doi:10.1017/S0003055408080210

The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting

STEPHEN ANSOLABEHERE *Massachusetts Institute of Technology*
JONATHAN RODDEN *Stanford University*
JAMES M. SNYDER, JR. *Massachusetts Institute of Technology*

A venerable supposition of American survey research is that the vast majority of voters have incoherent and unstable preferences about political issues, which in turn have little impact on vote choice. We demonstrate that these findings are manifestations of measurement error associated with individual survey items. First, we show that averaging a large number of survey items on the same broadly defined issue area—for example, government involvement in the economy, or moral issues—eliminates a large amount of measurement error and reveals issue preferences that are well structured and stable. This stability increases steadily as the number of survey items increases and can approach that of party identification. Second, we show that once measurement error has been reduced through the use of multiple measures, issue preferences have much greater explanatory power in models of presidential vote choice, again approaching that of party identification.

Classic theories of democracy as well as contemporary theories of voting behavior and political representation hold that voters assess politicians on the basis of their positions on issues of the day (Downs 1957; Key 1966). Candidates and parties announce positions on issues in order to win votes, and voters choose the alternatives that best represent their interests on those issues. Legislators and executives who are out of step with their constituents are voted out of office (e.g., Erikson and Wright 1993; Canes-Wrone et al. 2002). These assumptions undergird centuries of democratic theory and decades of spatial modeling.

This portrait of electoral politics, however, runs counter to much of American survey research over the last six decades. Converse (1964) issued the most stunning and most frequently cited critique. He found that voters exhibit little consistency in their attitudes over time and little ideological constraint from one issue to the next. Accordingly, the conventional wisdom portrayed in most American politics textbooks is that the vast majority of American voters do not have coherent issue preferences or even attitudes (e.g., Fiorina and Peterson 1998, 144–52). In his comprehensive literature review, Kinder (1998, 796) wrote: "Precious few Americans make sophisticated use of political abstractions. Most are mystified by or at least indifferent to standard ideological concepts, and not many express consistently liberal, conservative, or centrist positions on government policy."

This paper argues that such conclusions are driven largely by measurement error associated with the analysis of individual survey items. We revisit the American National Election Study panels with a relatively simple, transparent, and old-fashioned approach to reducing measurement error. We show that issue scales composed of multiple measures are much more stable than individual survey items and, with enough questions, can even approach the stability of party identification.

Following from the apparent incoherence and instability of voters' preferences, the conventional wisdom also holds that preferences over public policy issues have little independent impact on citizens' voting decisions. Yet once we implement a very simple correction for measurement error, we show that issue preferences have a large effect on voting in U.S. presidential elections, again approaching that of party identification.

Most research using individual-level survey data over the past several decades echoes the findings of *The American Voter* (Campbell et al. 1960): voters rely on their party identification and impressions about candidate image when deciding how to vote, and ideology or opinions on specific policies play a modest role at best. Although there are certainly prominent exceptions in the literature (e.g., Green et al. 2002; Abramowitz and Saunders 1998), these findings are recounted in many textbooks on American politics and elections. For example, Polsby and Wildavsky (2000, 15–17) write: "By the time we get down to those who know and care about and can discriminate between party positions on issues, we usually have a small proportion of the electorate . . . So while candidates matter sometimes and issues matter sometimes, and both are capable of affecting who wins, for most voters party matters almost all the time."

Yet an undercurrent of survey research, at least since the 1980s, argues that it is possible to identify "core values" or "predispositions" that are coherent and stable (McClosky and Zaller 1984; Feldman 1988; Heath et al. 1994.) When issues are framed in terms of such core values, survey respondents readily make sense of the choices at hand (Sniderman and Piazza 1993). Further research shows that scaled values of survey responses have considerable traction in predicting party identification, candidate evaluations, and vote choice (e.g.,

Stephen Ansolabehere is Professor, Department of Political Science, Massachusetts Institute of Technology, Building E53, 77 Massachusetts Ave., Cambridge, MA 02139 (sda@mit.edu).

Jonathan Rodden is Associate Professor, Department of Political Science, Stanford University, 616 Serra Street, Encina Hall West, Room 444C, Stanford, CA 94305-6044 (jrodden@stanford.edu), to whom correspondence should be addressed.

James M. Snyder, Jr., is Professor, Departments of Political Science and Economics, Massachusetts Institute of Technology, Building E53, 77 Massachusetts Ave., Cambridge, MA 02139 (millett@mit.edu).

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

Carmines and Stimson 1980; Miller and Shanks 1996; Erikson and Tedin 2007). Adding to the puzzle, averaging across large numbers of individuals produces aggregate public opinion on issues that is quite stable over time and strongly associated with public policy outcomes (e.g., Page and Shapiro 1982; Erikson et al. 2002).

One natural explanation for this pattern of findings is that the responses to individual issue questions in surveys are plagued with large amounts of measurement error. Measurement error in surveys arises for a variety of reasons, including vague question wordings, vague response categories or categories that do not reflect the individual's actual attitude, inattentiveness on the part of the respondent, and even typographical errors (see Mosteller 1968). Achen (1975) showed that measurement error in survey items is sufficient to explain the low correlations of individuals' issue preferences over time and the apparent lack of constraint. The correlations observed by Converse are easily reconciled with a model in which there is a high degree of measurement error and a high degree of stability in preferences (see also Erikson 1978, 1979). Mysteriously, however, this is not a dominant idea, at least not in the American context. We think it should be.

In this paper, we show that there is a large amount of measurement error in the responses to typical survey questions on policy issues, and we suggest a simple method for reducing the effects of this error. Our approach uses multiple measures. First, multiple measures allow us to estimate the relative amounts of signal and noise in survey items. Second, constructing scores by averaging several items together—either by taking the simple arithmetic mean or by using factor analysis—yields much better estimates of respondents' underlying issue preferences. Averaging multiple items reduces the variance of the measurement error at roughly the rate of one over the number of items used. The scores can then be used to study the stability of latent preferences and the extent of issue voting.

This technique is widely used in political science to construct legislative roll-call voting scores, and has been used for over 70 years in psychometrics to construct test scores (e.g., Kuder and Richardson 1937; Lord and Novick 1968). Outside the American context, Evans, Heath, and Lallje (1996) used this approach to cast doubt on Butler and Stokes's (1974) claims about the instability and incoherence of political attitudes in Britain. Yet in spite of useful methodological overviews by McIver and Carmines (1981) and Spector (1992), multiple measures are rarely used in American survey research.

A number of studies of the American public have used LISREL to estimate structural models that explicitly incorporate measurement error and use multiple survey items to help identify latent variables (e.g., Judd and Milburn 1980; Jackson 1983; Norpoth and Lodge 1985; Hurwitz and Peffley 1987; Moskowitz and Jenkins 2004). Virtually all of these papers demonstrate that the amount of measurement error in most survey items is large, and after correcting for this, the latent variables are highly stable over time. Yet such studies—harshly criticized by Converse (1980), Luskin (1987), and others—may obscure more than they reveal. Confronted with complex structural models with many layers and parameters, skeptical readers see an unintelligible black box and are left with the impression that the findings have been manufactured by technique.

As a result, the simple idea that multiple measures can reduce measurement error seems not to have taken hold in American survey research, and the LISREL studies appear to have had little impact on textbooks or the main trend in academic research. First, Converse's conclusions about the instability of issue preferences are still widely accepted. Second, although other branches of social science have developed entire literatures using multiple indicators to measure crucial concepts like happiness, empirical studies in American political science too often turn a blind eye to measurement problems and rely heavily on "attitudes" or "preferences" measured with the response to a single survey item.

The goal of this paper is to set the record straight with a very simple and intuitive approach to measurement error. We show that issue scales can approach the coherence and stability of party identification, and can be almost as powerful in predicting voting behavior. This is the case whether we focus on general questions that tap into "predispositions," or limit the analysis to questions that address specific policies. The same is true for respondents at all levels of education or "sophistication," and in fact, measurement error is especially pronounced among the least educated respondents.

Thus, correcting for measurement error leads to a radically different picture of citizens' issue preferences than that presented by Converse (1964) and much subsequent research. Our results encourage a fresh look at the role of issues in U.S. elections, and a fresh look at multiple measures in American survey research. Once measurement error is reduced, we can firmly reject the notion that the American voter holds no coherent or stable attitudes, or that issue positions play only a sporadic cameo role in vote choice. These findings relieve much of the tension between survey research and spatial models of elections, and they blunt the normative concerns for democratic theory raised by Converse and *The American Voter*.

THE MEASUREMENT ERROR PROBLEM IN THEORY

Theoretical models of measurement errors in surveys treat responses to individual questions, or items, as consisting of the true attitude plus random error. The standard model assumes that measurement error is inherent in the instrument or question and has the same structure throughout the population. Measurement error, then, originates from vague or confusing questions asked in the survey. This is the approach taken by Achen (1975) and in most of the literature on measurement error in statistics. In this section we consider the consequences of such error for correlations used to measure stability of attitudes and for regressions used to estimate the effects of issues on voting preferences.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

A more subtle concern involves heterogeneity of measurement error, such as Converse's (1964) "Black-White" model in which some people answer survey questions without error and others answer essentially at random. We treat this issue more formally in a separate paper. The results here may be thought of as applying "within groups" in that case.

A Standard Model of Common Measurement Error

The standard model of measurement error assumes that there is a true underlying preference or belief and an additive random error in the response. We consider the effects of such measurement error in two settings: (1) correlations among survey questions and (2) correlations over time of a single survey item. These are the problems of *constraint* and *stability*. Mathematically these have the same structure, though in practice they may be quite different. For simplicity we develop the measurement error model from the perspective of constraint.

The model of true beliefs or preferences holds that there are n dimensions, $X_1, \ldots, X_n$, corresponding to n different issues. Let ρ_{X_i,X_j} be the correlation between any pair of issues i and j. Lack of ideological constraint may take two distinct forms. One manifestation is that people have preferences on each dimension but the issues are unrelated. If there is no ideological constraint, then any survey item that taps a given dimension will be uncorrelated with any item that taps a different dimension. A second notion holds that most people have no attitudes or opinions at all on most issues; their answers are just noise. Both notions imply that ρ_{X_i,X_j} will be near zero.

Contrast this depiction with a high degree of constraint: only one ideological dimension underlies most issues preferences. In this case, one can think of survey items as different alternatives along a single dimension or line, X. For example, X may be preferences on economic redistribution, and individual questions ask about particular redistributive policies, such as government guaranteed jobs, minimum wages, or the alternative minimum tax, which are points along the dimension. In this case the true correlation between any two issues i and j will be quite high; 1, if there is a single dimension.[1]

Survey questions used to measure voters' preferences on issues, however, are imperfect. They are subject to random error because of format and survey context, errors made by respondents, and so on. This measurement error biases correlations among true attitudes toward zero.

The standard population model incorporates measurement error as an additive term on each question or item. There may be many items for each issue or over time (e.g., Lord and Novick 1968; Wiley and Wiley 1970; Achen 1975). Let W_i be the observed response on item i, $1 = 1, 2$. Assume $W_i = X_i + e_i$, where X_i is the true attitude on issue i, and e_i is a random error term with $E[e_i] = 0$ and $Var(e_i) = E[e_i^2] = \sigma_{e_i}^2$. The measurement error in each item is assumed to be uncorrelated with the true value of the item itself, uncorrelated with the true value of the other item, and uncorrelated with the measurement error in the other item; that is, $E[X_i e_j] = 0$ for $i = 1, 2$ and $j = 1, 2$, and $E[e_1 e_2] = 0$. Finally, let $Var(W_i) = \sigma_{W_i}^2$ and $Var(X_i) = \sigma_{X_i}^2$ for $i = 1, 2$.

As is well known, the square of the correlation between W_1 and W_2 is biased toward zero relative to the square of the correlation between the true attitudes. The square of the correlation coefficient between items 1 and 2 is

$$\rho_{W_1,W_2}^2 = \rho_{X_1,X_2}^2 \frac{\sigma_{X_1}^2 \sigma_{X_2}^2}{\left(\sigma_{X_1}^2 + \sigma_{e_1}^2\right)\left(\sigma_{X_2}^2 + \sigma_{e_2}^2\right)} < \rho_{X_1,X_2}^2 .$$

The amount of bias depends on the variance in X's relative to the variance in e's; that is, the signal-to-noise ratio. Note that we have assumed no autocorrelation in the e's. If the errors are positively autocorrelated, and that correlation is larger than the correlation between X_1 and X_2, then there is an upward bias: $\rho_{W_1,W_2}^2 > \rho_{X_1,X_2}^2$.[2]

Applying this result to the problems of constraint and stability reveals that measurement error in individual survey items will lead to underestimation of the true correlations. Suppose the individual's true preference on issue i at time t is X_{it}. Under a strong version of ideological constraint, where there is a single latent dimension, $W_i = X + e_i$ for $i = 1, 2$. In this case, $\sigma_{X_1}^2 = \sigma_{X_2}^2 = \sigma_X^2$ and $\rho_{X_1,X_2} = 1$, so

$$\rho_{W_1,W_2}^2 = \frac{1}{\left(1 + \sigma_{e_1}^2/\sigma_X^2\right)\left(1 + \sigma_{e_2}^2/\sigma_X^2\right)} < 1 .$$

Under a strong version of stability, where true preferences and beliefs do not change at all over time, $W_t = X_t + e_t$ for each survey respondent. Consider two periods, $t = 1, 2$. If preferences are perfectly stable over time, then $W_t = X + e_t$, and again $\rho_{W_1,W_2}^2 < 1$.

Critiques of Converse (1964) by Achen (1975), Erikson (1975), Feldman (1989), and others followed this line of reasoning. Using the specification of Lord and Novick (1968) in which the parameters of the measurement error model in a three-wave panel are exactly identified under a set of assumptions, Achen finds that the true correlations in attitudes over time are plausibly as high as .7 or .8, rather than .2 to .45 as Converse found. Under alternative assumptions, Achen (1983) arrives at lower estimates of the true correlations for

[1] Following the literature on legislative roll-call voting, one can formalize issue voting using a spatial model. Assume there is a single issue dimension, as would arise under strong ideological constraint. Let X be the issue scale, θ be the individual's ideal policy, and Q be the status quo. The extent to which a survey respondent prefers any point along X to Q is the distance of the respondent's ideal point from the point X relative to the status quo; that is, $d = -(X - \theta)^2 + (Q - \theta)^2$. Hence, each question asks about a particular distance: $d_i = -(X_i - \theta)^2 + (Q - \theta)^2$. The distance can be thought of as the respondent's, true attitude, under the assumption of strong ideological constraint. We may write this, further, as $d_i = -X_i^2 + Q^2 + 2(X_i - Q)\theta$.

[2] See footnote 3 for the proof.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

these data. In a five-wave panel, Feldman (1989) finds correlations closer to those in Achen (1975). These papers deal only with stability over time rather than constraint. With the exception of Feldman these studies are highly dependent on modeling assumptions and do not have the degrees of freedom to test the validity of assumptions. The upshot of this research is that measurement error *may* produce substantial bias. The challenge is to reduce measurement error without relying on exact identification.

Using Multiple Measures to Reduce Measurement Error

Multiple survey items, combined either as a simple average or as a score using factor analysis, reduce measurement error at the rate of approximately $1/K$, where K is the number of items. Here we show the statistical gain from constructing scores in the context of the measurement error model and use these results in the next subsection to show how to estimate the measurement model.

There are two sets of items, each with K elements, $\{W_{11}, W_{12}, \ldots, W_{1K}\}$ and $\{W_{21}, W_{22}, \ldots, W_{2K}\}$. These might be K repeated questions in a panel at periods 1 and 2, or two distinct sets of questions in the same survey. Suppose each item in the first set taps issue 1 and each item in the second set taps issue 2. That is, $W_{ik} = X_i + e_{ik}$, for each $i = 1, 2$ and $k = 1, \ldots, K$. X_i is the true attitude on issue i, and e_{ik} is a random error term with $E[e_{ik}] = 0$ and $Var(e_{ik}) = E[e_{ik}^2] = \sigma_{e_{ik}}^2$. As above, assume that the measurement error in each item is uncorrelated with the true value of the item itself, uncorrelated with the true values of all other items, and uncorrelated with the measurement error in all other items; i.e., $E[X_i e_{jk}] = 0$ for $i = 1, 2, j = 1, 2$, and $k = 1, \ldots, K$, and $E[e_{ik}e_{jl}] = 0$ unless $i = j$ and $k = l$. Finally, let $Var(X_i) = \sigma_{X_i}^2$ for $i = 1, 2$.

Consider the variables made by averaging the individual items:

$$\overline{W}_1 = \frac{1}{K}\sum_{k=1}^{K} W_{1k} \quad \text{and} \quad \overline{W}_2 = \frac{1}{K}\sum_{k=1}^{K} W_{2k}\,.$$

Let $\overline{\sigma}_{e_1}^2 = \frac{1}{K}\sum_{k=1}^{K}\sigma_{e_{1k}}^2$ and $\overline{\sigma}_{e_2}^2 = \frac{1}{K}\sum_{k=1}^{K}\sigma_{e_{2k}}^2$ be the average measurement error variance among the items in sets 1 and 2, respectively. It is straightforward to show that:

$$\sigma_{\overline{W}_1}^2 = \sigma_{X_1}^2 + \overline{\sigma}_{e_1}^2/K \qquad \textbf{(1)}$$

$$\sigma_{\overline{W}_2}^2 = \sigma_{X_2}^2 + \overline{\sigma}_{e_2}^2/K \qquad \textbf{(2)}$$

$$\rho_{\overline{W}_1,\overline{W}_2}^2 = \rho_{X_1,X_2}^2 \frac{\sigma_{X_1}^2\sigma_{X_2}^2}{\left[\sigma_{X_1}^2 + \left(\overline{\sigma}_{e_1}^2/K\right)\right]\left[\sigma_{X_2}^2 + \left(\overline{\sigma}_{e_2}^2/K\right)\right]} \qquad \textbf{(3)}$$

$$Cov(\overline{W}_1, \overline{W}_2) = Cov(X_1, X_2) \qquad \textbf{(4)}$$

As K becomes large, $\overline{\sigma}_{e_1}^2/K$ and $\overline{\sigma}_{e_2}^2/K$ become small, and $\rho_{\overline{W}_1,\overline{W}_2}^2$ increases towards ρ_{X_1,X_2}^2. If $K = 10$, then the contribution of measurement error to the correlation coefficient is roughly one tenth as large as when a single item is used. Also, for each $i = 1, 2$, $\sigma_{\overline{W}_i}^2$ decreases toward $\sigma_{X_i}^2$ as K becomes large.

Note that we have assumed no autocorrelation in the errors; that is, $E[e_{1k}e_{2k}] = 0$ for all k. Autocorrelation introduces an additional source of bias. However, it is still the case that $\rho_{\overline{W}_1,\overline{W}_2}^2$ tends toward ρ_{X_1,X_2}^2 as K increases.[3]

Using these results we can immediately address two design considerations that bear on the number of items used to construct the scales.

(1) How closely will the correlations using K items approximate the true correlations? The answer depends on the ratio of signal to noise in the items. Clearly, when there is less measurement error in each item, fewer items are required to achieve the same approximation to the true values. Achen's analysis suggests that the variance in the measurement error in the ANES is approximately equal to the variance of the true issue scales. If that is approximately right, then the true correlation is approximately equal to $(K+1)/K$ times the observed correlation (between the average measures).

(2) What is the value of adding another item to a scale? For large values of K, one should add another item if the variance of the measurement error in the item is less than twice the average measurement error in the existing set of items. The objective is to minimize mean squared error, that is, to minimize $\mathrm{E}[\overline{W}_i - X_i]^2 = \overline{\sigma}_{e_i}^2/K = (\sum_{k=1}^{K}\sigma_{e_{ik}}^2)/K^2$. Suppose we begin with K items and consider adding item $K+1$. Equations (1) and (2) also reveal conditions under which adding additional items improves matters and when it does not. Adding item $K+1$ changes the maximand to $(\sigma_{e_{i,k+1}}^2 + \sum_{k=1}^{K}\sigma_{e_{ik}}^2)/(K+1)^2$. This is less than $(\sum_{k=1}^{K}\sigma_{e_{ik}}^2)/K^2$ if and only if $\sigma_{e_{i,k+1}}^2 < [(2K+1)/K]\,\overline{\sigma}_{e_i}^2$. Note that for large K this reduces to the simple rule of thumb: Add an item if and only if the variance of the measurement error in that item is less than twice the measurement error of the existing items.

Even though the results have the flavor of the sampling theory, the reduction in measurement error does not necessarily occur uniformly. Adding more variables to a scale will tend to make for a better measure, but it is possible to make a scale worse by adding a variable that has an extremely high amount of measurement

[3] Suppose $E[e_{1k}e_{2k}] \neq 0$ for some $k = 1, \ldots, K$, and let $\overline{\gamma}_e = \frac{1}{K}\sum_{k=1}^{K} E[e_{1k}e_{2k}]$. Then $Cov(\overline{W}_1, \overline{W}_2) = Cov(X_1, X_2) + \overline{\gamma}_e/K$. Also, for simplicity assume $\sigma_{X_1}^2 = \sigma_{X_2}^2 = \sigma_X^2$, and $\overline{\sigma}_{e_1}^2 = \overline{\sigma}_{e_2}^2 = \overline{\sigma}_e^2$. Then $\rho_{\overline{W}_1,\overline{W}_2} = [Cov(X_1, X_2) + \overline{\gamma}_e/K]/[\sigma_X^2 + \overline{\sigma}_e^2/K]$. Note that $\rho_{\overline{W}_1,\overline{W}_2} > \rho_{X_1,X_2}$ if and only if $\overline{\gamma}_e/\overline{\sigma}_e^2 > \rho_{X_1,X_2}$ (i.e., the "average" autocorrelation in the errors exceeds the correlation in the underlying dimension). If this condition holds, then $\rho_{\overline{W}_1,\overline{W}_2}$ decreases in K to ρ_{X_1,X_2}, and if $\overline{\gamma}_e/\overline{\sigma}_e^2 < \rho_{X_1,X_2}$ then $\rho_{\overline{W}_1,\overline{W}_2}$ increases in K to ρ_{X_1,X_2}. There is some evidence of autocorrelation in the ANES panels—we find that $Cov(\overline{W}_1, \overline{W}_2)$ decreases with K. It is consistent with the second case since, as shown below, $\rho_{\overline{W}_1,\overline{W}_2}$ increases with K.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

error. Improvement in the measures depends on the quality of measures as well as their number.

Estimating Parameters of the Measurement Error Model

It is straightforward to estimate the main parameters of interest in the model above. We will focus on the case where $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2_X$ and $\sigma^2_{e_{1k}} = \sigma^2_{e_{2k}} = \sigma^2_{e_k}$ for all $k = 1, \ldots, K$. Although the model can be estimated under more general conditions, this case is sufficient for the purposes of our paper.

Note first that $\overline{\sigma}^2_{e_1} = \overline{\sigma}^2_{e_2} = \overline{\sigma}^2_e$, and equation (3) can then be written as

$$1/\rho_{\overline{W}_1,\overline{W}_2} = [1/\rho_{X_1,X_2}]\left[1 + \left(\overline{\sigma}^2_e/\sigma^2_X\right)(1/K)\right] \quad \textbf{(5)}$$

Next, note that the intertemporal correlation for each individual item k can be written as

$$1/\rho_{W_{1k},W_{2k}} = [1/\rho_{X_1,X_2}]\left[1 + \left(\sigma^2_{e_k}/\sigma^2_X\right)\right], \quad k = 1, \ldots, K. \quad \textbf{(6)}$$

Letting $\overline{1/\rho_{W_1,W_2}} = \frac{1}{K}\sum_{k=1}^{K}(1/\rho_{W_{1k},W_{2k}})$ and averaging over k in (6) yields

$$\overline{1/\rho_{W_1,W_2}} = [1/\rho_{X_1,X_2}]\left[1 + \left(\overline{\sigma}^2_e/\sigma^2_X\right)\right]. \quad \textbf{(7)}$$

So, solving (5) and (7) we have

$$\sigma^2_X/\overline{\sigma}^2_e = \frac{1/\rho_{\overline{W}_1,\overline{W}_2} - \overline{1/\rho_{W_1,W_2}}/K}{\overline{1/\rho_{W_1,W_2}} - 1/\rho_{\overline{W}_1,\overline{W}_2}} \quad \textbf{(8)}$$

$$\rho_{X_1,X_2} = \frac{K-1}{K/\rho_{\overline{W}_1,\overline{W}_2} - \overline{1/\rho_{W_1,W_2}}}. \quad \textbf{(9)}$$

Equation (8) gives a measure of the average signal to noise ratio, and (9) gives the true correlation between issues 1 and 2. These are both easily computed from the underlying correlations of the individual items and the variables made by averaging.[4]

MEASUREMENT ERROR AND PREFERENCE STABILITY

In this section we show that measurement error is a severe problem in individual survey items, but taking multiple items and averaging can dramatically reduces this error. The resulting scales based on multiple measures are highly stable over-time, and scales based on subsets of enough questions are highly correlated within survey. In many cases the over-time correlations are as high as the over-time correlation for the question addressing party identification.

To get a sense of the problem of measurement error, consider, for example, the abortion issue. The National Election Study and other surveys usually measure opinions on this complicated issue with just a single item. A standard form of the question asks:

There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view on this issue? (1) By law, abortion should never be permitted. (2) The law should permit abortion only in case of rape, incest or when the woman's life is in danger. (3) The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established. (4) By law, a woman should always be able to obtain an abortion as a matter of personal choice. (5) Other.

A critical reading of this question reveals considerable ambiguity. The categories do not fully cover all possible responses along the given dimension. One can imagine cases between (1) and (2) or between (3) and (4). Some of the categories, especially the third, may encompass a wide range of opinions. Correlation of this item over successive waves of surveys shows that responses to some of the categories (especially 2 and 3) are very unstable while others are not, indicative of measurement problems caused by question wording.[5] In addition, the categories leave out other very important dimensions to the issue, such as the time during the pregnancy, or the permissibility of in vitro fertilization or embryonic stem cell research. They force respondents into a choice that only partly reflects actual opinion. A stunning example is that 30% of those who chose option 4 also support a ban on partial-birth abortion. A single question seems insufficient to capture the range of preferences on this complicated issue. However, if we add together the several items on abortion (including general abortion attitudes, stem cell research, and partial birth abortion) the resulting score is much more stable over time than the individual items.

Data Sources and Factor Analysis Results

To establish this more generally, we use the 1956–1960, 1972–1976, 1990–1992, and 1992–1996 panel data sets from the National Election Studies. We study the first and last years of each panel, because these typically contain the largest batteries of repeated issue questions. We selected all available questions on public

[4] Another approach is to use information in the variances as well as the correlations. In fact, this information must be used in the more general case where $\sigma^2_{X_i}$ and $\sigma^2_{e_{ik}}$ may vary across i. Consider this more general case. The variances for the individual items are $\sigma^2_{W_{ik}} = \sigma^2_{X_i} + \sigma^2_{e_{ik}}$ for $i = 1, 2$ and $k = 1, \ldots, K$. Let $\overline{\sigma}^2_{W_i} = \frac{1}{K}\sum_{k=1}^{K}\sigma^2_{W_{ik}}$ for $i = 1, 2$. Averaging over k then yields $\overline{\sigma}^2_{W_i} = \sigma^2_{X_i} + \overline{\sigma}^2_{e_i}$ for $i = 1, 2$. Combining these with equations (1) and (2) and rearranging, we have $\sigma^2_{X_i} = (K\sigma^2_{\overline{W}_i} - \overline{\sigma}^2_{W_i})/(K-1)$ and $\overline{\sigma}^2_{e_i} = (\overline{\sigma}^2_{W_i} - \sigma^2_{\overline{W}_i})K/(K-1)$ for $i = 1, 2$. These equations provide us with average signal to noise ratios, $\sigma^2_{X_1}/\overline{\sigma}^2_{e_1}$ and $\sigma^2_{X_2}/\overline{\sigma}^2_{e_2}$. Also, substituting them into (3) and rearranging yields an estimate of the true correlation between issues 1 and 2: $\rho_{X_1,X_2} = \rho_{W_1,W_2}\sqrt{[1+(\overline{\sigma}^2_{W_1} - \sigma^2_{\overline{W}_1})\,/\,(K\sigma^2_{\overline{W}_1} - \overline{\sigma}^2_{W_1})][1+(\overline{\sigma}^2_{W_2} - \sigma^2_{\overline{W}_2})/(K\sigma^2_{\overline{W}_2} - \overline{\sigma}^2_{W_2})]}$. Note finally that the model is evidently overidentified and therefore may be tested.

[5] Compare, for instance, the question in the 2006 and 2007 Waves of the Cooperative Congressional Election Study. Only 34% of those who chose option 3 on the question in the 2006 wave chose it again in the 2007 wave. By comparison, 86% of those who chose option 4 in 2006 chose it again in 2007 and 70% of those who chose option 1 in 2006 chose it in 2007.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

policy issues, plus a few "feeling thermometer" items for particular groups—labor, big business, poor people, welfare recipients, and women's groups—to increase the number of items available for scaling. We grouped these according to issue area.

We consider all issue areas for which we have at least 10 survey items that are repeated in both years of the panel. In addition, because of its key role in Converse's original article and the early debate, we analyze one issue area from the 1956–60 panel—*Economic Issues*—even though we only have four repeated survey items for it. We also include *Law and Order Issues* from the 1972–76 panel, even though we have only six repeated items, to increase the variety of issue areas covered in our study. The items used are shown in the Appendix.

We scaled the items using principal factors factor analysis. In all cases we find a single dominant dimension. The eigenvalues and factor loadings for the 1992 and 1996 economic and moral issues are shown in Table A.1 in the Appendix. (The full set of results takes several pages to present and is available from the authors on request.) We then computed the factor scores for the first factor, and use these scores as our issue scales.[6]

There is nothing magical about the factor scores. In fact, up to an affine transformation they are almost exactly what we get by simply averaging the survey items. For each issue, we oriented each survey item so that higher scores mean more "conservative" positions on the issue, standardized them to have mean 0 and variance 1, and then took the simple unweighted average. The correlation between the simple averages and the factor scores are all .97 or higher, and 13 out of the 18 correlations are .99 or higher. This is not surprising, because the factor loadings are roughly similar across most survey items (see Table A.1). Also, we get essentially identical results if we normalize each item by setting the minimum value to 0 and the maximum value to 1 rather than standardizing.

We use factor analysis in order to separate the first dimension of preferences from any higher dimensions. Throughout we assume a one-dimensional model of preferences. Comparing results using the first factor and the simple average of individual items reveals that the two approaches are nearly identical. This further suggests that, at least in the ANES data, preferences on each issue are mainly one-dimensional.

Intertemporal Stability of Issue Scales

Analysis of issue scales, rather than single survey items, reveals a very high degree of stability in issue preferences. (1) Issue scales are *much* more highly correlated over time than are individual items. (2) With enough questions, some issue scales are *as stable as party identification*. (3) Adding more questions the scales become more stable over time, in a manner consistent with the simple measurement error model analyzed above.

Following Converse, we take simple correlations among scores over time or within surveys to gauge the degree of stability and constraint. For each issue area, we construct the factor scores in the first and last years of the panel. We then compute the correlation coefficient between the scores in the 2 years. These correlations are in columns 4 and 5 of Table 1. In column 5 we construct each issue scale using only the survey items that appear in both years of the panel. In column 4 we construct the issue scales using all available questions in each year—some of these appear in both years and some appear in just 1 year. The results are almost identical in both columns, so neither choice appears clearly superior.

The intertemporal correlations of the issue scales are quite high: the average correlation is .77 for the five issue areas of 1990–92 and 1992–96. This is much higher than the average correlation among individual survey items, which is just .46. No issue scale has a stability coefficient below .63.

General questions on ideology also exhibit more stability when scaled, even with just three items.[7] Single-item ideology measures have intertemporal correlations of approximately .55, whereas scales constructed from those items have intertemporal correlations in the neighborhood of .7 to .75. Providing further evidence of measurement error, the 7-point ideology questions are more stable than the 3-point ideology questions.

Perhaps most striking, the issues scales exhibit a degree of intertemporal stability on par with party identification. The average intertemporal correlation is .77 among the issue scales and .79 among party identification. None of the issue scales in 1972–72, 1990–92, or 1992–96 panels are markedly less stable than partisanship. This is an important result, because the relatively high level of stability found for party identification is often taken as evidence that party identification is something real, solid, and meaningful in public opinion.

Party identification may be measured with error as well, and the intertemporal stability of "true" party identification is probably even higher than .8 (Green and Palmquist, 1994; Green, Palmquist, and Schickler 2002). A three-item partisanship scale, however, does no better than the single-item scale. This suggests that the extra two items—feeling thermometers for Democrats and Republicans—have considerably more measurement error than the 7-point party identification scale.

Comparing the different studies, we see that the earlier panels produced scales with lower intertemporal correlations than later panels. This might reflect the rise of issue voting in the mid-1970s. We suspect it arises from the small number of items in the earlier panels.

Figure 1 shows the intertemporal stability of a scale as a function of the number of items in a scale. The analysis is of the 1990 and 1992 *Economic Issues* scales. We performed a Monte Carlo simulation in which we randomly chose an appropriate number of items (from

[6] For each issue area we used all respondents who answered at least 75% of the associated survey items, and imputed values for missing responses, via best-subset linear imputation.

[7] The multi-item Ideology scale is constructed from the self-placed 7-point item, plus two feeling thermometer questions, one for liberals and one for conservatives.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

TABLE 1. Correlations Between ANES Panel Waves, Issue Scales, and Individual Survey Items

Issue Area	Total Items Year 1	Total Items Year 2	Total Items in Common	Correlations Between: Scales Made with All Items	Correlations Between: Scales Made with Common Items	Correlations Between: Individual Common Items (Average)
1992, 1996						
Economic Issues	32	34	25	.77	.76	.42
Moral Issues	18	14	12	.84	.83	.52
Ideology (multi-item)	3	3	3		.74	.57
Ideology (1 item)	1	1	1			.57
Party ID (multi-item)	3	3	3		.80	.67
Party ID (1 item)	1	1	1			.79
1990, 1992						
Economic Issues	27	33	23	.78	.76	.41
Racial Issues	11	12	11	.77	.77	.51
Moral Issues	13	18	10	.69	.71	.45
Ideology (multi-item)	3	3	3		.69	.52
Ideology (1 item)	1	1	1			.51
Party ID (multi-item)	3	3	3		.79	.54
Party ID (1 item)	1	1	1			.79
1972, 1976						
Economic Issues	16	21	11	.65	.67	.42
Racial Issues	22	16	15	.76	.76	.42
Womens' Issues	20	12	10	.64	.65	.41
Law & Order	9	11	6	.71	.73	.52
Ideology (multi-item)	4	4	4		.69	.49
Ideology (1 item)	1	1	1			.56
Party ID (multi-item)	3	3	3		.75	.55
Party ID (1 item)	1	1	1			.76
1956, 1960						
Economic Issues	6	4	4	.62	.63	.43
Party ID (1 item)	1	1	1			.83

Note: The number of observations (N) varies slightly from cell to cell due to missing data. For the Economic Issues scales made using all common items, $N = 534$ in the 1992–96 panel; $N = 607$ in 1990–92; $N = 971$ in 1972–76; and $N = 953$ in 1956–60.

1 up to all 23), constructed the scales, and computed the intertemporal correlations of the scales.[8] The plot displays the median intertemporal correlations plotted against K, together with the interquartile range, in a box-and-whiskers plot.

The intertemporal stability of the issue scales rises smoothly and concavely with the number of survey items used. In this particular case (Economic Issues in 1990–92), the typical single item exhibits stability of .41, far below the actual correlation. That gap shrinks as the number of items increases, and at 23 items, the intertemporal correlation averages .76, close to its upper limit of .79.

The pattern in Figure 1 allows us to put the 1952–56 panel in context. That panel had only 4 items in common across all waves and a stability coefficient on the 4-item scale of .63. However, using 4 randomly chosen items in the 1990–92 economic issues to construct a scale produced a typical correlation of .57. The 1952–56 panel, then, shows as much stability as the later panels, but that study simply included too few questions to reduce the measurement error fully. The 1972–76 panel had 11 items concerning economic issues; the level of stability in that study is consistent with the graph as well and suggests that a larger number of items would produce even more stable measures of issue preferences.

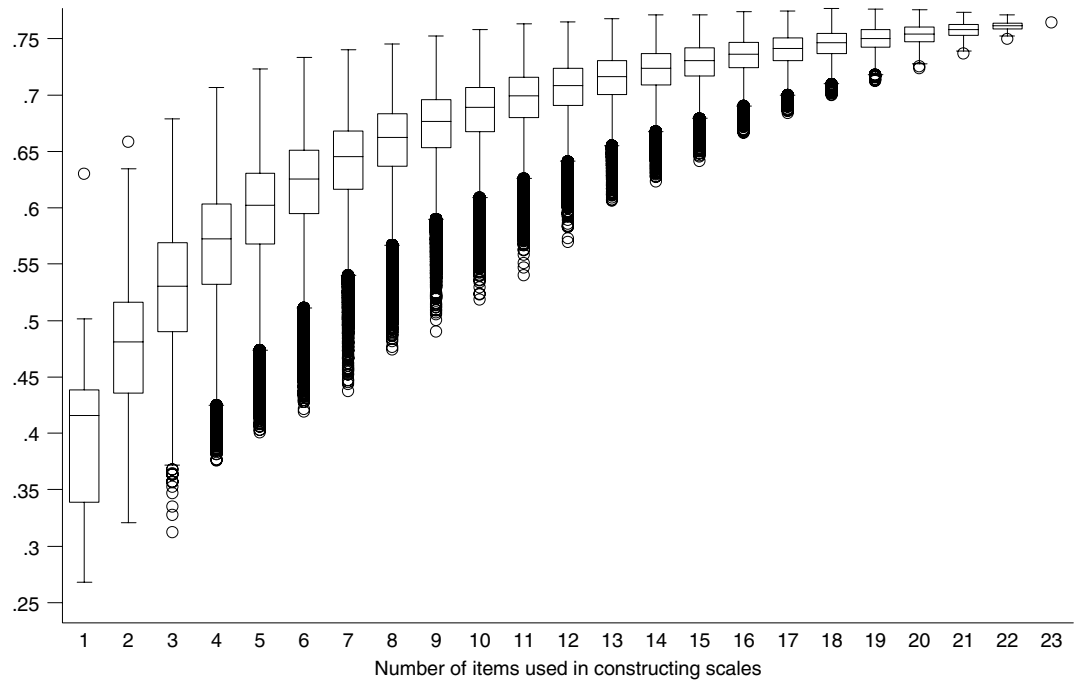
Equation (8) allows us to use the figures in the graph to calculate the amount of measurement error in the typical survey item. Recall that equation (8) gives an estimate of the average signal to noise ratio in the individual survey items, and the intertemporal stability of the underlying preferences.[9] Applying this to the items

[8] For each value of $K = 1, \ldots, 23$, there are $\binom{23}{K}$ distinct subsets of K survey items taken from the 23 available items. For each subset $\mathbf{S}$ of K survey items, we can construct a scale for 1990 and a scale for 1992, then calculate the correlation $\hat{\rho}_{\mathbf{S}}$ between these scales. Averaging over all of the $\binom{23}{K}$ correlations produces a measure of the average degree of intertemporal stability of scales constructed with K items. Since $\sum_{K=1}^{23}\binom{23}{K}$ is a large number (8,388,584), we used Monte Carlo techniques that iteratively drew 500,000 subsets of questions, each time drawing a subset $\mathbf{S}$ of survey items and constructing $\hat{\rho}_{\mathbf{S}}$. We mimic the frequency distribution given by the function $\binom{23}{K}$, but oversample in the tails to ensure coverage of all subsets for $K = 1, 2, 3, 4, 19, 20, 21, 22, 23$ (since there are fewer distinct subsets for these values of K). We discarded duplicates and then averaged the $\hat{\rho}_{\mathbf{S}}$ for each value of K.

[9] One caveat should be mentioned. Factor analysis produced the scales, but equations (8)–(9) are derived using simple averages. As

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

FIGURE 1. Correlation Between 1990 and 1992 Economic Issue Scales Box-and-Whiskers Plot



in the 1990 and 1992 *Economic Issues*, the average signal to noise ratio is about .95. This implies that on average about half of the variance in individual survey items is from actual issue positions and half is from measurement error. This is approximately what Achen (1975) found for the 1958 to 1960 panel using a very different methodology.[10]

The high degree of intertemporal stability of the scales, displayed in Figure 1, need not have emerged. Little relationship between the number of items and the correlations of the scales would have emerged if either (1) responses to the questions were primarily noise or had little stability, or (2) there were many dimensions underlying the data. A high degree of heterogeneity among the respondents, as with Converse's Black-White model discussed below, would also have prevented the correlations from approaching the true correlations in the asymptotes, because a large fraction of people would have 0 correlations in their answers. Rather correlations increase monotonically in the number of items in the scale; this holds true for the entire sample and for high and low sophistication people.

noted previously, however, the factor scores follow the scores made from simple averages quite closely. Not surprisingly, the estimates are almost identical when we use scales made from simple averages rather than the factor scores.

[10] Using the equations outlined in footnote 3 implies there is even more measurement error in the individual items. We suspect the differences are due to autocorrelation in the error terms of the individual items, as discussed in footnote 3. We will pursue this in future work.

Within-Survey Stability of Issue Scales

We can also examine the within-survey, cross-sectional stability of issue scales. This is what Converse (1964) and others call "constraint." Consider a given issue area in a given year, and suppose the total number of associated survey items is M, with M even. We can divide these items into two disjoint subsets, each with $M/2$ items, then scale each subset separately and calculate the correlation between these scales. We calculate the average correlation, because there are many possible partitions of the M items.[11]

The results of this analysis are shown in Table 2, which includes all issue areas and years for which we have 14 or more total survey items (so that each scale is made with at least 7 items). As expected given the pattern of eigenvalues shown in Table A.1, the correlations between the scales are quite high—in most cases higher than .70. Moreover, they are much larger than the average correlations among the individual items

[11] Specifically, divide the items into sets $\mathbf{S}$ and $\mathbf{S}'$, each containing $K = M/2$ items. Scale each subset separately, and calculate the correlation between the resulting scales, $\hat{\rho}_{\mathbf{SS}'}$. There are $\binom{M}{M/2}$ distinct subsets of $M/2$ items. Averaging over all of these subsets produces a measure of the average cross-sectional stability of scales made with $M/2$ items. Again, since $\binom{M}{M/2}$ is large when M is, and enumerating all possible subsets with $M/2$ items is difficult, we use Monte Carlo techniques to estimate this average. We iterated through 500,000 loops, each time drawing a subset $\mathbf{S}$ of survey items and its complement $\mathbf{S}'$ and constructing $\hat{\rho}_{\mathbf{SS}'}$. We discarded duplicate subsets and then averaged the $\hat{\rho}_{\mathbf{SS}'}$. If the number of items is odd, then $\mathbf{S}$ contains $(M-1)/2$ items and $\mathbf{S}'$ contains $(M+1)/2$ items.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

TABLE 2. Correlations Between Issue Scales of Disjoint Subsets of Half of the Items

Issue Area	Total Items	Average Correlations: Issue Scales	Average Correlations: Pairs of Items
ANES			
Economic Issues, 1996	34	.84	.24
Moral Issues, 1996	14	.65	.25
Economic Issues, 1992	32	.79	.21
Moral Issues, 1992	18	.74	.23
Economic Issues, 1990	27	.74	.19
Economic Issues, 1976	21	.68	.18
Racial Issues, 1976	18	.75	.27
Economic Issues, 1972	16	.61	.16
Racial Issues, 1972	22	.80	.27
Womens' Issues, 1972	20	.73	.20

that went into the construction of the scales, which range from .16 to .27.

Figure 2 shows how the correlations depend on the number of items scaled, analogous to Figure 1. The figure examines the 1996 ANES economic issues. There are $M = 34$ relevant items in the survey, which allows us to study the stability of scales made from up to 17 items, $K = 1, \ldots, 17$. For each value of K, we implement the following procedure. Draw two disjoint subsets, each of size K, from the 34 questions, scale each subset separately, and calculate the correlation between the resulting scales. We calculate the average correlation, since there are many possible disjoint subsets for each value of K. As in the construction of Figure 1 and Table 2, we use Monte Carlo techniques to construct our estimates.

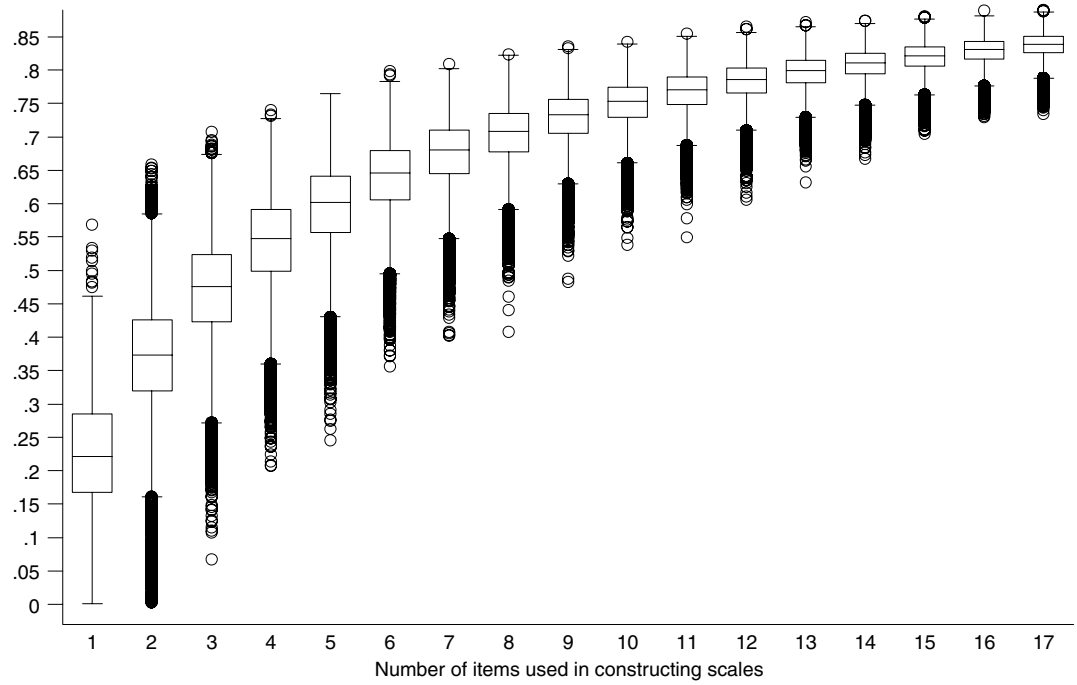
Figure 2 shows the resulting median correlations plotted against K, together with the inter-quartile range, in a box-and-whiskers plot. As in Figure 1, the average degree of stability (constraint) rises smoothly and concavely as the number of survey items increases. With 17 questions the average correlation is .84, fairly close to its theoretical upper limit of 1.00.

The analysis reflected in Figures 1 and 2 suggests the enormous amount of measurement error in single survey items used to capture issue preferences. The analysis points to a specific solution as well. Survey researchers should use multiple items—and the more the better.

Policy Positions versus Policy-Related Predispositions

As noted in the introduction, numerous scholars have argued that while voters may not have stable preferences over particular policies or issues, they do possess "core values" or "predispositions" that are coherent and stable (McClosky and Zaller 1984; Feldman 1988; Zaller 1992). Since the 1970s the ANES has included several subsets of items designed to tap into these predispositions. We included these items in the batteries used to construct the issue scales analyzed above. It is therefore possible that most of the stability we find is due to the inclusion of these items.

FIGURE 2. Correlation Between Various 1996 Economic Issue Scales Box-and-Whiskers Plot



https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

TABLE 3. Correlations Between ANES Panel Waves, Policy Positions vs. Policy-Related Predispositions

Issue Area	Total Items in Common	Correlations Between: Scales Made with Common Items	Correlations Between: Individual Common Items (Average)
1992, 1996			
Policy Positions			
Lower Income Assistance	8	.70	.45
Moral Regulation & Rights	4	.69	.53
Policy Related Predispositions			
Beliefs About Equality	6	.59	.40
Limited Government	3	.53	.40
Traditional Morality	4	.61	.43
1990, 1992			
Policy Positions			
Lower Income Assistance	6	.67	.43
Moral Regulation & Rights	4	.71	.48
Affirmative Action & Desegregation	5	.72	.53
Policy Related Predispositions			
Beliefs About Equality	6	.48	.33
Limited Government	3	.57	.41
Traditional Morality	4	.57	.38
Race-Related Attitudes	4	.68	.51

To examine this possibility, following Miller and Shanks (1996), we constructed "policy related predisposition" scales for the 1990–92 and 1992–96 panels. These are Beliefs About Equality, Limited Government, Traditional Morality, and Race-Related Attitudes. We also constructed several "policy position" scales based solely on questions that refer to specific policies or government activities. We call these Lower Income Assistance; Moral Regulation & Rights; and Affirmative Action & Desegregation. The Appendix gives the exact survey items used for each scale.

Table 3 shows the results. Once again, in each row the issue scales show substantially more stability than the average for the individual items. More importantly, when we focus on particular issues using purely policy questions, the resulting scales are *even more stable* over time than the "predisposition" scales. For instance, for the 1990–92 panel, the intertemporal correlation for a scale constructed from six questions about assistance for low-income individuals was .67, whereas that for six questions tapping beliefs about equality was .48. The correlation for a scale based on policies related to moral regulations and rights was .71, while that for predispositions related to traditional morality was .57. The story is similar for the 1992–96 panel. Only in the area of race and affirmative action do policy positions and predispositions show roughly similar levels of stability.

Heterogeneity in Political Sophistication

The standard measurement error model laid out above treats all individuals similarly. Survey researchers, however, have long argued that citizens vary widely in their degree of "political sophistication" (e.g., Luskin 1987; Saris and Sniderman 2004). For example, Converse (1964) rationalizes the low overall correlations in survey responses as reflecting a heterogeneous population, of which a small fraction understand complicated public policy matters, have well formed preferences, and answer survey questions without error. The remaining majority—the unsophisticated—appear to answer questions almost at random. Converse terms this the Black-White model.

Here we examine this issue in the context of issue scales. Although we do find evidence of heterogeneity, consistent with Feldman (1989), measurement error is the larger issue.

A first step is to examine how measurement error varies across levels of political sophistication. Table 4 reports intertemporal stability correlations analogous to those in Table 1, broken down by high versus low education levels (following Converse 1964), and high versus low levels of political information (using the interviewer assessments favored by Zaller (1992) and others). The high-education group consists of individuals with college degrees or more. The low-education group consists of those with a high-school degree or less. Interviewers were asked to rate each respondent's "general level of information about politics and public affairs" on a 5-point scale ranging from 1 = "very high" to 5 = "very low." The high-information group consists of individuals rated "very high" or "fairly high," whereas the low-information group consists of those rated "very low" or "fairly low."

There is clear evidence of some heterogeneity. The intertemporal correlations for low-education and low-information individuals are uniformly lower than the corresponding correlations for high-education or high-information individuals. This is true for the issue scales,

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

TABLE 4. Correlations Between ANES Panel Waves, Issue Scales, and Individual Survey Items by Education and Political Information Level

	High-Educ. Respondents		Low-Educ. Respondents		High-Info. Respondents		Low-Info. Respondents	
Issue Area	Issue Scales	Indiv. Items	Issue Scales	Indiv. Items	Issue Scales	Indiv. Items	Issue Scales	Indiv. Items
1992, 1996								
Economic Issues	.81	.47	.71	.35	.78	.45	.68	.30
Moral Issues	.86	.58	.75	.42	.84	.54	.71	.38
Ideology		.84		.31		.73		.19[1]
Party ID		.77		.79		.81		.76
1990, 1992								
Economic Issues	.81	.50	.74	.36	.78	.46	.68	.31
Racial Issues	.86	.57	.73	.48	.82	.54	.68	.43
Moral Issues	.85	.57	.58	.35	.82	.53	.47	.31
Ideology		.79		.39		.67		.32
Party ID		.83		.76		.83		.57
1972, 1976								
Economic Issues	.73	.46	.66	.39	.71	.45	.60	.35
Racial Issues	.80	.44	.72	.37	.80	.47	.69	.32
Womens' Issues	.76	.47	.59	.36	.70	.45	.60	.36
Law & Order	.82	.61	.63	.45	.78	.58	.58	.36
Ideology		.66		.48		.63		.56
Party ID		.87		.71		.83		.61
1956, 1960								
Economic Issues	.62	.47	.58	.39				
Party ID		.92		.79				

Note: In each case, the issue scales are made using only the survey items common to both years of the panel. The number of items for each issue area can be found in Table 1, column 3.

and also for the individual survey items underlying the scales. It is also true for self-placed ideology, and for party identification (except in one case).

It is equally clear, however, that the responses of both high-information and low-information individuals are plagued with a large amount of measurement error. In the 1992–96 panel, the intertemporal stability of the Economic Issues Scale for high-education individuals is .81, whereas for low-education respondents it is .71. But the average intertemporal stability of the individual items that comprise the scale is .47 for high-education individuals and only .35 for low-education individuals. Using equation (8), we estimate a signal to noise ratio of 1.22 for high-education individuals and 0.81 for low-education individuals. These are intuitive, implying a bit more signal than noise in the responses of the high-education group, and a bit more noise than signal in the responses of the low-education group. Also, although different, they are in the same ball park.

Note also that for both high-education and low-education individuals, the intertemporal stability of the Economic Issues Scale is much higher than the stability of self-placed ideology, and nearly as high as the stability of party identification. These general patterns hold for other scales and other years as well.

One can criticize the analysis above on the grounds that measurement error also plagues our measures of political sophistication. We may understate the importance of heterogeneity if we have misclassified many high-information respondents as having low-information, and vice-versa. In fact, we can construct better measures of political sophistication by combining multiple items. We constructed one such measure for the 1992–96 panel using education, the interviewer assessment, and the answers to any factual questions asked in the panel, such as which party controls the U.S. House, which party controls the Senate, who is William Rehnquist, and who is Boris Yeltsin. Interestingly, the resulting scale is much more stable over time than any of the individual items. Even the interviewer assessment is fairly unstable—the intertemporal correlation is just .55. By contrast, the intertemporal correlation of the scale is .77. Replicating the analysis discussed above using the scaled measure of political sophistication leads to qualitatively similar results to those reported in Table 4. For example, dividing the sample at the mean value of the sophistication scale, the intertemporal stability of the Economic Issues Scale for "high-sophistication" individuals is .81, while for "low-sophistication" respondents it is .66.

Finally, we can refine the analysis further by switching the focus from correlations to variances. When we do this the bottom line is the same. On one hand there is evidence of heterogeneity with respect to levels of sophistication. We cannot tell if this is due to differences in measurement error or differences in the distribution of underlying preferences, but there are clearly differences of some sort. Consider, for example, Economic Issues Scale for 1992.[12] The variance of the

[12] We use simple averages of the various survey items ($\overline{W}_i = \frac{1}{K}\sum_{k=1}^{K} W_k$) for each respondent rather than factor scores. Also,

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

average of the 23 survey items, $\sigma^2_{\overline{W}_i}$, is .21 in the entire sample. Examining subsets, we see that $\sigma^2_{\overline{W}_i}$ is almost monotonic in the level of sophistication, declining from .39 for the group with a "very high" level of political information to .18 for the group with a "very low" level of information.

Perhaps more importantly, the variances provide further evidence that even the least sophisticated are not merely "guessing" when answering issue questions. If these respondents were simply guessing, so their responses were purely random noise, the variance of the average of 23 items would be $1/23 = .043$. As noted, the variance among the group with "very low" political information is .18, more than four times larger than the predicted value under a null hypothesis of pure noise. Thus, the answers of even this least sophisticated group of respondents have much more content than Converse asserted.

ISSUE VOTING

Perhaps the most damning empirical result for issue voting is the fact that individual issue items rarely show large, statistically significant, and robust effects in explaining voting preference or approval of elected officials. By contrast, party identification is always found to be extremely important in predicting vote choice and approval. Kinder (1998) surveys the vast number of studies on this subject and aptly characterizes issue voting as of marginal importance in the literature.

In this section we argue that measurement error is a major reason for this conclusion. If individual survey items contain a large amount of measurement error, then it is likely that most of the literature on issue voting has underestimated the effects of issues on electoral decisions. It is well known in statistics that measurement error in independent variables biases estimates of regression coefficients. Measurement error has two effects. First, measurement error creates bias and inconsistency in the independent variable that is measured with error. Typically, the effect of measurement error is to shrink the estimated effect toward zero. If only one variable is measured with error, then the coefficient on that variable will be biased toward zero. When many variables contain measurement error, biases may take more general forms, including incorrect signs (e.g., Klepper and Leamer 1984). Second, measurement error on one variable spills over onto other variables. Because all coefficient estimates depend on the variances and covariances of all variables, measurement error in one variable necessarily affects the estimated coefficients on other variables.

Similar to Erikson and Tedin (2007), we show that correcting for measurement error by averaging together multiple measures produces a radically different conclusion about the relative importance of issue positions on vote choice. Issues have statistically significant and substantively large effects, comparable in magnitude to the effects of party identification.

To do this, we contrast two sorts of estimates. First, we regress voting decisions on party identification, ideology, and a battery of individual issue items. Second, we construct issue scales from the individual issue items, and use these as independent variables in place of the items. If issues matter for voting but measurement error in the individual issue items is a substantial problem, then the estimated coefficients on these variables will tend to be attenuated and few will be statistically significant. The coefficients on the scales will be much larger than the coefficients on the individual items, and the scales will be highly significant.

We have followed this approach with a number of surveys and obtained similar results, but due to space constraints, we present the results from the American National Election Studies of 1992 and 1996. For each of these surveys, we identified two to three dozen issue questions likely to matter directly to the vote. We constructed two scales—an Economic Issues Scale using all items on economic issues, and a Moral Issues Scale using all items on moral and social issues.[13]

Table 5 contrasts the evidence of issue voting using individual issue items and the scaled issue measures. The top panel displays the results of probit regression analyses predicting the Republican versus the Democratic vote for president as a function of party identification, an ideology scale, and issue positions on all the relevant questions. In order to aid comparisons, we standardized all of the regressors. The vote is coded as 1 for Republican and 0 for Democrat (we drop respondents who voted for minor candidates or did not vote). Rather than show a large number of small and insignificant coefficients, we display the average and median of the absolute values of the coefficients, and the fraction that are statistically significant at the .05 level. The bottom panel shows the results using the three issue scales in place of the individual issue items.

The results are broadly consistent with the conjecture that there is substantial measurement error bias in regressions that use individual items to capture voters' issue positions. When we use individual issue items, the coefficients are small and few are statistically significant. Consider, for example, the 1992 ANES. When we use the individual issue items, only one in ten economic issue coefficients, and one in five moral issue coefficients are statistically significant at the .05 level. The average absolute values of the coefficients are .09 and .14. In 1996 the average absolute values of the coefficients are slightly lower, although a slightly larger fraction are statistically significant.

When we combine the individual items to construct issue scales, however, the picture changes dramatically. Issue preferences suddenly appear to have much larger and more robust effects on the vote. In both analyses, the coefficients on the "economic issues" and "moral issues" scales are large and statistically significant at the .01 level. In the 1996 ANES, the coefficient on the

in order to obtain as large a sample as possible, we study the entire sample from 1992, including both the panel and non-panel respondents, which gives us 1,217 observations. Recall that each survey item is standardized to have a mean of 0 and a variance of 1.

[13] We also tried adding a Foreign Policy Scale but it was never statistically or substantively significant in 1992 or 1996.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

TABLE 5. Effects of Party and Issues on Presidential Voting

	ANES, 1992 1992 Vote	ANES, 1996 1996 Vote
Party ID	1.18 (.12)**	.989 (.08)**
Ideology Scale	.58 (.12)**	.54 (.10)**
Individual economic issue items:		
number of items	33	34
average |coefficient|	.09	.07
median |coefficient|	.06	.05
fraction significant (.05 level)	.09	.15
Individual moral issue items:		
number of items	18	14
average |coefficient|	.14	.12
median |coefficient|	.08	.06
fraction significant (.05 level)	.22	29
N	653	995
Pseudo-R^2	.695	.672
Party ID	.99 (.09)**	.86 (.08)**
Ideology Scale	.40 (.10)**	.53 (.09)**
Economic Issues Scale	.33 (.08)**	.52 (.09)**
Moral Issues Scale	.43 (.09)**	.29 (.07)**
N	653	995
Pseudo-R^2	.622	.629

Note: All variables standardized.
** = significant at the .01 level.
** = significant at the .05 level.

Economic Issue Scale is more than twice as large as the largest coefficient on any of the individual economic issue items. In both regressions, the combined effect of changing both the Economic Issues Scale and the Moral Issues Scale by 1 standard deviation (SD) is nearly as large, or even larger than, a change of 1 SD in party identification. The coefficient on the Moral Issues Scale is the same size as the coefficient on the Economic Issues Scale in the 1992 ANES, but smaller in the 1996 ANES.[14] Moral issues are significant in all three analyses, and the scale always has a larger coefficient than any of the individual moral issue items.[15]

Heterogeneity, along the lines of Converse's Black-White model, would manifest itself as interactions in these specifications. Specifically, high information and interest individuals would have definite issue preferences, and their expressed opinions would translate directly into votes. However, low information and interest individuals are thought to have no well defined attitudes. According to the Black-White model, low-information voters' responses to issue questions are largely noise and, thus, bear no relation to their vote choice. This model is surely an extreme case, but substantial heterogeneity may still operate if those with higher levels of information are noticeably more issue-oriented in their voting.

What, then, is the relationship between level of information or education and issue voting? The answer, evidently, is none. We reanalyzed the specifications in Table 5 including measures of information and interactions between the issue scales and respondents' information or education.[16] Education has a significant main effect on the vote, reflecting the fact that higher

[14] If we include the items one at a time rather than in the same regression, then many more coefficients are statistically significant (roughly half). However, the substantive significance of issues still appears small. The following table summarizes the results:

Economic issue coefficients:	1992	1996
number of items	33	34
median |coefficient|	.13	.16
fraction significant	.42	.74
median Pseudo-R^2	.574	.594
Moral issue coefficients:		
number of items	18	14
median |coefficient|	.18	.08
fraction significant	.61	. 36
median Pseudo-R^2	.580	.591

Thus, compared to the results in Table 5, the median effect of the individual items is only about one third as large as the corresponding Economic Issues Scale or Moral Issues Scale. In terms of R-square the results are even more striking. In a "baseline" model in which Party ID and Ideology Scale are the only independent variables, the probit Pseudo-R^2 is .570 for 1992 and .590. Thus, on average the individual items only increase the Pseudo-R^2 by about 1% in 1992 and only one-half of 1% in 1996. By contrast, the Economic Issues Scale and Moral Issues Scale together increase the Pseudo-R^2 by over 9% in 1992 and by almost 7% in 1996.

[15] Note that in all these analyses we control for party identification. There is evidence within the data that party identification is more strongly correlated with economic issues than with moral and foreign policy preferences. We are agnostic about the causal relations among the variables on the right-hand side of the equation. However, if one were to measure the total effect of issues, not conditioning on party identification, economic issues become much more important. See Ansolabehere, Rodden, and Snyder 2006.

[16] Because the results are entirely null findings, we do not present the full analysis here.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

educated Americans tend to be more Republican. None of the interactions are statistically significant, and the magnitudes of the estimated interactions are small and substantively unimportant. In the 1996 ANES, the estimated interaction between high education (college or more) and the Economic Issues Scale is −.08 with a standard error (SE) of .18, and the estimated interaction between high education and the Moral Issues Scale is −.07 (SE = .19). By contrast the main effects are .55 for the Economic Issues Scale and .44 for the Moral Issues Scale. When we used the interviewer's rating of respondents' information levels to measure sophistication, the interactions are .16 (SE = .17) and −.18 (SE = .17) for economic issues and moral issues, respectively. In neither survey does the likelihood ratio test reject the hypothesis that all of the interactions equal 0. There is, then, little evidence that sophistication magnifies the extent to which issue preferences affect the vote.

This is a striking conclusion in light of the large literature that looks for such interactions (e.g., Delli Carpini and Keeter 1996; Goren 1997). The fact that the issue scales do not interact with sophistication in predicting the vote carries the implication that such interactions in studies using single items to measure issue preferences are likely capturing measurement error, not substance. Survey respondents with high education and high information may simply be better test takers, not better citizens.

CONCLUSIONS

The implications of our analysis are at once methodological and substantive. At least since Achen 1975, political scientists have recognized the potentially serious biases introduced by measurement error in single survey items. We have documented that those errors are indeed large, and the consequences for our discipline are far-reaching, owing to the centrality of survey research to political science.[17] We have focused on just one important line of inquiry, and discovered that measurement error has led the field to under estimate the extent to which the electorate holds and relies on their policy preferences.

Three basic methodological lessons emerge. First, measurement error is substantial, even in high-quality surveys such as the ANES and GSS, where approximately half of the variance in responses across individuals within the typical issue item is attributable to measurement error, and the other half reflects true differences in preferences. This finding suggests caution in drawing inferences from analyses of surveys that focus on single items introduced to capture issue preferences.

Second, averaging offers a simple and time-honored way to reduce the measurement error long known to corrupt individual survey items and the resulting bias in regression models. Others have used factor analysis to uncover latent dimensions of preferences, but political scientists have not appreciated the degree to which factor analysis and other methods of averaging can fix measurement error of the sort that arises from vague questions and categories that attempt to capture the complexity of public opinion on a topic with a single question, or from distraction or boredom of the respondent. Multiple measures should, of course, be used with care. Among other things, it is important to examine the consistency and reliability of one's scale and to justify the inclusion of items (see, e.g., McIver and Carmines 1981). Moreover, it may be necessary to address the problem of nonrandom measurement error (see, e.g., Green and Citrin 1994). How best to use multiple measures is an important avenue for further development.

Third, we encourage those engaged in the design of surveys to focus on the quality of batteries of questions, not just individual items. There are two ways to reduce measurement error: ask better questions and ask more questions. These two are not at odds. Although the ANES and GSS pilot and vet their questions carefully, substantial measurement error remains. Further efforts to improve the quality of questions are valuable, but often, no single survey question can capture the complexity of opinion on a given issue or issue cluster. The work here on Economic and Moral issues is a start, but a more in-depth consideration of these and other aspects of issue voting (notably race, civil liberties, and foreign affairs) is in order. We recommend a systematic effort to develop batteries of questions designed to capture issue preferences rather than focusing on single items.

The lessons here also extend to other survey research problems, especially measuring ideology and citizen knowledge. Congressional studies have long used roll call vote scores to measure ideology, but survey researchers have used a single item in which respondents use terms such as "liberal" and "conservative" to describe their ideology. The approach of congressional scholarship to the measure of ideology seems particularly promising, as the terms liberal and conservative may not fit people's belief systems well. Measurement of knowledge has seen the most use of multiple measures in the survey world, but there is still considerable disagreement about what is an appropriate measure. Zaller (1992) discusses the advantages of using multiple measures to capture knowledge, but in the end, settles on a single best indicator based on the interviewer's evaluation. However, a preliminary analysis of the measures Zaller examined suggests that averaging together multiple measures of knowledge produces a measure with higher construct validity than any single item. These themes await more concerted examination in their own right.

Sometimes the value of improved methods lies in convincing us more firmly of an accepted wisdom. Other times the value is in allowing us to see things that were previously obscured from view. The empirical analysis in this paper comes closer to the latter than the former. Although others have come before us, our rather simple approach to measurement error yields a markedly different picture of belief systems in mass

[17] According to one accounting almost half of all articles in the APSR analyze survey data, especially the ANES (King et al. 2001).

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

publics and issue voting than the consensus view that still holds sway over the textbook descriptions of the American public.

First and foremost, we find strong evidence that voters have stable policy preferences. Converse (1964) famously concluded that the mass of the electorate holds nonattitudes on important policy questions, such as race relations, foreign affairs, and economic redistribution. His evidence was twofold. The correlation among individual items on each of these questions was relatively small, suggesting little "constraint," and individual's responses to a given item were unstable over time. The correlations he found ranged from .2 to .4. Upon averaging items we find much more evidence of constraint and stability. The correlations among the averages of half-samples of items approaches .8, as do the over-time correlations of the individuals' issue policy scales. If the vast majority of people have nonattitudes, as Converse conjectured, averaging would have again returned low correlations.

Second, issues matter to the electorate. The vast majority of research on the topic finds weak or no evidence that issue preferences explain the vote. But, those studies rely almost entirely on single items to measure voters' policy preferences. A small minority construct scales, analogous to roll-call voting scores, and that branch of the literature finds substantively important effects of issues (e.g., Carmines and Stimson 1980; Erikson and Tedin 2007). Ansolabehere, Rodden, and Snyder (2006) show that issue scales like those used here for presidential elections are also very good predictors of vote choice in U.S. Congressional elections, and Heath et al. (1994) construct issue scales that are powerful predictors of vote choice in British parliamentary elections. Issue scales constructed from questions in the World Values Survey have also been used to understand vote choice and other forms of political behavior in cross-national studies such as Inglehart (1990) and more recently, De La O and Rodden (2008).

We argue that the tension in the empirical literature stems from measurement error in individual issue items. When we use individual items to measure policy preferences, we find relatively little evidence of issue voting. However, consistent with a simple model of measurement error, we find strong evidence of issue voting when vote preference is regressed on issue scales, controlling for party identification and ideology. In fact, contrary to the bulk of the literature, we find that issue voting may rival party in explaining the vote. The combined effects of issue preferences are about as large as party identification in the multivariate analyses predicting the vote.

Third (and certainly worthy of deeper investigation), heterogeneity of sophistication among the public is not as important a phenomenon as the line of inquiry from Converse through Zaller would suggest. Converse distinguished two sorts of citizens—a relatively small group with high levels of education and political sophistication, and a great mass of unsophisticated voters whose responses to policy-related survey questions are essentially guesses. Yet we find that the over-time correlations of policy preference scales among low sophistication respondents are typically in the range of .6 to .7. These are much larger even than the levels Converse observed for single items among high sophisticates, and not much different than the intertemporal stability found by Jennings (1992) on three issue items in a study of a group of highly sophisticated elites—delegates to national party conventions.

The least educated or sophisticated voters do appear to have somewhat less stable policy opinions than those with high education or sophistication. But, the difference between these two groups is small compared with the enormous amount of measurement error in individual survey items on policy issues. Correcting for the measurement error, all groups exhibit considerable policy content to their opinions.

Moreover, what heterogeneity we do observe has little direct effect on vote choice. Models such as Converse (1964) and Zaller (1992) imply that we should observe strong interactions between sophistication and policy preferences in predicting the vote. We find no such interactions. Neither education nor political information magnifies the extent of issue voting. All respondents draw equally strongly on their policy preferences in choosing elected officials.

Our findings are encouraging for theoretical models of elections and representation that rely on spatial representations of preferences. Converse's nonattitudes finding has been something of a show stopper for the spatial theory of elections. If voters do not have policy preferences, how could such models possibly characterize elections? Erikson, MacKuen, and Stimson (2002) suggest that a meaningful policy "signal" is produced by averaging many millions of votes. Our findings imply an even simpler story: there is substantial policy content to individual voters' preferences and behaviors, but the survey instrument has not been sufficiently finely tuned to detect it. Using multiple measures to capture issue preferences offers a way to solve longstanding puzzles concerning the nature of belief systems in mass publics.

APPENDIX

Here we list the variable numbers of the survey items used in constructing each of scales in the paper. The items in parentheses appear in only one survey of the panel. These are used for the scales made with *all* items but not for the scales made with *common* items.

1992 Economic Issues: 923701, 923716, 923718, 923725, 923726, 923727, 923728, 923730, 923811, 923813, 923815, 923817, 923818, 925316, 925318, 925320, 925729, 925730, 925731, 926024, 926025, 926026, 926027, 926028, 926029, (923717, 923729, 923745, 923746, 923816, 923819, 926137)
1996 Economic Issues: 960450, 960479, 960483, 960496, 960497, 960498, 960500, 960501, 960560, 960561, 960562, 960564, 960565, 961033, 961035, 961036, 961229, 961230, 961231, 961232, 961233, 961234, 961144, 961145, 961146, (960537, 961219, 961220, 961226, 961227, 961283, 961320, 961322, 961324)
1992 Moral Issues: 923732, 923801, 925324, 925335, 925338, 925924, 925926, 925945, 926115, 926116, 926117, 926118, (923736, 923738, 923740, 925928, 926020, 926119)

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

TABLE A.1. Factor Analysis Results

1992 Economic Issues						eigenvalues = 5.43, 1.11, 1.02, 0.56					
variable	load	uniq	variable	load	uniq	variable	load	uniq	variable	load	uniq
923718	.54	.71	923716	.47	.78	925730	.48	.77	925729	.57	.67
925731	.48	.76	923701	.60	.65	923727	.41	.83	923725	.50	.75
923726	.54	.70	923730	.51	.74	923811	.40	.84	923813	.53	.72
923817	.60	.64	923815	.39	.84	923818	.43	.81	923728	.30	.91
926024	.34	.88	926025	.40	.84	926029	.56	.68	926026	.39	.85
926027	.35	.87	926028	.48	.77	925316	.43	.82	925318	.41	.83
925320	.36	.87									
1996 Economic Issues						eigenvalues = 6.50, 1.00, 0.78, 0.58					
variable	load	uniq	variable	load	uniq	variable	load	uniq	variable	load	uniq
960483	.61	.63	960479	.49	.76	961145	.55	.70	961144	.61	.63
961146	.53	.72	960450	.68	.53	960498	.46	.79	960496	.52	.73
960497	.54	.70	960501	.59	.65	960560	.45	.80	960564	.61	.63
960565	.66	.56	960561	.41	.83	960562	.51	.74	960500	.43	.81
961229	.46	.78	961230	.45	.80	961231	.52	.73	961232	.36	.87
961233	.30	.91	961234	.51	.74	961033	.50	.75	961036	.44	.81
961035	.34	.89									
1992 Moral Issues						eigenvalues = 3.33, 0.57, 0.53, 0.26					
variable	load	uniq	variable	load	uniq	variable	load	uniq	variable	load	uniq
923732	.54	.71	925945	.34	.89	926118	.56	.69	926115	.40	.84
926117	.48	.77	926116	.56	.69	923801	.48	.77	925924	.61	.63
925926	.63	.60	925324	.49	.76	925335	.68	.54	925338	.47	.78
1996 Moral Issues						eigenvalues = 3.18, 0.68, 0.38, 0.12					
variable	load	uniq	variable	load	uniq	variable	load	uniq	variable	load	uniq
960503	.52	.73	961214	.31	.90	961247	.52	.73	961248	.41	.83
961249	.51	.74	961250	.54	.71	960543	.49	.76	961194	.59	.66
961196	.59	.66	961039	.50	.75	961042	.67	.55	961038	.44	.80

Note: Here we present summary statistics from the factor analyses used in constructing some of the scales using common items.

1996 Moral Issues: 960503, 960543, 961038, 961039, 961042, 961194, 961196, 961214, 961247, 961248, 961249, 961250, (961043, 961199)
1992 Multi-Item Party ID: 923634, 923317, 923318
1996 Multi-Item Party ID: 960420, 960292, 960293
1992 Multi-Item Ideology: 923509, 923511, 923512, 925326, 925319
1996 Multi-Item Ideology: 960365, 960366, 961032, 961031
(Note, 7-point ideology scale item was constructed from underlying branching questions)
1990 Economic Issues: 900157, 900159, 900162, 900331, 900333, 900335, 900377, 900379, 900380, 900382, 900383, 900384, 900385, 900386, 900426, 900427, 900428, 900429, 900430, 900431, 900446, 900488, 900452, (900332, 900487, 900490)
1992 Economic Issues: 925316, 925318, 925320, 925729, 925730, 925731, 926024, 926025, 926026, 926027, 926028, 926029, 923701, 923716, 923717, 923718, 923725, 923727, 923729, 923730, 923745, 923811, 923813, 923815, 923818, (923726, 923728, 923746, 923816, 923817, 923819, 926133, 926137)
1990 Racial Issues: 900386, 900447, 900464, 900466, 900470, 900518, 900519, 900520, 900521, 900522, 900523
1992 Racial Issues: 923724, 923729, 925929, 925930, 925932, 925936, 925948, 926126, 926127, 926128, 926129, (925938)
1990 Moral Issues: 900158, 900330, 900467, 900479, 900481, 900483, 900500, 900501, 900502, 900503, (900154, 900163, 900471, 900472
1992 Moral Issues: 923732, 923736, 923738, 925324, 925945, 926115, 926116, 926117, 926118, 926020, (923740, 923801, 925335, 925338, 925924, 925926, 925928, 926119)
1990 Multi-Item Party ID: 900320, 900151
1992 Multi-Item Party ID: 923634, 923317, 923318
1990 Multi-Item Ideology: 900406, 900407, 900161, 900156
1992 Multi-Item Ideology: 923509, 923511, 923512, 925326, 925319
(Note, 7-point ideology item was constructed from underlying branching questions)
1972 Economic Issues: 720208, 720629, 720707, 720708, 720722, 720737, 720738, 720744, 720753, 721067, 721068, (720214, 720843, 720848, 720856, 720688, 720690, 720693, 721111, 721112)
(Note, 720670, 720686, 720687, 720691, 720692, 720694, 720689 were not used because they were only given to a subsample of respondent's.)
1976 Economic Issues: 763241, 763264, 763273, 763566, 763567, 763573, 763582, 763751, 763752, 763753, 763754, 763755, 763757, 763758, 763767, 763779, 763821, 763822, 763836, (763353, 763589, 763562)
1972 Racial Issues: 720106, 720112, 720115, 720118, 720202, 720727, 720729, 720745, 720845, 720847, 720849, 720855, 720859, 720862, (720104, 720110, 720113, 720114, 720119, 720851, 720857)
1976 Racial Issues: 763211, 763213, 763214, 763217, 763257, 763570, 763803, 763805, 763806, 763807, 763812, 763814, 763841, 763843, (763588)
1972 Womens' Issues: 720232, 720238, 720239, 720725, 720844, 720846, 720850, 720852, 720858, 720860, (720240, 720242, 720243, 720244, 720245, 720246, 720247, 720248, 720249, 720251)
1976 Womens' Issues: 763787, 763796, 763798, 763802, 763804, 763808, 763809, 763811, 763813, 763839, (763574, 763799)

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

1972 Law & Order & Military: 720196, 720621, 720713, 720714, 720717, 720726, (720678, 720728, 720733)
1976 Law & Order & Military: 763248, 763772, 763827, 763828, 763831, 763840, (763561, 763563, 763565, 763909, 763913)
1972 Multi-Item Party ID: 720140, 720719, 720721
1976 Multi-Item Party ID: 763174, 763833, 763835
1972 Multi-Item Ideology: 720652, 720709, 720724
1976 Multi-Item Ideology: 763286, 763823, 763838
1956 Economic Issues: 560032, 560038, 560053, 560059, (560029, 560047)
1960 Economic Issues: 600618, 600620, 600624, 600632
1992 Beliefs About Equality: 926024, 926025, 926026, 926027, 926028, 926029
1996 Beliefs About Equality: 961229, 961230, 961231, 961232, 961233, 961234
1992 Limited Government: 925729, 925730, 925731
1996 Limited Government: 961144, 961145, 961146
1992 Traditional Morality: 926115, 926116, 926117, 926118
1996 Traditional Morality: 961247, 961248, 961249, 961250
1992 Lower Income Assistance: 923716, 923718, 923725, 923726, 923730, 923813, 923817, 923818
1996 Lower Income Assistance: 960479, 960483, 960496, 960497, 960501, 960562, 960564, 960565
1992 Moral Regulation & Rights: 923732, 925924, 925926, 925945
1996 Moral Regulation & Rights: 960503, 961194, 961196, 961214
1990 Beliefs About Equality: 900426, 900427, 900428, 900429, 900430, 900431
1992 Beliefs About Equality: 926024, 926025, 926026, 926027, 926028, 926029
1990 Limited Government: 900331, 900333, 900335
1992 Limited Government: 925729, 925730, 925731
1990 Traditional Morality: 900500, 900501, 900502, 900503
1992 Traditional Morality: 926115, 926116, 926117, 926118
1990 Race-Related Attitudes: 900520, 900521, 900522, 900523
1992 Race-Related Attitudes: 926126, 926127, 926128, 926129
1990 Lower Income Assistance: 900382, 900383, 900384, 900385, 900446, 900488
1992 Lower Income Assistance: 923718, 923725, 923730, 923745, 923813, 923818
1990 Moral Regulation & Rights: 900467, 900479, 900481, 900483
1992 Moral Regulation & Rights: 923732, 923736, 923738, 925945
1990 Affirmative Action & Desegregation: 900386, 900447, 900464, 900466, 900470
1992 Affirmative Action & Desegregation: 923724, 925929, 925932, 925936, 925948

REFERENCES

Abramowitz, Alan, and Kyle Saunders. 1998. "Ideological Realignment in the U.S. Electorate." *Journal of Politics* 60, 3 (August): 634–52.

Achen, Christopher H. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69 (Dec.): 1218–23.

Achen, Christopher H. 1983. "Toward Theories of Political Data." In *Political Science: The State of the Discipline*, ed. A.W. Finifter. Washington, DC: American Political Science Association.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder, Jr. 2006. "Purple America." *Journal of Economic Perspectives* 20 (Spring): 97–118.

Butler, D. E., and D. Stokes. 1974. *Political Change in Britain.*. 2nd ed. London: Macmillan.

Campbell, Angus, Warren Miller, Philip Converse, and Donald Stokes. 1960. *The American Voter*. New York: Wiley.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96 (March): 127–40.

Carmines, Edward G., and James A. Stimson. 1980. "The Two Faces of Issue Voting." *American Political Science Review* 74 (March): 78–91.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. D. E. Apter. New York: Free Press.

Converse, Philip E. 1980. "Comment: Rejoinder to Judd and Milburn." *American Sociological Review* 45 (August): 644–46.

De La O, Ana, and Jonathan Rodden. 2008. "Does Religion Distract the Poor? Income and Issue Voting Around the World." *Comparative Political Studies* 41, 4/5 (April/May): 437–76.

Delli Carpini, Michael X, and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters.* New Haven, CT: Yale University Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Erikson, Robert S. 1978. "Analyzing One-Variable Three Wave Panel Data: A Comparison of Two Models." *Political Methodology* 5 (2): 151–61.

Erikson, Robert S. 1979. "The SRC Panel Data and Mass Political Attitudes." *British Journal of Political Science* 9 (Jan.): 89–114.

Erikson, Robert S., and Kent Tedin. 2007. *American Public Opinion: Its Origin, Content, and Impact.* 7th ed. New York: Wiley.

Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2002. *The Macro Polity.* New York: Cambridge University Press.

Erikson, Robert S., and Gerald C. Wright. 1993. "Voters, Candidates, and Issues in Congressional Elections." In *Congress Reconsidered*, 5th ed., ed. Lawrence C. Dodd and Bruce I. Oppenheimer. Washington: Congressional Quarterly Press.

Evans, Geoffrey, Anthony Heath, and Mansur Lalljee. 1996. "Measuring Left-Right and Libertarian-Authoritarian Values in the British Electorate." *British Journal of Sociology* 47 (March): 93–112.

Feldman, Stanley. 1988. "Structure and Consistency in Public Opinion: the Role of Core Beliefs and Values." *American Journal of Political Science* 32 (May): 416–40.

Feldman, Stanley. 1989. "Measuring Issue Preferences: The Problem of Response Instability." *Political Analysis* 1 (1): 25–60.

Fiorina, Morris, and Paul Peterson. 1998. *The New American Democracy*. Needham Heights, MA: Allyn and Bacon.

Goren, Paul. 1997. "Political Expertise and Issue Voting in Presidential Elections." *Political Research Quarterly* 50: 387–412.

Green, Donald, and Jack Citrin. 1994. "Measurement Error and the Structure of Attitudes: Are Positive and Negative Judgments Opposites?' *American Journal of Political Science* 38, 1 (Feb.): 256–81.

Green, Donald, and Bradley Palmquist. 1994. "How Stable is Party Identification?" *Political Behavior* 16, 4 (Dec.): 437–66.

Green, Donald, Bradley Palmquist, and Eric Schickler. 2002. *Partisan Hearts and Minds.* New Haven, CT: Yale University Press.

Heath, Anthony, Geoffrey Evans, and Jean Martin. 1994. "The Measurement of Core Beliefs and Values: The Development of Balanced Socialist/Laissez Faire and Libertarian/Authoritarian Scales." *British Journal of Political Science* 24 (Jan.): 115–32.

Hurwitz, Jon, and Mark Peffley. 1987 (Dec.). "How are Foreign Policy Attitudes Structured? A Hierarchical Model." *American Political Science Review* 81: 1099–1120.

Inglehart, Ronald. 1990. *Culture Shift in Advanced Industrial Society.* Princeton, NJ: Princeton University Press.

Jackson, John E. 1983. "The Systematic Beliefs of the Mass Public: Estimating Policy Preferences with Survey Data." *Journal of Politics* 45 (Nov.): 840–65.

Jennings, M. Kent. 1992. "Ideological Thinking Among Mass Publics and Political Elites." *Public Opinion Quarterly* 56 (Winter): 419–41.

Judd, Charles M., and Michael A. Milburn. 1980. "The Structure of Attitude Systems in the General Public: Comparisons of Structural Equation Models." *American Sociological Review* 45 (Aug.): 627–43.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press

Key, V. O. 1966. *The Responsible Electorate: Rationality in Presidential Voting, 1936–1960.* Cambridge, MA: Harvard University Press.
Kinder, Donald R. 1998. "Opinion and Action in the Realm of Politics." In *Handbook of Social Psychology.* Oxford: Oxford University Press.
King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomeplete Political Science Data: An Alternative Algorithm for Multiple Imputation" *American Political Science Review* 95, 1 (March): 49–69.
Klepper, Steven, and Edward E. Leamer. 1984. "Consistent Sets of Estimates for Regressions with Errors in All Variables." *Econometrica* 52 (Jan.): 163–84.
Kuder, G. F., and M. W. Richardson. 1937. "The Theory of the Estimator of Test Reliability." *Psychometrika* 2 (Sept): 151–60.
Lord, L. M., and M. R. Novick. 1968. *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.
Luskin, Robert C. 1987. "Measuring Political Sophistication." *American Journal of Political Science* 31 (Nov.): 856–99.
McClosky, Herbert, and John Zaller. 1984. *The American Ethos: Public Attitudes Toward Capitalism and Democracy.* Cambridge, MA: Harvard University Press.
McIver, John, and Edward Carmines. 1981. *Unidimensional Scaling* Sage University Paper, Quantitative Applications in the Social Sciences Series, 24.
Miller, Warren E., and J. Merrill Shanks. 1996. *The New American Voter.* Cambridge MA: Harvard University Press.
Moskowitz, Adam N., and J. Craig Jenkins. 2004. "Structuring Political Opinions: Attitude Consistency and Democratic Competence among the U.S. Mass Public." *Sociological Quarterly* 45 (May): 395–419.
Mosteller, Frederick. 1968. "Non-Sampling Errors." In *International Encyclopaedia of Statistics*, Volume 1. ed. William Kruskal and Judith Tanur. New York: Crowell Collier and Macmillan, 208–29.
Norpoth, Helmut, and Milton Lodge. 1985. "The Difference between Attitutes and Nonattitudes in the Mass Public: Just Measurement?" *American Journal of Political Science* 29 (May): 291–307.
Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric Theory*. New York: McGraw-Hill.
Page, Benjamin I., and Robert Y. Shapiro. 1982. "Changes in Americans' Policy Preferences, 1935–1979." *Public Opinion Quarterly* 46 (Spring): 24–42.
Polsby, Nelson W., and Aaron Wildavsky. 2000. *Presidential Elections: Strategies and Structures of American Politics.* Tenth ed. New York: Chatham House.
Saris, Willem, and Paul Sniderman. 2004. *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change.* Princeton, NJ: Princeton University Press.
Sniderman, Paul, and Thomas Piazza. 1993. *The Scar of Race.* Cambridge MA: Harvard University Press.
Spector, Paul. 1992. *Summated Rating Scale Construction: An Introduction* Sage University Paper, Quantitative Applications in the Social Sciences Series, 82.
Wiley, D. E., and J. A. Wiley. 1970. "The Estimation of Measurement Error in Panel Data." *American Sociological Review* 35 (Feb.): 112–16.
Zaller, John R. 1992. *The Nature and Origins of Mass Opinion.* New York: Cambridge University Press.

https://doi.org/10.1017/S0003055408080210 Published online by Cambridge University Press