Decomposing Public Opinion Variation into Ideology, Idiosyncrasy and Instability [*]

Benjamin E Lauderdale   *London School of Economics and Political Science*
Chris Hanretty   *Royal Holloway, University of London*
Nick Vivyan   *Durham University*

We propose a method for decomposing variation in the issue preferences that US citizens express on surveys into three sources of variability that correspond to major threads in public opinion research. We find that, averaging across a set of high profile US political issues, a single ideological dimension accounts for about 1/7 of opinion variation, individuals' idiosyncratic preferences account for about 3/7, and response instability for the remaining 3/7. These shares vary substantially across issue types and the average share attributable to ideology doubles when a second ideological dimension is permitted. We also find that (unidimensional) ideology accounts for almost twice as much response variation (and response instability is substantially lower) among respondents with high, rather than low, political knowledge. Our estimation strategy is based on an ordinal probit model with random effects, and is applicable to other data sets that include repeated measurements of ordinal issue position data.

Introduction   Since Converse (1964) political scientists have debated the extent to which citizens have organized preferences about political issues, or indeed whether they have such preferences at all (Achen 1975; Zaller and Feldman 1992; Zaller 1992). In this short paper, we exploit recent survey data from Broockman (2016) to decompose variation in citizens' expressed issue preferences into three sources of variability. First, how much is *ideological*, by which we mean predictable from a citizen's preferences on other issues given the way preferences are typically organized in the population?[1] Second, how much is *idiosyncratic*, by which we mean preferences that are not predictable from a citizen's preferences on other issues, but which are nonetheless stable in repeated measurement? Third, how much is *instability*, by which we mean preference variation that is not stable in repeated measurement and which combines measurement error and non-attitudinal response behaviours? Although scholars have at various points argued for the relative importance of each of these sources of variability in opinion, to our knowledge none have provided a quantitative summary of how much variation is attributable to each one.

Context   There is a long-running debate within political science about the extent to which most citizens have "reasonably well-formed attitudes on major political issues" (Zaller and Feldman 1992, 579) which exist independently of the survey instruments designed to measure such attitudes. On one view, survey respondents may be perfectly capable of responding to questions designed to elicit political preferences, but these responses merely reflect the result of a sample of considerations bearing on the matter, which are context-biased in ways that lead them to give different answers at different points in time (Zaller and Feldman 1992, 579). Evidence in favour of this view comes from the low correlation between many measures of attitudes at different points in time; from open-ended survey questions eliciting voter considerations; and from the extensive literature on priming. Against this evidence, it has been argued that low over-time correlation is an indication that survey responses are subject to large measurement error such that stable underlying attitudes can result in volatile survey responses (Achen 1975), and that many priming effects are better understood as information effects (Lenz 2009).

If citizens do not have stable preferences on major political issues at all, then there is no point arguing about whether those attitudes can be effectively summarized by a position in some latent

---

[1]This is a thin definition of *ideology*—what Converse refers to as *constraint*.

ideological space. However, if political preferences are at least partially stable or real, then they may or may not have a low dimensional structure (ideology) that explains much of the variation in preferences across different issues. There is disagreement about whether preferences can be represented using just one left-right dimension (Jacoby 1995; Jost, Federico, and Napier 2009) or whether they must instead be represented using two dimensions (Duckitt and Sibley 2009; Feldman and Johnston 2014; Evans, Heath, and Lalljee 1996). The more that preferences can be represented in terms of a small number of underlying dimensions, the less role there is for idiosyncratic preferences, understood as 'non-ideological' preferences which cannot be predicted on the basis of preferences about other issues, but which are nonetheless real, stable views about particular issues (Broockman 2016).

This paper argues that we can usefully understand these two debates—about the extent to which preferences are real and about the extent to which they are ideological—as part of a larger question about the composition of survey response variation. Our decomposition of preferences into components corresponding to ideology, idiosyncrasy and instability can be mapped onto these previous debates. The debate over whether preferences are real is a debate about the relative contribution of ideology plus idiosyncrasy versus instability. The debate over whether preferences are ideological is a debate about the relative contribution of ideology versus idiosyncrasy (either with or without instability).

In our reading, no one has directly assessed the three-way decomposition quantitatively. Converse evaluates both the correlation structure across issues as well as the stability of responses within issues, but the synthesis of the evidence from those analyses is qualitative (Converse 1964). Subsequent studies, including the Broockman study whose data we re-analyse, have similarly done a series of analyses and then summarized the results qualitatively. The costs of a purely qualitative synthesis of the evidence are apparent in the subsequent multi-decade debate spurred by Converse's study, a debate that tended to get simplified into binary questions about whether preferences were real or not, or ideological or not, rather than decompositional questions about how much of observed preference variation is real and how much is ideological (Converse 2000; Feldman 2013).

Data   To quantitatively decompose response variation into ideology, idiosyncrasy and instability, we need data with several features. Repeated measurements on the same respondents must be taken in order to distinguish stable preferences from response instability. Questions on a range of political

issues must be asked in order to assess whether preferences are ideologically structured or not. Anchored response scales are desirable in order to limit measurement error associated with differential scale use, both across respondents and within respondents across repeated measurements.

Fortunately, data satisfying all of these criteria are provided by Broockman (2016). Broockman surveyed US citizens in early January and again in late February 2014. He asked eight questions as part of a political knowledge battery and thirteen ordinal questions on particular policies. The questions covered both economic issues (health care, taxes, Medicare, unions, education, social security) and social/cultural issues (gun control, immigration, abortion, the environment, gay rights, affirmative action, and contraception). The responses to the ordinal policy questions were constructed so that the middle option represented the status quo, and the next least extreme response options captured the median position of the Democratic and Republican parties respectively. Remaining options represented more extreme positions.[2] Giving such specific response options mitigates the risk that variation in manifest survey responses will result from the way respondents use response scales rather than from respondents' varying underlying preferences.

Though we focus on the policy questions, we also use Broockman's latent trait measure of political knowledge. Correct or incorrect answers to eight knowledge questions were aggregated allowing for variation in question difficulty and the ability of each question to discriminate between respondents with otherwise similar levels of knowledge. For simplicity, we categorise respondents as 'high' or 'low' political knowledge using the median score political knowledge score as the cutoff. 1137 respondents participated in the first wave, of whom 513 also participated in the second wave. Average levels of political knowledge were were modestly (0.15 SD) higher among wave two respondents than among respondents who only responded to wave one.

Methods Decomposing variation in ordinal response data requires us to make some scaling and modelling assumptions. We assume an ordinal probit response model, where respondents have latent continuous responses (Y^*) which map on to categorical responses (Y) according to cutpoints. Latent responses are turned into manifest responses according to per-question cutpoints which are drawn from a uniform prior subject to the ordering constraint. Our model for the latent response

---

[2]Response options were ordered from left- to right-wing for some questions, and right- to left-wing for others.

takes the form:

$$Y_{ijt}^* = \beta_j \theta_i + \nu_{ij} + \epsilon_{ijt}$$

where i indexes respondents, where j indexes questions, where t indexes the survey wave ($t = 1, 2$) and where the parameters in the model are drawn $\theta_i \sim N(0,1)$, $\nu_{ij} \sim N(0, \omega_j^2)$, and $\epsilon_{ijt} \sim N(0, \sigma_j^2)$.

The first term in the model ($\beta_j \theta_i$) is the *ideological* component, operationalised as a unidimensional spatial model where β_j describes how the respondent's latent ideal point θ predicts her latent response on issue j. The signs of θ and β are both arbitrary and unimportant for the variance decomposition. The second term in the model (ν_{ij}) is a respondent-by-issue random effect. This term captures *idiosyncratic* variation in preferences which is not attributable to ideology, but which is nonetheless stable across repeat measurements t. The third term in the model (ϵ_{ijt}) is an error term capturing *instability*, variation in an individuals' manifest responses across repeated measurements. This *instability* could include response changes from several hypothesized mechanisms, including instability as sheer error (Achen 1975), as sampling from top-of-the-head considerations (Zaller and Feldman 1992), as "flipping a coin" (Converse 1964), or 'real' opinion change in the brief time between survey waves (which we discuss further in the conclusion).

We are not interested in the sign or magnitude of these terms for any individual respondent, but rather in their variance per survey-item across respondents. In the case of the respondent-by-issue random effect and the error term, we use ω_j^2 and σ_j^2 to refer to these variances. The variance of the ideological component is β_j^2, which follows from the identity $Var(cx) = c^2 Var(x)$ and the fact that $\theta_i \sim N(0,1)$.

Because these sources of variation are independent by assumption, the total variance per survey-item can be expressed as $\Sigma_j = \beta_j^2 + \omega_j^2 + \sigma_j^2$. The share of the total variance attributable to ideology can therefore be expressed as β_j^2/Σ_j; the share attributable to idiosyncrasy as ω_j^2/Σ_j, and that attributable to instability σ_j^2/Σ_j. This three-way decomposition of latent scale variance is similar in spirit to the McKelvey and Zavoina (1975) pseudo-R^2 statistic for the ordinal probit regression model, which estimates the R^2 that would be recovered via linear regression on the unobserved latent variable.

The priors for β_j, ω_j and σ_j are standard normal, half-normal, and half-normal, respectively. These priors are chosen so that the prior over the decomposition ($\beta_j^2/\Sigma_j, \omega_j^2/\Sigma_j, \sigma_j^2/\Sigma_j$) for each issue j is

a symmetric Dirichlet distribution with parameters equal to 1/2, which is Jeffrey's non-informative prior.[3] These priors correspond to a weak prior expectation for an equal split of the variance between ideology, idiosyncrasy and instability, and any deviation from this in our results reflects evidence from the data. The model was estimated using Stan 2.16.0 (Carpenter et al. 2017).

Results   Figure 1 shows the estimated composition of opinion variation for each issue in the data set that we examine. Averaging across issues, we estimate that the fraction of variation accounted for by ideology is low (14%, 95% interval: 11-17) when compared to that explained by idiosyncrasy (42%, 95% interval: 38-45) or instability (45%, 95% interval: 41-48). Most of the stable preference variation across individuals and issues is not organised along a single political dimension.

The four most ideological issues are those related to the size of government: Medicare, Taxes, Social Security and Health. However, the four issues with the most stable preferences are mostly different issues: Immigration, Marijuana, Gay Rights and Health. This highlights the distinction between citizens having stable preferences on issues and having ideological preferences. Respondents to this study have real and measurable views about marijuana and immigration (relatively low instability) but those views are almost entirely unpredictable on the basis of a unidimensional summary of their other positions. Stated preferences regarding unions (and to a lesser extent education and the environment) are both lacking in ideological structure and highly unstable; these are the issues on which most respondents seem to lack real preferences.

We have conducted a large number of robustness checks on variant forms of the model and data set (a full tabulation of these is provided in an online appendix). When we simulate data from the estimates and data generating process of the model and refit the model, we recover the estimated parameter values and decomposition almost exactly, so there is negligible estimator bias. If we omit the spatial term and decompose preference variation into only idiosyncrasy and instability, we get nearly identical estimates for instability, and the estimates of idiosyncrasy absorb the variation associated with ideology in the full model. If we omit the idiosyncrasy term instead, the share estimated for ideology absorbs some of the stable variation, increasing from 14% to 18%.

If we analyse only the first wave of data, we get nearly the same estimate for the share of variation

---

[3]If β_j, ω_j and σ_j have prior distributions that are standard (half-)normal, their squares are distributed χ^2_1, which is equivalent to $\Gamma(1/2, 2)$. The simplex constructed by dividing three $\Gamma(1/2, 2)$ random variables by their sum is distributed $D(1/2, 1/2, 1/2)$.

associated with ideology, but we can no longer distinguish between idiosyncrasy and instability. For comparison purposes, this one-wave model can also be applied to Broockman's accompanying survey of state legislators. Averaging across the policy issues about which both state legislators and citizens were asked, we estimate that ideology accounts for 55% of the variation among state legislators, versus 15% for citizens. This contrast illustrates that it is possible to estimate high shares of variation for ideology under this approach, as well as quantifying the well-known, very large difference between the degree of spatial structure in the positions of elites and citizens.

Adding a second spatial dimension to the ideological component of the model increases the fraction of variation explained by ideology from 14% to 24% (95% interval: 21-28) while leaving that due to instability unchanged. The fact that adding a second dimension nearly doubles the variance explained by ideology reinforces the point that a modest fraction of the variation in public opinion is captured by a single dimension, despite a two-party system that one might expect to organize opinion unidimensionally.

Our estimates are somewhat sensitive to collapsing the extreme policy categories, with the ideology component rising to 16% when we collapse the two most extreme categories on each side to generate five point scales, and 20% when we collapse all categories on each side of the central status quo category to generate three point scales. The increase in the ideology component is entirely at the expense of the inconsistency component, indicating that there is more structure and stability in the *direction* of change that respondents would like than in exactly *how far* they would like policy to move. Finally, we have analysed a version of the model where θ_i is allowed to change at the individual-level between the two periods as an alternative to the unstructured change captured by the instability term of the model, but we see very little evidence of ideologically coherent movement between these two survey waves.

Previous research has suggested that individuals fall into different groups which exhibit different levels of opinion stability (Hill 2001; Hill and Kriesi 2001) and opinion structure. Figure 2 shows that, when we split the sample into low and high knowledge individuals, and allow for different ideological dimensions in each half of the sample, we find that high knowledge individuals have more ideologically structured preferences on all but two issues and have less response instability on all issues. Averaging across issues, ideological structure explains 23% of variation among high knowledge respondents versus 12% among low knowledge respondents, with a 95% posterior interval for

Idiosyncrasy

100

80

20

40

Immigration ●
Marijuana ●

60

Gay Rights ●

● Guns

Health ●
Abortion ●
Social Security ●
Contraception

● *Average*

60

40

● Education

Taxes ●
Environment ●

Unions

80

Medicare ●

20

Ideology
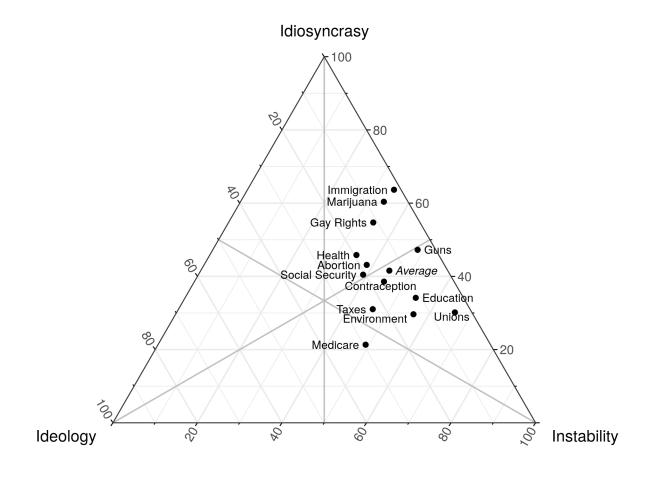
20

40

60

80

100

100

Instability

Figure 1: Ternary plot showing the composition of the variance for each issue. Medicare is the most ideological issue, immigration is the most idiosyncratic, unions is the most unstable.
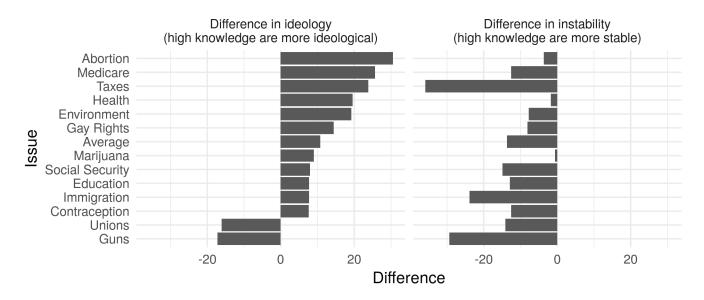
Figure 2: Difference in the percentage of opinion variation explained by ideology and idiosyncracy for high-knowledge respondents versus low-knowledge respondents.

the difference running from 6% to 16%. Response instability accounts for 38% of response variation among high knowledge respondents versus 51% among low knowledge respondents, with a 95% posterior interval for the difference running from -19% to -8%. The single ideological dimensions that best predict responses among low and high knowledge respondents are different: among low knowledge respondents, attitudes on guns and unions are the attitudes that most strongly predict other issue preferences, whereas they are among the least predictive among high knowledge respondents.

Discussion and Conclusion   These estimates will not come as a great shock to public opinion researchers. As we noted at the outset, the ideas that citizens do not organise their views as strongly as political parties and elites, and are sometimes unstable in the preferences they express, are hardly novel. The value of this paper is in providing a quantitative decomposition: *how much* of the preference variation we see is due to each of these factors? On average, in a particular survey instrument querying a particular set of issues at a particular moment in US history, about 1/7 of cross-sectional variation reflects a common ideological dimension of variation, 3/7 is real but idiosyncratic variation in individuals' views, and 3/7 is response instability.

The substantial variation across political issues in the degree to which citizens have "real preferences" (either ideological or idiosyncratic), plus the further variation in the extent to which those

preferences are ideological rather than idiosyncratic, indicates that our results might change substantially with a different sample of issues. Our results might also change with different response categories or different wordings of those response categories – and we note Broockman's own cautionary remarks regarding the choice of policy options (Broockman 2016, 187) – but we have no reasons to believe that differently (or better-) worded response options would affect our results in any particular direction. The issues chosen include a variety of types of issues (some primarily economic, some primarily sociocultural), and all of the issues are sufficiently prominent for researchers to be able to identify the preferred positions of elite actors (Broockman 2016). If lower profile issues were chosen instead, we would expect the estimated ideological and idiosyncratic shares of the decomposition to decline and instability to increase.

The variation we observe by political knowledge levels suggests that the results might also vary in a more representative sample. Those who responded to the second survey wave had somewhat higher levels of knowledge than those who only responded to the first wave. This suggests that our findings, if anything, over-estimate the portion of variation accounted for by ideology, and under-estimate the portion of variation accounted for by instability.

In addition to the ways that our estimates are context dependent because of the structure of the survey data we work with, they are also model dependent because of the way we structure the analysis. While the model we specify enables an elegant variance decomposition, it makes certain functional form assumptions in order to enable that decomposition. While we think these are fairly innocuous with respect to the general conclusions, the exact estimates would surely change under a different approach to decomposition. Because the data we work with has only two waves, we also cannot distinguish instability from real opinion change, which is a fourth potential source of variation in response, albeit one that is likely to be more important over longer time scales than six weeks. The framework described above can be easily extended to perform a four-way decomposition given additional survey waves. The key to doing so is the logic used by Converse (1964): if opinion change is real, opinions in adjacent waves will be more similar than those in non-adjacent waves. In the context of our model, this would most naturally take the form of a normally distributed random-walk process, for each respondent, for each item. We have also set aside possible opinion processes that would fundamentally undermine the logic of this sort of decomposition, such as the possibility that opinion on one issue might predict changes in opinion on other issues across waves (dynamic

constraint).

In sum, our findings put quantitative figures behind the argument made by Broockman (2016) that real idiosyncratic preference variation is much greater than real ideological preference variation among the US public. This is important not only for adjudicating between theories of public opinion, but also because of the widespread use of unidimensional models, both formal and informal, to understand interactions betweens elites and the public. The fact that roughly 3/4 of real public opinion variation does not reflect a unidimensional spatial dimension, and the fact that a second spatial dimension is nearly as powerful as the first, does not necessarily undermine the use of such models. There is, by construction, always a single dimension that captures more opinion variation than any other, and positions on that dimension will therefore always be important to electoral interactions. However, conclusions drawn from models focusing exclusively on a single dimension may or may not be robust to the fact that this dimension captures only a small proportion of the real opinion variation on which basis citizens may make choices and to which elites may have opportunities to respond.

References   Achen, Christopher H. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69 (4): 1218–31.

Broockman, David E. 2016. "Approaches to Studying Policy Representation." *Legislative Studies Quarterly* 41 (1): 181–215.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1).

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Its Discontents*, edited by David Apter. New York: Glencoe Free Press.

———. 2000. "Assessing the Capacity of Mass Electorates." *Annual Review of Political Science* 3: 331–53.

Duckitt, John, and Chris G Sibley. 2009. "A Dual-Process Motivational Model of Ideology, Politics, and Prejudice." *Psychological Inquiry* 20 (2-3): 98–109.

Evans, Geoffrey, Anthony Heath, and Mansur Lalljee. 1996. "Measuring Left-Right and Libertarian-

Authoritarian Values in the British Electorate." *British Journal of Sociology*, 93–112.

Feldman, Stanley. 2013. "Political Ideology." In *The Oxford Handbook of Political Psychology*, edited by L. Huddy, D.O. Sears, and J.S. Levy. Oxford University Press.

Feldman, Stanley, and Christopher Johnston. 2014. "Understanding the Determinants of Political Ideology: Implications of Structural Complexity." *Political Psychology* 35 (3): 337–58.

Hill, Jennifer L. 2001. "An Extension and Test of Converse's 'Black-and-White' Model of Response Stability." *American Political Science Review* 95 (02): 397–413.

Hill, Jennifer L, and Hanspeter Kriesi. 2001. "Classification by Opinion-Changing Behavior: A Mixture Model Approach." *Political Analysis* 9 (4): 301–24.

Jacoby, William G. 1995. "The Structure of Ideological Thinking in the American Electorate." *American Journal of Political Science*, 314–35.

Jost, John T, Christopher M Federico, and Jaime L Napier. 2009. "Political Ideology: Its Structure, Functions, and Elective Affinities." *Annual Review of Psychology* 60: 307–37.

Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53 (4): 821–37.

McKelvey, Richard D, and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4 (1): 103–20.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zaller, John, and Stanley Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions Versus Revealing Preferences." *American Journal of Political Science* 36 (3): 579–616.