

PARTY POLICY IN MODERN DEMOCRACIES

Kenneth Benoit
Trinity College, Dublin
kbenoit@tcd.ie

Michael Laver
New York University
ml127@nyu.edu

|

April 11, 2006
(For publication by Routledge)

FINAL VERSION

Chapter 3

Empirical policy spaces

The spatial models of political competition we discuss in the previous chapter can be constructed either at the level of pure theory, or in a form that allows them to address politics in the real world. Many formal models are in effect self-contained and stylized “Platonic” systems of assumed motivations, institutions and rules of interaction. Yet, at the same time, the authors of these models typically name the concepts they use in ways that refer suggestively to the real world. Thus modeled agents such as “voters” and “politicians” are given these sobriquets precisely because the claim is being made, at least implicitly and often explicitly, that these theoretical abstractions from reality do bear some meaningful resemblance to wet-life human “voters” and “politicians” who can actually be observed and touched. This in turn is often taken to imply, even if only rhetorically, that analytical implications of the theoretical model do have something to do with reality. Indeed the authors of such models are only very rarely satisfied to present us with a purely Platonic system adorned with “political” labels – however beautiful this system might be. They are typically concerned to argue, often surprisingly informally, that their model addresses some aspect of real political competition. This brings us back to Reichenbach’s distinction, discussed in the Chapter 1, between mathematical and physical geometries – with the latter an essentially empirical enterprise that involves measuring the real world (Reichenbach 1956). In the present context this means that the development of physical, as opposed to mathematical, spatial models of political competition depends upon the creation of a set of measuring rods that can be used to describe real world spatial locations. In this chapter, we consider alternative types of measuring rod, and the extent to which different types of measurement instrument may be suitable for different types of spatial models.

MEASUREMENT CHOICES AND TRADEOFFS

In the empirical measurement of policy spaces, as in any practical activity requiring the use of tools, we select our instruments based on their fit to our objectives. Selecting tools typically implies making tradeoffs. In this section we explore some of these tradeoffs, before discussing specific tools for measuring empirical policy spaces.

In order to operationalize the models of political competition described in the previous chapter, we need to estimate both the positions of political parties on various policy dimensions,

and the relative importance of each dimension for each party. While party policy positions, as well as the weights attached by particular parties to particular policy dimensions, are fixed at any given point in time, they may change over time. Because these policy positions and weights are fundamentally abstract concepts, they *cannot be observed directly*. However there are many directly observable *manifestations* of these in the activities of parties, including political statements and speeches, election manifestos, legislative speeches, and voting patterns. Furthermore, citizens, elites, as well as experts, also gather “*second-hand*” information about party policy positions using numerous sources, notably the various communications media, though also via direct personal communication with others. Hence, while a party’s position on some policy dimension may be fundamentally unobservable, a strong consensus may arise about its location – which in this sense may be treated as “common knowledge”.

Nearly all observers of British politics, for example would probably agree that the Labour party moved towards the center of the political space under Tony Blair’s leadership. Yet no purely physical evidence of this shift, beyond the types of manifestation we have just mentioned, will ever measure this shift directly. Thus, while we take the location of the British Labour party’s policy position as something fixed and real – manifested as common knowledge – no amount of archeological excavation, brain scanning, satellite imaging or anything else will allow us to measure the party’s “real” location directly. Instead, we must rely on either forming our own judgments based on first-hand manifestations of party policy positions, or on somehow accessing judgments of this formed by others. To do this scientifically, of course, we need a systematic method, a well-specified and properly tested procedure.

Given the theoretical and substantive importance of measuring real-world policy spaces, it should come as no surprise to find a considerable body of research devoted to precisely this problem – and coming at it from a range of quite different perspectives. These perspectives can be distinguished according to their answers to several key questions.

1. *Is the relevant evidence about policy positions “second-hand” or “behavioral”?*

Behavioral evidence includes phenomena directly observable by the analyst – such as party statements, election manifestos, political speeches, and voting records. Second-hand evidence, on the other hand, typically consists of *evaluations* of evidence about policy positions by third-party analysts of these observables. These analysts may be real voters or politicians, whose views are collected in public opinion or elite surveys. Or they may be “professional” observers of politics of some sort, whose views may be collected using systematic methods such as expert surveys, or indeed in a more haphazard manner.

2. *What a priori assumptions are made, prior to measurement, about the nature of policy space being measured?* Two key questions arise here:

- a. *Do we assume the policy space to be high- or low-dimensional?* We have already noted that there is no such thing as the “true” underlying dimensionality of any policy space. We can define issues for consideration at a very fine-grained level and think in terms of an issue space of very high dimensionality. Or we can see most of the important features of the political system under investigation as being captured by agents’ positions on a single underlying dimension. The type of space we want to measure depends upon the type of model we want to operationalize. The decisions we make on this important matter drive all subsequent measurement decisions, and thus the type of empirical data we generate.
- b. *Do we have a priori knowledge of the substance of key policy dimensions before we set out to measure agents’ positions on these?* It may be that we know, in advance of measurement, the key substantive policy dimensions of the political decision-making space in which we are interested. Alternatively, we may be engaging in empirical research precisely to find out what these dimensions are. In the first situation we can adopt an *a priori* approach. Our task is to estimate agents’ unknown positions on “known” dimensions that we explicitly specify when designing the research. Essentially, this approach assumes we know more about key substantive policy dimensions than we do about the positions of key agents on these dimensions. In the second situation we do not want to assume in advance that we know number and substantive meanings of key policy dimensions, but instead want to treat these as open empirical questions. This *a posteriori* and quintessentially inductive approach sets its essential empirical task as finding the best-fitting empirical representation of the policy space under investigation, using techniques of dimensional analysis to infer latent policy dimensions and then

interpreting the substantive meaning of these dimensions in terms of relative locations of key political agents on these. The approach thus assumes that we know more about the positions of key political actors, relative to each other, than we know about the substantive meaning of key policy dimensions. In a nutshell, the *a priori* approach assumes the substantive meaning of the dimensions and investigates the relative locations of agents, while the *a posteriori* approach assumes substantive meaning in the relative locations of key agents and uses this information to investigate the dimensional structure of the political space.

3. *What are the statistical properties of the estimated policy positions generated?*

Estimation is a problem treated most formally in the field of statistics, where estimators have well-defined different properties that are used to evaluate alternatives. However, estimators of party policy positions also have properties that may be compared when evaluating alternatives. *Precision* is one such property, which includes whether discrete or continuous scales, and/or whether 5-, 7-, 20-, or 200-point scales are used. *Accuracy*, of course, is another key criterion, concerned with whether the measured position reflects a presumed “true” party position. *Reliability* is another important property of any measure, and concerns whether repeated measurements of the same party position would be likely to yield similar results. Finally, estimators may be distinguished by whether they come with accurate assessments of their associated *uncertainty*, providing some measure of confidence in a particular point estimate of the quantity being measured, in this case a party policy position.

This list articulates some of the fundamental scientific issues we confront when evaluating tools we might use to measure empirical policy spaces and the location of political actors within these spaces. These choices and tradeoffs are themes to which we return many times during the rest of this chapter, in which we discuss various means of measuring policy spaces and locations.

SURVEYS OF CITIZENS AND POLITICIANS

On the face of things, perhaps the most obvious way to estimate the policy positions of either citizens or politicians is to ask them directly. As far as citizens are concerned, there is certainly a huge and rapidly expanding database of surveys dealing with the attitudes of citizens in different countries at different times to many different matters. These data take several forms.

First respondents may be presented with particular “synthetic” policy scales and asked to locate themselves on these, possibly also being asked to locate named politicians or organizations on the same scale. Such synthetic scales include for example, a left-right scale or a scale that contrasts protecting the environment with encouraging economic growth. The danger inherent in this approach is that analytical scales that are meaningful to the researcher as ways of describing a policy space may not be meaningful to survey respondents – who out of politeness nonetheless answer the question rather than responding that the questions they have been asked are meaningless.

Although these are not directly related to politics, the findings of cognitive scientists on human perceptions of color are instructive in this context. Most models of the human color space (with which any reader who has used a computer graphics package will be familiar whether they realize this or not) are three-dimensional – describing the set of all humanly perceived colors in terms of hue (typically using a color circle), saturation and brightness (Gärdenfors 2000). While there are different models of the human color space, all are three dimensional and broadly analogous. Extensive empirical research on color matching by humans has confirmed that such models fit human perceptions of color very well indeed (Gärdenfors 2000). Yet even a highly intelligent real human would be incapable of describing a color he or she is looking at using the three analytical dimensions of hue, saturation and brightness. These dimensions are constructed by analysts of human color perception. They do very systematically describe how ordinary humans behave but they themselves are not perceived by ordinary humans. We might for the same reason treat with circumspection survey data that are derived from asking respondents to place themselves on synthetic analytical policy scales – such scale may not actually mean anything to the respondents, or may mean quite different things to different respondents. Indeed there is strong evidence that this latter point is a significant problem. When Irish citizens were asked in an Irish election study to locate parties on synthetic policy scales, for example, respondents located parties in positions that were strongly influenced by their own views on the issues at stake. In addition, survey respondents’ use of the scales bunched all parties toward the midpoint and typically avoided placing parties in either extreme quartile (Benoit and Laver 2005).

The way that psychological researchers draw maps of the “cognitive spaces” of real humans is to collect data on human perceptions of similarity and difference. Such perceptions are taken in some sense to be “natural” or “primitive”, in contrast to the synthetic dimensional structure used by analysts to describe cognitive spaces. This synthetic dimensional structure is then inferred by analysts from data on human perceptions of similarity and difference, using techniques of multidimensional scaling. Very striking in this context is the fact that the analytical technique used to infer the dimensional structure of human perceptions of similarity and distance must inevitably make assumptions about the metric that best describes “distances” in this conceptual space. Thus the choice of metric is not just the abstract theoretical matter we discussed in Chapter 1, but a practical empirical concern – albeit one that is either ignored or not recognized at all by many empirical analysts of politics.

Cognitive scientists, when engaging in the multidimensional scaling of such data, typically see the choice as between a Euclidean and a City Block metric. As we saw in Chapter 1, a considerable body of empirical psychological research suggests that the City Block metric fits human behavior best used when the dimensions of difference are “separable” and the Euclidean metric fits best used when they are “integral” (Attneave 1950; Shepard 1991; Gärdenfors 2000). For example, an empirical finding about human color perception is that the hue of an object cannot be assigned without also assigning its brightness, so these two dimensions of perception are integral. In contrast the weight of an object can be assigned quite independently of its temperature, so these two dimensions of perception are separable. It also seems likely that the ability to perceive an increasing number of *separable* dimensions of similarity and difference is an important part of human cognitive development during childhood (Gärdenfors 2000).

These distinctions should also be important when we think about measuring the positions of real humans in political spaces. To do this involves scaling, which involves making assumptions about metrics. When we estimate real political spaces by analyzing survey data, for example, we can follow the cognitive scientists and take perceptions of similarity and distance as psychological primitives, inferring from data on such perceptions the dimensional structure of the underlying political space. Or we can attempt in some more direct way to estimate respondents’ positions on substantive analytical policy dimensions – taking these positions as primitive and using an assumption about metrics to draw inferences about distances between points in the underlying space. Data are available in election surveys that facilitate both approaches.

Thus some survey questions ask respondents “close” they feel to named and well-known politicians. These can be taken as revealing primitive perceptions of political similarity and can be subjected to multidimensional scaling using some assumption about the most appropriate

distance metric. Opinion surveys typically also contain batteries of attitude questions. These in effect locate the positions of respondents in a high-dimensional attitude space, taken to be primitive, with one dimension for each question to which a response is given. This high-dimensional attitude space can then be subjected to some form of dimensional analysis to discover whether patterns in respondents' positions on coherent sets of attitudinal dimensions can be explained by a limited number of underlying "policy" dimensions. What is perhaps not fully appreciated by some scholars using this approach is that assumptions about cognitive metrics are inevitably embedded in such dimensional analyses.

For example, one scaling approach is to take battery of attitude questions on matters considered on *a priori* grounds to be substantively related – attitudes on a series of "moral" issues, for example, or on the economy – and to analyze the extent to which respondents' answers to each these questions can be combined into a single reliable additive (Likert) scale. Such a scale might be used to measure the "conservatism" of respondents on moral issues, for example, or of left-right positions on economic policy, and thereby to reduce positions in a high dimensional attitude space to points on one synthetic analytical dimension. Note that the additive combination of survey items into a single Likert scale, common when survey data are used to estimate the scale positions of citizens, makes the implicit cognitive assumption about respondents that they use the City Block metric when judging similarity and distance between agents in the attitude space. Cognitive scientists would tell us that this is the right thing to do if we think the component parts of the scale are separable – if we feel a person can state a position on stem cell research, for example, without intrinsically needing to condition this position on their simultaneous positions on capital punishment and/or gay marriage. An alternative approach would be to use a data reduction technique such as factor analysis to search for latent dimensions, with which answers to batteries of attitude questions are correlated. Note that most factor analysis is based on least squares algorithms that minimize Euclidean rather than City Block distances between latent dimensions and the observed measures from which they are constructed. Using factor analysis to derive respondents' positions from a high dimensional attitude space thus makes different cognitive assumptions about how agents perceive political distance. Cognitive scientists would tell us that this is behaviorally more suitable for "integral" sets of issue dimensions, for which positions on one dimension intrinsically depend on positions on other dimensions – for example preferences on the relative proportion of public spending to be devoted, respectively, to education, defense, health and welfare. Micro-economic theorists like Milyo, however, would as we saw in the previous chapter disagree with this on pure theoretical grounds (Milyo 2000). Indeed, estimating distinct policy positions on a set of policy dimensions

that we have good *a priori* reasons to believe are non-separable raises complex methodological issues that have not to our knowledge been addressed in the analysis of real public opinion survey data. By default, such data are typically analyzed as if dimensions of the attitude space are cognitively separable – suggesting the use of a City Block distance metric and the resulting additive scaling when sets of survey items are combined to give estimated positions on a single underlying dimension.

Overall, carefully designed and analyzed mass survey research remains the only practicable way to derive estimates of the policy positions of members of mass electorates. Hence “election studies”—mass surveys—carefully designed for explicit social and political research purposes—form the main source of data on policy positions for spatial models that use the ideal points of *citizens* as input – typically models found “above” the electoral system line in the “big” model of party competition set out in Figure 2.1. For accurate and reliable measurements of political actors such as politicians and parties, however, citizen surveys are beset by a number of problems. Mass survey research is useful for telling us how citizens *perceive* parties, but inherently problematic when used in estimating where these parties are *actually positioned* in relation to different dimensions of policy.

Turning now to the need to make inferences about the positions of politicians and political parties, the most obvious thing to do might seem to be to extend the logic of citizen surveys to estimate citizen positions, and use “elite” surveys that ask politicians directly about their own policy positions. This approach, however, is beset by problems when the goal is obtain accurate and reliable estimates of party positions on policy. One key problem is the strong incentives for politicians to characterize party policy positions—both their own and those of others—in a non-sincere fashion.¹ Politicians from more centrist parties, for instance, are more likely to rate extreme parties as extreme, while politicians from extreme parties are more likely to place such parties as being less extreme. In other contexts, centrist parties may have electoral incentives to attempt to differentiate themselves from close policy neighbors even when their actual differences are negligible. Indeed, many political parties tend to view representations of their policy positions as something to be carefully controlled. For political reasons, some political organizations explicitly forbid their members from participating in academic surveys. The British Labour Party, for instance, has allegedly issued instructions to its MPs not to respond to political questionnaires without explicit permission from the party leadership.

The other main problem with politician or elite surveys is purely practical, although just as limiting. It is nearly impossible to get high levels of cooperation from the members of any

¹ “Perish the thought!” do we hear someone say?

significant legislature in the time-consuming completion of an academic survey. In a 2000 survey of Members of the European Parliament (MEPs), for instance, less than one third of MEPs responded. Moreover, this low participation rate occurred despite the high expertise and reputation of the survey's principal investigators, and despite the fact that, more so than most legislatures by its nature and the backgrounds of its members, the European Parliament is favorably predisposed towards academic research.²

To estimate the policy positions of politicians and parties, therefore, we need to turn to alternative sources of data, of which there are essentially two in addition to expert surveys. These involve the systematic analysis either of political text generated by politicians or of their legislative voting behavior.

ANALYZING POLITICAL TEXT

If we want to estimate the policy positions of politicians, one of the main sources of information at our disposal is political text. There are of course many different types of political text, but one authoritative source of information about the stated electoral policy positions of political parties is the official party manifesto. It might be argued that very few real voters read any party manifesto at all, while almost no sane voter peruses all party manifestoes on offer and conducts an in-depth comparative analysis of these, basing her voting decision on the results of this analysis. Nonetheless the party manifesto is the official statement of party policy, to which the party can be held accountable – by critics, journalists and expert observers of the political scene. In this sense, positions outlined in the party manifesto can be taken as “official” party policy.

The longstanding Comparative Manifestos Project (CMP) has conducted a systematic analysis of party manifestos over a long period of time, using trained human readers to code, into a predefined 56-category coding scheme, every sentence of every manifesto investigated. Coverage extends to almost every party manifesto issued at every democratic election since World War 2. This has generated a time series of the *electoral* party policy positions that spans the post-war era for most parties in most democratic states (Budge, Klingemann, Volkens et al. 2001). The comprehensive coverage of this dataset has made it a popular choice with researchers, despite the fact that, theoretically, the CMP actual coding scheme very explicitly impounds a

² See Scully and Farrell (2003). The MEP Survey 2000 was co-authored by Simon Hix and Roger Scully in 2000. Details are available from <http://www.lse.ac.uk/depts/eprg>.

particular “saliency” theory of politics that is relatively far from the mainstream of recent spatial modeling and despite the fact that, methodologically, each CMP text is coded once and once only by a human coder, so that no policy position that is generated comes with any estimate whatsoever of associated error.

The theoretical basis of the CMP dataset is set out very clearly in its book *Mapping Policy Preferences*, hereafter MPP (Budge, Klingemann, Volkens et al. 2001). The authors of MPP are unequivocal in arguing that the CMP data are fundamentally grounded in a “saliency theory” of party competition. “Not only coding categories but also rules for assigning textual units to them are shaped by theory . . .” (Budge et al 2001: 12). “[The] saliency theory of party competition is the one the manifesto codes and estimates are based on” (Budge et al.: 76). For researchers who want to know about the theory that so firmly structures the data they are using, the authors of MPP could not be clearer about their assumptions. The first of these is “1. Party strategists see electors as overwhelmingly favouring one course of action on most issues. Hence *all party programmes endorse the same position*, with only minor exceptions” (Budge et al: 2001 82, emphasis added). This clear statement lies at the heart of the saliency theory of party competition and therefore, according to the authors of MPP, at the heart of the CMP data that derive from this. The CMP data are not at all, according to the CMP, about party *positions* on particular policy dimensions; these party positions are all very explicitly assumed to be the same. Rather, the CMP data are ostensibly about the party-specific *saliency weights* of different policy dimensions.

The good news for those who have used CMP data to operationalize theoretical models, assuming that these were about party *policy positions* as opposed to *dimension weights*, is that the CMP did not use saliency theory when designing its own coding scheme for party manifestos. As a result of debates within the research team when the coding scheme was originally devised (Budge et al 2001: 82-83), about half of the CMP coding categories are explicitly positional and do not derive directly from saliency theory. MPP openly admits that the inclusion of these categories “undermined the pure saliency nature of our framework” (Budge et al 2001: 83). Reading the actual definitions of the remaining CMP coding categories, it quickly becomes clear that most of the remaining “saliency” coding categories are also explicitly positional rather than saliency-driven. Most people, for example, would regard “centralization” and “decentralization” as opposite ends of the same policy continuum despite the fact that they do not have “pro” and “con” attached to their labels. The pattern is very striking and the reader is referred to Appendix III of MPP. Take the first coding category, “anti-imperialism”, defined as “negative references to exerting strong influence (political, military or commercial) over other states; negative references

to controlling other states as if they were part of an Empire; favourable mentions of decolonialization ...” and so on (Budge et al 2001: 222). This is self-evidently not a saliency category dealing with, for example, attitudes to imperialism (pro and con). The issue *position* content in this definition could hardly be more explicit. This coding category is in fact one end of a bipolar variable for which the designers of the coding scheme have simply predicted that the other end is unlikely to be populated because imperialism is not very popular these days. The result is that any hostile mention of, or regret about, decolonialization that might be observed in a text does not contribute at all to the salience of the “imperialism” issue area, but must either be left uncoded or coded into some other category. The prediction that the other side of the issue will not be populated by party manifesto references can never be tested, furthermore, because the coding scheme actively prevents information from being collected on it.

This pattern has been systematically demonstrated by two of MPP’s authors, albeit writing in another context (McDonald and Mendes 2001). McDonald and Mendes analyze the positional content of each of the MPP coding categories and conclude that “[i]t is not as if the MRG coding is non-positional in policy terms ... Fifty-four of the fifty-six categories involve clear value statements.” (McDonald and Mendes, 2001: 91). One of the remaining two, “nationalisation, is actually directional” (McDonald and Mendes, 2001: 92). There is thus actually just one of the 56 coding categories in the CMP coding scheme, “economic goals”, that does *not*, according to the logic of MPP, undermine the assumptions of “saliency theory”. In short, and probably much to the relief of the many third-party analysts who have used the CMP data to measure party *positions*, the CMP coding scheme is not in practice a “pure salience” scheme, as claimed so explicitly in MPP’s theoretical discussions. It is a *positional* coding scheme in which many of the potential positional categories have been censored in advance on the basis of the empirical expectations of the scheme’s designers.

Human-coded content analysis is a painfully resource-intensive activity. Nowhere is this more true than for the CMP project, which is reported by MPP to have covered 2,347 manifestos, issued by 632 parties, in 52 countries (Budge et al 2001: 95). And the CMP has coded many more manifestos since the publication of that report. Given the huge expense involved, the vast majority of the manifestos that form the basis of the CMP dataset were coded once only by a single human coder. A crucial consequence of this is that every single number in the CMP dataset, as in almost all other datasets generated by human-coded content analysis, is presented as a single point estimate with no estimate of associated error. But there is surely error in these, as in all other, data. We get inklings of this in MPP’s discussion of Italian party positions. When the

Italian manifestos were recoded by a different coder, the net result was different left-right policy positions for the Italian Communist Party (Budge et al 2001 50 fn 2). The crucial implication of having no estimate of associated error is that, when evaluating the difference between the estimated positions of two parties (or the same party at two points in time), we have no way of knowing, systematically, whether these positions are “the same” or “different”. The same two positions might be judged to be the same if they had large standard errors, or different if their standard errors were small. It is impossible, in the CMP data, to distinguish measurement error from “real” underlying change in the policy positions under investigation.

Such problems, while almost inevitable given the huge cost of human coded content analysis, may soon be overcome by computerized text analysis. For example, Laver, Benoit and Garry recently proposed a language-blind computerized technique for political text analysis that retrieves valid party positions on *a priori* policy dimensions and reports associated standard errors (Laver, Benoit and Garry 2003). Essentially this technique estimates for one or more policy dimensions the (unknown) positions of a set of “virgin” texts under investigation, stating these positions in relation to the (known or assumed) positions of a particular set of “reference” texts. The reference texts are chosen at least in part because their positions on the policy dimensions under investigation are either known, or can uncontroversially be assumed. Patterns of relative word frequencies are first analyzed for the reference texts, allowing the computation, for every word in these texts, of a key conditional probability – the probability you are reading reference text R_i , given you are reading word w . Knowing or assuming the position of each reference text on each dimension under investigation, this conditional probability allows the computation of a dimension score for each unique word in the set of reference texts. This score reflects the expected position of a text on the dimension in question, given only that one is reading word w . Having computed these word scores, relative word frequencies for all of the virgin texts, whose policy positions are to be estimated, are then computed. This allows the computation of an expected position of each virgin text on each dimension under investigation, with associated standard errors, given the words it contains. The application of this technique was recently extended from manifestos to speeches in the Italian legislature, retrieving estimated policy positions for individual Italian legislators (Giannetti and Laver 2005). Sooner rather than later, therefore, systematic research programs using computerized text analysis will provide alternative sources of information about the policy positions of individual legislators and, crucially, because they are statistically based will provide estimates of the error associated with each policy position estimated.

The word scoring technique for computational text analysis, despite its alluring potential to analyze immense volumes of text written in languages not understood by the researcher, is not however a magic bullet for those who are in the business of estimating policy positions. In the first place it requires good reference texts and valid estimates of the policy positions of these texts on the dimensional under investigation. These are not always available. In the second place computer word scoring runs into significant problems when it comes to generating long time series of the policy positions of particular texts authors. Essentially this is because words change their political associations over time, which makes it difficult for us to know, if we estimate the positions of the same author of different texts, issued at different time points, whether any movement we observe can be attributed to a changing meaning of the words, or to a changing underlying policy position of the author.

This is not a problem that applies only to computer word scoring and is actually a particularly explicit manifestation of a rather deep problem in estimating any time series of policy positions. This has to do with whether any change we observe results from a change in the underlying policy positions under investigation or a change in the calibration of our measuring instrument. Thus, when we measure the height or the weight of human beings, we have no reason to suppose that our measuring instruments are themselves changing over time. We can therefore infer, when our measurements of height and weight change over time, that the height or weight of the subject has in fact changed.

We can for obvious reasons be much less confident in assuming that the calibration of any scale for measuring important features of the political system remains fixed over time. Thus the CMP has devised a very widely-used left-right scale for measuring party policy positions. This scale was devised in the early 1980s, from an inductive analysis of party manifesto content between about 1945 and 1985; the average manifesto used in this inductive analysis was thus written about 1965. The CMP's left-right scale's substantive content has remained fixed since it was devised. This scale, for example, did not include party manifesto positions on the environment as part of the left right-scale. As we shall see later in this book, there is very strong evidence that party positions on the environment are indeed now part of our contemporary understanding of left and right in politics. The meaning of left and right has almost certainly changed over time to encompass attitudes on the environment. If a party in the CMP dataset starts talking more about the environment and less about other content categories in the CMP left-right scale, then it will appear to be becoming more centrist – measuring this using the CMP left-right scale. But we might well feel that this party is not “really” becoming more centrist at all – but rather that the scale we are using is getting progressively more “out of date”.

We appear to be stuck between a rock and a hard place. If we do not change the definition of some policy scale in which we are interested, then it remains frozen at a particular time point and becomes progressively less valid. If we do change the content to reflect the changing political reality the scale is intended to measure, then how do we compare scale positions before and after the scale's definition has been changed? In short, how DO we measure a valid time series using some scale whose meaning we have good reason to suspect is changing over time? We are aware of no definitive solution to this problem on offer in the professional political science literature, although a directly analogous problem confronts economists when defining the bundle of good they use to measure "inflation", for example. But this problem is nonetheless something to be kept firmly in mind by analysts who intend to lay theoretical or empirical emphasis on observed movements over time, measured on some synthetic political scale.

The analysis of legislative voting behavior

As far as the US Congress is concerned, the primary source of information on the policy positions of legislators has, as a matter of research practice, become the systematic analysis of roll call voting data, with a remarkable recent coordination among scholars on using the NOMINATE technique (Poole and Rosenthal 1997). This very explicitly *assumes* a spatial model of party competition in which differences between the policy positions of legislators can be represented as Euclidean distances. Conditional on these assumptions, the spatial policy positions of legislators are retrieved by analyzing roll call voting records, assuming that two legislators with more similar voting records can be seen as being closer to each other than are two legislators with more different voting records. The NOMINATE technique is a form of multidimensional scaling, and using this with a Euclidean metric comes close to the approach of cognitive scientists who feel that the underlying dimensions being analyzed are not separable. Given this setup, Poole and Rosenthal (1997: 22) give themselves the job of retrieving "the 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". This is no mean feat, and the promise of generating extended time series of the estimated policy positions of every individual US legislator, given the partially overlapping memberships of consecutive legislatures, has made the Poole and Rosenthal approach extremely attractive to many other legislative scholars. What is particularly striking in an environment in which levels of party discipline, as we have seen, are relatively low, is that this approach allows us to plot what is happening inside legislative parties, and thereby to investigate the structure of intra-party coalitions and factions. Indeed this is actually the main thing that the NOMINATE

technique does allow us to plot – so it is an appropriate data source of the positions of members of the US Congress in models that assume endogenous party discipline. If there are two opposing parties with rigid discipline in the sense that all Party X legislators always vote in the same way as each other and in the opposite way to all legislators from Party Y, then NOMINATE will simply tell us the two parties and their legislators are different, which we already know, and will add no metric information to this.

This highlights the fundamental problem that arises if we want to export techniques of Congressional roll-call analysis to political systems in which there are very high levels of legislative party discipline. In a multiparty system where party discipline is close to 100 percent, roll-call analysis might conceivably allow us to retrieve the positions of *parties*, rather than individual *legislators* – since legislators would be revealing their party membership in their voting record, rather than their ideal policy positions. However, since multi-party systems also beget coalition cabinets, and since members of such cabinets are bound together by constitutional rules of collective cabinet responsibility, it is likely that all parties in the executive coalitions will vote in the same way, despite having different policy positions. It is also possible that all members of a diverse opposition to the government will vote in the same way, and against the government. High levels of party discipline combine with the parliamentary government system, therefore, to undermine quite fundamentally the potential of roll-call analysis to give us useful information about the policy positions of either individual legislators or, indeed, of legislative parties.

The converse situation also presents a problem associated with measuring party positions on policy from roll-call voting analysis. When party discipline is low and voting sincere, estimates of legislator positions *from the same party* will vary significantly, often in both of the dimensions typically produced by NOMINATE. For investigating many questions about the behavior of individual legislators, such results may yield valuable insights into intra-party politics. One setting to which NOMINATE has been successfully exported, for example, is the European Parliament, where the absence of an executive sustained in office by the legislature, and consequently looser discipline among European party groups, combine to allow roll-call analysis to yield fruitful insights into party group cohesion and discipline (Hix 2001; Hix, Noury and Roland Forthcoming). For producing point estimates of the policy positions of political *parties*, however, variance in NOMINATE scores raises the question of where the party's official position lies, within the cloud of points described by the positions of its individual legislative members. Substantively, it may not be warranted in all cases to assume that a party's position is simply the

mean of its legislator's estimated ideal points. Methodologically, the assumption of party-as-mean also makes strong assumptions about underlying Euclidean distances, and also ignores the inherently spherical nature of NOMINATE estimates.

As with any empirical method of measuring positions in policy spaces, of course, the analysis of roll-call votes also confronts practical limitations regarding data availability. The great advantage of roll-call votes is that their availability requires no action or consent of the actors generating observable data. Their great disadvantage, however, is that roll-call votes are unavailable or only selectively available in many contexts. In some countries—for instance Ireland and Hungary—no roll-call votes are taken or recorded, making it simply impossible to in these contexts to obtain estimates of the positions of either legislators or parties using this method. In other contexts where roll-call votes are available, they are frequently highly selective, called strategically, perhaps in order to verify discipline among a party's own members, or to expose a lack of discipline among a rival party's members. In the much analyzed European Parliament, for instance, any vote can be made into a roll-call vote by request of a party group or 32 MEPs (from a current total of 732), and there is evidence that roll call votes tend to be called on some issues more than others (Carruba et. al. 2004). Consequently, roll call votes tend to be called selectively and strategically, as well as only for only certain issues, especially in the European Parliament (Kreppel 2002) p129). In addition to the selection bias this produces, the policy scales which roll-call vote analysis yields must be interpreted *a posteriori*, and these interpretations are by no means self-evident (see McElroy and Benoit forthcoming; Hix et al 2005).

Expert surveys

What we have just described might appear to be a considerable armory of techniques for estimating the policy positions, of both large numbers of voters and smaller numbers of “elite” politicians or political parties. So why do we need expert surveys? In this last section we address this question, discussing epistemological and practical advantages of the expert survey method.

Expert survey methodology has been used in such diverse fields as cognitive psychology, decision analysis, statistics, sociology, cultural anthropology, and knowledge acquisition. It entails the use of specific procedures to identify experts, define the problem, before eliciting and characterizing the experts' collective judgment (Meyer, Butterfield, Murray et al. 2002). Expert judgment data is typically used as a means of providing information when other sources, such as direct measurements, observations, or experiments are unavailable. In these situations, expert

judgment may be useful to (a) to provide estimates of complex or difficult to observe phenomena, (b) to forecast future events, and (c) to integrate or interpret existing data (Meyer and Booker 2001), 4). Substantive problems to which expert judgment has been applied include:

- assessing the social and economic impact of the spread of HIV/AIDS (CINSSA Ltda 1997);
- studies of climate change;
- economic forecasts;
- risk assessment (e.g. (van Dorp, Merrick, Harrauld et al. 2001); and
- estimating empirical policy spaces in political science (Castles and Mair 1984; Laver and Hunt 1992; Huber and Inglehart 1995; McElroy and Benoit, forthcoming; Marks and Steenbergen 2004).

Moving beyond these types of application, expert judgment is frequently used to estimate qualities and quantities that are unobservable directly. Nearly everyone will be familiar with the scoring of gymnastic, skating, and diving events in the Olympic Games, for example, where performance quality is estimated by panels of expert judges applying pre-defined criteria. In academia, expert surveys (peer review) form the basis for evaluating many important decisions, ultimately determining whether grant proposals are funded and whether individuals are granted tenure or promotion. Indeed, the anonymous peer review process by which journal editors solicit opinions on the publishability, or otherwise, of almost any academic paper, is in itself a form of expert survey.

Turning to the more specific task at hand, using expert surveys to measure empirical policy spaces involves the *a priori* identification of salient dimensions of policy competition and the location of political parties on these dimensions, based on systematically collected judgments of political experts. To locate a party's economic policy position, for instance, an expert survey might present a ten or twenty-point scale anchored by two short characterizations of an extreme left position on one end, and an extreme right position on another, and ask respondents to locate each party at a position on the scale. The expert scorings of each party on the economic left-right economic policy dimension are then summarized statistically in order to measure the party positions, as well as the degree of certainty and consensus over these positions. In this particular context, "experts" are defined as people with expertise in party politics in their own national contexts, having considerable knowledge about the policies positions of those parties. Deciding

which experts to select for such surveys is extremely important, of course, and we return to this matter in the next chapter. When chosen properly, however, experts who are knowledgeable in a field are much more capable of rendering accurate information than those who are less knowledgeable, and have also been shown to be less prone to overestimate the confidence of their answers (Lichtenstein and Fischhoff 1977). In addition, research results show that experts are less subject to biases than non-experts; not only do they possess superior knowledge, but they also ignore factors such as ordering effects and other irrelevant factors (Haan, Dijkstra and Dijkstra 2003). For these reasons, when using “second-hand” sources of data as opposed to direct behavioral manifestations of the underlying constructs to be measured, expert surveys are viewed as superior to the alternative of asking mass publics to identify parties’ locations in policy spaces.

Expert surveys are fundamentally grounded in the *a priori* approach to measuring empirical policy spaces, since they almost invariably proceed by presenting predefined scales and asking respondents to use their best judgments to locate specified political actors on these scales. The underlying assumption is that the key substantive policy dimensions in the a particular context can be identified in advance of the location of party groups, based on substantive expert understanding of potentially salient policy issues. The unknowns which experts are then asked to estimate are the locations of each party group on these *a priori* dimensions. The estimates of party group positions are then taken to be the statistically aggregated judgment of the experts, on each pre-defined dimension. Unlike factor analytic scorings, constructed scales, or locations in a purely inductive space from multi-dimensional scaling analyses, expert survey summaries eliminate the need for subjective and often ad hoc *a posteriori* interpretation of results in terms of substantive policy scales.

Expert surveys provide measures that are explicitly second-hand in their approach to observation, rather than attempting to measure policy positions based on observable first-hand evidence. Especially in fluid political situations, this may indeed be one of the most attractive properties of expert surveys. Any inductive technique ultimately relies on expert judgment to judge the validity of *a posteriori* interpretations of results but, in fluid or new political contexts, this judgment may be hard to form, or may overwhelm the evidence, especially if first-hand evidence is scant. In a very real sense with all *a posteriori* interpretation, the analyst sets up him- or herself as an expert panel of one. Expert surveys, on the other hand, collect the best knowledge and wisdom of a population of experts, based on their evaluation of all the evidence at their disposal, and summarizes their consensus in a set of tractable estimates. Indeed, when trying to resolve which method of estimating party positions is “best” in a given context, we typically fall back on the expertise and wisdom of political experts. By extension, then, we see systematic

collection of judgments of political experts on party locations as the best way systematically to harvest this wisdom, which takes into account all relevant information about a party group's position, including voting behavior, political speeches, debates, expressed opinions of party leaders, and so on. Studies of different forecasting techniques, for instance, have demonstrated that surveys of expert opinion, in addition to being efficient and economic, do very well in terms of accuracy because they reflect the most up-to-date core assumptions (Ascher 1978), 203). Even though experts will vary in their judgments, we can combine and summarize these judgments as a substantive indication of a party's likely set of policy locations. In short, our best estimate of party positions on policy resides in the collective wisdom of experts, available through systematically collected and summarized expert judgments.

Some measure of the ultimate authority of expert judgments may be seen in their use to cross-validate the other means of measuring empirical policy spaces reviewed in this chapter. Once we have used some measurement instrument, such as the coding of text or the scaling of roll call voting behavior, to estimate the policy positions of political actors, it is of course vital to assess the validity of our estimates. At its simplest, validity in this context has to do with the correspondence between our measurements, for example, of some party policy position and the actual "real" policy position held by the party. Because parties' "real" policy positions are intangible and ultimately unknowable, however, we must proceed more indirectly. The two most commonly used approaches we find in published work in this field are to assess the "face validity" of the resulting estimates, and to compare new estimates with authoritatively published and cited alternatives.

A conclusion that a set of estimated policy positions does indeed have face validity essentially relies upon an informal judgment that these estimates have certain familiar and expected properties – for example that the positions of prominent actors are more or less in their "correct" places. But who is to say what is, and what is not, "correct"? Such judgments of face validity must inevitably be grounded, explicitly or implicitly, in the accumulated wisdom of country specialists. Thus, if we were to challenge a judgment that a particular set of estimates has face validity, we would almost certainly be referred to published work by a relevant country specialist. A clear example of this approach can be found in more recent publications by the Comparative Manifestos Project, in particular their book *Mapping Policy Preferences* (Budge, Klingemann, Volkens et al. 2001). In this, the face validity of the CMP's left–right ideological scale is investigated by assessing the substantive plausibility of its measured movements of party left–right positions in a series of country-by-country discussions. The ultimate arbiters of this

“plausibility” are in practice selected specialists in the politics of each country, whose published work is cited in MPP’s footnotes as validating the CMP estimates.

A conclusion that an estimated set of policy positions has been cross-validated against an independent published alternative is, of course, only convincing if this independent source is itself widely accepted as valid – as some sort of benchmark against which to measure alternatives. And who are the arbiters of a valid benchmark in this case? Once more, either explicitly or implicitly, we are likely to be referred to the judgments of selected country specialists. When two sets of published estimates differ in substantively significant ways, then how do we resolve the situation? We can either engage in a methodological investigation of the potential sources of this difference, or we can appeal to some independent third-party arbiter of substantive validity. Once more, we will be led to rely on the substantive knowledge of country specialists.

All of this is rather obvious, but the implications are rather deeper than many have appreciated. We have no access to “real” policy positions against which to validate any particular estimate of these. We therefore rely, in assessing the validity of such estimates, on some professionally established benchmark. The validity of the benchmark ultimately derives from substantive judgments by specialists in the politics of the political system under investigation. The obvious problem when doing this is that, since country specialists are never in complete agreement on any matter, and since their discursive writing is always subject to *a posteriori* interpretation by the reader, there is a danger that the work of country specialists will be used at best haphazardly and at worst rhetorically when resolving any issue about the content validity of particular empirical estimates. What expert surveys of a population of country specialists do, on the other hand, is to summarize their accumulated wisdom in a systematic way, seeking an unbiased estimate of their judgments on particular matters that are specified *a priori*. As such, expert surveys may well be the most systematic source of reference on questions that might arise about the validity of estimates derived using other methods.

Beyond the methodological advantages of expert surveys, there is also an enormously compelling practical reason for their use: Expert surveys may be deployed quickly and inexpensively, to derive systematic estimates of policy positions, on a similar basis, across a wide range of countries. Given their relative ease of setup, it is a fairly simple matter to survey experts at any given time point, without the fixed costs of a huge data-gathering project, detailed document coding, time-consuming interviews, or costly opinion surveys. This is a far less “noble” justification than the scientific need for systematic benchmark estimates, but it is no less persuasive. Using the expert survey method, we were able in a relatively short period of time to estimate the positions of all significant political parties in 47 countries, on a common set of policy

dimensions, as well as on a series of country specific dimensions. This is a task that would have been extraordinarily resource intensive and time consuming for any of the other techniques we have been discussing. There is also a considerable demand from cross-national researchers for multidimensional estimates of party positions on a standard set of policy dimensions, a demand attested to by the wide range of usage and citation of this current book's precursor, *Policy and Party Competition* (Laver and Hunt 1992). In many settings, therefore, expert surveys may simply be the most efficient way to generate reliable dataset and, for the reasons we have discussed in the previous paragraphs, it is certainly arguable that they also provide the most scientific benchmark against which to assess the content validity of other estimates of the policy positions of key political actors.

CONCLUSION: ADVANTAGES OF THE EXPERT SURVEY METHOD

To summarize the preceding discussions, several cogent justifications follow for the use of expert surveys when compared to other methods. To begin at a practical level, one huge advantage of expert surveys is an attractive combination of economy and access. Not only do expert surveys typically require little time and expense to conduct, but they also enable researchers to explore policy spaces in almost any context, regardless of the availability of more problematic first-hand data. Expert surveys may thus be used as a research tool to explore empirical policy spaces independently of the systematic practical availability of behavioral data such as speeches, elections, party manifestos, or roll call votes. Put very crudely but nonetheless realistically, data for many of the countries covered in the expert surveys reported in this book would probably have been unavailable using any other means. This is no small advantage.

Another compelling advantage of expert surveys arises from their very explicit use of the *a priori* approach to estimating key political parameters. First, informal surveys of expert judgment may be used to identify key policy dimensions in each country, ensuring that substantive expertise guides the selection of the precise scales to be applied in each context. Following on this, the use of pre-defined scales provides complete flexibility for designers of expert surveys to treat policy spaces as high-dimensional or low-dimensional, depending on expert knowledge in context. Other more explicitly behavioral methods, such as the analysis of legislative roll call votes, do not offer such choices. Finally, the use of pre-defined scales eliminates any ambiguity or guesswork from interpreting final results, minimizing ad hoc interpretation of results by the researcher and the consequent (quite possibly unconscious)

temptation to read substantive meaning into the tea leaves that appear to form patterns when in fact there are none.

Another advantage relates to the statistical character of expert sampling, namely the property that gathering more information increases our certainty in the accuracy of our estimates (something we explore in detail in the next chapter). Researchers using expert surveys have control over their sample sizes, furthermore, and it is fairly economical to add more experts to a survey as informational requirements dictate. Moreover, it is possible, using well-understood statistical rules, to quantify the uncertainty associated with any estimates of policy location that expert survey scores produce.

A final, deeper justification for using expert surveys lies in the ability of expert consensus, summarizing the state of the art of knowledge and expertise, to provide the ultimate means of estimating intangible and ultimately unobservable empirical policy spaces. Political scientists may with some justification consider what they do to be science, but we nonetheless lack the physical measurement benchmarks of many of the empirical natural sciences. In politics, there is no equivalent of the atomic clock against which to calibrate alternative measurements of time, and no universally accepted measuring stick against which to calibrate alternative measurements of distance. When discussing alternative approaches to measuring the key parameters of political spaces, we often make reference to the need to appeal to the “accumulated wisdom of country specialists”, whether used implicitly or explicitly, when assessing the substantive content validity of particular empirical measurements. The obvious problem when doing this is that, since country specialists are never in complete agreement on any matter, and since their discursive writing is always subject to *a posteriori* interpretation by the reader, there is a danger that the work of country specialists will be used at best haphazardly and at worst selectively when resolving any issue about the content validity of particular empirical estimates. What expert surveys of a population of country specialists do, on the other hand, is to summarize their accumulated wisdom in a systematic way, seeking an unbiased estimate of their judgments on particular matters that are specified *a priori*. As such, they may well be the most systematic source of reference on questions that might arise about the validity of estimates derived using other methods. Instead of referring to a haphazard selection of country specialists when assessing the “face validity” of some estimate, our best estimate of the collective wisdom of the population of country specialists is available in more systematically collected and summarized expert survey results.

Table 3.1. Comparison of Measurement Approaches of Party Policy Positions

Characteristic	<i>Expert surveys</i>	<i>Mass surveys</i>	<i>Politician surveys</i>	<i>Roll Call Votes</i>	<i>Manifesto analysis - CMP</i>	<i>Manifesto analysis - Wordscores</i>
Evidential basis	Second-hand	Second-hand	Second-hand	Behavioral	Behavioral	Behavioral
Dimensionality level	Flexible	Flexible	Flexible	Low	Typically Low	High
Scale epistemology	A priori	A priori	A priori	A posteriori	A posteriori	A priori
Estimand-observation link	Direct	Indirect	Indirect	Indirect	Direct	Direct
Observability of data	Complete	Partial – depends on response frequency per party	Selective – depends on response rate	Selective or incomplete – depends on application and existence of RCVs	Complete	Complete
Researcher resource commitment	Low	High	Low	Low	High	Low
“Sincerity” of observed behavior	High	High	Low	Medium	High	High
Reliability	High	Low	Low	High	Medium	High
Estimation uncertainty	Known	Known	Known	Knowable	Unknown	Known
Examples	This book; Laver and Hunt (1992); Castles and Mair (1984); Huber and Inglehart (1995); Marks and Steenbergen (2004)	Thomassen and Schmidt (1997); Van der Eijk	Scully and Farrell (2003)	Poole and Rosenthal (1997); Clinton, Jackman and Rivers (2004); Hix (2001) Hix, Noury and Roland (2005)	Budge et. al. (2001); McDonald and Mendes (2001)	Laver, Garry and Benoit (2003); Benoit and Laver (2003)

REFERENCES

- Ascher, W. (1978). Forecasting: an appraisal for policy-makers and planners. Baltimore, Johns Hopkins University Press.
- Attneave, F. (1950). "Dimensions of similarity." American Journal of Psychology **63**: 546-554.
- Benoit, K. and M. Laver (2003). "Estimating Irish Party Positions Using Computer Wordscoring: The 2002 Elections." Irish Political Studies **18**(1): 97-107.
- Benoit, K. and M. Laver (2005). "Mapping the Irish policy space: voter and party spaces in preferential elections." Economic and Social Review **36**(2): 83-108.
- Budge, I., H.-D. Klingemann, A. Volkens, et al. (2001). Mapping Policy Preferences: Parties, Electors and Governments: 1945-1998: Estimates for Parties, Electors and Governments 1945-1998. Oxford, Oxford University Press.
- Castles, F. M. and P. Mair (1984). "Left-Right Political Scales: Some 'Expert' Judgements." European Journal of Political Research **12**: 73-88.
- Clinton, J., S. Jackman and D. Rivers (2004). "The Statistical Analysis Of Roll Call Voting: A Unified Approach." American Political Science Review **98**(2): 355-70.
- Gärdenfors, P. (2000). Conceptual spaces: the geometry of thought. Cambridge, Mass., MIT Press.
- Giannetti, D. and M. Laver (2005). "Policy positions and jobs in the government." European Journal of Political Research **44**(1): 91-120.
- Haan, M., G. Dijkstra and P. Dijkstra (2003). Expert Judgment versus Public Opinion: Evidence from the Eurovision Song Contest, University of Groningen.
- Hix, S. (2001). "Legislative behaviour and party competition in the European Parliament: an application of Nominate to the EU." Journal of Common Market Studies **39**(4): 663-688.
- Hix, S., A. Noury and G. Roland (2005). "Power to the Parties: Cohesion and Competition in the European Parliament, 1979-2001." British Journal of Political Science **35**(2): 209-234.
- Hix, S., A. Noury and G. Roland (Forthcoming). Democracy in the European Parliament.
- Inglehart, R. and J. Huber (1995). "Expert Interpretations of Party Space and Party Locations." Party Politics **1**: 73-112.

Kreppel, A. (2002). The European Parliament and Supranational Party System: a study in institutional development. Cambridge; New York, Cambridge University Press.

Laver, M., K. Benoit and J. Garry (2003). "Estimating the policy positions of political actors using words as data." American Political Science Review **97**(2): 311-331.

Laver, M. and W. B. Hunt (1992). Policy and party competition. New York, Routledge.

Lichtenstein, S. and B. Fischhoff (1977). "Do those who know more also know more about how much they know?" Organizational Behavior and Human Performance **20**: 159-183.

McDonald, M. and S. Mendes (2001). The policy space of party manifestos. Estimating the Policy Positions of Political Actors. M. Laver. London, Routledge: 90-114.

Meyer, M. A. and J. M. Booker (2001). Eliciting and analyzing expert judgment: a practical guide. Philadelphia, Pa., Society for Industrial and Applied Mathematics and American Statistical Association.

Meyer, M. A., K. B. Butterfield, W. S. Murray, et al. (2002). Guidelines For Eliciting Expert Judgment As Probabilities or Fuzzy Logic. Fuzzy logic and probability applications: bridging the gap. T. J. Ross, J. M. Booker and W. J. Parkinson. Philadelphia, Society for Industrial and Applied Mathematics and the American Statistical Association.

Milyo, J. (2000). "Logical deficiencies in spatial models: A constructive critique." Public Choice **105**: 273-289.

Poole, K. and H. Rosenthal (1997). Congress: A Political-Economic History of Roll Call Voting. New York, Oxford University Press.

Reichenbach, H. (1956). The Philosophy of Space and Time. New York, Dover Publications Inc.

Scully, R. and D. Farrell (2003). "MEPs as Representatives: Individual and Institutional Roles." Journal of Common Market Studies **41**(April): 269-288.

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. The Perception of Structure: Essays in Honor of Wendell R. Garner. J. R. Pomerantz and G. L. Lockhead. Washington, DC, American Psychological Association: 53-71.

Thomassen, J. and H. Schmidt (1997). "Policy Representation." European Journal of Political Research **32**(2): 165-184.

van Dorp, J. R., J. R. W. Merrick, J. Harrauld, et al. (2001). "A Risk Management Procedure for the Washington State Ferries." Risk Analysis **21**(1): 127-142.

ADDITIONAL REFERENCES

Carrubba, Clifford J., Matthew Gabel, Lacey Murrah, Ryan Clough, Elizabeth Montgomery and Rebecca Schambach. 2004. "A Second Look at Legislative Behavior in the European Parliament: Roll-Call Votes and the Party System." Working Paper, Vienna: Institute for Advanced Studies.

CINSSA Ltda. 1997. *HIV/AIDS: Its Social and Economic Impact In Colombia*. Report from http://www.iaen.org/files.cgi/8703_COLOM6.pdf.

Marks, Gary and Marco Steenbergen. 2004. Marks/Steenbergen Party Dataset, Chapel Hill, NC: University of North Carolina Chapel Hill, available from: <http://www.unc.edu/~gwmmarks/data.htm>.

McElroy, Gail and Kenneth Benoit. Forthcoming. "Party Groups and Policy Positions in the European Parliament." *Party Politics*.