How Should We Choose Survey Questions to Measure Citizens' Policy Preferences?

Chris Tausanovitch
ctausanovitch@stanford.edu

Christopher Warshaw
cwarshaw@stanford.edu

Department of Political Science
Stanford University

February 2012

Abstract:
The effect of variation in citizens' policy preferences on salient political outcomes lies at the heart of a number of research agendas in political science. Little attention, however, has been paid to the quality of our measures of policy preferences. Even less attention has been paid to how we might improve our measures of policy preferences by asking better questions. In this paper, we address these questions using a unique survey of the American public with over 100 policy questions on economic and social issues. First, we evaluate the dimensionality of citizens' policy preferences. We show that one dimension captures the policy preferences of the American public. Although this does not imply that the public is "unidimensional" per se, "off-dimensional" opinions are probably idiosyncratic for most voters. Next, we show that different sets of survey questions yield substantively different estimates of policy preferences for the same set of respondents. A small number of policy questions or inadequate question quality can lead to incorrect inferences for research questions on polarization and issue voting. Finally, we examine how to select survey questions to optimize our measure of policy preferences, and provide recommendations for survey designers.

I. Introduction

The effect of variation in citizens' policy preferences on salient political outcomes lies at the heart of dozens of questions in political science, including research agendas on polarization (e.g. Fiorina, Abrams & Pope 2008, Fiorina & Abrams 2008, Abramowitz & Saunders 2008, Abramowitz 2010), issue voting (Ansolabehere, Rodden & Snyder 2006), and legislative representation (Clinton 2006). As a result, it is crucial for scholars to accurately measure citizens' policy preferences. However, little attention has been paid to the quality of our measures. Even less attention has been paid to how we might improve our measures of policy preferences by asking better questions.

Scholars of public opinion often use single policy questions to measure aggregate views about policy. However, these policy views are measured with error (Ansolabehere, Rodden and Snyder 2008). Moreover, citizens' views on individual policies are likely to be based on their general policy preferences.[1] In our model, we assume that responses to policy questions are a function of the spatial distance between a person's most preferred point in the underlying space and the policy alternatives on offer, with people tending to prefer the "closer" alternative (see Clinton, Jackman, and Rivers 2004; Poole and Rosenthal 2007). In one dimension, each voter can be thought of as having an ideal point such that individuals with liberal views have ideal points smaller in a numerical sense than those with politically conservative views (Bafumi and Herron 2010). In our setup, an ideal point is a sufficient statistic for an individual's policy preferences, so we will refer to ideal points simply as "policy preferences."[2]

There are many reasons to improve our measurements of citizens' policy preferences. First, a fine-grained, continuous, and precise measurement is required to fully and accurately describe the distribution of a continuous latent variable such as policy preferences. Second, better measures give us a better sense of effect size, and a reduced chance that closely related concepts (such as party and preferences) will be confounded in our analysis. Finally, as we explore the higher-dimensional structure of citizens' policy preferences, better measurement becomes more difficult and even more important. Most importantly, we want to know how many dimensions are needed to accurately characterize people's political choices, what these dimensions consist of, and what their relative importance is for political behavior such as voting, donating, and volunteering.

Fortunately, the importance of accurate measurement is not confined to political science and the measure of policy preferences, and so we can look to developments in other fields for guidance. In particular, the challenge of selecting good policy questions on surveys is akin to the challenge of selecting good standardized test questions. To address this challenge, the fields of educational testing and psychometrics have been concerned with improving the measurement of latent variables at the level of items or questions since the influential work of Rasch (1960) and Lord, Novick & Birnbaum (1968). The Rasch model is similar to the Two Parameter Item Response model for measuring citizens' and legislators' ideal points introduced into political science by Clinton, Jackman & Rivers (2004). As a result, it is relatively straightforward to use techniques developed in the education literature to assess standardized test questions to examine the quality of the policy questions asked on political science surveys (Van der Linden 2005). We can then choose the questions that best measure individuals' latent policy preferences.

In this paper, we show that treating policy preferences as a latent variable and analyzing the measurement properties of this variable can improve our ability to answer important questions in political science. One such question concerns the dimensionality of the mass public's policy preferences. A number of scholars have argued that citizens' policy preferences are "multi-dimensional," by which it is usually meant that multiple distinct measures are needed to accurately describe voters' shared structure of beliefs (e.g., Treier and Hillygus 2009). We find, however, that moving from one to two dimensions only yields about a one-and-a-half percentage point increase in correct prediction. Using the same criteria commonly applied to legislators, this indicates that the structure of voter preferences is one-dimensional with respect to the questions we analyze. This is not to say that no voters have highly structured beliefs that do not fall neatly within a single liberal-conservative dimension. But such beliefs are rare enough or idiosyncratic enough that higher dimensions add little in terms of explaining the vote choices of the larger voting public.

Next, we show that it matters for substantive questions what policy questions are included on surveys. For instance, the debate about whether or not the American public is polarized is primarily a debate about the shape of the distribution of the mass public's policy preferences. We show that inferences about the shape of this distribution drawn from any particular survey are extremely sensitive to the specific questions included on the survey. Not surprisingly, surveys fare better when they include more policy questions that separate people in different regions of the ideological space and discriminate well across individuals.

In addition to demonstrating the benefits of good measurement and examining the structure of policy preferences, we show that it is possible to use the Two Parameter Logistic Item Response Model to pick the best policy questions to estimate latent policy preferences. With only 12 questions, it is possible to obtain reliable estimates of 1-dimensional policy preferences that are almost as accurate as estimates obtained from a set of over 20 questions asked on a recent large sample survey.

More broadly, the optimal question selection approach described in this paper points the way towards better survey design and analysis. In large sample surveys, in which scaling a large number of questions is computationally intensive, it provides a methodology for analysts to choose a relatively small number of survey questions that accurately measure citizens' policy preferences in multiple dimensions. For surveys where interview time is scarce, optimal question selection enables survey designers to optimize their measurement of policy preferences given the number of questions available to them. For instance, a survey on the Tea Party could use our approach to pick a handful of questions that measure conservatives' policy preferences well.

Our findings also have implications beyond survey design. Latent variables are prevalent throughout political science. Concepts such as democracy (Treier & Jackman 2008), the strength of political institutions (Lau 2005), and the levels of support for political candidates over the course of an election campaign (e.g. Green, Gerber & Suzanna N.d.) are all latent variables. As Treier & Jackman (2008) point out, "in each instance, the available data are manifestations of the latent quantity and the inferential problem can be stated as follows: given observable data y, what should we believe about latent quantities x?" However, the observable data y available to scholars is rarely fixed.

For instance, scholars of democracy could invest resources into obtaining new observable data to improve their measures of democracy. Optimal question selection provides a method for choosing which data provide the biggest bang for the buck in terms of improving scholars' measurement of democracy or other latent variables of interest.

The following section explains how we conceptualize policy preference, and how optimal question selection can be applied to improve our measures of policy preferences. Next, we examine the dimensionality of the mass public's preferences. In the following section, we analyze sets of questions asked on existing surveys. We show that different question sets yield substantively different answers to major research questions. Then, we use optimal question selection to propose sets of questions that are particularly effective for measuring policy preferences. We also present recommendations for survey designers. In conclusion, we discuss avenues for future research.

II. Theory

Our approach derives from the spatial theory of voting. We assume that both citizens and legislators have a unique set of policies that they "prefer" to all others. This point in the policy space is called an "ideal point" or "policy preference" (see Poole and Rosenthal 1985; Clinton, Jackman, and Rivers 2004). We assume that on any given dimension, people prefer policies that are closer to their ideal point over policies that are farther away.[3] An ideal point is a convenient summary of how far to the "left" or the "right" a person's policy preferences are on each policy dimension. In one dimension, individuals with liberal views will have ideal points smaller in a numerical sense than those with politically conservative views.

This spatial voting model has now been employed in a number of different institutional contexts (e.g., Bafumi and Herron 2010; Bailey 2007; Clinton, Jackman, and Rivers 2004; Martin and Quinn 2002, 2007; Poole and Rosenthal 1997, 2007). It also been applied to measure citizens' policy preferences in one (Bafumi and Herron 2010; Jesse 2009) and two dimensions (Treier and Hillygus 2009).

The view that citizens have underlying preferences that form the foundation of their choices on specific policies has a long history in political science. Converse (1964, 207) argues a "belief system" can be defined as "a configuration of ideas and attitudes" where the elements are "bound together by some form of constraint or functional interdependence." For instance, Converse points out, "if a person is opposed to the expansion of social security, he is probably a conservative and … opposed as well to any nationalization of private industries, federal aid to education, sharply progressive income taxation, and so forth" (Converse 1964, 207). Similarly, Shafer and Claggett (1995) refer to "policy preferences" and "policy predispositions" within an "issue context." They capture the common definition of this concept, albeit loosely, when they say that it is "the 'deep preferences' to political opinion."

Unfortunately, measuring citizens' policy preferences is difficult for two reasons. The first is that survey respondents often make "errors" in their responses. For a variety of reasons they fail to choose the response that is most associated with their underlying values. In the simplest case, the respondent may check the wrong box, or mishear a question asked by a telephone interviewer. If this were the only type of error, it would not be of great concern. But it has been demonstrated that respondents will often chose different answers to the same survey question at different times, or change their response in light of irrelevant information (i.e. Zaller 1992). Although some take this as evidence of "non-attitudes," an alternative hypothesis is that survey responses are simply noisy. In order to remover the signal from the noise, repeated measurements are required (Ansolabehere, Rodden & Snyder 2008). The solution is to design surveys with an adequate number of questions that are highly related to individuals' policy preferences. In other words, we want questions where the probability of mistakes is low.

The second difficulty is that survey questions are dichotomous or categorical. Each question separates respondents into categories of responses. However, the variable that we really want is continuous: how "liberal" or "conservative" is the respondent? If we take individual questions at face value, we can't separate respondents who give the same response. However, what we can do is ask multiple questions about policy, each of which separates respondents into finer groups. Given that the rate of errors is low, we require questions that separate different groups: liberals from extreme liberals, moderates from liberals, conservatives from moderates, and so on.

A. Measuring Citizens' Preferences

To measure citizens' policy preferences, we use a 2-Parameter Logistic Item Response Model (Clinton, Jackman, and Rivers 2004).[4] We start by assuming a one-dimensional model, which will capture the dominant dimension that explains answers to our questions. Let x_i denote person *i's* latent policy preferences, and let y_{ij} denote person i's response to question j, where $y_{ij} = 1$ indicates a "yes" response and $y_{ij} = 0$ indicates a "no" response[5]. Then the probability of person i answering "yes" to a given question is given by:

$$\Pr(y_{ij} = 1) = \wedge(\beta_j x_i - \alpha_j)$$

where α_j and β_j are the item parameters. $\wedge$ is the logistic cumulative distribution function. In the education literature, α_j is referred to as the “difficulty parameter" because a higher value of α indicates a lower probability of a “correct" answer (in our case, a yes answer). It is easier to think in terms of the “cut point" α_j / β_j which is the value of x_i at which the probabilities of answering yes or no to a question are 50-50. β_j is referred to as the “discrimination" parameter because it captures the degree to which the latent trait affects the probability of a yes answer. If β is 0, then question *j* tells us nothing about the respondent's level of liberalism or conservatism. We would expect β to be close to 0 if we ask a completely irrelevant question; for instance, a question about the respondent's favorite flavor of ice cream.

The complete likelihood is simply the product of all of the individual likelihoods for each vote choice:

$$\prod_{i \in I} \prod_{j \in J} (\wedge(\beta_j x_i - \alpha_j))^y (1 - \wedge(\beta_j x_i - \alpha_j))^{1-y}$$

where I is the set of all people and J is the set of all items. In practice, we will exclude any pair *ij* if that choice is missing or the respondent answered “Don't Know". In our data, we have a very low rate of voluntary missingness. Most missingness is due to the fact that in order to ask more questions, we did not ask every question to every person, but randomly mixed them.

Although the parameters are identified relative to one another, they lack a scale. We establish an arbitrary scale by normalizing the policy preferences, x_i's. A more difficult issue is how to achieve parameter values that are comparable across different scalings. Our goal is to compare the x_i estimates using different sets of question. It would be a mistake to normalize each scaling, because this would artificially impose a common variance. Instead, for each scaling other than the one that uses the entire survey we will set the item parameters for one item equal to the estimates we obtain when we scale the entire survey, and transform the other parameter values accordingly. This will allow us to compare the distribution of the x_i's relative to that item. However, the measurement error in this “reference" item will lead to some variance in the position of the scale.

B. Assessing Individual Survey Items

Our approach to assessing survey questions draws heavily from Van der Linden (2005), the authoritative text in the education literature. We can assess the contribution of a given item to our level of certainty at a value of x_i by evaluating Fisher's Information for that item at that value (Bimbaum 1968):

$$IIF_j(x) = \beta_j^2 pq$$

where $p = \wedge(\beta_j x_i - \alpha_j)$ and $q = 1 - p$

This is referred to as the Item Information Function. In frequentist statistics, use of the IIF is motivated by the fact that the IIF is inversely proportional to the variance of the maximum likelihood estimator of x. In other words, higher values of the IIF indicate that we would estimate x more precisely in a frequentist context. For convenience we will estimate this model using Bayesian methods, with very similar results. A Bayesian may justify the use of the IIF on the basis that it is increasing in the gradient of the likelihood, and that under the (non-informative) priors used here, the gradient of the posterior will be increasing in information.

A convenient feature of the IRT model is that the Fisher information for the entire likelihood, the Test Information Function (TIF) (so called because the theory was developed in the literature on educational testing), is simply the sum of the individual Item Information Functions. We will use the IIF as means of selecting items, and the TIF as a way of comparing sets of items. Our first task will be to select items that maximize the TIF across the range of values for x. Van der Linden (2005) shows that this can be done by maximizing the TIF at 3 to 5 uniformly distributed points in the range of x. Since the TIF is an additive function of the IIFs, this requires only that we calculate the values of the IIF at each of 5 points, and choose the items with the highest sum of these values.

However, we may not only be interested in maximizing information across the range of x. If, for instance, we are only interested in discriminating between, say, conservatives and liberals, we could maximize the TIF at a single value in the middle of the spectrum. One subject of interest might be the location of ideological extremists. Very few surveys ask questions that have small margins, and apply to only extremely conservative or extremely liberal individuals. As a result, it is difficult to differentiate between the views of, say, liberals who associate with the left wing of the Democratic party, and Marxists. Using this as an example, we will find optimal sets of items both for the range of x and for the 5% and 95% quantiles of x.

C. Data

We measure citizens' policy preferences using a module we placed on the 2010 Cooperative Congressional Election Survey (CCES). Here, we asked approximately 175 policy questions to 1,300 respondents. These questions include almost every policy question from size recent large-sample national surveys (the 2000, 2004, and 2008 National Annenberg Election Surveys (NAES) and the common content from the 2006, 2008, and 2010 Cooperative Congressional Election Studies (CCES)). They also include a large set of additional policy questions, which enables us to estimate more precise ideal points than was possible with the smaller sets of questions on earlier surveys.

III. Evaluating the Dimensionality of the American Public's Policy Preferences

Before we proceed to selecting optimal questions, we need to know how many dimensions are required to describe citizens' policy preferences. If the model is mis-specified with respect to the number of dimensions, then estimates may be biased. There is wide disagreement among scholars regarding the dimensionality of the American public's policy preferences. In contrast to the uni-dimensional Congress, many scholars contend that the American public's policy preferences are multi-dimensional (Ansolabehere, Rodden, and Snyder 2006; Treier and Hillygus 2009). However, others contend that the electorate is fundamentally one-dimensional. For instance, Jessee (2009) finds that citizens' policy preferences can be characterized by "a dominant first dimension, with further dimensions contributing little explanatory power." Finally, other scholars assume that the electorate is one-dimensional but they do not actually prove this assumption (e.g., Bafumi and Herron 2010). In this section we will show that higher-dimensional models provide surprisingly little increase in model fit.

As in the one-dimensional case, we need to make some identifying assumptions to identify the scale. We identify the first dimension by requiring a question about social security privatization to discriminate on one dimension and not the other. We establish the second dimension by requiring the person parameters to be orthogonal to the person parameters on the first dimension. Estimates on both dimensions are normalized. This is an "exploratory" two-dimensional IRT model. The value of this approach is that it allows us to examine the content of the second dimension without making any assumptions about what the dimensions should consist of.[6]

Table 1 about here

Table 1 shows the discrimination parameters for the highest discriminating items in each dimension. The top discriminating items on dimension 1 are a mix of issues, from gay marriage to inequality and climate change. In contrast, dimension 2 consists more exclusively of social issues. However, the most notable difference between the two dimensions is that the discrimination parameters for dimension one are generally higher, even for questions that discriminate highly on the 2nd dimension. Dimension 1 has roughly as much social issue content as dimension 2. Dimension 2 helps the model fit in cases where predictions on social issues questions might be slightly off, but for the most part these questions, and all others, are fit by dimension 1.

Table 2 about here

Next, we evaluate whether adding dimensions improves our ability to predict citizens' responses to individual survey questions. Table 2 shows model fit statistics for 4 different models. The first is the "naïve model," in which each respondent selects a given response with probability equal to the population proportion that selected that response. This is equivalent to an item response model in which all respondents are constrained to have exactly the same ideal point. The second model is the 1-dimensional item response model analyzed above followed by the 2-dimensional and 3-dimensional models. Each of these models is fit to the entire set of questions asked to all respondents on our module.

The first row of table 2 shows the log likelihood of each model. The second row shows the percent of the responses that are correctly predicted. The third row shows the average of the geometric mean probabilities, calculated for each question separately. This statistic effectively gives higher weight to fitting a broad range of questions- it is each question, rather than each response, that counts equally.

Do additional dimensions decrease model errors enough to justify their inclusion? We find that the addition of a second dimension increases the percent of responses correctly predicted by 1.4 percentage points. This is very similar to the improvement in model fit from a 2nd dimension in recent Congresses (Poole and Rosenthal 2007). Poole and Rosenthal argue that this minimal improvement in model fit implies that Congress is essentially one-dimensional. Based on this logic, the mass public's policy preferences are also one-dimensional. This is not to say that no voters have highly structured beliefs that do not fall neatly within a single liberal-conservative dimension. However, such beliefs are rare enough or idiosyncratic enough that higher dimensions add little in terms of explaining the vote choices of the larger voting public.

What explains our finding that higher dimensions do not significantly improve the fit of the model when past research finds that the electorate is at least two-dimensional? First, we are one of the first to apply measurement strategies from the congressional literature to the public opinion literature. Aside from Jesse (2009), we are aware of no paper that examines the fit statistics of IRT models in the way that we have done here.

Most of the existing research uses separate scalings or indexes, and examines the correlations between these measures.[7]

Second, unlike the estimates from past work, the estimates presented here do not presuppose the dimensional structure of policy preferences. When the model is constrained, the resulting estimates fit the data worse than they would otherwise. The estimates themselves are measured with more error. Likewise, estimates from separate scalings include fewer highly informative questions. Error attenuates the correlation between the dimensions. Thus, the literature is mistaken in taking low correlations as evidence of high dimensionality.

IV. How much does question selection matter for estimating policy preferences?

Having shown that one dimension is adequate to describe the space of policy preferences we proceed to show that better question selection can improve our measurement of this latent variable. Policy preferences are one of the most important variables in political science. However, the typical approach to measuring it is *ad hoc*. Survey designers usually select a set of questions with little attention to their measurement properties. Questions that are more recent and salient are more likely to be included in surveys, whereas questions that evaluate long-standing attitudes are more likely to be passed up. In this section we demonstrate that the choice of questions has significant consequences for the accuracy of our measurement of citizens' policy preferences.

In order to show the impact that questions selection can have on the resulting scaling, we individually scale our sample using 6 different subsets of questions, each of which corresponds to the set of economic and social policy questions asked on one of the National Annenberg Election Surveys or Cooperative Congressional Election Studies. This correspondence is not exact because there is some ambiguity about which questions should be classified as “policy" questions. To be classified as a policy question, the question had to address choices that are understood to be in the governmental sphere. Questions on partisanship, ideological self-image, personal philosophy, or judgments about people or groups were not considered policy questions. We were lenient with regard to classifying issues as economic or social, usually neglecting to include questions about foreign wars and trade agreements, but little else. In one case, a widely-used question was not included because of a transcription error.[8] Because 2010 was a banner year for healthcare, we estimated our model both with and without healthcare questions, avoiding the possibility that we are estimating a “healthcare scale[9].” The results in this section exclude healthcare questions, however, they are similar in either case. We replicate the results with healthcare questions in Appendix B, and discuss this choice further. The purpose of this section is not to criticize the set of questions included on any particular survey, but to show that question selection matters.

Figure 1 about here

Figure 1 shows the test information function for a one-dimensional scaling done using all of the questions on our module. Because the TIF is additive in the information functions for the individual items, this function dwarfs the TIFs for the small subsets of items that are commonly used. The TIFs for the question subsets are shown in Figure 2. Note that the scale of the y-axis for Figure 2 is substantially smaller. A common feature of the test information functions is that there is usually higher information in the center of the distribution. This is sensible, since most questions asked have margins not too far from 50-50, and questions with smaller margins are estimated with less precision. The 2010 CCES dominates all of the other surveys in terms of test information. In other cases, it is unclear which survey performs better. For instance, the 2008 CCES fares better in the left of the distribution and the 2000 NAES fares better in the right. The 2008 NAES lags far behind any other survey.

Figure 2 about here

What determines which surveys provide more information in which regions of the space? There are three crucial factors. Firstly, and most obviously, is the number of items. The 2010 CCES asked more policy items than any of the other surveys, at 22 items. The curves are roughly in order of the number of questions asked, although not quite. The 2006 CCES does squarely better than the 2004 NAES despite asking one fewer question. We will show that careful item selection can vastly improve the informativeness of a survey in section 4. For now, we note that the number of items is important, but it isn't everything.

The second factor is discrimination, the β parameter of the items. Better discriminating items show a stronger relationship between the response to the item and the latent variable. In the case of policy preferences this may mean that the question is more apropos of policy preferences, easier for respondents to understand, or simply taps more clearly into the primary cleavage of politics rather than more idiosyncratic concerns.

The third factor is the cut point, α / β. When a cut point is closer to a given point on the parameter space, it is more informative about respondents whose ideal point is also close to that point. This is very intuitive: a question that has a 50-50 margin and is highly discriminating should have a cut point close to the center, and be most useful for separating liberals from conservatives. A question with a small margin that asks respondents to agree or disagree with an extreme position should be better for separating moderate liberals from extreme liberals or moderate conservatives from extreme conservatives. In order to discriminate well for a certain x, there must be a cut point nearby.

Figure 3 about here

Figure 3 shows the kernel density estimates of the distribution of preferences using the scaling from each subset of questions, weighted using survey weights,[10] labeled by the survey they are drawn from. The dashed vertical lines are the cut points for each item asked, with the darkness of the line in proportion to its discrimination. The solid vertical lines are the cut points of the items that are used to identify the scale. They are set to be identical to the values for the full module scaling on the top left. The item chosen to identify the space is always the most informative item. The fact that these items are estimated with uncertainty within each scaling will lead to some error in the comparability across scalings, with more error in more inaccurate scalings or when the most informative item is less informative. A table of items used for identification and their informativeness rank can be found in the online appendix. Each scaling uses one of the top 10 most informative items except for the 2008 CCES and the 2008 NAES, which use the 22nd and 25th most informative items, respectively, reflecting the fact that items in these subsets are not very informative.

The first thing to note about these distributions is that their shapes vary significantly across the set of items used. In particular, we tend to observe modes in the distribution in regions where there are few or low discriminating items. These modes are not the product of any actual bunching in the population, but simply uncertainty about where in a region the respondent should be placed. The length of the tails of the distribution is determined in part by the overall informativeness of the items. Because the default, uninformed estimate of any individual's policy preferences is 0, in less informative tests the estimates will tend to be closer to 0, *ceteris paribus*.

A. Application to Polarization

A large literature in political science is concerned with the shape of the distribution of preferences (e.g. Fiorina, Abrams & Pope 2008, Fiorina & Abrams 2008, Abramowitz & Saunders 2008, Abramowitz 2010). A couple of features of these distributions stand out. The estimation from the full module is by far the smoothest, lacking many of the “interesting" features that the analyst may be tempted to remark about the other scalings. This scaling is an accurate reflection of two quantities that scholars struggle to estimate: the modality of the distribution of preferences, and its variance. Modality is difficult to estimate because a lack of granularity and discrimination leads to the observation of modes that aren't there. Variance is difficult to estimate because few scalings discriminate well in the extreme of the distribution. Indeed, it is often the case that the most extreme individuals are to the right or left of every cut point. In a maximum likelihood framework, the estimated position of these individuals would be positive or negative infinity. In a Bayesian model, their position outside the most extreme cut point is determined entirely by the prior and the identifying restrictions.

Variance and modality are both fundamental to understanding the polarization of the American electorate. The claim that America is a polarized nation is often summarized in terms of a change in one or the other statistic (e.g. Fiorina, Abrams & Pope 2008, Fiorina & Abrams 2008, Abramowitz & Saunders 2008, Abramowitz 2010). We cannot speak to over time changes here, however we can say that there is one mode in the density of American policy preferences. There are many people who are distant from this mode, but they are few as a percentage of the public. This is a different conclusion than one might come to if one were to examine a density based on fewer questions, where there are other modes.

Figure 4 about here

The assertion that bunching in the distributions is “artificial" is backed up in figure 4, which plots the posterior variance of estimates against the value of the estimates. The y-axis of each panel differs to reflect the different levels of posterior certainty for each scaling. The panel that corresponds to the full set of questions looks like a smooth quadratic curve; almost the inverse of the TIF. This makes sense if we recall that information is the inverse of the standard error of the maximum likelihood estimate. In the other panels, however, the relationship is much messier. There are clumps of points that share a common variance, and spaces between these points and other points. The panel for the 2006 CCES shows an extreme version of a common pattern. Here, the plot looks like a smiley face, with a smooth quadratic in the middle and two high-variance bunches on the outside. To put it simply, this scaling does not allow us to tell the difference between people on the outside of the distribution. We can locate people in the center of the distribution, but people in the outside of the distribution may be even more extreme than the graph indicates- we don't have enough information to place them. High variance is associated with clumping points into a “default" location.[11]

B. Application to Issue Voting

We have shown that different items lead to substantially different distributions, with the full module yielding the most convincing measure. How do the measures stack up in terms of their predictive power regarding phenomena that political scientists care about? We take voting for president as an example. Two major questions in political science are the extent of voting based on policy preferences (often framed in terms of the rationality of the vote choice) and the comparative power of policy preferences vis-a-vis party identification (e.g., Ansolabehere, Rodden & Snyder 2006). We find that poor measurement leads to an underestimate of the importance of policy preferences in voting behavior.

Table 3 summarizes the results of 7 logistic regressions of retrospective vote choice on party identification and policy preferences. In each case, the measure of policy preferences uses the scaling from the set of items corresponding to the survey at the top of each column. The dependent variable is coded as 1 if the respondent voted Republican for President and 0 otherwise. Non-voters are dropped. The first row is the difference in the predicted probability of voting Republican for respondents who identify as Independents rather than Democrats. The second row shows this difference for Republican respondents versus independents. The third row shows the predicted change in probability of voting Republican resulting from a one standard deviation change in policy preferences in the conservative direction. The fourth row shows the log likelihood, and the fifth row shows the percent correctly predicted.

Table 3 about here

Table 3 shows that the relative strength associated with policy preferences relative to party, and the predictive accuracy of the model, are increasing in the goodness of the measure of policy preferences, with the full CCES module outperforming all other measures. Presidential vote is an easy thing to predict since most voters vote the party line, so the changes are not staggering, especially in the percent correctly predicted. However, the inferences about the relative importance of policy preferences change greatly depending on which set of questions is used to estimate respondents' policy preferences. When the full module is used for scaling, a one standard deviation change in policy preferences has an effect size 64% of the effect of switching parties. When the 2008 NAES is used, a one standard deviation change in policy preferences has an effect size that is 35% as big. Clearly, poor measurement can lead to a vast underestimate of the importance of policy preferences in voting behavior. The 2010 CCES and the full module statistically outperform all of the other surveys.

V. Optimal Questions to Estimate Policy Preferences

Can we do better by carefully selecting items, even under constraints with regard to the number of questions we can put on a given survey? To examine this question, we scale subsets of items that are selected to optimize information criteria. We use three different criteria: greatest total information, greatest "liberal" information, and greatest "conservative" information, each of which we will consider in turn.

A. Optimal Questions to Differentiate Liberals from Conservatives

First, we evaluate the optimal set of questions to differentiate liberals from conservatives. The "greatest total information" criterion is calculated by using the sum of information for all of the items at five points across the distribution. For each criterion, we select the 12 items that maximize the criterion. Recall that information is additive in the information of each item. Therefore, we can simply rank each item according to each criterion, and choose the top *n* for scaling. We use the 2010 CCES scaling as a baseline. Recall that there were 22 policy items in the 2010 CCES. Figure 5 shows the test information function for the 2010 CCES and the optimal item sets. With 12 items, we do better than the 2010 CCES does with 22 items in the middle of the distribution, and only slightly worse over the support of the rest of the data.[12]

Figure 5 about here

Table 4 about here

Table 4 shows the results for logit models of presidential vote share on party and preferences measured using the full module and the "greatest total information" question sets with 12 items. Both perform as well as survey item sets that use many more items, with the 12 item set outperforming all other item sets except the 2010 CCES (see tables 3 and 4). The coefficients on policy preferences for the 2010 CCES and the 12 item set are statistically indistinguishable. The most informative questions for the maximum total information criteria are given in table 5, above, in decreasing order of informativeness.

Table 5 about here

Table 6 about here

Table 7 about here

B. Optimal Questions to Differentiate *Among* Liberals and Conservatives

Whereas we calculated the “greatest total information” by focusing on informativeness at a range of points, we obtain better estimates of the policy preferences of extreme liberals and conservatives by focusing on one point near the extreme of the distribution. In order to calculate greatest “liberal" information, we simply calculate information at the 5% quantile of the data. Likewise, for greatest “conservative" information, we calculate information at the 95% quantile of the data. Tables 6 and 7 show the most informative questions for the maximum liberal information criteria and the maximum conservative information criteria, respectively. Note that these sets of questions have very little overlap with the maximum total information criteria questions. This demonstrates that cut points matter. Questions that have cut points closer to a certain group measure that group better. Existing surveys do an especially poor job discriminating between extremists. This is difficult to do with few questions, but by selecting questions with respect to informativeness, we can do much better.

C. Recommendations For Question Selection

In this paper we have presented several lists of “optimal” sets of questions. Which ones should survey designers use in future surveys? As the above discussion indicates, the optimum depends on what the survey designer seeks to optimize. Optimizing total information will yield the best estimate over the range of the space, but a survey designer may wish to focus on one region of the space. In two dimensions, this question is further complicated by the trade-off between the dimensions.

What determines which surveys provide more information in which regions of the space? There are three crucial factors. Firstly, and most obviously, is the number of items. More items yield better estimates of policy preferences. The second factor is discrimination, the β parameter of the items. Better discriminating items show a stronger relationship between the response to the item and citizens' underlying policy preferences. In the case of policy preferences this may mean that the question is more apropos of policy preferences, easier for respondents to understand, or simply taps more clearly into the primary cleavage of politics rather than more idiosyncratic concerns. The third factor is the cut point, α / β. When a cut point is closer to a given point on the parameter space, it is more informative about respondents whose ideal point is also close to that point. This is very intuitive: a question that has a 50-50 margin and is highly discriminating should have a cut point close to the center, and be most useful for separating liberals from conservatives. A question with a small margin that asks respondents to agree or disagree with an extreme position should be better for separating moderate liberals from extreme liberals or moderate conservatives from extreme conservatives.

Sets of optimal items and item information statistics are posted on our websites so that survey designers can use them in informing their question selection. We also have posted descriptions of the items that we used. A significant concern is raised by the fact that the informativeness of some items may be dependent on their current salience. For instance, it is unlikely that the choice of Congress to confirm Justice Kagan to the

Supreme Court will continue to be as salient an issue as it was in 2010. As a result, we predict that the informativeness of this question for policy preferences will decline. If Justice Kagan takes a prominent role in a controversial decision, this question may remain informative, but with changed item parameters. We do not deny the role of the survey designer's judgment in making good question selection decision. However, the fact that many of the questions asked on the surveys studied here are lacking in informativeness is evidence that survey designers can benefit from these analyses.

VI. Conclusion

In this paper, we have used a unique survey instrument with over 100 policy questions to evaluate the policy preferences of the American public. First, we find that one dimension can characterize citizens' policy preferences. Additional dimensions yield little improvement in model fit. This finding resolves a long debate in the literature about the dimensionality of American policy preferences, and allows us to focus on improving our measurements in a single dimension.

Next, we show that treating policy preferences as a latent variable and analyzing the measurement properties of this variable can improve our ability to answer important questions in political science. We have shown that sets of policy questions that were asked to tens of thousands of respondents have extremely heterogeneous accuracy across surveys. Analyzing the properties of many items on a small survey with respect to their informativeness allows us to design surveys that perform much better with fewer questions. These "optimal" question sets can be designed to be informative in different parts of the political spectrum, and perform particularly well relative to *ad-hoc* survey designs in two dimensions.

The approach in this paper has implications for both survey design and for political science more broadly. For survey designers, it provides some heuristics as well as a formal methodology for choosing a relatively small number of survey questions that accurately measure policy preferences in multiple dimensions. While the formal method presented here is the best method for choosing questions from a pre-existing pool, the heuristics are useful for formulating original questions. Proposed questions should aim for high discrimination, but also for cutpoints that are spread throughout the target region of the political preference space.

These more accurate measurements of policy preferences will enable scholars to improve their inferences on polarization, issue voting, and the structure of policy preferences in the American electorate, among others. Moreover, the method outlined here provides a tool for survey designers to measure particular slices of the ideological continuum well with a very small number of survey items.[13]

Future research should consider other possible issue dimensions. While our analysis is based on a very large number of issue questions, we have only scratched the surface of the issue areas that one could ask about. Few of our questions address immigration, and almost none relate to foreign policy. We suggest items that are particularly informative among the ones we ask, but an obvious avenue for future research is to ask more questions to larger samples in order to find even better question sets and evaluate whether there is higher dimensionality in citizens' policy preferences with regard to issues that were not sufficiently addressed in our question pool. There is ample room for methodological innovation as well, from the use of polythomous models to more complete treatment of “don't know" answers. Fox (2010) has shown that greater precision can be obtained by modeling the relationship between the discrimination and ability parameters. We have used simple criteria for informativeness, but there are many other useful criteria, some of which require complex linear programming to optimize. Given that surveys are increasingly computerized, even better information may be gained than that offered here by using adaptive methods for question selection (e.g., similar to those used on the GRE exam).

Our findings also have implications for other areas of political science. Latent variables are used for research agendas throughout political science. Our findings imply that small variations in the data used to measure latent variables, such as the strength of democracy or political institutions, could have a large impact on substantive inferences. Moreover, optimal question selection could be used to better measure latent attitudes from partisan attachment to racial resentment to authoritarianism. Our ability to choose which questions we ask in a survey context is the distinctive feature that optimal question selection takes advantage of. However, to the extent that political scientists get to choose what they measure, it can apply even more broadly. For instance, comparative politics scholars could use our technique to examine which data on indicators related to democracy have the largest impact on measurements of democracy as a latent variable. Treier and Jackman (2008) demonstrate how to take advantage of the available information on a variable of interest. Our contribution is to show how to better design measurement instruments in the first place.

REFERENCES

Abramowitz, A. 2010. The disappearing center: engaged citizens, polarization, and American democracy. Yale University Press.

Abramowitz, A.I. & K.L. Saunders. 2008. “Is polarization a myth?” The Journal of Politics 70(02):542–555.

Ansolabehere, Stephen, Jonathan Rodden & James M. Snyder. 2006. “Purple America.” The Journal of Economic Perspectives 20(2):97–118.

Ansolabehere, Stephen, Jonathan Rodden & James M. Snyder. 2008. “The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting.” American Political Science Review 102(02):215–232.

Bafumi, Joseph & Michael Herron. 2010. “Leapfrog Representation and Extremism: A Study of American Voters and their Members in Congress.” American Political Science Review 104:519-542.

Baldassarri, D. & A. Gelman. 2008. “Partisans without constraint: Political polarization and trends in American public opinion.” American Journal of Sociology 114(2):408–446.

Bimbaum, A. 1968. “Some latent trait models and their use in inferring an examinee’s ability.” Statistical theories of mental test scores pp. 395–479.

Clinton, J.D. 2006. “Representation in Congress: constituents and roll calls in the 106th House.” Journal of Politics 68(2):397–409.

Clinton, Joshua D., Simon Jackman & Douglas Rivers. 2004. “The Statistical Analysis of Roll Call Data.” American Political Science Review 98(02):355–370.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In Ideology and Discontent.

Fiorina, M.P., S.A. Abrams & J.C. Pope. 2008. “Polarization in the American public: Misconceptions and misreadings.” The Journal of Politics 70(02):556–560.

Fiorina, M.P. & S.J. Abrams. 2008. “Political polarization in the American public.” Annual Review of Political Science 11:563–588.

Fox, J.P. 2010. Bayesian Item Response Modeling: Theory and Applications. Springer Verlag.

Frank, T. 2005. What’s the matter with Kansas?: how conservatives won the heart of America. Holt Paperbacks.

Green, D.P., A.S. Gerber & L. Suzanna. N.d. "De Boef. 1999. Tracking opinion over time: A method for reducing sampling error." Public Opinion Quarterly. Forthcoming.

Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." American Political Science Review. 103:59-81.

Lau, O. 2005. "Measuring Institutional Strength: The Correlates of Growth." Working Paper.

Lord, F.M., M.R. Novick & A. Birnbaum. 1968. "Statistical theories of mental test scores."

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." Political Analysis 10: 134–53.

Martin, Andrew D., and Kevin M. Quinn. 2007. "Assessing Prefer- ence Change on the US Supreme Court." Journal of Law, Eco- nomics, and Organization 23 (2): 365–85.

Poole, Keith & Howard Rosenthal. 1997, 2007 Ideology and Congress. Transaction Publishers.

Rasch, G. 1960. "Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.".

Schattschneider, E.E. 1960. The semisovereign people: A realist's view of democracy in America. Holt, Rinehart and Winston.

Snyder, J.M. 1992. "Artificial extremism in interest group ratings." Legislative Studies Quarterly 17(3):319–345.

Treier, S. & D.S. Hillygus. 2009. "The nature of political ideology in the contemporary electorate." Public Opinion Quarterly 73(4):679.

Treier, Sean & Simon Jackman. 2008. "Democracy as a latent variable." American Journal of Political Science 52(1):201–217.

Van der Linden, W.J. 2005. Linear models of optimal test design. Springer Verlag.

Tables and Figures

Table 1

Top Dimension 1 Items			Top Dimension 2 Items		
Item	**Beta 1**	**Beta 2**	**Item**	**Beta 1**	**Beta 2**
Gay marriage	-2.378	-1.694	Gay marriage (NPAT)	-2.378	-1.694
Recovery Act	-2.124	-0.015	Gay marriage amendment (CCES)	1.628	1.554
State gay marriage laws	-1.943	-1.465	State gay marriage	-1.943	-1.465
Climate change	-1.881	-0.268	Gay marriage (NAES)	-1.162	-1.427
Red. income inequality	-1.873	0.533	Abortion (NAES)	1.495	1.086
Gov. Should Provide Min. Standard of Living	-1.811	0.793	Gay marriage amendment (NAES)	0.879	0.964
Kagan nomination	-1.786	-0.289	Gov. should take over bad companies	-1.159	0.94
Labor organizing	-1.689	0.197	Stem cell funding	-1.462	-0.937
Stem cells	-1.657	-0.861	School prayer	0.613	0.93
Gov. effort on environm. initiatives	-1.642	0.034	Abortion (CCES)	1.102	0.923

Table 2

	Naïve Model	**1 Dimension**	**2 Dimensions**	**3 Dimensions**
Log Likelihood		-54884.64	-51548.08	-49489.65
PCP	66.0%	78.8%	80.2%	81.3%
GMP		0.645	0.665	0.675

Table 3: Effect of Question Choice on Relationship between Policy Preferences and Presidential Vote Choice

	NAES2000	NAES2004	NAES2008	CCES2006	CCES2008	CCES2010	Module
Δ I	0.38	0.37	0.43	0.38	0.40	0.34	0.30
Δ R	0.46	0.47	0.45	0.44	0.45	0.43	0.45
Preferences	0.37	0.38	0.31	0.42	0.36	0.47	0.48
Log likelihood	493	488	535	430	493	387	387
% Correct	0.91	0.91	0.90	0.93	0.91	0.93	0.93

This table summarizes 7 logistic regressions where the dependent variable is coded as 1 if the respondent voted Republican for President and 0 otherwise. Non-voters are dropped. The key independent variable is political preferences as measured using items that correspond to the surveys above each column. The first row is the difference in the predicted probability of voting Republican for respondents who identify as Independents rather than Democrats. The second row shows this difference for Republican respondents versus independents. The third row shows the predicted change in probability of voting Republican resulting from a one standard deviation increase in political preferences. The fourth row shows the log likelihood, and the fifth row shows the percent correctly predicted.

Table 4: Effect of Question Choice on Relationship between Policy Preferences and Presidential Vote Choice

	CCES Module	GI12
Δ I	0.30	0.33
Δ R	0.45	0.46
Preferences	0.48	0.45
Log likelihood	387	413
% Correct	0.93	0.93

This table summarizes 2 logistic regressions where the dependent variable is coded as 1 if the respondent voted Democrat for President and 0 otherwise. Non-voters are dropped. The key independent variable is political preferences as measured using items that correspond to the item sets above each column. The first row is the difference in the predicted probability of voting Democrat for respondents who identify as Independents rather than Republicans. The second row shows this difference for Democrat respondents versus independents. The third row shows the predicted change in probability of voting Democrat resulting from a one standard deviation increase in political preferences. The fourth row shows the log likelihood, and the fifth row shows the percent correctly predicted.

Table 5: Top 12 questions according to the maximum test information criteria

Source	Question
CCES 2010	Recovery act
CCES 2006, 2010	Climate change Opinion
CCES 2010	Kagan Nomination
NAES 2000	Gov.'s Environmental Protection Effort
Original	Regulation: Plants Should Reduce Emissions
CCES 2006	Immigration Reform
NAES 2000, 2004	Gov. Should Reduce Inequality
NPAT	Affirmative Action
NAES 2000, 2004	Universal Healthcare for Children
NAES 2004	Minimum wage
NPAT	Stem Cell Funding
NAES 2004	State Gay Marriage Laws

This table shows the top 12 questions according to the maximum test information criteria, in descending order of informativeness. The table excludes healthcare questions.

Table 6: Top 12 questions according to the maximum liberal information criteria

Source	Question
Original	Regulation: Plants Should Reduce Emissions
NPAT	Gov. Regulation of Private Sector
Original	Redist: Rich Pay More Taxes
CCES 2008	Carbon Tax
CCES 2008	Housing Assistance
NAES 2000	Gov. Should Restrict Abortion
Original	Redist: Gov. Should Guarantee Minimum Standard of Living
NPAT	Employment Discrimination Laws Should Protect Gays
NAES 2000, 2004, CCES 2010	Gun Control
NPAT	Employment Discrimination Laws Should Protect Women
NPAT	Workers Should be Able to Unionize
NAES 2008	Tax Policy

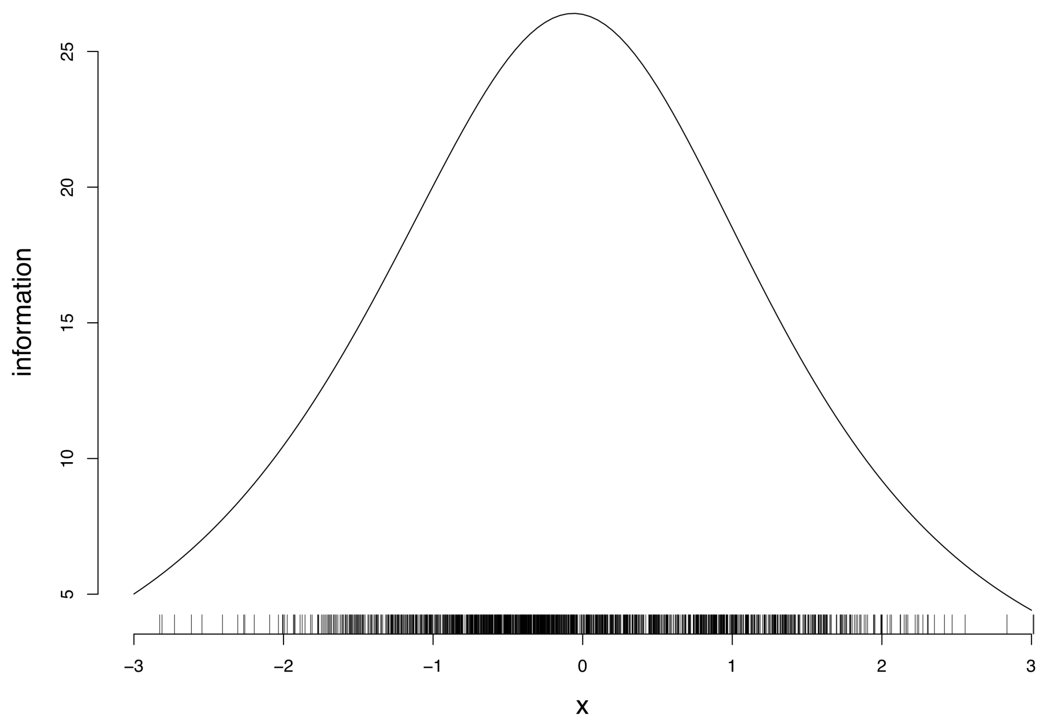
This table shows the top 12 questions according to the maximum liberal information criteria, in descending order of informativeness.

Table 7: Top 12 questions according to the maximum conservative information criteria

Source	**Question**
CCES 2006	Minimum Wage
NAES 2004	Labor Organizing
CCES 2010	Clean Energy Bill
CCES 2010	Financial Reform Bill
NAES 2004	Minimum Wage
CCES 2006, 2008, 2010	Stem Cell Funding
CCES 2010	Kagan Nomination
NAES 2004	Stem Cell Funding
NPAT	Gay Marriage
CCES 2006, 2010, NPAT	Climate Change Opinion
CCES 2006	Immigration Bill (Path to Citizenship)
Original	Eliminate EPA

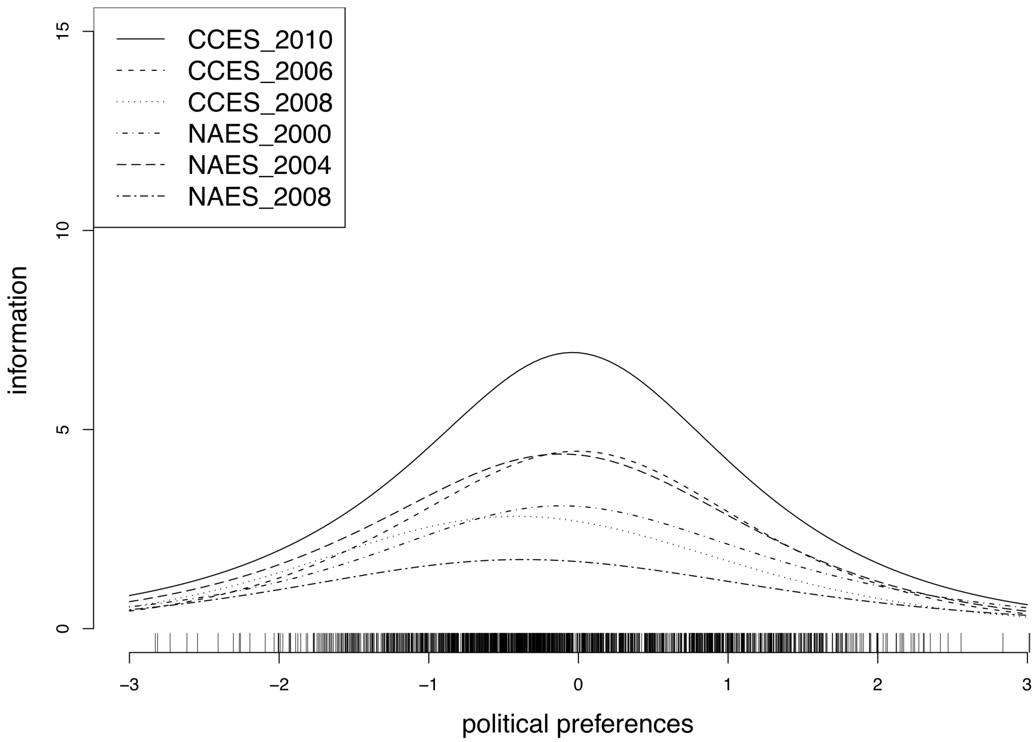
This table shows the top 12 questions according to the maximum conservative information criteria, in descending order of informativeness.

Figure 1: Test Information Function for full module



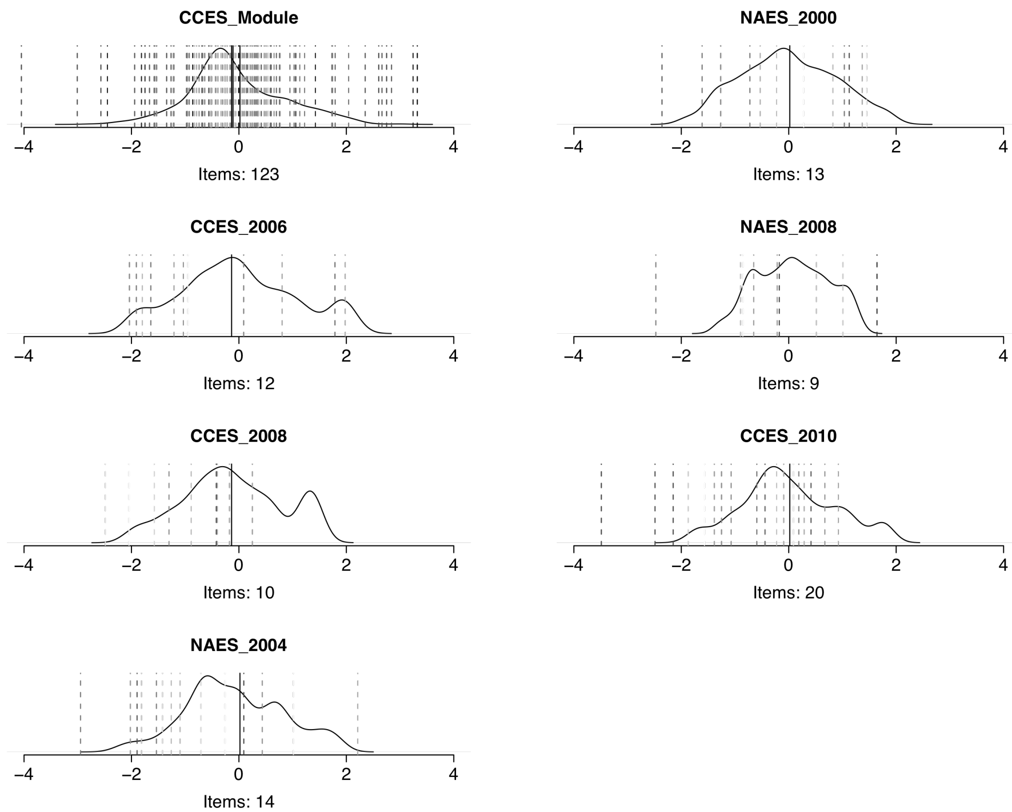
This graph shows information as a function of the latent policy preferences of the respondent. The rug at the bottom of the graph shows the distribution of the estimates.

Figure 2: Test Information Functions for estimates from models for item subsets



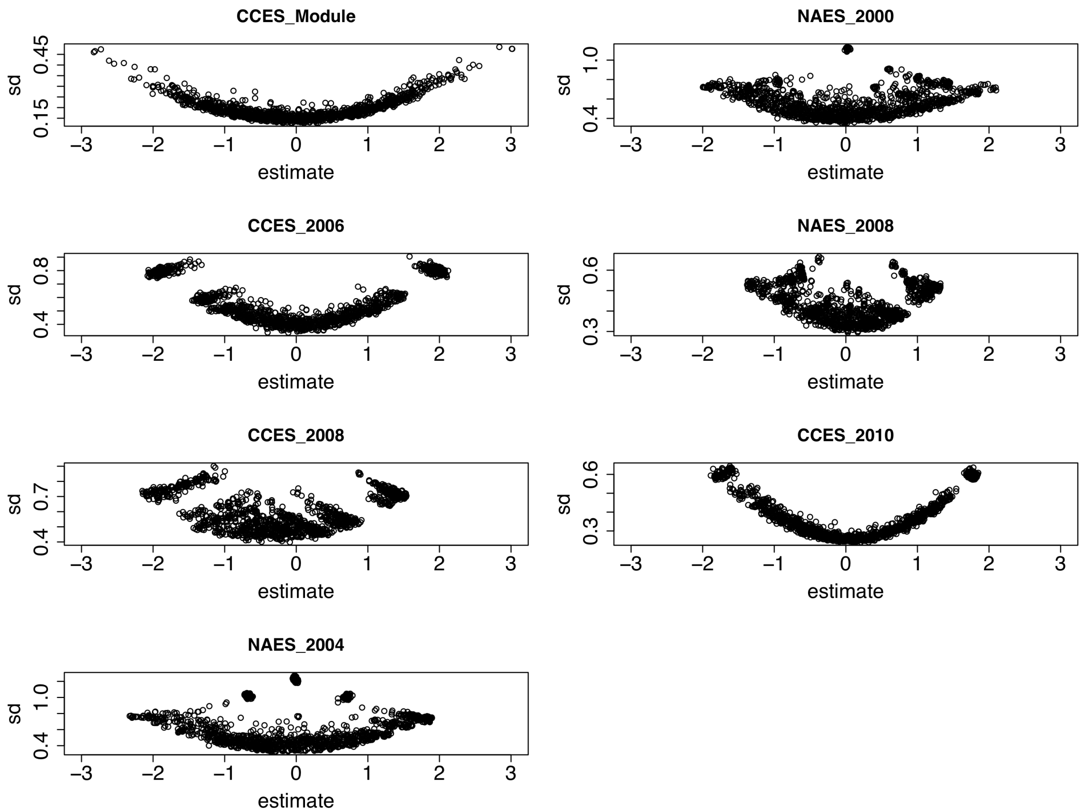
This graph shows the test information function for the item subsets that correspond to the items used on each of these large sample surveys. The rug at the bottom of the graph shows the distribution of the estimates.

Figure 3: Policy preferences



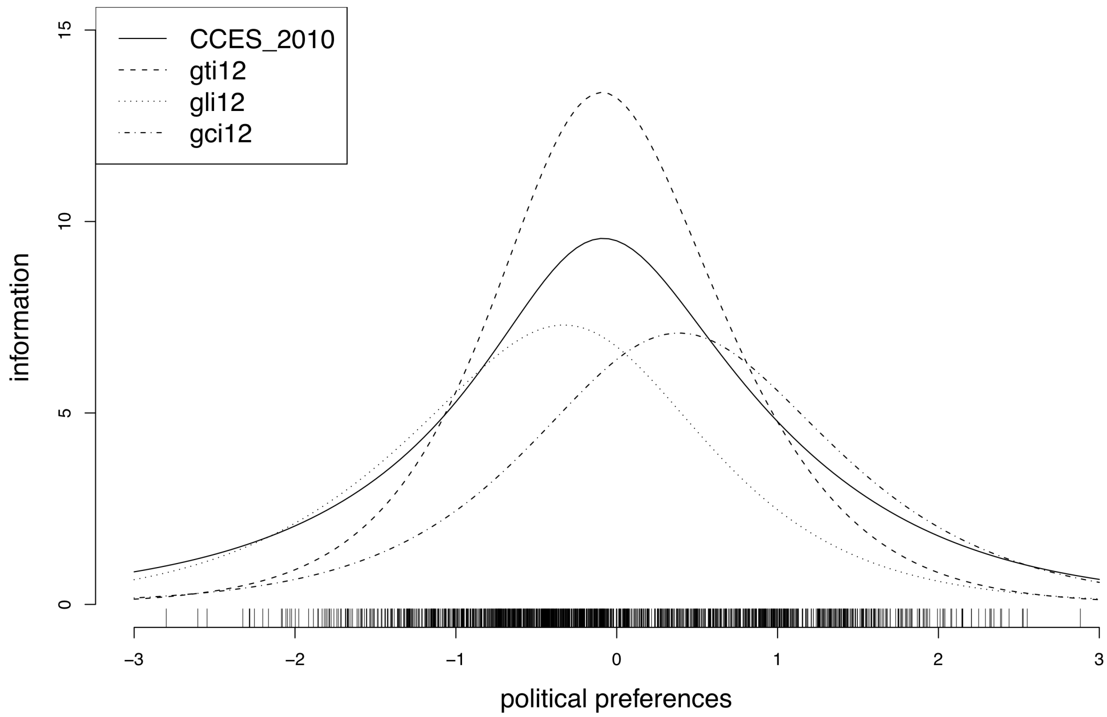
Densities of estimates resulting from scalings based on the item subsets corresponding to items on each of these surveys. The solid lines indicate the reference item used to identify the space. The parameters of each reference item are set to match the parameters of that item in the CCES Module estimation, which uses all questions asked. All non-reference items are represented by dashed line, with darkness in proportion to the magnitude of the discrimination parameter for that item.

Figure 4: Scatterplot of posterior standard deviation and policy preferences



These scatterplots show the posterior standard deviation and estimate of each person's policy preferences in each scaling. In a well-estimated space, there should be a quadratic relationship between the position of the estimate and its posterior standard deviation.

Figure 5: Information Functions for optimal subsets and the 2010 CCES.



This graph shows the test information function for the subset corresponding to the 2010 CCES as well as the test information functions corresponding to 6 "optimal" subsets. For instance, "gti12" corresponds to a 12-question set of items that is optimized to be informative across the parameter space. In the middle of the space, it discriminates better than the 20-question 2010 CCES set.

[1] Why are measures of citizens' policy preferences useful? We agree with Ansolabehere, Rodden and Snyder (2006) that it is useful to construct continuous measures of citizens' policy preferences because "it is difficult to get an overall sense of the distribution of preferences, or to analyze their impact on voting behavior, by sifting through a battery of individual questions." They continue: "Picking particular survey items leaves too much room for interpretation and manipulation. For example, the wording of questions related to abortion or homosexual rights may have a dramatic impact on the appearance of polarization. Another problem with analyzing individual survey questions is that many items are plagued with measurement error (Achen, 1975)... To address these problems, [it is useful to] aggregate as many questions as possible [to estimate citizens' policy preferences]."

[2] Our conception of citizens' policy preferences draws on a long literature in political behavior. Rodden and Snyder (2006, 2008) use the terms "issue preferences" and "policy preferences" to describe a factor score that summarizes a set of responses to policy questions. Treier and Hillygus (2009) use, alternatively, "policy attitudes," "ideology", "political ideology" and "policy preferences" to describe citizens' two-dimensional ideal points. Bafumi and Herron (2010) refer to citizens' "preferences" and "ideal points" more or less interchangeably. In an earlier work, Shafer and Claggett (1995) referred to "policy preferences" and "policy predispositions" within an "issue context." They capture the common definition of this concept, albeit loosely, when they say that it is "the 'deep preferences' to political opinion." All of this draws from initially from Converse (1964).

[3] We assume a quadratic utility function, as is common in both the spatial voting literature and the measurement literature. A convenient feature of the spatial model is that it leads directly to an estimable econometric model.

[4] The results will be indistinguishable from using the Normal model used by Clinton, Jackman, and Rivers (2004), but this model will be more mathematically convenient when it comes to choosing "optimal" items.

[5] Most of the questions used are dichotomous. Where they are not dichotomous, we choose a sensible cut off point and code all prior items as "yes" and all later items as "no."

[6] This "exploratory" approach is commonly used for models that estimate legislators ideal points in multiple dimensions (see Poole and Rosenthal 2007).

[7] For instance, Ansolabehere, Rodden, and Snyder (2006) factor analyze two different sets of questions, one of which consists of "economic" policy questions, and the other of which consists of "social" policy questions. They find a correlation between these scales of .28 for one survey and .04 on another. Treier and Hillygus (2009) use an IRT model to examine the structure of social and economic policy preferences. They find a correlation

of .3 between the two dimensions. However, they constrain their model much more than necessary, requiring most of the items to discriminate on only one dimension.

[8] The question asked respondents to classify their views on several items about abortion. A “none of the above" category was accidentally added, which lead some respondents to view the question as a “check one" question when in fact multiple boxes could be checked.

[9]

[10] The use of survey weights does not drive any of the results here, including the prominent features of the shape of the distribution. Using weights give a slightly more accurate picture of the population based on the sample.

[11] Because the data generating process that selects items in a survey is determined by the analyis, artificial bunching does not always occur in the extremes only. This is in contrast to the literature on measuring Congressional ideology, in which the agenda process is thought to produce artificial extremism (Snyder 1992).

[12] These TIFs exclude health care questions. With health care questions included, we can do even better, with the 6 item scaling approaching the informativeness of the CCES 2010 at 0.

[13] On our website, we have posted data on the informativeness of the questions used here in order to guide future research.