance
8. Testing Spatial Voting Using Group-Based Estimates