The European Common Space: Extending the Use of Anchoring Vignettes

Ryan Bakker   University of Georgia
Seth Jolly   Syracuse University
Jonathan Polk   University of Gothenburg
Keith Poole   University of Georgia

In this article, we combine advances in both survey research and scaling techniques to estimate a common dimension for political parties across the member states of the European Union. Most previous scholarship has either ignored or assumed cross-national comparability of party placements across a variety of dimensions. The 2010 wave of the Chapel Hill Expert Survey includes anchoring vignettes which we use as "bridge votes" to place parties from different countries on a common space. We estimate our dimensions using the "blackbox" technique. Our results demonstrate both the usefulness of anchoring vignettes and the broad applicability of the blackbox scaling routine. Further, the resulting scale offers a cross-nationally comparable interval-level measure of a party's left/right ideological position with a high degree of face validity. In short, we argue that the left/right economic dimension travels well across European countries.

From the right-wing Danish Liberal Alliance (LA) to the leftist German Die Linke (LINKE), European political parties span the economic left/right ideological space.[1] Since the economic left/right dimension is "the one dimension that arguably allows for meaningful cross-national comparisons" (Adams 2012, 402), much of the research that incorporates placements of political parties on the left/right dimension assumes, at least implicitly, comparability of the measures across national contexts (cf. Adams et al. 2006; Adams, Ezrow, and Somer-Topcu 2011; Aspinwall 2007; Hooghe, Marks, and Wilson 2002; Marks et al. 2006; Tavits 2007; Tavits and Letki 2009).

While proponents of expert surveys find that they provide cross-nationally valid and reliable measures of party positioning for scholars of European political parties (Bakker et al. 2012; Benoit and Laver 2006; Hooghe et al. 2010; Ray 1999; Steenbergen and Marks 2007), critics point out that central concepts, such as economic left/right, lack a secure anchor in the surveys, leaving the expert to determine their own frame of reference (Budge 2000; McDonald, Mendes,

and Kim 2007). Without a shared anchor, scholars are less certain whether a centrist party in Sweden would be considered a centrist or a left-wing party in the United Kingdom. Taken to its logical extreme, this problem could confound cross-country comparison, implying that researchers using expert surveys risk missing potential differences between countries. In other words, we do not have a common ideological space for European political parties to use for cross-national comparisons.

McDonald, Mendes, and Kim recommend that future "expert surveys might be able to create anchors by identifying for respondents a common reference" (2007, 74). In this article, we address this criticism by combining recent methodological advances in survey research with well-known scaling techniques to estimate cross-nationally comparable ideological positions of political parties across a range of countries. We use anchoring vignettes embedded in cross-national surveys to estimate a common scale for economic left/right ideological positions of political parties. The technique we employ is generalizable to nearly

---

[1]An online appendix with supplementary material for this article is available at http://dx.doi.org/10.1017/S0022381614000449. Data and supporting materials necessary to reproduce the numerical results are available at chesdata.eu/replication.html.

any survey-based perception data that include some sort of bridging information (i.e., some question(s) in the survey that bridge across different national or institutional contexts). In our application, we use the anchoring vignettes included in the 2010 wave of the Chapel Hill Expert Survey (Bakker et al. 2012) as bridges to place political parties in the European Union on a common ideological space. We estimate the positions using blackbox scaling, which was developed by Poole (1998). The scaling routine fits the data well and returns estimates that have strong face validity.

The issue of cross-context comparability is relevant to a wide variety of substantive areas. One of the first, and most prominent, attempts to overcome this problem was the work of Aldrich and McKelvey (1977), who developed a scaling routine to address the fact that survey respondents' perceptions of the ideological positions of presidential candidates were often substantially different from one another. If such differences were real, then placing candidates and respondents on a common scale would be highly problematic. Their technique assumes the candidate positions are fixed and treats differences in respondents' placements of candidates as different interpretations of the underlying scale and then corrects for these differences. Poole (1998) generalized their solution to allow for missing data and multiple dimensions with his blackbox technique applied to a transposed matrix of perception data.

Applying scaling techniques such as Aldrich-McKelvey or Blackbox requires some sort of bridging information. Bridging information can come in many forms (e.g., votes or actors) and allows placements of different stimuli (e.g., parties or legislators) to be located on the same latent scale, thus facilitating cross-context comparisons. Bridge votes are commonly employed in American politics literature to place Senators and House members (Clinton and Meirowitz 2003; Poole 1998), actors from different institutions (Bailey 2007; Epstein et al. 2007), or legislators from different states on a common scale (Shor, Berry, and McCarty 2010).

More recently, King et al. (2004) introduced the idea of anchoring vignettes to survey research in political science. These authors demonstrate how to correct for cross-cultural differences in the respondents' perceptions of difficult to measure concepts. Providing concrete examples of these concepts in the survey helps to overcome the biases in self-placement that individuals in varying cultural contexts often display. King and Wand (2007) develop software that helps diagnose, explain, and correct for this differential item functioning (DIF) in perception.

We combine these two ideas, drawing from the strengths of both techniques. Including anchoring vignettes in the survey allows us to measure the degree to which the experts perceive the left/right economic scale of political competition in the same way. By applying blackbox scaling, using the anchoring vignettes as bridges, we generate a new variable that has highly desirable properties. Namely, the resulting measure is interval-level, cross-nationally comparable, and exportable for use in subsequent analyses. Finally, we estimate bootstrapped standard errors for the placements in order to summarize the uncertainty in the results.

The article proceeds as follows. The second section introduces the problem of cross-national comparability as it pertains to the study of ideological positioning on the left/right economic scale. We then provide an overview of the use of anchoring vignettes to address DIF in survey-based measures. The next section considers the use of blackbox scaling applied to perception data using anchoring vignettes as bridges. The following section describes the Chapel Hill Expert Survey (CHES) and provides descriptive information regarding the anchoring vignettes. We then provide the results of our estimation. To preview the conclusions, we present a common scale of European parties on the economic left/right dimension. Further, we find that experts conceive of the left/right economic dimension in similar ways across Europe, lending confidence to scholars interested in cross-national comparisons of party positions.

Comparable Cross-National Party Position Measures

Aldrich and McKelvey sought to resolve "the inability to obtain good empirically based measurements of the positions of the candidates and citizens in a common issue space." (1977, 111): Our article draws on the insights of those authors as well as innovations in scaling and survey methods (King et al. 2004; King and Wand 2007; Poole 1998; Wand 2013) to address a comparable problem in the measurement of political party positioning. Scholars of European political parties often require cross-nationally valid and equivalent measures of party positioning.

European Parliament (EP) research provides a specific example. Marsh and Mikhaylov argue that voter policy considerations and political party competition must be structured by one overarching dimension for the development of a European representation

system, and that "[t]he left/right dimension has been most often identified as the dominant dimension of political competition both on the national and European levels, thus serving as a necessary link between these two levels" (2010, 7) They go on to point out that this dimension has been found to structure the behavior of parties and voters in EP elections (Van Der Brug et al. 2009), as well as party group organization and competition within the European Parliament (Hix, Noury, and Roland 2006), which demonstrates that comparable measurements of left/right positions are important to our understanding of European multilevel governance.

Whether drawn from the Comparative Manifesto Project (CMP) data (Budge et al. 2001; Klingemann et al. 2006), expert surveys (Benoit and Laver 2006; Hooghe et al. 2010; Ray 1999; Steenbergen and Marks 2007), or public opinion surveys such as the European Election Studies, research that incorporates information on the positioning of political parties on economic left/right policies assumes, at least implicitly, comparability of the measures of positioning across a variety of national and cultural contexts (Adams et al. 2006; Adams, Ezrow, and Somer-Topcu 2011; Aspinwall 2007; Hooghe, Marks, and Wilson 2002; Marks et al. 2006; Tavits 2007; Tavits and Letki 2009). With few exceptions, such as Gschwend, Lo, and Proksch (2014) and König, Marbach, and Osnabrügge (2013), the majority of this research does not test the comparability of these data across countries nor attempt to correct for systematic biases that may be present.[2]

When relying on expert surveys or public opinion measures of party positioning, it becomes important to consider the potential obstacle that different groups of survey respondents may answer questions with ordinal response categories in systematically different ways because "some people obviously differ in optimism,

agreeability, mood, propensity to use extreme categories, and other characteristics" (King and Wand 2007, 46–47). "Response-category differential item functioning," or DIF, seems especially likely in cross-national surveys. The presence of DIF, whether because of individual or broader cultural context, in expert or mass survey respondents potentially compromises the comparability of the measurements produced by these instruments. In other words, it may complicate our ability to assert that an expert on the Swedish party system thinks of and applies concepts such as "left" and "right" the same way as an expert on British parties when they are asked to place the parties of their respective countries on an economic left/right scale.

The Comparative Manifesto Project's (CMP) data are also not exempt from these concerns. CMP researchers have analyzed and coded postwar election manifestos for a large number of parties in many industrialized democracies, providing information on various parties' support for several political issues. Although CMP coders are provided with identical coding specifications, the coders are still country-specific individuals subject to the same possible differences in optimism, mood, or likelihood to use extreme categories as the individuals that compose expert or public opinion surveys. It is difficult to know whether or not the country-specific coders are prone to DIF, but it is plausible to assume that some degree of subjectivity is present in the scoring of party manifestos and that these perceptions would differ from country to country (Bakker et al. 2011; Mikhaylov, Laver, and Benoit 2012).

Recent research by König, Marbach, and Osnabrügge (2013) attempts to directly address the question of cross-national and temporal comparability with the comparative manifesto data. The authors use Bayesian dynamic latent variable models to evaluate the comparability of the left/right positions of political parties as derived from their manifestos. In order to test for such differences, the authors employ an innovative strategy to create bridge observations with which to construct a common scale across years and countries. Much like Gschwend, Lo, and Proksch (2014), though, the validity of the bridging observations is based upon some strong assumptions, namely that left/right scores based on the European Party Manifesto (EMP) data are necessarily cross-nationally comparable and that a party's left/right score at its first European election is the same as its left/right score at the preceding national election. As the authors acknowledge (König, Marbach, and Osnabrügge 2013, 21), these strong assumptions are not directly tested.

[2]Gschwend, Lo, and Proksch (2014) use 2009 European Election Study surveys to place parties and the public on a common scale within country. They then cleverly use a party's European Parliament party affiliation to bridge across countries. This technique has the advantage of being able to be applied retroactively. In contrast, our technique is only applicable when anchoring vignettes have been included on a survey, thus it cannot be applied retroactively. However, their technique has the disadvantage of relying on EP party groups to act as reliable bridges, but research suggests that European party groups can include parties that are quite ideologically different. As McElroy and Benoit (2010, 20) argue, a national party and a European party group can diverge in policy to some extent before that policy incongruence prompts the national party to switch to another group. Crucially, some parties' EP group membership may seem inappropriate from the perspective of policy alone. Given that we control the design of the anchoring vignettes, however, our bridges should be reliable.

Even so, the König, Marbach, and Osnabrügge (2013) corrected scale correlates highly with the unadjusted CHES left/right scores, suggesting that their measure has a high degree of face validity. At the same time, they highlight some bias issues with the manifesto data that could affect scholars conducting cross-national or time-series analysis with certain groups of countries. This fact further highlights the importance of applying similar tests to the expert survey data. Although limited by time coverage, our technique using anchoring vignettes has the advantage of being premised on a single assumption, vignette equivalence, which we are able to directly test in our analysis.

In order to account for potential differences in perception of the left/right economic dimension across countries, it is necessary to have experts, voters, or text coders each "place" a common party on the scale. Given that experts and voters are only asked to place parties in their own countries and that CMP coders only code party manifestos from one country, this has been a seemingly insurmountable task. If, instead, there were a common party placed by all experts, then this problem would be tractable.

Anchoring Vignettes

In a recent series of articles, King, Wand, and co-authors (King et al. 2004; King and Wand 2007; Wand 2013) introduced anchoring vignettes to political science as a way to identify whether survey respondents use ordinal response categories differently. Some respondents, for example, may have higher standards for what "strongly agree" means, thereby complicating comparisons with other respondents. In their articles, King and his colleagues used a cross-national survey item focusing on political efficacy to illustrate the problem. In this survey, respondents in Mexico and China answered the following question, "How much say do you have in getting the government to address issues that interest you?" Respondents use the following scale (1) "No say," (2) "Little say," (3) "Some say," (4) "A lot of say," (5) "Unlimited say." Contrary to expectations, Chinese respondents had much higher levels of political efficacy than their Mexican counterparts (King et al. 2004). Given that Freedom House rates China as not free (6.5/7, where 7 is least free) while Mexico is rated as partly free (3/7),[3] most observers would expect that Mexicans would have higher levels of efficacy.

While this surprising finding may be a result of actual differences in perceived efficacy, King et al. (2004) expect that the result simply reflects differing standards for efficacy between Mexico and China. To ascertain whether respondents perceived the efficacy question scale the same way, King et al. (2004, 193) included several vignettes describing the sense of efficacy of several hypothetical people alongside the self-placement to anchor a shared reference point. Using these vignettes as anchors across individuals and contexts, the researchers can correct the raw responses, thereby obtaining a comparable DIF-free measure of political efficacy. After correcting for DIF, the survey showed that Mexicans have much higher levels of efficacy than Chinese respondents. As King and Wand report, "In fact, although the raw responses had the Chinese judging themselves to have considerably more efficacy than the Mexicans judged themselves to have, the DIF-corrected responses indicate the reverse . . . "(2007, 3). This result clearly demonstrates that ignoring DIF in cross-national survey research could be seriously misleading to researchers and policy makers.

Turning to the use of expert surveys to identify ideological placements of political parties, we are interested in whether the ideological placements of parties on a left/right scale are cross-nationally comparable. In short, do we know whether a "4" in Sweden on an 11-point ideological scale is the same as a "4" in the United Kingdom? For example, would the British Labour Party, a center-left party in the United Kingdom in 2010 (4 on an 11-point scale), be perceived as center-right by Swedish experts? As with the efficacy example, we need to evaluate whether the raw responses are cross-nationally comparable. To do so, we first need to consider alternative scaling techniques.

Aldrich-McKelvey and The Blackbox

The Aldrich-McKelvey scaling procedure is widely viewed as one of the best solutions for correcting DIF in perceptual data (King et al. 2004; Saiegh 2009). Palfrey and Poole (1987) show that the Aldrich-McKelvey (A-M) solution does an exceptional job of recovering the positions of the stimuli (e.g., presidential candidates) even in the presence of heteroskedastic errors over the stimuli.

The A-M solution was developed for scaling individuals' reported perceptions of the locations of stimuli (e.g., political candidates) along a scale with labeled end points. A-M assume that the individual

reports a noisy linear transformation of the true location of the stimulus. The A-M solution, then, recovers both the "true" underlying configuration of the stimuli as well as a set of parameters that map the noisy perceptions onto the "true" scale. This mapping corrects for differences in location as well as differences in relative distances between points. That is, the A-M solution produces a scale where a "4" in Sweden is the same as a "4" in the United Kingdom and where a one-unit change in the scale means the same thing across different contexts. It is these individual linear maps that allow both respondents and stimuli to be placed on a common scale. Formally,

$$\alpha_i + \beta_i z_{ij} = z_j + u_{ij}, \tag{1}$$

Where α and β are the linear maps, z_{ij} is the perceived location of stimulus j by individual i, z_j is the "true" location of stimulus j, and u_{ij} meets Gauss-Markov assumptions.[4]

Thus, the solution for the individual transformation parameters is simply the "least-squares regression of the reported on the actual (unknown) positions of the candidates" (Aldrich and McKelvey 1977, 115). For our purposes, though, the major limitations to the A-M solution are that it only works for unidimensional solutions, and it cannot handle missing data.

Poole (1998) generalizes A-M to allow for missing data and the estimation of higher-dimensional solutions. Poole's blackbox scaling procedure can handle both perception and preference data and has been shown to recover valid and precise estimates of both stimuli and respondents. In order to estimate positions of the stimuli, political parties in our application, in a common space, it is required that the respondents answer some of the same questions. In many applications, this has been a self-placement question combined with placements of the stimuli. Like A-M, the model assumes the stimuli have true positions, but the respondent perceptions of these positions are linear transformations of these true positions. The model then estimates the parameters of these linear transformations (constants and weights) for each respondent and corrects the placements of the stimuli and respondents appropriately.[5]

In our application, the stimuli are the parties and the respondents are the experts. Party experts in any given country, however, only place parties from their country. When compiled together, the matrix of all experts by all parties is a block diagonal matrix with missing values on the off diagonals. For example, British experts do not place Belgian parties, so there is an abundance of missingness. This is not true, however, for the vignette parties on the survey. Every expert, regardless of country of expertise, was asked to place the three vignette parties and, as will be described in more detail below, were largely in agreement as to the ordering of the vignette positions. We use this information to correct for any potential DIF in the perception of the left/right economic scales and correct the perceived positions of the stimuli in exactly the same way as A-M.[6]

Our resulting data matrix has parties down the rows and experts across the columns—which is the key to allowing common-space scaling. Given the demands on the data, four times as many placements of each stimuli as there are dimensions is appropriate (i.e., we need a minimum of eight experts per party to conduct the scaling). This results in the loss of several parties from our sample because we have fewer than eight experts in some of the countries in the survey. With these limitations, the resulting data matrix has 118 parties (rows) and 224 experts (columns) across 14 countries. Given that a vast majority of the experts places each of the three vignette parties, there are over 160 "bridging" observations (experts) in the data which can be used to place all stimuli (parties) on a common scale. In the following section, we describe the anchoring vignettes in the 2010 Chapel Hill Expert Survey data.

Vignettes in the 2010 Chapel Hill Expert Survey

The 2010 Chapel Hill Expert Survey (CHES) measures party positioning on European integration, ideology, and policy issues for 227 national parties in a variety of European countries, extending a time series dating back to 1996 (Hooghe et al. 2010; Ray 1999; Steenbergen and Marks 2007). In addition to the general left/right, economic left/right, european integration, and social left/right scales, the survey includes a battery of 18 issue-specific questions that are designed to capture a number of potentially

---

[4]For more technical information about our use of the Aldrich and McKelvey (1977) and (Poole 1998) procedures, see the online appendix.

[5]For additional details on estimation, see Poole (1998) and the online appendix.

[6]This technique provides better estimates of the stimuli (parties) than the respondents (experts), but given our interest in the stimuli only, this poses no problem.

salient dimensions and are informed by prior scholarship on party competition in Europe (Hooghe and Marks 2009; Hooghe, Marks, and Wilson 2002; Inglehart 1977, 1990; Kitschelt 1994; Lijphart 1999; Lipset and Rokkan 1967).[7]

Researchers have cross-validated the 2010 CHES data with both party manifesto data and public opinion data (Bakker et al. 2013). Other research finds that Chapel Hill expert judgments correlate highly with measures drawn from different datasets (Netjes and Binnema 2007; Steenbergen and Marks 2007). Furthermore, expert survey-based party positions are more consistent with the evaluations of voters and parliamentarians than data currently available from party manifestos (Marks et al. 2007). However, previous reliability tests of expert survey data (CHES or Benoit and Laver) have not tested for DIF bias, so the question of cross-national comparability remains.

In the survey, experts place political parties on a basic economic left/right dimension focusing on the government's role in the economy, taxes, and redistribution. After the experts answer all the substantive questions about the parties, the survey presents a series of vignettes and asks experts to place these hypothetical parties on the same 0–10 scale. Below, we present the vignettes from the survey.

> **Party A** views the provision of a social safety net as important, but believes there is a sharp trade-off between welfare spending and economic competitiveness. It favors limiting government regulation to instances of market failure and prefers cuts in social spending over new taxes to meet rising social needs.
>
> **Party B** views the equalization of life chances for all citizens as an important goal of government. It favors active government in regulating domestic and international markets and supports steeply progressive taxes to fund redistributive social programs.
>
> **Party C** believes in small government. It favors minimal regulation of domestic and international markets, supports the privatization of many government operations, and opposes high taxes to fund redistributive social programs.

In order to achieve vignette equivalence, the experts need to order the parties in a consistent way across countries. According to King et al., vignette equivalence is the "assumption that the level of the variable represented in any one vignette is perceived by all respondents in the same way and on the same

unidimensional scale, apart from random measurement error" (2004, 194). We expected these vignette parties to follow a consistent pattern along the economic left/right scale, ranging from left-wing (Party B) to right-wing (Party A) and further right-wing (Party C). Empirical analysis confirmed that most experts perceived the vignettes in the intended order (Bakker et al. 2011).[8] These findings demonstrate that the *vignette equivalence* assumption has been met, suggesting that we can assert that the economic left/right dimension as described by the vignettes has a "logically coherent and consistent meaning in different cultures" (King et al. 2004, 194).

Given the large degree of agreement on the part of the experts regarding the placements of the vignette parties, these placements can be used to correct for any potential DIF in the expert perceptions of the left/right economic scale. Although the anchoring vignette software developed by King and Wand (2007) makes it simple to diagnose and correct for DIF using either a parametric or nonparametric solution, their technique does not yield an exportable, continuous, DIF-corrected variable nor does it yield estimates of uncertainty for the DIF-corrected placements.

Their nonparametric method can create a DIF-corrected variable, which could then be exported and used in subsequent models, but it is rather coarsely measured and is highly influenced by the distances between vignette parties. If two vignette parties are relatively far apart from one another, then the distribution of parties across the categories of the DIF-corrected variable will be highly skewed, with one category dominating the others. The researcher can attenuate this problem by including more vignettes in the survey and designing them to be as evenly spaced from one another as possible. Conditional on the length of the survey and the number of ordered categories in the variable(s) of interest, this can add a nontrivial number of items to the survey, thus increasing its expense. Regardless of the number of vignettes included on the survey, the DIF-corrected variable will

---

[7]The Chapel Hill expert survey incorporates most of the questions that were used for European countries in Benoit and Laver's (2006, 2007) 47 country expert survey from 2002 to 2003. The survey questions and data can be found at chesdata.eu.

[8]Nearly 89% of the experts did not violate the expected order, while another 10% had only one violation. The frequency table showing all expert orderings of the experts is available in the online appendix. Bakker et al. (2011) also used the rank-order correlations of the vignette parties across all countries to ascertain whether the vignettes fall on a cross-nationally comparable-ordered scale. With the exception of Greece, the mean placements of the vignette parties on the left/right scale are perceived in the same order across all countries. The Greece experts perceived the vignette parties A and C as equals ($B < A = C$). Given the lack of vignette equivalence in Greece, we had to remove Greece from the data.

have a set number of categories ($2J + 1$ where J is the number of vignettes).

The parametric model treats the placement as the response variable and corrects for DIF in the estimation of the *placement variable*. Compared to the nonparametric method, this approach has the advantage of needing only a few vignettes (technically only one) to complete the estimation (King et al. 2004; King and Wand 2007). This approach estimates a censored ordered-probit model with random cut points that can vary as a function of respondent-specific covariates. The estimates of the cut points for the ordered categories may often be of interest. This approach is ideal if the researcher is interested in diagnosing and explaining the causes of DIF as the estimated effects for the covariates on the random cut points provide this information. Although the results from this routine can be highly informative, it does restrict the variable of interest to the left-hand side of the regression and does not yield a new, exportable, DIF-corrected variable. If a researcher wants to use a DIF-corrected measure of ideological placements as an explanatory variable and/or desires an interval-level measure of the concept, other scaling techniques are necessary.

The European Common Space

Using anchoring vignettes to help bridge across countries, we apply the blackbox method described above to the 2010 round of the Chapel Hill expert survey. We estimate a two-dimensional solution for the economic left/right position of the parties in our sample. The decomposition of the data provides a diagnostic for the dimensionality of the underlying scale.

The data consist of 118 parties (rows) and 224 experts (columns) in 14 countries. We have responses from over 160 experts per vignette party and a range of 8 to 17 expert placements for each actual political party in the data. Given the structure of the data (experts on the columns), this gives us ample information with which to bridge across countries. That is, Equation (1) is identified, and the individual transformation parameters are estimable. These parameters are the "linear maps" that allow the stimuli to be placed in a common space.

The first dimension explains approximately 87% of the variance in the party placements, with the second dimension adding an additional 6%. Figure 1 displays the plot of the first common-space dimension versus the second common-space dimension, with the raw, unscaled CHES data represented with shapes.

FIGURE 1 **Two-Dimensional Left/Right Economic Solution, with Party Labels (Left = Circles, Moderate = Squares, Right = Triangles)**

Source: http://chesdata.eu/replication.html.
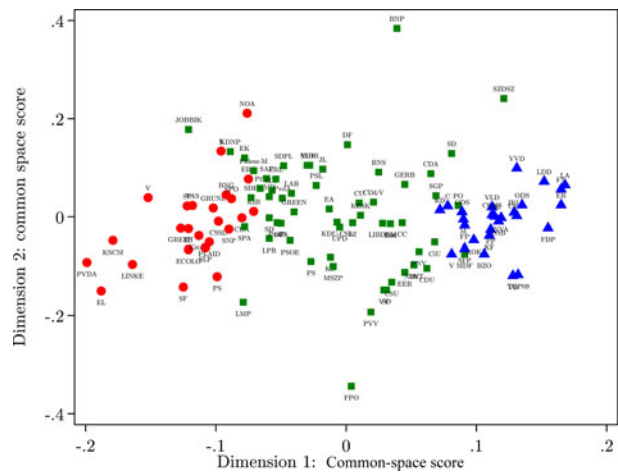
Parties with a mean expert placement on the left-end of the scale (< 3) are circles, those with mean expert placements between 3 and 7 are squares, and those greater than 7 are triangles.

To further assess the face validity of the results, we include party labels for the markers. In Figure 1, the parties to the furthest left and right are the Worker's Party of Belgium (PVDA/PTB) and the Danish Liberal Alliance, respectively. These are also the two most extreme parties according to the unadjusted mean placements by the experts. The ordering of the other parties along the X-axis strongly suggests that the first common-space dimension is the classic economic left/right dimension.

The meaning of the second dimension is less obviously clear from this plot. The British National Party and the Austrian FPÖ both appear as extreme outliers on this dimension in Figure 1. In contrast, the mean expert left/right placement of these parties is approximately in the middle of the scale, which is where the common-space solution placed them as well.

Upon further inspection, however, a commonality between the two parties is the extreme disagreement among the experts in their placements. In fact, experts in the United Kingdom and in Austria place these parties from 0 to 10 along the economic left/right scale, spanning the entire range of the scale. This indicates that expert uncertainty in the placement of parties on the economic left/right scale could explain the parties' positions on the second dimension.

In order to better characterize the uncertainty in the party placements, we estimated bootstrap standard errors for the parties on both dimensions. We sampled 1,000 replicate data sets by sampling from the experts with replacement. This required a bit of additional "record keeping" given the missingness discussed above. That is, in order to maintain the common-space scores, we had to sample the appropriate experts with replacement from the appropriate countries. We then compute standard errors by summarizing the results of the 1,000 blackbox estimations. Figure 2 displays the standard errors for each party in the data for each country individually. It is immediately apparent that the parties with more extreme values on the second common-space dimension also have considerably larger standard errors in terms of their placements on the first dimension. Given this result, it appears that the second dimension captures a degree of uncertainty around the party placements. Since the second dimension only explains 6% of the variance in party placements, we concentrate attention on the first dimension.

Using the first-dimension scores, Figure 3 presents party ideological positions and their associated 95% confidence intervals. Here again, we see the large confidence bounds for parties such as the British BNP and the Austrian FPÖ.[9]

Notice that Figure 3 places all European parties in a single common space, so we can compare the left/right positions (and rank order) of parties from different countries with far more confidence than previous estimates of party ideological positions. By simply looking at the placements and confidence intervals, readers can quickly evaluate whether a French left-wing party is further right-wing than a Swedish or Hungarian left-wing party.

This comparability has practical value for European party scholars. Whitaker and Lynch (2013), for example, use the weighted positions of national parties as measured in the 2010 CHES data to calculate the ideological diversity of party groups in the European Parliament. Their analysis assumes cross-national comparability in the left/right positioning of political parties across the European Union; our results lend credibility to this important, as yet untested, assumption.

Finally, we conduct analyses to assess the face validity of the common-space scores. In Figure 4,

FIGURE 2 Bootstrapped Standard Errors for First and Second Common-Space Dimension
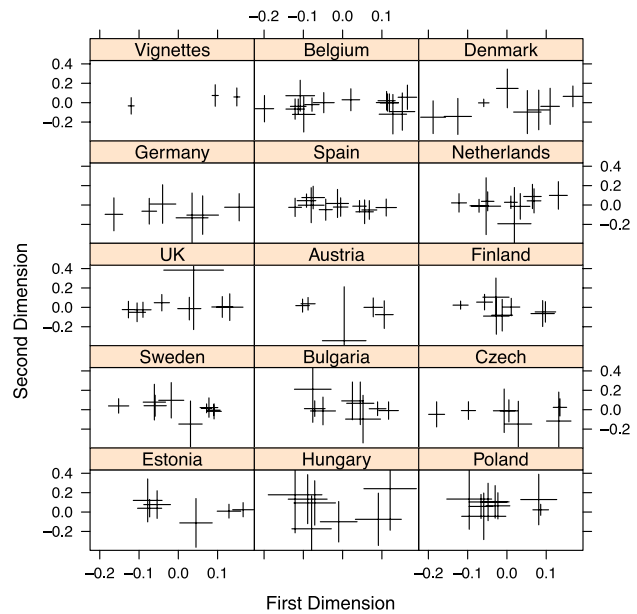
Note: The intersection of the line segments plots the two-dimensional placement, and the length of each segment is the standard error in that dimension. That is, the horizontal line segment represents the standard error of the first dimension whereas the vertical segment represents the standard error of the second dimension.

we present country-specific graphs of the common-space first dimension against the mean economic left/right placements from the experts.

Figure 4 shows that in most countries, the ordering from left to right does not change in interesting ways. In fact, in very few countries is the rank-order correlation between the two scales less than 0.90, indicating that the rank ordering is extremely similar across the two measures. There are some differences in rank orders in the newer member states of the EU, but the smallest correlation is still 0.81, indicating a large amount of agreement.[10]

In addition to comparing with the raw expert scores, we also compared the common-space scores with two other commonly used measures of party placements, drawing on the European Election Survey and the Comparative Manifesto Project database.[11]

[9]The blackbox scaling routine provides additional information in terms of the fit of the model for specific parties. In the online appendix, we present the number of expert placements and the R^2 for each party in the data. In the analysis, the R^2 is a measure of how much of the variance in a given party's placement scores is explained by the latent dimensions.

[10]We present the rank-order correlations along with the number of parties in each country in the online appendix.

[11]For the EES and manifesto data, see http://eeshomepage.net and https://manifesto-project.wzb.eu. The EES placements are the public's average left/right placement of the parties while the CMP measure is the left/right summary placement ("rile").

FIGURE 3   European Parties in a Common Ideological Space
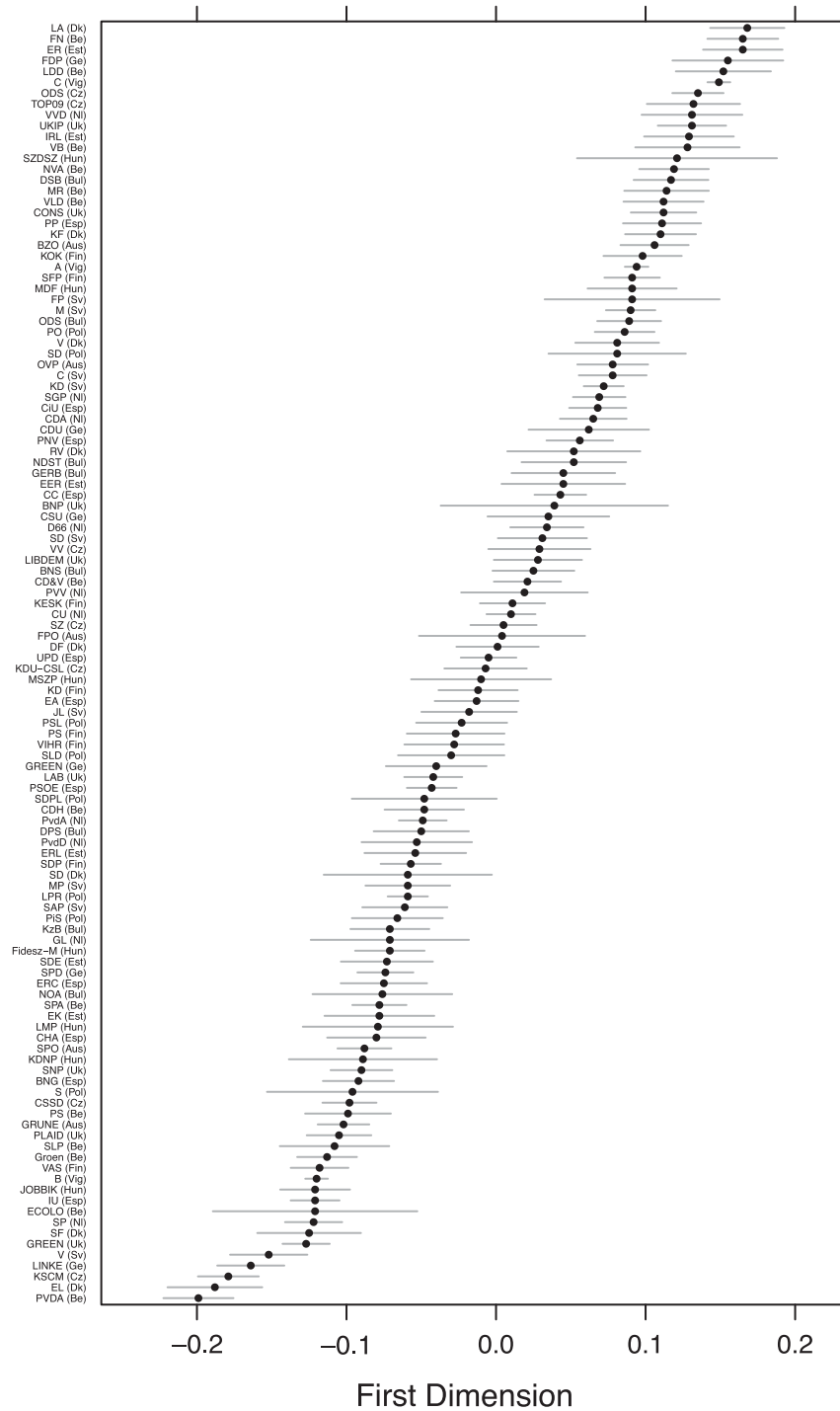
Note: In this plot, the dots represent the placement of each party along the first common-space dimension, and the bars are 95% confidence intervals built using bootstrapped standard errors.

The common-space correlations for EES and CMP are 0.85 and 0.51, respectively. Figure 5 presents the relationships between the common-space scores and the other two measures, with CMP on the left-hand axis and EES on the right-hand side. Despite a few cases off the diagonal, especially in the CMP comparison, Figure 5 shows a strong relationship between the indicators.

Figures 4 and 5 offer evidence that the estimates from the blackbox technique produce valid,

FIGURE 4 First Common-Space Dimension and Mean-Expert Placements, by Country
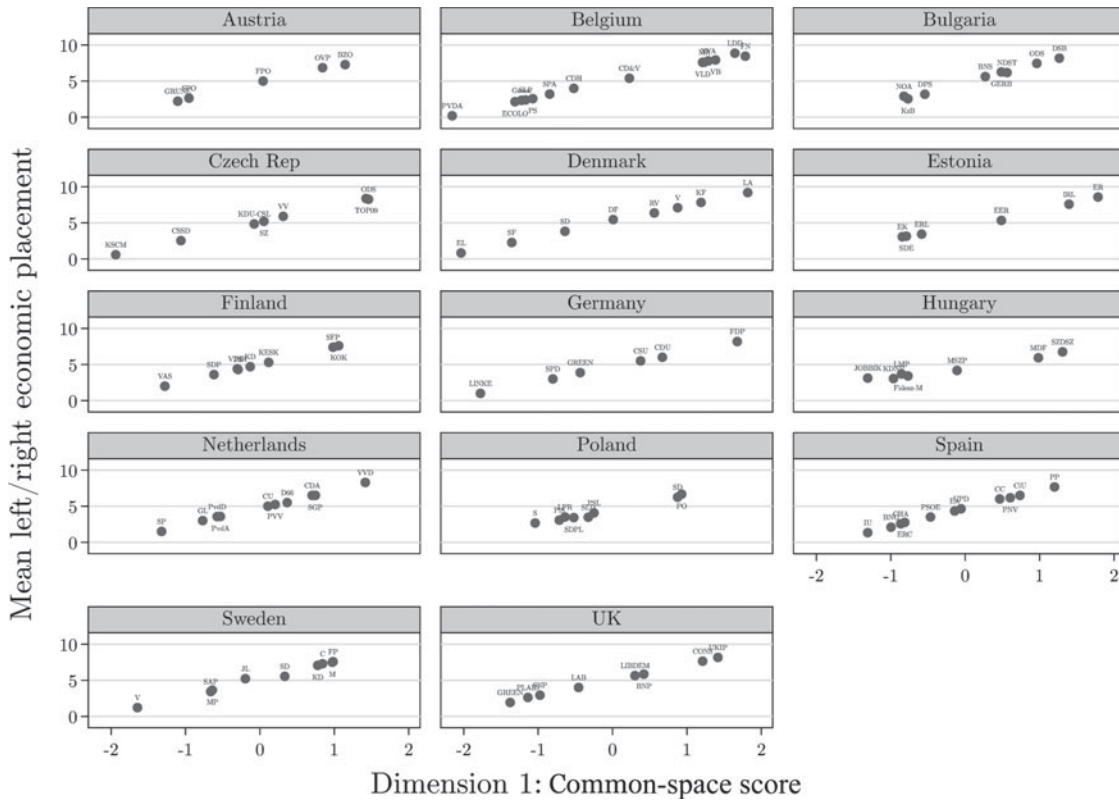
Note: This figure displays the relationship between the first common-space dimension and the unadjusted mean-expert placements from the 2010 Chapel Hill Expert Survey, by country.

cross-nationally comparable placement of parties on the economic left/right dimension. In addition, this consistency between the common-space scale and the unadjusted expert mean placements should be seen as providing confidence to researchers interested in previous waves of the CHES surveys.[12] Based on these analyses, we assert that previous waves of the survey can be treated as providing valid, cross-nationally comparable measures of party positions.

Furthermore, these analyses offer an internal validity check of the common-space scale. Intracountry party-position rank orderings are not as susceptible to DIF as cross-country comparisons; because the common-space scale matches the intracountry expert rankings, we have more confidence in the validity of our measure. Returning to Figure 3, then, we offer for the first time a reliable ideological rank ordering of all European parties—within the 14 countries analyzed—on a common ideological space.

---

[12]Similarly, the fact that König, Marbach, and Osnabrügge'o (2013) bias-corrected manifesto data correlates highly with the unadjusted CHES positions is encouraging.

Discussion

To study political parties cross-nationally, researchers need reliable estimates of party positions. While manifestos offer a potential solution with obvious advantages in data availability across time and space, this approach is not without limitations. Scholars cannot ignore the fact that manifestos are strategic documents written by the parties themselves. Due to campaign priorities or intraparty dissent, manifestos do not always discuss all policy areas, which limits analysis of any extant but unmentioned positions. Quite simply, what parties say in manifestos does not necessarily equate to objective policy positions.

In the last few decades, expert surveys have emerged as a viable alternative to the manifestos (Bakker et al. 2012; Benoit and Laver 2006, 2007; Hooghe et al. 2010; Ray 1999; Steenbergen and Marks 2007), a complementary data source that meets requests for an emphasis on observable party behavior rather than party promises as measures of party positions (Adams, Ezrow, and Leiter 2012; Adams, Ezrow, and Somer-Topcu 2011; Fortunato and

FIGURE 5 **First Common-Space Dimension, EES, and CMP Party Placements**
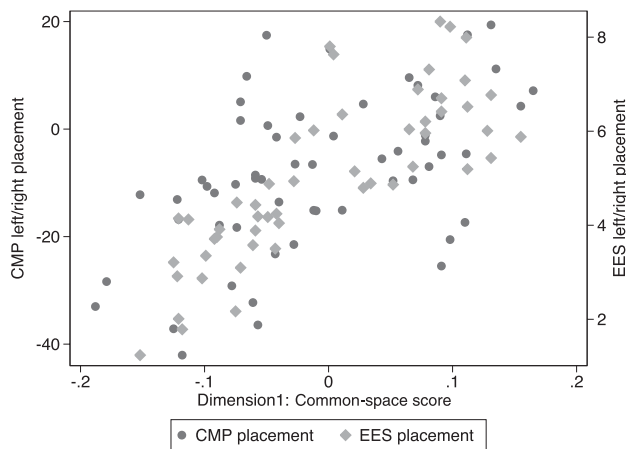
Note: This figure displays the relationship between the first common-space dimension and the left/right party placements from the Comparative Manifesto Project (circles) and the first common-space dimension and left/right placements based on the European Election Study (diamonds).
Source: https://manifesto-project.wzb.eu/publications/marpor and http://eeshomepage.net/ees-2009-study/

Stevenson 2013). Prior to this article, however, researchers could not know with certainty whether experts in different countries perceived the ideological questions in the same way. In short, as McDonald, Mendes, and Kim (2007) have argued, expert survey proponents could not confidently assert that a "2" in Sweden would be a "2" in the United Kingdom or Estonia without using anchors. With the anchoring vignettes embedded in the 2010 CHES survey and the innovative use of existing scaling techniques in this article, we have addressed this concern. As expert surveys become more widely used in other subfields of political science (Adams et al. 2011; cf. Stone and Simas 2010), establishing the extent of comparability in expert evaluations takes on additional importance for the broader discipline.

A conventional application of the anchoring vignettes technique is appropriate and necessary for detecting and/or diagnosing the causes of DIF. If, however, a researcher's interest lies in estimating a latent position and then using this scale as either a right- or left-hand side variable in subsequent work, the anchoring vignette technology must be combined with other methods. We have shown that anchoring vignettes can be used as bridge votes in a set of scaling algorithms designed to place respondents and stimuli on common scales. A-M scaling yields such a result as long as some sort of self-placement score is available. If, however, there is interest in comparing the

placements of stimuli and/or respondents across contexts, or if the scaling solution is multidimensional, then blackbox scaling is required as there will be missing data, i.e., respondents in one context that do not place stimuli in a different context.

In such settings, anchoring vignettes provide the bridges that allow the estimation of a common-space solution. Our technique does not diagnose the causes of DIF, as does the anchoring vignette software, but it does create an exportable variable that is free of DIF. Our proposed strategy easily applies to any survey-based perception data that includes anchoring vignettes or some other bridging information. The combination of anchoring vignettes and blackbox scaling creates a powerful tool to test for and overcome issues of cross-contextual comparisons with broad applicability to a variety of research agendas. While the application in this article focuses on the cross-national comparability of expert survey data, the methodological contributions of this article are useful to most research involving survey-based placements of positions of people, parties, and institutions across different contexts. For instance, Kitschelt and Wilkinson (2007) argue for the need for cross-nationally comparable measures of clientelism using expert surveys; the techniques developed here would work well to advance that agenda.

In using the three hypothetical parties included in the CHES as the bridge items to jointly scale the expert perceptions data for the country-specific parties, this article presents an innovative application of anchoring vignettes. Including the anchoring vignettes on expert surveys of party ideological positions facilitates the estimation of an interval-level, two-dimensional scale. The resulting estimates fit the data well and have strong face validity, given the high correlation between the first dimension of the common-space solution and other common measures of ideological positions of the parties in our sample.

It is true that the present technique cannot be applied to previous waves of the survey that lacked anchoring vignettes, but the relatively consistent group of countries, stable pool of experts surveyed, and centrality of the economic left/right dimension to European political competition support the cross-national validity of expert placements in previous rounds of the survey. The increased confidence we have in asserting the comparability of expert placements of parties on the economic left/right scale across the wide range of countries included in the CHES data is relevant for scholars interested in party politics (Schumacher, De Vries, and Vis 2013), the causes and electoral consequences of party policy

shifts in multiparty democracies (Adams, Ezrow, and Somer-Topcu 2011), and coalition behavior (Fortunato and Stevenson 2013), along with scholars interested in adopting best practices for survey-based research.

Significantly for expert survey data users, our results suggest that party experts in Europe view the left/right economic dimension of party competition in largely the same way across countries. The left/right economic dimension travels well throughout Europe.

Acknowledgments

References

Adams, James. 2012. "Causes and Electoral Consequences of Party Policy Shifts in Multiparty Elections: Theoretical Results and Empirical Evidence." *Annual Review of Political Science* 15: 401–19.

Adams, James, Lawrence Ezrow, and Debra Leiter. 2012. "Partisan Sorting and Niche Parties." *West European Politics* 35 (6): 1272–94.

Adams, James, Lawrence Ezrow, and Zeynep Somer-Topcu. 2011. "Is Anybody Listening? Evidence That Voters Do Not Respond to European Parties' Policy Statements During Elections." *American Journal of Political Science* 55 (2): 370–82.

Adams, James, Michael Clark, Lawrence Ezrow, and Garrett Glasgow. 2006. "Are Niche Parties Fundamentally Different from Mainstream Parties? The Causes and the Electoral Consequences of Western European Parties' Policy Shifts, 1976–1998." *American Journal of Political Science* 50 (3): 513–29.

Adams, James, Samuel Merrill III, Elizabeth N. Simas, and Walter J. Stone. 2011. "When Candidates Value Good Character: A Spatial Model with Applications to Congressional Elections." *Journal of Politics* 73 (1): 17–30.

Aldrich, John H., and Richard D. McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *American Political Science Review* 71 (1): 111–30.

Aspinwall, Mark. 2007. "Government Preferences on European Integration: An Empirical Test of Five Theories." *British Journal of Political Science* 37 (1): 89–114.

Bailey, Michael A. 2007. "Comparable Preference Estimates Across Time and Institutions for the Court, Congress, and Presidency." *American Journal of Political Science* 51 (3): 433–48.

Bakker, Ryan, Catherine de Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. 2012. "Measuring Party Positions in Europe: The Chapel Hill Expert Survey Trend File, 1999–2010." *Party Politics* Forthcoming. Published online before print November 29, 2012. doi: 10.1177/1354068812462031.

Bakker, Ryan, Erica Edwards, Seth Jolly, Jon Polk, Jan Rovny, and Marco Steenbergen. 2011. "Anchoring the Experts: Using Vignettes to Compare Party Ideology across Countries." Working paper, University of Georgia.

Benoit, Kenneth, and Michael Laver. 2006. *Party Policy in Modern Democracies*. London: Routledge.

Benoit, Kenneth, and Michael Laver. 2007. "Estimating Party Policy Positions: Comparing Expert Surveys and Hand-Coded Content Analysis." *Electoral Studies* 26: 90–107.

Budge, Ian. 2000. "Expert Judgements of Party Policy Positions: Uses and Limitations in Political Research." *European Journal of Political Research* 37 (1): 103–13.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tanenbaum. 2001. *Mapping Policy Preferences. Estimates for Parties, Electors, and Governments 1945-1998*. New York: Oxford University Press.

Clinton, Joshua D., and Adam Meirowitz. 2003. "Integrating Voting Theory and Roll Call Analysis: A Framework." *Political Analysis* 11: 381–96.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland. 2007. "The Judicial Common Space." *Journal of Law, Economics, and Organization* 23 (2): 303–25.

Fortunato, David, and Randolph T. Stevenson. 2013. "Perceptions of Partisan Ideologies: The Effects of Coalition Participation." *American Journal of Political Science* 57 (2): 459–77.

Gschwend, Thomas, James Lo, and Sven-Oliver Proksch. 2014. "A Common Left-Right Scale for Voters and Parties in Europe." *Political Analysis* 22 (2): 205–23.

Hix, Simon, Abdul Noury, and Gérard Roland. 2006. "Dimensions of Politics in the European Parliament." *American Journal of Political Science* 50 (2): 494–511.

Hooghe, Liesbet, and Gary Marks. 2009. "A Postfunctionalist Theory of European Integration: From Permissive Consensus to Constraining Dissensus." *British Journal of Political Science* 39 (1): 1–23.

Hooghe, Liesbet, Gary Marks, and Carole J. Wilson. 2002. "Does Left/Right Structure Party Positions on European Integration?" *Comparative Political Studies* 35 (8): 965–89.

Hooghe, Liesbet, Ryan Bakker, Anna Brigevich, Catherine de Vries, Erica Edwards, Gary Marks, Jan Rovny, Marco Steenbergen, and Milada Vachudova. 2010. "Reliability and Validity of Measuring Party Positions: The Chapel Hill Expert Surveys of 2002 and 2006." *European Journal of Political Research* 49 (5): 687–703.

Inglehart, Ronald. 1977. *The Silent Revolution: Changing Values and Political Styles Among Western Publics*. Princeton, NJ: Princeton University Press.

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

King, Gary, Christopher J. L. Murray, Joshua A. Salmon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207.

King, Gary, and Jonathan Wand. 2007. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes." *Political Analysis* 15 (1): 46–66.

Kitschelt, Herbert. 1994. *The Transformation of European Social Democracy*. New York: Cambridge University Press.

Kitschelt, Herbert, and Steven I. Wilkinson. 2007. A Research Agenda for the Study of Citizen-Politician Linkages and Democratic Accountability. In *Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition*, ed. Gary Marks and Marco R. Steenbergen. New York: Cambridge University Press, 322–43.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge, and Michael McDonald. 2006. *Mapping Policy Preferences II. Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990–2003*. New York: Oxford University Press.

König, Thomas, Moritz Marbach, and Moritz Osnabrügge. 2013. "Estimating Party Positions across Countries and Time—A Dynamic Latent Variable Model for Manifesto Data." *Political Analysis* 21 (4): 468–91.

Lijphart, Arend. 1999. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven, CT: Yale University Press.

Lipset, Seymour Martin, and Stein Rokkan. 1967. Cleavage Structures, Party Systems, and Voter Alignments. In *Party Systems and Voter Alignments: Cross-National Perspectives*, ed. Seymour Martin Lipset and Stein Rokkan. New York: Free Press, 1–64.

Marks, Gary, Liesbet Hooghe, Marco R. Steenbergen, and Ryan Bakker. 2007. "Cross-Validating Data on Party Positioning on European Integration." *Electoral Studies* 26 (1): 23–38.

Marks, Gary, Liesbet Hooghe, Moira Nelson, and Erica Edwards. 2006. "Party Competition and European Integration in East and West. Different Structure, Same Causality." *Comparative Political Studies* 39 (March): 155–75.

Marsh, Michael, and Slava Mikhaylov. 2010. "European Parliament Elections and EU Governance." *Living Reviews in European Governance* 5 (4): 1–30.

McDonald, Michael D., Silvia M. Mendes, and Myunghee Kim. 2007. "Cross-Temporal and Cross-National Comparisons of Party Left-Right Positions." *Electoral Studies* 26: 62–75.

McElroy, Gail, and Kenneth Benoit. 2010. "Party Policy and Group Affilation in the European Parliament." *British Journal of Political Science* 40 (2): 377–98.

Mikhaylov, Slava, Michael Laver, and Kenneth Benoit. 2012. "Coder Reliability and Misclassification in the Human Coding of Party Manifestos." *Political Analysis* 20 (1): 78–91.

Netjes, C., and H. A. Binnema. 2007. "The Salience of the European Integration Issue: Three Data Sources Compared." *Electoral Studies* 26 (1): 39–49.

Palfrey, Thomas R., and Keith T. Poole. 1987. "The Relationship between Information, Ideology, and Voting Behavior." *American Journal of Political Science* 31 (3): 511–30.

Poole, Keith T. 1998. "Recovering a Basic Space From a Set of Issue Scales." *American Journal of Political Science* 42 (3): 954–93.

Ray, Leonard. 1999. "Measuring Party Orientation Towards European Integration: Results From an Expert Survey." *European Journal of Political Research* 36 (6): 283–306.

Saiegh, Sebastian M. 2009. "Recovering a Basic Space from Elite Surveys: Evidence from Latin America." *Legislative Studies Quarterly* XXXIV (1): 117–45.

Schumacher, Gijs, Catherine De Vries, and Barbara Vis. 2013. "Why Do Parties Change Position? Party Organization and Environmental Incentives." *Journal of Politics* 75 (2): 464–77.

Shor, Boris, Christopher Berry, and Nolan McCarty. 2010. "A Bridge to Somewhere: Mapping State and Congressional Ideology on a Cross-Institutional Common Space." *Legislative Studies Quarterly* 35 (3): 417–48.

Steenbergen, Marco, and Gary Marks. 2007. "Evaluating Expert Surveys." *European Journal of Political Research* 46 (3): 347–66.

Stone, Walter J., and Elizabeth N. Simas. 2010. "Candidate Valence and Ideological Positions in US House Elections." *American Journal of Political Science* 54 (2): 371–88.

Tavits, Margit. 2007. "Principle vs. Pragmatism: Policy Shifts and Political Competition." *American Journal of Political Science* 51 (1): 151–65.

Tavits, Margit, and Natalia Letki. 2009. "When Left is Right: Party Ideology and Policy in Post-Communist Europe." *American Political Science Review* 103 (4): 555–69.

Van Der Brug, Wouter, Mark Franklin, Marina Popescu, and Gabor Toka. 2009. "Towards a European Electorate: One Electorate or Many?" In *The Legitimacy of the European Union After Enlargement*, ed. Jacques Thomassen. Oxford University Press, 65–93.

Wand, Jonathan. 2013. "Credible Comparisons Using Interpersonally Incomparable Data: Nonparametric Scales with Anchoring Vignettes." *American Journal of Political Science* 57 (1): 249–62.

Whitaker, Richard, and Philip Lynch. 2013. "Understanding the Formation and Actions of Eurosceptic Groups in the European Parliament: Pragmatism, Principles and Publicity." *Government and Opposition* December: 1–32 doi: 10.1017/gov.2013.40.

Ryan Bakker is an Assistant Professor at the University of Georgia, Athens, GA 30606.

Seth Jolly is an Assistant Professor in the Department of Political Science at Syracuse University, Syracuse, NY 13244.

Jonathan Polk is a Post-doctoral Research Fellow at the University of Gothenburg's Department of Political Science and Centre for European Research, Gothenburg, 405 30, Sweden.

Keith Poole is the Philip H. Alston, Jr. Distinguished Chair in Political Science at the University of Georgia, Athens, GA 30606.