Measuring Policy Positions in Political Space

Michael Laver

Department of Political Science, New York University, New York, New York 10003; email: michael.laver@nyu.edu

Annu. Rev. Polit. Sci. 2014. 17:207–23

First published online as a Review in Advance on February 10, 2014

The *Annual Review of Political Science* is online at polisci.annualreviews.org

This article's doi: 10.1146/annurev-polisci-061413-041905

Keywords

metrics, scaling, dimensionality, surveys, text, roll-call voting

Abstract

Spatial models are ubiquitous within political science. Whenever we confront spatial models with data, we need valid and reliable ways to measure policy positions in political space. I first review a range of general issues that must be resolved before thinking about how to measure policy positions, including cognitive metrics, a priori and a posteriori scale interpretation, dimensionality, common spaces, and comparability across settings. I then briefly review different types of data we can use to do this and measurement techniques associated with each type, focusing on headline issues with each type of data and pointing to comprehensive surveys of relevant literatures—including expert, elite, and mass surveys; text analysis; and legislative voting behavior.

INTRODUCTION

Useful models of politics offer realistic yet simple accounts of human thought, behavior, and interaction. A common simplification describes politics in terms of "dimensions" of similarity and difference, such as left–right and liberal–conservative. Not only are spatial metaphors such as these pervasive in everyday political discourse, but spatial models are also ubiquitous within political science, following seminal contributions by Hotelling, Black, and Downs (Black 1948, Downs 1957, Hotelling 1929). Whenever we confront these spatial models with data, we need valid and reliable ways to measure policy positions in political space.

A person may hold several different types of policy position. We can think of his or her possibly subconscious "true" needs; likely conscious sincere private desires; publicly stated positions; policies the person might actually implement, given the chance; and policies actually implemented. Every model of strategic behavior, for example, is based on the fundamental distinction between unobservable private desires and observable public behaviors. Measuring published policy positions is in principle straightforward. Drawing valid inferences about fundamentally unobservable private desires, however, requires a behavioral model that links private desires to public behavior (Benoit et al. 2009[1]). I first review general issues we must resolve before we even think about how to measure any policy position. I then briefly review different types of data we can use to do this and the measurement techniques associated with each, focusing on headline issues with each type of data and pointing beyond these to comprehensive surveys of relevant literatures.

STRUCTURE AND DIMENSIONALITY OF POLICY SPACES

Are Policy Spaces Theoretical Artifacts or Behavioral Models of the Human Brain?

Jack and Jill have similar preferences; Jack and Jolene have divergent preferences. We say Jack is closer to Jill than to Jolene because Jack is more similar to Jill. Conceptually, our notion of similarity/closeness uses point locations to describe objects of interest, a set of dimensions to describe these locations, and a metric for measuring distances between points. This gives us a space spanned by dimensions that describe relative positions of, and movement by, people who interest us (Laver & Benoit 2006). Methodologically, given a matrix of similarities and differences between pairs of people, we can use some scaling technique to generate a spatial representation of relationships between them. But are the resulting policy spaces just ways to describe such relationships, as we describe positions in a chess game by listing the coordinates of each piece on the board? Or are they behavioral models of how real humans think? Do the chess pieces have brains of their own? Humphreys & Laver (2010) distinguish "weakly" spatial models, which merely characterize people's preferences, from "strongly" spatial models, which set out a behavioral theory of how real people think about politics. The notion of conceptual spaces as models of how humans think is grounded in a long research tradition in cognitive science (Aisbett & Gibbon 2001; Attneave 1950; Gärdenfors 2000, 2004; Shepard 1987, 1991).

Look at the 16 squares in the **Figure 1*a***. What similarities and differences do you see? I notice the attributes of rotation and density. **Figure 1*b*** organizes these squares using my two dimensions of perceived similarity and difference. The horizontal dimension analyzes density; the vertical dimension analyzes rotation. To me, these dimensions appear separable, by which I mean that my perception of the density of each object is independent of its rotation. Rotating any square does not

[1] See especially figure 1.

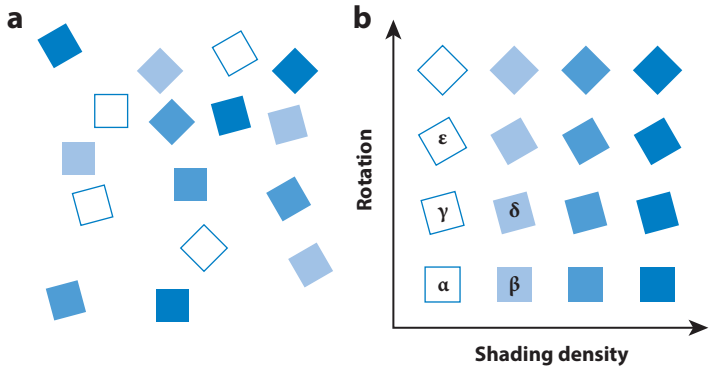

Figure 1

Haphazard arrangement (*a*) and conceptual space (*b*) with 16 objects. Suppose α is a voter and β, γ, δ, and ε are political parties. By construction, since rotation is a dimension of difference, α is more like γ than like ε. But is α more like δ than like ε? See discussion in text.

in itself make the square look darker to me—if it did, these two dimensions would be nonseparable. Look again at **Figure 1*b*** and think about the relative distance from object α of objects β, γ, δ, and ε. Perhaps α is a voter and β, γ, δ, and ε are political parties. By construction, since rotation is a dimension of difference, α is more like γ than like ε. But is α more like δ than like ε? It is certainly closer to δ on the page, but do you yourself actually see α as being more like δ than like ε? I think α looks more like ε than δ. Why? Apparently I attach more weight to shading than to rotation. If you like everything lined up the same way, you may find rotation more salient than shading. If our conceptual spaces are more than one-dimensional, the similarity/distance we see between pairs of objects depends on the relative salience of the dimensions of difference.

Even assuming the two dimensions of difference in **Figure 1** are equally salient for everyone, it may still be that somebody in position α does not see δ as being closer than ε. Both δ and ε differ from α by two units—ε by two units on the rotation dimension, δ by one unit of rotation and one unit of shading. Why would we assume that real humans measure the distance between α and δ using Euclidean geometry and Pythagoras' Theorem, as the square root of the sum of the squares of the difference between them on each dimension? Why would they not simply add up the differences between α and δ on each dimension of difference, using a "city block" conceptual geometry? In the latter case, δ and ε are equidistant from α. The choice of distance metric is another purely empirical question concerning human perceptions of similarity and difference. It is important theoretically because majority rule equilibria in multidimensional settings are more likely if agents perceive similarities and differences in city-block rather than Euclidean terms (Humphreys & Laver 2010, Kadane 1972, Kats & Nitzan 1977, McKelvey & Wendell 1976, Rae & Taylor 1971, Wendell & Thorson 1974). Empirical research by cognitive scientists on perceptions of nonpolitical objects suggests strongly that, if dimensions of perceived difference are separable in the sense I defined above, then the city-block metric is a better measure of human perceptions (Attneave 1950; Gärdenfors 2000; Shepard 1987, 1991). When the dimensions of difference are nonseparable—if your perception of the darkness of the squares in **Figure 1** does indeed depend on their rotation—then Euclidean distances are a better way to measure how you see things. The limited empirical research on perceptions of similarity and difference between political stimuli has tended to conclude that city-block distances better fit the data (Enelow et al. 1988, Grynaviski & Corrigan 2006, Westholm 1997). Notwithstanding all of this, city-block metrics

are rarely used by political scientists, either in theoretical modeling or in empirical measurement (Eguia 2013 provides a comprehensive synthesis of the issues).

CMP: Comparative Manifestos Project

Substantive Meanings of Policy Dimensions: A Priori or A Posteriori Interpretation?

The next question is whether policy dimensions have any a priori substantive meaning, independent of the particular setting under investigation, or are merely dimensions of similarity and difference within a specified set of agents. I return to text analysis below but for now illustrate this difference by comparing two different ways to measure party policy positions using election manifestos.

One is the left–right scale developed and called RILE by the Comparative Manifestos Project (CMP) (Budge et al. 2001, 1987; Klingemann et al. 1994, 2006; Laver & Budge 1992). Sentences in manifestos are assigned by expert coders to one of 56 substantive policy categories, specified a priori as part of the design of the project. The RILE scale is defined by assigning some coding categories to the left, other categories to the right. The substantive meaning of this scale is in this way fixed a priori by the content of the categories assigned to right or left. Two features of a priori scales are crucial. First, substantive scale definitions remain constant across different settings. Second, adding or subtracting parties does not affect the scale positions of other parties. In a sharp contrast with the a priori approach, Gabel & Huber (2000, p. 96) proposed a different way of analyzing the same data: ". . .the left-right dimension is defined inductively and empirically as the 'super issue'. . .that best accounts for the observed covariation among the. . .policy categories across parties." This type of policy dimension derives no substantive meaning from its policy content. It is simply the dimension best distinguishing a given set of parties. Unlike an a priori scale, the substantive content of which is specified absent any knowledge of stimuli to be measured, the substantive meaning of this inductive scale may change from setting to setting, can only be supplied a posteriori, and depends entirely on which stimuli are analyzed. A different set of party manifestos likely generates a different dimension of difference, hence a different scale. Crucially, adding or subtracting a party changes estimated scale positions of other parties.

The distinction between a priori and a posteriori approaches to specifying policy dimensions arises whenever we set out to measure agents' policy positions, whether we use surveys, text, or roll-call votes. We can specify substantive meanings of policy dimensions in advance and then find the best way to measure positions on these. Or we can measure generic similarities and differences between agents in some way, scale the resulting distance matrix, and then interpret the substantive meaning of the resulting inductive "policy" dimensions.

Comparability of Scale Meanings Across Settings

The identity and meaning of most policy dimensions change across space and time, an argument closely identified with Ronald Inglehart (Inglehart 1997, 2008; Inglehart & Flanagan 1987; Inglehart & Welzel 2005). "Left" and "right," in particular, mean different things in different settings, as argued by Inglehart & Klingemann (1976) with regard to voter surveys and supported by subsequent studies (Zechmeister & Corral 2012). Laver & Benoit (2006), for example, document big differences between countries in the substantive policy scales best predicting expert judgments of positions on a general left–right dimension. In some countries, such as the Netherlands, Norway, and Sweden, economic policy is the best predictor of left and right. In others, such as Austria, Hungary, and Japan, the best predictor is the liberal–conservative dimension of social policy. We must therefore be very careful when making transnational comparisons of left–right scale positions because the "same" scale may have different meanings in different countries. What looks like a center-right position on health or social welfare policy to Northern European voters, for example,

may look like left-wing socialism in certain parts of North America. Similar problems apply to more specific policy dimensions, in particular those dealing with controversial social issues such as immigration or gun control, where mainstream social norms differ substantially between settings.

The argument that scale meanings, hence policy positions, should not be compared across different political settings is a smaller problem than it might seem at first sight. Few research designs actually require using the same scale to measure policy positions in different countries. We may, for example, wish to test a spatial model of party competition in different settings. Confronting model predictions with empirical data requires us to compare measured policy positions within a country at a given time, but not between countries. Shifts in scale meanings are much more serious for dynamic models of politics, however, which intrinsically concern changes in policy positions within the same setting. If scale meanings change over time, we cannot distinguish changes in measured policy positions on some scale from changes in the scale itself.

Common Spaces Containing Different Classes of Agent

Empirical evaluations of spatial models often require measuring positions of different classes of political actor in a common space. For example, if we want to make an empirical statement about whether Party A is closer to Voter V than is Party B, then we need A, B, and V in the same space. We need the "same" dimension to mean the same thing for different classes of agent, voters and party leaders in this case, a difficult issue even when we use the same method for estimating their policy positions. The problem is compounded if, as is often the case, different types of data are suitable and/or available for different classes of agent. For example, we may want to use mass opinion surveys to measure the positions of voters but use party manifestos or legislative speeches to measure those of party leaders. Putting two classes of political actor on the same scale may involve critical but heroic rescaling assumptions, for example that voters and party leaders have distributions of policy positions with the same means and variances. One possibility for voters and parties is to use both self-placements and party placements by the same respondents in mass opinion surveys (Alvarez et al. 2000). The obvious problem with this is one of endogenous party placement, whereby voters place parties they support close to their own positions, making it difficult to separate cause from effect when empirically modeling the voting decision. Poole & Rosenthal (1984) make the strong assumption of a proximity voting model and place presidential candidates and voters in the same policy space by scaling survey-derived thermometer scores of how respondents feel about candidates. Schofield & Sened (2005) adopt the ingenious approach of placing voters in a space using mass survey responses, then placing parties in the same space by asking experts to use manifesto content to complete the election study questionnaire on behalf of author parties. This approach avoids endogeneity problems by generating independent estimates of party and voter positions in the same space. It is labor intensive and has not yet been widely adopted, however.

Dimensionality of Policy Spaces

Another deep and difficult question concerns the dimensionality of any policy space we might want to use. How many dimensions of policy do we need for a good representation of some political setting that interests us? There is no definitive answer, no "one true dimensionality" for any setting (Benoit & Laver 2012). The political actors who interest us have preferences on many different matters, and the political spaces that describe these preferences have many different dimensions. Because we always seek a model of politics realistic enough to be substantively interesting yet simple enough to be tractable and intuitive, we seek the lowest dimensional representation that retains substantive plausibility while enabling tractable analysis. This is an exercise in data reduction, and

the data-reduction techniques of factor and principal components analysis give us good intuitions about dimensionality.

We can give substantive meaning to a low-dimensional representation of some setting if, despite a large number of issue dimensions, we find bundles of issues such that preferences on one issue can reliably be predicted from positions on others in the bundle. Think about the legal status of abortion, marijuana, and same-sex marriage. Empirically, if in some setting we know peoples' positions on abortion and marijuana, we may be quite successful at predicting their positions on same-sex marriage. We may also find this makes theoretical sense if we can think of a common thread tying these issues together, for example, a liberal-conservative dimension of social policy. The latent dimension is not only a principal component of the higher-dimensional issue space but also gives *substantive* meaning to the bundle of issues it represents. We in effect treat observed preferences on a bundle of issues as if these were all correlated to some unobserved dimension (Converse 1964). This involves a data-reduction exercise followed by a posteriori interpretation of the dimensions that emerge. It also gives us clues about the underlying dimensionality of the high-dimensional issue space under investigation. For example, our multidimensional scaling exercise may suggest that we can explain most of the variance in a 20-variable dataset on issue positions using two latent dimensions, and that adding a third dimension explains hardly any more variance. This may lead us to conclude, inductively, that the original 20-dimensional space can be characterized as having an underlying two-dimensional structure. We will be particularly encouraged if, a posteriori, we can interpret the set of variables that are highly correlated to the first latent dimension as having something substantively meaningful in common, and even more encouraged if the same applies to the second dimension.

When, on the other hand, we specify the set of policy dimensions of interest a priori, before engaging in any data analysis, we at the same time specify the dimensionality of the space that interests us, using whatever independent sources inform our a priori substantive judgments. Having done this, we in effect run the scaling exercise in reverse. We specify unobservable latent dimensions of a priori interest and then seek ways to estimate positions on these using unidimensional scales. We can use many different observable quantities if we want to enhance scale reliability—for example, batteries of related attitude questions in opinion surveys, or political text coded into a range of predefined substantive categories—but the exercise is fundamentally one of unidimensional scaling.

SURVEYS OF EXPERTS, ELITES, AND VOTERS

I now change focus from important conceptual issues that underlie any research project that involves measuring policy positions in political space to a brief overview of the types of data and method that can be used to do this. I begin with various types of survey.

Expert Surveys

Whenever we measure policy positions, our measures must be satisfactory in terms of both reliability and validity. Reliability is straightforward to define and assess. Does the same method used in the same setting tend to generate the same results, even when deployed independently by different scholars? Validity is more difficult to assess. How can we know if our method really measures what we claim it measures, when what we claim it measures cannot be observed? There is considerable room for argument about the content validity of many of the well-known measures that are part of the mainstream of modern political science. This is something of a rhetorical exercise, but one way to make progress is to solicit the wisdom of experts. Aggregated expert judgments can be used as systematic benchmarks for the content validity of other scales, or can

even be used as scales in their own right. This is the core rationale for using "expert surveys" to measure policy positions in political space.

The first modern expert survey of party policy positions was reported in an unpublished PhD thesis (Morgan 1976). The first published expert survey was by Castles & Mair (1984), who asked "leading political scientists" to locate political parties on a single left–right dimension in 17 countries of which they had knowledge. The scale of such expert surveys was expanded by Laver & Hunt (1992), who approached many more experts with a much longer survey asking for judgments of party positions on a range of policy dimensions in 25 countries. This work was followed by Huber & Inglehart (1995), then by Benoit & Laver (2006), all the time expanding expert and country coverage. The expert-survey project has been taken on by the Chapel Hill Expert Survey (CHES) program, which has expanded with each iteration, incorporating scale definitions from the Laver-Hunt and Benoit-Laver surveys, to become a prime source of updated expert-survey data on party policy positions (Bakker et al. 2012, Hooghe et al. 2010, Steenbergen & Marks 2007).[2]

Despite great practical convenience and a claim to substantive authority derived from the collective wisdom of country specialists, expert-survey data are subject to several criticisms (Budge 2003, Mair 1991, Volkens 2007). Some refer to fixable matters concerning wording of survey questions, and the precise time point to which expert judgments apply. A more telling criticism refers to the intrinsic endogeneity of expert judgments. As Budge (2003, p. 109) succinctly puts it: "Do expert judgements reflect intention or behavior?" This is critical because expert-survey estimates are often used as inputs to models that predict behavior conditional on policy positions. It could well be, however, that the expert judgments are conditioned on the very behaviors they are then used to predict. This is a particular manifestation of a more general problem of endogenous data coding in social science: coded data on latent variables, such as levels of democracy or corruption, are likely conditioned by strong priors brought to the coding exercise by the experts who generate the data. This problem is illustrated by a striking example. Benoit (2005, 2010) extended Benoit-Laver expert surveys to British elections of 1995 and 2010. Estimated Labor, Liberal Democrat, and Conservative positions on the 20-point economic policy scale in 1995 were (7.6, 4.9, 14.2), and in 2010 (6.8, 11.5, 12.4). The expert surveys indicate a substantively big and statistically significant rightward shift by the Liberal Democrats. There is no corresponding shift in estimates of Liberal Democrat positions derived independently from content analyses of party manifestos by the CMP and by Benoit et al. (2012). Note that, although manifestos were issued before the 2010 election, expert judgments were collected not only after the election but also after the unprecedented formation of a government coalition between Liberal Democrats and Conservatives. It is clearly possible that experts judged the Liberal Democrats to have moved sharply rightward because they went into government with the Conservatives—not that they went into government with the Conservatives because they had moved sharply rightward. It is unclear in this case what, precisely, expert-survey estimates are measuring, and this may be a more general problem.

Surveys of Legislators

Why not measure policy positions of politicians by asking the politicians themselves? "Elite" surveys of legislators, especially of US state legislators, are a considerable scholarly project. Maestas et al. (2003) review 73 articles reporting surveys of US state legislators, published in top journals

[2]There have also been ad hoc expert surveys for particular elections using Benoit-Laver scales in Britain (Laver 1998a; Benoit 2005, 2010); Ireland (Laver 1994, 1998b); Japan (Kato & Laver 1998, 2003); Italy (Benoit 2008, Giannetti & De Giorgi 2006); the Netherlands (Laver 1995, Laver & Mair 1999); Norway (Ray & Narud 2000); and the European Parliament (McElroy & Benoit 2007, 2012).

between 1975 and 2000. They contrast face-to-face interviews, which have an average response rate of 84%, with telephone and mail surveys, which have an average response rate of 49%. Fisher & Herrick (2013) review another 25 publications on the same theme between 2000 and 2010. They document steeply declining response rates and contrast traditional postal surveys (average post-2000 response rate 32%) with the now-typical internet surveys (11%). The project of surveying legislators (at least in US states) has thus evolved from traditional face-to-face interviews with response rates typically over 80% to internet surveys with response rates barely over 10%. However, Fisher & Herrick find no significant representational bias in low-response internet surveys, implying the core methodological issue is the classic problem of research budgeting. For a fixed research budget, many more legislators can be contacted and asked to complete internet surveys, and the response rate will be much lower. If we indeed find no bias arising from low response rates associated with internet surveys, it might be that internet surveys yield more information per research dollar than face-to-face interviews or postal surveys.

Turning to Europe, there is no overview of legislator surveys across the range of European countries, although an ongoing research program of the European Parliament Research Group (EPRG) fielded surveys of Members of the European Parliament (MEPs) in 2000, 2006, and 2010 (Farrell et al. 2011, Hix et al. 2011).[3] The EPRG consolidated surveys of earlier directly elected European Parliaments: 1979–1984 (Bardi 1989), 1984–1989 (Hrbek & Schweitzer 1989), 1989–1994 (Bowler & Farrell 2008), and 1994–1999 (Katz &Wessels 1999). Although data arising from more recent MEP surveys are readily available from the EPRG, relatively little use has been made of MEPs' surveyed preferences to date, in contrast to uses made of MEP roll-call data, discussed below. To the extent that MEPs are representative of their national parties, these data have the potential to form the basis of a dataset on European party positions, a potential undermined somewhat by the tiny numbers of surveyed MEPs who belong to smaller European parties. Summary tables in Hix et al. (2011) hint at what is possible.

Saiegh (2009) provides a comprehensive overview of legislator surveys in 18 Latin American countries. He analyzed nine of these to recover a common space for legislators as well as country-specific spaces, cross-validating these against analogous estimates derived from roll-call voting data (**http://www.uiowa.edu/~lsq/Saiegh_Appendix**). He also references surveys of national legislators in 26 non-Latin America countries (pp. 124–25), and concludes with strong advocacy for the use of legislator surveys (pp. 137–38).

There is an understandable suspicion within the scholarly community that those few legislators who do respond to surveys are atypical and possibly strategic in their responses. Nonetheless, data derived from surveys of legislators' policy positions—analyzed carefully in a way that takes account of these possible problems—are underutilized by scholars who confront theoretical models with positions of real politicians. Despite low response rates, such surveys are not in short supply. Time, effort, and expense have gone into fielding them and compiling results; these results bear directly on the project of measuring policy positions. With the exception of the EPRG data on MEPs, however, results of legislator surveys do not seem to have been consolidated into a single easily accessible resource, and this may explain their underuse.

Surveys of Voters

If we want to measure policy positions of voters, the dominant data resources are mass opinion surveys, whether commercial polls or full-dress academic election studies. The University of

[3] **http://www2.lse.ac.uk/government/research/resgroups/EPRG/MEPsurveyData.aspx**.

Michigan has long been at the center of this enterprise, and good windows on what is available can be found at two Michigan-based projects: the Inter-university Consortium for Political and Social Research (ICPSR) and the Comparative Study of Electoral Systems (CSES). The ICPSR is a well-known, authoritative, and widely used data depository of record, offering a large and searchable collection of survey datasets (**http://www.icpsr.umich.edu/icpsrweb/landing.jsp**). The CSES website (**http://www.cses.org**) also links to a huge array of other survey sources: Afrobarometer, Arab Barometer, AsiaBarometer, Latinobarometro, Eurobarometer, World Values Study, European Values Study, International Social Survey Program, European Social Survey, European Election Studies, and many more besides. The survey data resource is thus vast and sprawling before we even start to look at commercial polling. Researchers who want to measure voters' policy positions in a comparative context face the problem of managing and exploiting huge existing deposits of data. The obvious danger is that key research findings may not be robust to detailed design decisions about how to specify, extract, and organize a particular dataset from the huge range of resources on offer.

Assuming we have a nice clean dataset of voters' survey responses to thoroughly tested attitude questions or voter self-placements on synthetic scales, what comes next? The well-known problem with voter self-placements concerns "centering" the scale across space and time, lucidly expressed by Kim & Fording (1998). "Since the 'left' and 'right' exist in the mind of the respondent with respect to the location of the 'center', self-placement scores can only be compared across countries and time periods if the location of the middle can be assumed to be the same in some absolute sense. Because it is likely that this is not the case, aggregate country measures of left-right self-placement are likely to be very inadequate measures of ideology" (Kim & Fording 1998, p. 75). The Kim-Fording solution is to ignore voter surveys, estimate party positions (for example, using manifesto data), assume a naive spatial-proximity model of voting, and infer the left–right positions of segments of voters from the midpoints of parties' scale positions. Recall, however, that many research designs do not require comparability of scale meanings across settings, so that context-specific survey-based estimates of voters' policy positions may still be valuable and are independent of any assumed model of voting behavior.

All surveys contain batteries of items seeking respondents' positions on a range of specific issues, many of which are associated with widely used latent dimensions such as left–right economic policy, liberal–conservative social policy, and so on. Opinion surveys also typically contain items asking respondents to place themselves on more general synthetic scales such as the life-right scale. The advantage of building scales from responses to sets of precise survey questions on particular issues, such as same-sex marriage, legalization of marijuana, or capital punishment, is that the substance of the scale is explicitly defined a priori. The disadvantage is that combining responses to different survey items typically requires a brute-force assumption that any given item has the same salience for every respondent—or at least that the aggregation of respondents' positions on different items does not distort their relative positions on the latent dimension being estimated. Asking survey respondents to place themselves directly on latent dimensions such as left–right avoids such potential problems. The sore point is now that we do not know a priori what each respondent has in mind when thinking of left and right. As always, there is no "one true way" to do things, only a requirement for scholars to think hard about why they want to measure policy positions and what precisely they want to measure.

TEXT

Communication is fundamental to human interaction. Communication generates text, which may be written down by some author or arise from transcribed speech. Text is everywhere,

in extraordinary volumes. Much of it is now easy to capture and analyze using high-powered information-processing systems that can add massive leverage to human expertise. Fundamental implications of this for modern political science are only beginning to be realized. Analyzing gigantic volumes of text requires automated systems, but all valid automated systems require expert human input at some point.

If we want to measure policy positions of political parties, a natural way to start is to analyze official statements of these positions in election manifestos. If we want to measure policy positions of individual legislators, we can analyze their authored political texts or spoken words. Given the widespread adoption of various social media platforms and the comprehensive self-publication opportunities these offer, we can measure policy positions of individual citizens who use social media by analyzing the bodies of text they generate. In each of these settings, we can analyze texts by using computational methods or human coders, although the exploding volume of text generated as we move from parties to legislators to citizens means that automated methods become the only feasible means to begin analysis. This is particularly evident in recent work on text generated by social media such as Twitter.

The CMP is a longstanding research program measuring party policy positions using human experts to code party manifestos (Budge et al. 1987, 2001; Klingemann et al. 1994, 2006; Laver & Budge 1992). As of July 2013, its five books had generated about 4,000 citations according to Google Scholar. The Policy Agendas Project reports more recent work in the same tradition, based on human expert coding of party policy documents (Baumgartner et al. 2007).

The first job in any text-coding project is "text unitization," the specification of the text units to be coded, whether these units are words, sentences, paragraphs, or entire documents. CMP researchers unitize party manifestos by parsing them into "quasi-sentences," each a "verbal expression of one political idea or issue" (Volkens 2001). Judging what is "one political idea" is of course a fundamentally subjective exercise. Having unitized the text, a single trained CMP coder assigns each text unit to one category in a 56-category coding scheme that derives from the project's theoretical underpinnings,[4] of which Dolezal et al. (2014) offer a recent review. The resulting 56-dimensional data matrix can then be scaled by various means to locate party manifestos on latent policy dimensions, by far the most widely used of which is the CMP's left–right scale, RILE, which uses 26 of the dataset's coding categories. CMP data are widely available (**http://manifesto-project.wzb.eu**), are without doubt a prime source of time-series party policy positions within the profession, and form the empirical core of many articles published in top journals.

Scholars have noted a number of problems with CMP data. Parsing texts into quasi-sentences has been shown to be a significant source of intercoder unreliability in CMP data and, furthermore, to generate results with no better content validity than using objectively defined "natural" manifesto sentences—text strings separated by one of a prespecified set of punctuation marks (Däubler et al. 2011). Intercoder reliability has been shown to be relatively low in tests deploying multiple coders using the same CMP coding scheme on the same document (Mikhaylov et al. 2008). Another significant consequence of using a single coder is the lack of standard errors for point estimates of individual variables and aggregate scales (Benoit et al. 2009). The additive scaling model used to derive the RILE scale has also been criticized (Lowe et al. 2011). Although there are retroactive fixes for some of these problems, many could only be resolved by generating multiple human codings for each party manifesto in the CMP text corpus. This solution is

[4]Examples of these categories include "Internationalism: positive," "Incentives: positive," and "Nationalization: positive."

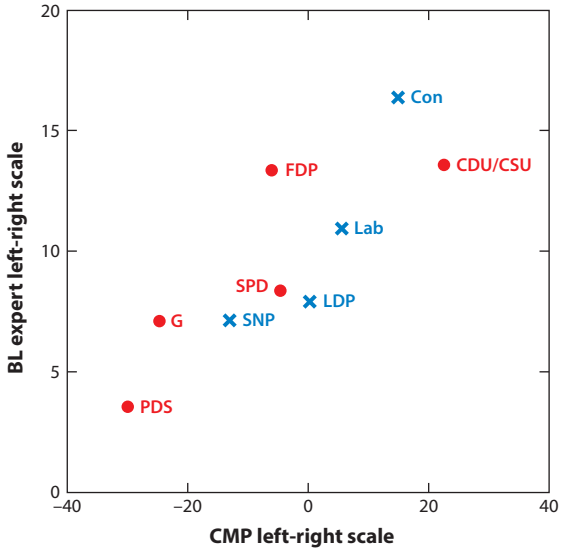

Figure 2

Party positions in Germany 2002 (*red circles*) and Britain 2001 (*blue crosses*), estimated using Comparative Manifesto Project (CMP) and Benoit-Laver (BL) expert-survey scales. Abbreviations: CDU/CSU, Christian Democratic Party; Con, Conservative Party; FDP, Free Democratic Party; G, Green Party; Lab, Labour Party; LDP, Liberal Democratic Party; PDS, Party of Democratic Socialism; SNP, Scottish National Party; SPD, Social Democratic Party.

unrealistic owing to its huge costs, given the large number of manifestos, all the time using a coding scheme that may well not fit the theoretical concerns of many potential end users.

It is of course intrinsic to good science that any dataset or empirical inference is continuously subject to vigorous review and criticism. As we have just seen, the CMP data are no exception to this rule, but they have been fiercely defended by the originators of the project (Budge 2013), on the grounds, first, that they are widely cited and used and, second, that scale positions estimated using CMP data typically correlate highly with alternatives. This led Budge (2013, p. 5) to ask, notwithstanding vigorous methodological critiques of the CMP data, "why bother" with alternatives? Pragmatically, the comprehensive party coverage of the CMP time series, combined with their heavy citation and ready availability, have made them attractive to many third-party scholars whose prime concern is with using an accepted dataset and not with how these data were estimated. At an aggregate level, furthermore, the CMP RILE scale does tend to correlate with independent left–right scales.

Figure 2 plots left–right scale estimates from the CMP against Benoit-Laver expert-survey estimates of left–right party positions in both Britain and Germany around 2002, the time of the Benoit-Laver expert survey. We do see a high aggregate correlation between the different sets of estimates, but we also see important substantive differences, especially for the German parties. The CMP estimates do not distinguish between left–right scale positions of the German Social Democrats (SPD) and Free Democrats (FDP), whereas the experts place the FDP to the right of the SPD. Conversely, the experts do not distinguish between left–right positions of the FDP and the Christian Democrats (CDU/CSU), whereas the CMP data place the CDU/CSU to the right of the FDP. These substantive differences could well arise from differences of scale interpretation rather than error in measuring positions on "one true" latent dimension. Both CMP and expert left–right scales mix aspects of social and economic policy, which experts may mix in different

proportions for Germany in 2002 than does the CMP's uniform 26-category scale definition, which applies to all countries and all times. We need not judge which estimates are "better," but we do see that they differ in substantively important ways, despite their high aggregate correlation. Different conclusions about German politics in 2002 might be drawn, depending on the scale used. We should indeed "bother" about which scale we use, although we must of course be rigorous in justifying our choice of scale on firm a priori grounds rather than "fishing" for the scale that delivers the nicest results (Humphreys et al. 2013).

Carefully conducted expert text coding will always be very resource intensive, however, and is typically not scalable to large research problems, whether these involve multiple human codings of the same texts to generate robust estimates of uncertainty, or the coding of massive text databases. One possible way forward involves crowd-sourced text coding—a human intelligence task that can be deployed, using a much-simplified coding scheme, on crowd-sourcing platforms such as Mechanical Turk (**https://www.mturk.com/mturk/welcome**). This offers a potential way to generate multiple human codings of the same political texts quickly and cheaply, scalable to a huge text corpus, with well-behaved measures of uncertainty (Benoit et al. 2012).

Leveraging human coding, automated methods for analyzing (political) texts have recently been a source of substantial progress within the general endeavor of measuring policy positions. This is a huge and highly interdisciplinary project with obvious payoffs for scholarly, commercial, and intelligence communities, all focused on extracting meaningful and unbiased signal from gigantic volumes of noisy text. Early publications in mainstream political science journals offered simple methods that could easily be deployed by nonspecialists (Laver et al. 2003, Slapin & Proksch 2008), but there is now a large and complex technical literature on "text as data"—the use of automated methods to measure policy positions using political texts. Grimmer & Stewart (2013) provide an excellent and authoritative review of this literature that should be a starting point for scholars thinking of moving into this area. Their table 1 sets out "four principles of quantitative text analysis" that are well worth repeating here. "(1) All quantitative models of language are wrong—but some are useful. (2) Quantitative methods for text amplify resources and augment humans. (3) There is no globally best method for automated text analysis. (4) Validate, Validate, Validate." This is an expanded statement of the well-known "garbage in, garbage out" principle that applies to all computational work. Grimmer & Stewart emphasize the crucial point that automated text analysis is no substitute for deep thought when designing and implementing any research project. In particular, they emphasize that all automated methods require human calibration, validation, and interpretation. These issues become critical in designing projects to analyze vast volumes of social media text, generated for example by Twitter feeds (Barbera 2013, King et al. 2011), because automation is the only feasible way forward and substantial efforts must be made to ensure reliable human calibration and validation of results.

BEHAVIOR: LEGISLATIVE VOTING AND CAMPAIGN CONTRIBUTIONS

The most widely used way to measure policy positions of US legislators derives from roll-call voting data. Prominently published work relies heavily on the NOMINATE method for scaling these (Poole 2005, Poole & Rosenthal 1997). NOMINATE assumes sincere proximity voting in legislatures, which implies that ideal points of two legislators whose voting records are relatively similar are closer together than ideal points of two legislators whose voting records are relatively different. Similarities and differences between voting records can be used to construct a matrix of interlegislator distances, a matrix that can be scaled to generate relative positions of individual legislators in low-dimensional spaces. Relative locations of well-known politicians are then

used to give a posteriori policy interpretations, such as liberal–conservative, to the dimensions of these spaces. Clinton et al. (2004) specify a fundamentally Bayesian and more flexible alternative method for scaling roll-call voting records while also taking account of additional information and modeling assumptions. The two approaches are compared by Carroll et al. (2009) and Clinton & Jackman (2009)—the latter noting that, notwithstanding fundamental differences in the method of estimation, the same assumption about the underlying voter utility function has the effect that "estimates of ideal points are frequently near-identical" (p. 597);. This is why the Poole & Rosenthal (1997, p. 22) estimates of "locations of 11,000 legislators and 70,000 roll calls from the 11,000,000 recorded individual decisions of Congresses stretching from 1789 to 1985" have proved attractive to many end users who have little or no intrinsic interest in methods of ideal-point estimation.

Low levels of party discipline found in the United States mean that researchers can use NOMINATE to measure policy positions of individual politicians. Relatively low levels of "party" discipline in the European Parliament[5] mean that NOMINATE can be used to measure policy positions of MEPs in two-dimensional spaces (Hix et al. 2006, 2007). However, when parties exert strong discipline over their legislators, typically true in European parliamentary systems, NOMINATE mostly tells us which party each legislator belongs to and does not distinguish policy positions of legislators within the same party (Spirling & McLean 2006). This is not the end of the story, however, because party discipline is never perfect and "rebellions" against party discipline can be informative. Spirling & McLean (2007) leverage rebellions using optimal classification analysis to score Members of Parliament (MPs) on a left–right scale. Spirling & Quinn (2010) use a sophisticated clustering algorithm to identify voting blocs of MPs within the British Labor Party, notwithstanding very high levels of party discipline. Although these clusters are not in themselves measurements of policy positions, the authors are then tempted into applying a posteriori interpretations of their relative positions, using positional language such as "leftist" and "socialist."

As with any method for measuring policy positions, analyses of roll-call votes depend on data availability. Roll-call votes are completely absent in some countries—Ireland and Hungary, for example. In countries that do have roll-call votes, as well as in the European Parliament, roll calls may be highly selective and strategic, fundamentally endogenous to the political setting under investigation. Carrubba and coauthors (2006, 2008) review potential biases introduced by "roll call selection," concluding (2006) such biases are "substantively important" and "very substantial."

One way forward in this context is to look beyond roll-call votes for other behavioral indicators of legislators' policy positions. A promising new avenue of research has been opened up by Bonica (2013), who builds on work by McCarty & Poole (1998) using contributions by political action committees to estimate policy positions of both successful and unsuccessful candidates. The obvious potential here is that political contributions more generally, if carefully analyzed, are sources of information about the preferences of both givers and receivers well beyond the set of people who cast roll-call votes.

CONCLUSION

The project of measuring policy positions is vast and sprawling, with many types of position we may want to measure and many methods we might use to do this. It is impossible to resist the worthy platitude that fancy measurement techniques and seductive prepackaged datasets are no substitute for deep thought about what, precisely, we want to measure and why we want to

[5] "Party" in this context refers to European party groups that comprise several national party delegations.

measure it. Too many prominently published papers use inappropriate datasets or measurement techniques because these are readily available and have something to do with policy positions, and it is difficult to assess the extent to which important empirical results are driven by poor choices on these matters. An ideal published empirical finding, of course, will measure policy positions in ten very different ways and always come up with the same result, but this is rarely an option that is open to us. What is always available to us is a careful consideration and justification of the data sources and measurement techniques we deploy, combined with a rigorous emphasis on specifying our measurement choices in advance, rather than fishing in the large pool of available data for the results that best suit us.

Looking beyond the wagging finger, we face extraordinarily exciting prospects as the profession expands its abilities to mine and analyze the enormous data deposits that lie all around us. There is a vast and rapidly growing corpus of electronic text to analyze. There is an explosively growing body of proximity measures generated by social media. These data increasingly have GPS coordinates that allow them to be linked to many other datasets. There is no reason, for example, why we should not set out to measure the policy positions of every single person who uses social media and, with appropriate modeling, to make inferences from these positions about people who do not use social media. Large parts of this endeavor will inevitably involve automated methods. It should be clear from this review, however, that these methods will leverage but never replace the human inputs we will always need to calibrate and interpret their results.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Aisbett J, Gibbon G. 2001. A general formulation of conceptual spaces as a meso level representation. *Artif. Intell.* 133:189–232

Alvarez RM, Nagler J, Bowler S. 2000. Issues, economics, and the dynamics of multiparty elections: the British 1987 general election. *Am. Polit. Sci. Rev.* 94:131–49

Attneave F. 1950. Dimensions of similarity. *Am. J. Psychol.* 63:546–54

Bakker R, de Vries C, Edwards E, Hooghe L, Jolly S, et al. 2012. Measuring party positions in Europe: the Chapel Hill expert survey trend file, 1999–2010. *Party Polit.* doi: 10.1177/1354068812462931

Barbera P. 2013. *Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data*. Presented at Annu. Meet. Midw. Polit. Sci. Assoc., Chicago

Bardi L. 1989. *Il Parlamento della Comunità europea: legittimità e riforma*. Vol. 259. Bologna: Il Mulino

Baumgartner FR, Green-Pedersen C, Jones BD. 2007. *Comparative Studies of Policy Agendas*. London: Routledge

Benoit K. 2005. *Policy positions in Britain 2005: results from an expert survey*. Work. pap., Lond. School Econ.

Benoit K. 2008. *Party positions in Italy 2008: results from an expert survey*. Work. pap., Lond. School Econ.

Benoit K. 2010. *Party policy positions in Britain, 2010: results from an expert survey*. Work. pap., Lond. School Econ.

Benoit K, Conway D, Laver M, Mikhaylov S. 2012. *Crowd-sourced data coding for the social sciences: massive non-expert coding of political texts*. Presented at New Directions in Analyzing Text as Data, Harvard Univ., Oct. 5–6, Cambridge, MA

Benoit K, Laver M. 2006. *Party Policy in Modern Democracies*. London: Routledge

Benoit K, Laver M. 2012. The dimensionality of political space: epistemological and methodological considerations. *Eur. Union Polit.* 13(2):194–218

Benoit K, Laver M, Mikhaylov S. 2009. Treating words as data with error: uncertainty in text statements of policy positions. *Am. J. Polit. Sci.* 53(2):495–513
Black D. 1948. On the rationale of group decision-making. *J. Polit. Econ.* 56(1):23–34
Bonica A. 2013. Ideology and interests in the political marketplace. *Am. J. Polit. Sci.* 57(2):294–311
Bowler S, Farrell DM. 2008. Legislator shirking and voter monitoring: impacts of European Parliament electoral systems upon legislator-voter relationships. *J. Common Market Stud.* 31(1):45–70
Budge I. 2003. Expert judgements of party policy positions: uses and limitations in political research. *Eur. J. Polit. Res.* 37(1):103–13
Budge I. 2013. *The 'paradox of the manifestos'—satisfied users, critical methodologists*. Work. pap., Univ. Essex
Budge I, Klingemann H-D, Volkens A, Bara J, Tannenbaum E, et al. 2001. *Mapping Policy Preferences: Parties, Electors and Governments: 1945–1998: Estimates for Parties, Electors and Governments 1945–1998*. Oxford, UK: Oxford Univ. Press
Budge I, Robertson D, Hearl D. 1987. *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*. Cambridge, UK: Cambridge Univ. Press
Carroll R, Lewis JB, Lo J, Poole KT, Rosenthal H. 2009. Comparing NOMINATE and IDEAL: points of difference and Monte Carlo tests. *Legis. Stud. Q.* 34(4):555–91
Carrubba C, Gabel M, Hug S. 2008. Legislative voting behavior, seen and unseen: a theory of roll-call vote selection. *Legis. Stud. Q.* 33(4):543–72
Carrubba CJ, Gabel M, Murrah L, Clough R, Montgomery E, Schambach R. 2006. Off the record: unrecorded legislative votes, selection bias and roll-call vote analysis. *Br. J. Polit. Sci.* 36(04):691–704
Castles FG, Mair P. 1984. Left–right political scales: some 'expert' judgments. *Eur. J. Polit. Res.* 12(1):73–88
Clinton JD, Jackman S. 2009. To simulate or NOMINATE? *Legis. Stud. Q.* 34(4):593–621
Clinton J, Jackman S, Rivers D. 2004. The statistical analysis of roll call data. *Am. Polit. Sci. Rev.* 98(2):355–70
Converse PE. 1964. The nature of belief systems in mass publics. In *Ideology and Discontent*, ed. DE Apter, pp. 206–61. London: Free Press of Glencoe
Däubler T, Benoit K, Mikhaylov S, Laver M. 2012. Natural sentences as valid units for coded political text. *Br. J. Polit. Sci.* 42(4):937–51
Dolezal M, Ennser-Jedenastik L, Müller WC, Winkler AK. 2014. How parties compete for votes: a test of saliency theory. *Eur. J. Polit. Res.* 53:57–76
Downs A. 1957. *An Economic Theory of Democracy*. New York: Harper
Eguia JX. 2013. Challenges to the standard Euclidean spatial model. In *Advances in Political Economy*, ed. G Caballero, DM Kselman, N Schofield, pp. 169–80. Heidelberg: Springer-Berlin
Enelow JM, Mendell NR, Ramesh S. 1988. A comparison of two distance metrics through regression diagnostics of a model of relative candidate evaluation. *J. Polit.* 50:1057–71
Farrell D, Hix S, Scully R. 2011. *EPRG MEP survey dataset: 2011 release*, ed. EPR Group. Lond. School Econ.
Fisher SH, Herrick R. 2013. Old versus new: the comparative efficiency of mail and internet surveys of state legislators. *State Polit. Policy Q.* 13(2):147–63
Gabel M, Huber J. 2000. Putting parties in their place: inferring party left-right ideological positions from party manifesto data. *Am. J. Polit. Sci.* 44:94–103
Gärdenfors P. 2000. *Conceptual Spaces: the Geometry of Thought*. Cambridge, MA: MIT Press
Gärdenfors P. 2004. Conceptual spaces as a framework for knowledge representation. *Mind Matter* 2(2):9–27
Giannetti D, De Giorgi E. 2006. The 2006 Italian general elections: issues, dimensions and policy positions of political parties. *J. Mod. Ital. Stud.* 11(4):494–515
Grimmer J, Stewart BM. 2013. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Polit. Anal.* 21(3):267–97
Grynaviski JD, Corrigan BE. 2006. Specification issues in proximity models of candidate evaluation (with issue importance). *Polit. Anal.* 14:393–420
Hix S, Noury A, Roland G. 2006. Dimensions of politics in the European Parliament. *Am. J. Polit. Sci.* 50(2):494–520
Hix S, Noury AG, Roland G. 2007. *Democratic Politics in the European Parliament*. Cambridge, UK: Cambridge Univ. Press
Hooghe L, Bakker R, Brigevich A, de Vries C, Edwards E, et al. 2010. Reliability and validity of the 2002 and 2006 Chapel Hill expert surveys on party positioning. *Eur. J. Polit. Res.* 49(5):687–703

Hotelling H. 1929. Stability in competition. *Econ. J.* 39(153):41–57
Hrbek R, Schweitzer C-C. 1989. Die deutschen Europa-Parlamentarier: Ergebnisse einer Befragung der deutschen Mitglieder des Europäischen Parliaments. *Aus. Polit. Zeitgeschichte B* 3(89):3–18
Huber J, Inglehart R. 1995. Expert interpretations of party space and party locations in 42 societies. *Party Polit.* 1(1):73–111
Humphreys M, de la Sierra RS, van der Windt P. 2013. Fishing, commitment, and communication: a proposal for comprehensive nonbinding research registration. *Polit. Anal.* 21(1):1–20
Humphreys M, Laver M. 2010. Spatial models, cognitive metrics, and majority voting equilibria. *Br. J. Polit. Sci.* 40:11–30
Inglehart R. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*, Vol. 19. New York: Cambridge Univ. Press
Inglehart R. 2008. Changing values among western publics from 1970 to 2006. *West Eur. Polit.* 31(1–2):130–46
Inglehart R, Flanagan SC. 1987. Value change in industrial societies. *Am. Polit. Sci. Rev.* 81:1289–319
Inglehart R, Klingemann H-D. 1976. Party identification, ideological preference and the left-right dimension among western mass publics. In *Party Identification and Beyond*, ed. I Budge, I Crewe, D Farlie, pp. 243–73. London: Wiley
Inglehart R, Welzel C. 2005. *Modernization, Cultural Change, and Democracy: the Human Development Sequence*. New York: Cambridge Univ. Press
Kadane JB. 1972. On division of the question. *Public Choice* 13:47–54
Kato J, Laver M. 1998. Party policy and cabinet portfolios in Japan, 1996. *Party Polit.* 4(2):253–60
Kato J, Laver M. 2003. Policy and party competition in Japan after the election of 2000. *Jpn. J. Polit. Sci.* 4(1):121–33
Kats A, Nitzan S. 1977. More on decision rules and policy outcomes. *Br. J. Polit. Sci.* 7(3):419–22
Katz RS, Wessels B. 1999. *The European Parliament, the National Parliaments, and European Integration*. Oxford, UK: Oxford Univ. Press
Kim H, Fording RC. 1998. Voter ideology in Western democracies, 1946–1989. *Eur. J. Polit. Res.* 33(1):73–97
King A, Orlando F, Sparks D. 2011. *Ideological extremity and primary success: a social network approach*. Presented at Annu. Meet. Midw. Polit. Sci. Assoc., Chicago
Klingemann H-D, Hofferbert RI, Budge I. 1994. *Parties, Policies, and Democracy*. Boulder, CO: Westview
Klingemann H-D, Volkens A, Bara J, Budge I, McDonald M. 2006. *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990–2003*. Oxford, UK: Oxford Univ. Press
Laver M. 1994. Party policy and cabinet portfolios in Ireland 1992: results from an expert survey. *Irish Polit. Stud.* 9(1):157–64
Laver M. 1995. Party policy and cabinet portfolios in the Netherlands, 1994. *Acta Polit.* 30(1):3–28
Laver M. 1998a. Party policy in Britain 1997: results from an expert survey. *Polit. Stud.* 46(2):336–47
Laver M. 1998b. Party policy in Ireland 1997: results from an expert survey. *Irish Polit. Stud.* 13(1):159–71
Laver M, Benoit K. 2006. *Party Policy in Modern Democracies*. London: Routledge
Laver M, Benoit K, Garry J. 2003. Estimating the policy positions of political actors using words as data. *Am. Polit. Sci. Rev.* 97(2):311–31
Laver M, Budge I. 1992. *Party Policy and Government Coalitions*. New York: St. Martin's
Laver M, Hunt WB. 1992. *Policy and Party Competition*. New York: Routledge
Laver M, Mair P. 1999. Party policy and cabinet portfolios in the Netherlands, 1998: results from an expert survey. *Acta Polit.* 34:49–66
Lowe W, Benoit K, Mikhaylov S, Laver M. 2011. Scaling policy positions from hand-coded political texts. *Legis. Stud. Q.* 36(1):123–55
Maestas C, Neeley GW, Richardson LE Jr. 2003. The state of surveying legislators: dilemmas and suggestions. *State Polit. Policy Q.* 3(1):90–108
Mair P. 1991. Searching for the positions of political actors: a review of approaches and a critical evaluation of expert surveys. In *Estimating the Policy Positions of Political Actors*, ed. M Laver, pp. 33–49. London/New York: Routledge
McCarty NM, Poole KT. 1998. An empirical spatial model of congressional campaigns. *Polit. Anal.* 7(1):1–30

McElroy G, Benoit K. 2007. Party groups and policy positions in the European Parliament. *Party Polit.* 13(1):5–28

McElroy G, Benoit K. 2012. Policy positioning in the European Parliament. *Eur. Union Polit.* 13(1):150–67

McKelvey RD, Wendell RE. 1976. Voting equilibria in multidimensional choice spaces. *Math. Oper. Res.* 1:144–58

Mikhaylov S, Laver M, Benoit KR. 2012. Coder reliability and misclassification in the human coding of party manifestos. *Polit. Anal.* 20(1):78–91

Morgan M-J. 1976. *The modeling of government coalition formation: a policy-based approach with interval measurement*. PhD thesis, Dep. Polit. Sci., Univ. Mich.

Poole KT. 2005. *Spatial Models of Parliamentary Voting*. New York: Cambridge Univ. Press

Poole KT, Rosenthal H. 1997. *Congress: a Political-Economic History of Roll Call Voting*. New York: Oxford Univ. Press

Poole KT, Rosenthal H. 1984. US presidential elections 1968–80: a spatial analysis. *Am. J. Polit. Sci.* 28(2):282–312

Rae DW, Taylor M. 1971. Decision rules and policy outcomes. *Br. J. Polit. Sci.* 1(1):71–90

Ray L, Narud HM. 2000. Mapping the Norwegian political space: some findings from an expert survey. *Party Polit.* 6(2):225–39

Saiegh SM. 2009. Recovering a basic space from elite surveys: evidence from Latin America. *Legis. Stud. Q.* 34(1):117–45

Schofield N, Sened I. 2005. Multiparty competition in Israel, 1988–96. *Br. J. Pol. Sci.* 35(4):635

Scully R, Hix S, Farrell, DM. 2012. National or European parliamentarians? Evidence from a new survey of the Members of the European Parliament. *J. Common Market Stud.* 50(4):670–83

Shepard RN. 1987. Toward a universal law of generalization for psychological science. *Science* 237:1317–23

Shepard RN. 1991. Integrality versus separability of stimulus dimensions: from an early convergence of evidence to a proposed theoretical basis. In *The Perception of Structure: Essays in Honor of Wendell R. Garner*, ed. JR Pomerantz, GL Lockhead, pp. 53–71. Washington, DC: Am. Psychol. Assoc.

Slapin JB, Proksch S-O. 2008. A scaling model for estimating time series policy positions from texts. *Am. J. Polit. Sci.* 52(3):705–22

Spirling A, McLean I. 2006. The rights and wrongs of roll calls. *Gov. Opposition* 41(4):581–88

Spirling A, McLean I. 2007. UK OC OK? Interpreting optimal classification scores for the UK House of Commons. *Polit. Anal.* 15(1):85–96

Spirling A, Quinn K. 2010. Identifying intraparty voting blocs in the UK House of Commons. *J. Am. Stat. Assoc.* 105(490):447–57

Steenbergen MR, Marks G. 2007. Evaluating expert judgments. *Eur. J. Polit. Res.* 46(3):347–66

Volkens A. 2001. Quantifying the election programmes: coding procedures and controls. In *Mapping Policy Preferences: Estimates for Parties, Electors and Governments 1945–1998*, ed. I Budge, H-D Klingemann, A Volkens, J Bara, E Tannenbaum, et al., pp. 93–107. Oxford/New York: Oxford Univ. Press

Volkens A. 2007. Strengths and weaknesses of approaches to measuring policy positions of parties. *Electoral Stud.* 26(1):108–20

Wendell RE, Thorson SJ. 1974. Some generalizations of social decisions under majority rule. *Econometrica* 42(5):893–912

Westholm A. 1997. Distance versus direction: the illusory defeat of the proximity theory of electoral choice. *Am. Polit. Sci. Rev.* 91:865–83

Zechmeister EJ, Corral M. 2012. Individual and contextual constraints on ideological labels in Latin America. *Comp. Polit. Stud.* 46(6):675–701