Party Positions from Wikipedia Classifications of Party Ideology*

Michael Herrmann
University of Konstanz

Holger Döring
University of Bremen

July 2019

Abstract

We develop a new measure of party position based on a scaling of ideology tags supplied in infoboxes on political parties’ Wikipedia pages. Assuming a simple model of tag assignment, we estimate the locations of parties and ideologies in a common space. We find that the recovered scale can be interpreted in familiar terms of ‘left vs. right’. Estimated party positions correlate well with ratings of parties’ positions from extant large-scale expert surveys, most strongly with ratings of general left-right ideology. Party position estimates also show high stability in a test-retest scenario. Our results demonstrate that a Wikipedia-based approach yields valid and reliable left-right scores comparable to scores obtained via conventional expert coding methods. It thus provides a measure with potentially unlimited party coverage. Our measurement strategy is also applicable beyond Wikipedia.

*Contact: michael.herrmann@uni-konstanz.de and doering@uni-bremen.de. We thank panelists at the 2019 Annual MPSA Conference, and participants in research colloquia at the Universities of Bremen, Konstanz, and Mannheim for their feedback, especially Christopher Hare, Christopher Wratil, Samuel Baltz, Markus Tepe, Dominic Nyhuis, Paul Bederke, Sascha Göbel, Benjamin Guinaudeau, Susumu Shikano, Christian Breunig, Thomas Gschwend, and Hans-Dieter Klingemann.
1 Introduction

The idea that parties occupy different positions on an ideological continuum is fundamental to theories of the political process. The need for scoring parties’ on such a continuum is evidenced by the sheer diversity of approaches and the creativity political scientists devote to obtaining such scores, which range from expert codings of party manifestos (Volkens et al. 2013), to left-right placements of parties by expert judges (Kitschelt 2014) or voters (Lo et al. 2014), to supervised scaling of word frequencies in party manifestos (Laver et al. 2003), to unsupervised scaling approaches that extract positions from expert-coded manifesto content (Däubler & Benoit 2017) or raw word frequencies (Slapin & Proksch 2008) in political manifestos, to analyses of roll call votes (Hix et al. 2006) and parliamentary speech (Lauderdale & Herzog 2016, Peterson & Spirling 2018). Together, these approaches enable the measurement of party positions across time, space, levels of government, and policy areas.1

We propose a new measure of party left-right position, a measure that is unique in its broad coverage. The measure is derived from semi-standardized information about party ideology available on the English Wikipedia.2 In particular, we draw on ideological keywords (e.g., socialism) that Wikipedia authors use to tag parties and to link them with Wikipedia articles on those ideologies. We develop an ideal point model of how these tags are assigned and use it to scale over 2,000 parties and their associated ideologies on a latent dimension. Our scaling approach is based on the idea that co-occurrences of ideological keywords across parties reveal information about the closeness of parties and ideologies. Keywords that often occur in the same parties (e.g., socialism vs. social democracy) should be closer in space than keywords that rarely occur together (e.g., socialism vs. conservatism). Likewise, parties sharing the same ideological keywords should occupy more similar positions in political space than parties sharing few keywords. To capture these dependencies, we formulate a model in which parties are more likely to get tagged with keywords that are close to them. In a second step, we extend the model to address the fact that some keywords on Wikipedia are inherently ordered.

Our analysis demonstrates that keyword summaries provided by Wikipedia authors enable

---

1 Similar efforts have been made in order to obtain left-right scores for actors other than parties, for example voters and political candidates (Jesse 2016), bureaucrats and agencies (Clinton et al. 2011), political action committees (Bonica 2013), and occupational groups and industries (Bonica 2014).

2 Wikipedia is the largest knowledge database in the world. The English Wikipedia alone has over 5.8 million articles (https://en.wikipedia.org/wiki/wikipedia accessed March 2019). Wikipedia is also the most used knowledge database in the world with over 500 million unique visitors each month, ranking among the top 10 most visited websites for years (http://alexa.com/topsites, accessed March 2019).
valid and reliable inferences about party left-right position. Based on our model of keyword assignment, which allows for misclassification and differences in keyword informativeness, we recover a scale from Wikipedia classifications that conforms with common intuitions of left vs. right. We show that estimates of party position on this scale correlate with ratings of party position from the largest available expert surveys, and most strongly with ratings of general left-right position. We further demonstrate the reliability of our estimates over repeated measurements with Wikipedia data collected months apart. Together, our results indicate that Wikipedia classifications allow for extracting left-right scores comparable to scores obtained via conventional expert coding methods. These findings are in line with studies showing that political information on Wikipedia is often factually correct\(^3\) (Brown 2011), and that Wikipedia can be used for extending political science measurement to a much larger universe of cases (Munzert 2018).

To get a sense of the potential scope afforded by our measure, note that the largest available databases in political science together include over 4,000 unique parties (Döring & Regel 2019). Extant measures of party position cover only a minority of these. For example, the largest data source on party positions in terms of countries covered, the Democratic Accountability and Linkages Project (DALP), provides expert ratings of left-right position for 506 parties from 88 electoral democracies (see Table 1). The data source with the largest party coverage, the Manifesto Project, includes 1,170 parties from 60 countries. The use of Wikipedia allows us to expand that coverage considerably: based on the largest available compilation of parties, drawn from major political

---

\(^3\)A number of studies examines the accuracy of Wikipedia vis à vis traditional expert sources (for a review, see Mesgari et al. 2014). Unlike the majority of these studies, which employ a “small-n, every-detail approach” (see Brown 2011, 340) in which one or several field experts scrutinize the entire content of a small and often selective sample of Wikipedia articles, our analysis is more akin to Brown’s “large-n, specific-facts approach” (340).
science datasets (Party Facts: Döring & Regel 2019), we identify about 3,900 parties (third column) that possess a Wikipedia article. Of these, about 2,100 (fourth column) contain scalable information on party ideology. With the steady expansion of Wikipedia, this number is likely to grow.

While our focus in this paper is on Wikipedia data, the measurement approach we pursue is applicable more generally. For example, selected experts (e.g., political scientists, historians, etc.), rather than anonymous Wikipedia authors, could be tasked with the indexing of parties with pre-defined keywords. Other applications include situations in which researchers might wish to scale entities other than parties for which they have obtained a set of classifications. One example could be policies (entities) scaled according to the policy areas (keywords) with which they are associated. Generally, our approach should be applicable whenever political entities are being indexed with keywords that relate them to some unobserved latent dimension.

2 Party ideology classifications on Wikipedia

We focus on information from ‘party infoboxes’ (see Figure 1 for an example). Infoboxes are placed at the top of a Wikipedia article and give a quick summary of facts on a topic. According to Wikipedia guidelines, they

“contain important facts and statistics of a type which are common to related articles. For instance, all animals have a scientific classification (species, family and so on), as well as a conservation status. Adding an [infobox] to articles on animals therefore makes it easier to quickly find such information and to compare it with that of other articles.”

Unlike the textual format of a basic Wikipedia page, infoboxes restrict authors to submitting information for a defined set of categories; for parties, these are, for instance, the party’s name, its founding year, or the name of its leader. The categories of a party infobox are fixed and cannot be altered by authors. Their content should be “comparable”, “concise”, “relevant to the subject”, and “already cited elsewhere in the article”. While the categories of a party infobox are pre-

5The party infobox template is protected, see: https://en.wikipedia.org/wiki/Template:Infobox_political_party (accessed March 2019).
defined, their use is optional. Hence, some categories, or even the entire infobox, may be missing from a party’s Wikipedia article.⁴

Of interest to us are the infobox entries (henceforth tags) provided in the categories “ideology” (henceforth ideology) and “political position” (henceforth lr-position). As can be seen from Figure 1, these categories enable Wikipedia authors to index parties with political philosophies and inclinations, and to establish links between their respective Wikipedia pages. For example, the French party La République En Marche!, formed by Emmanuel Macron in the run-up to his bid for presidency, is tagged with the ideologies liberalism, social liberalism, and pro-Europeanism, with links given for all tags. In terms of lr-position, the party is tagged as centre, including a link to an associated Wikipedia page.

There are no restrictions (that we know of) on which or how many ideology tags a party can have. Our data suggests, however, that authors use tags sparingly, and that they often draw on existing tags rather than generating new ones. The large majority of parties in our data only receive between one and four ideology tags. Furthermore, some tags are assigned much more often than others (see Figures 2 and 3).

In the political position category, authors mostly draw upon a set of seven tags for classifying a party as far-left, left-wing, centre-left, centre, centre-right, right-wing, or far-right (see Online Appendix E for details). Compared to ideology tags, lr-position tags tend to get used somewhat
less often: among all parties that exhibit a tag, only 4 out of 5 have an lr-position tag, while nearly all (97 percent) have an ideology tag. Roughly 3 out of 5 parties receive one lr-position tag and roughly 1 out of 5 parties receive two such tags (see Figure 2). In the latter case, the assigned tags nearly always represent adjacent positions on the political spectrum. Only very few parties receive more than two lr-position tags.

3 A model of tag assignment

We assume that tag assignment is driven by similarity: a party gets tagged if its platform is perceived to be in agreement with the tagged ideology’s basic tenets. We further assume that agreement between party platforms and ideologies can be represented as distances on a latent dimension. Let \( i = 1, 2, \ldots, N \) be an index of parties, let \( j = 1, 2, \ldots, J \) be an index of ideology tags, and let \( y \) be an \( N \) by \( J \) matrix with binary entries \( y_{ij} \) indicating the occurrence of tag \( j \) in party \( i \)’s Wikipedia article. We represent these occurrences by an ideal point model similar to the one suggested in Lowe (2008),

\[
\Pr(y_{ij} = 1) = F(\alpha_j - \beta_j(o_j - x_i)^2),
\]

where \( F \) is the inverse logit transformation, \( x_i \) and \( o_j \) are the positions of party \( i \) and ideology \( j \) on the latent dimension, \( \beta_j \) is a tag-specific discrimination parameter, and \( \alpha_j \) is a tag-specific constant.

As can be seen, the probability of observing tag \( j \) in party \( i \) is maximized when their positions on the latent dimension coincide (i.e., when \( x_i = o_j \)). Thus \( o_j \) is the modal value of the response function and the maximum probability at \( o_j \) is \( F(\alpha_j) \). The discrimination parameter \( \beta_j \) measures how strongly the probability of observing tag \( j \) in party \( i \) depends on their closeness on the latent dimension, i.e., how rapidly the probability of observing \( j \) decreases as we move away from \( o_j \). A high value of \( \beta_j \) implies a peaked response curve in which a party’s probability of exhibiting tag \( j \) decreases quickly with increasing distance from \( j \); a low value implies a wide response curve in which the probability that a party exhibits tag \( j \) does not depend much on its position on the latent dimension. Tags with low \( \beta \) are thus less informative about party position. If \( \beta_j = 0 \), tag \( j \) occurs with probability \( F(\alpha_j) \), regardless of party position. Parameters \( \beta \) and \( \alpha \) thus allow tags to differ in their informativeness and prevalence.
The above ideal point model is closely connected to extant approaches to scaling word frequencies in political manifestos. Lowe (2008, 365f.) discusses how the Wordscores and Wordfish approaches can be subsumed under an ideal point model with the same parametric structure as Eq. (1). Our application of the model is different in that we only observe binary word presences and absences. The main substantive difference to the commonly-used Wordfish approach is that Eq. (1) allows words to have distinct locations. As discussed in Lowe (2008), Wordfish implements a reduced version of the ideal point model that allows word usage only to rise or fall along the latent dimension. By contrast, we assume that keywords (i.e., ideologies) occupy positions on the underlying dimension such that their occurrence may rise and fall, i.e., parties may be too far to the right as well as too far to the left for being tagged with a particular ideology. Since the frequency of words in party manifestos is a direct function of manifesto length, word frequency scaling approaches typically also include constant terms for the manifestos. We do not include such terms as the number of tags observed in an infobox is not obviously a function of the length of its corresponding Wikipedia article. Inspection of the distribution of tags per party (see Figure 2) also does not lead us to suspect any systematic differences in parties’ baseline probabilities of exhibiting an ideology tag. We therefore assume that, all else equal, parties have the same probability of being tagged with an ideology.

As explained in the previous section, Wikipedia offers two kinds of tags: ideology and lr-position. Unlike ideology tags, lr-position tags directly encode regions (i.e., far-left, left-wing, centre-left, etc.) on the underlying scale. We shall treat these tags as coarse, graded indicators of party position in one of seven consecutive intervals along the latent dimension and assume that a party gets tagged if its position is perceived to fall within the interval implied by the tag. To achieve this, we represent the assignment of lr-position tags by an ordered logit model with unknown $x$ (Caughey & Warshaw 2016, Treier & Jackman 2008). Numbering tags from left to right, and letting $z_i = k$ denote the presence of tag $k = 1, 2, ..., 7$ in party $i$'s Wikipedia article,

$$
\Pr(z_i = k) = F(\tau_k - \gamma x_i) - F(\tau_{k-1} - \gamma x_i),
$$

where $\gamma$ is a discrimination parameter and $\tau$ are tag-specific cut points with $\tau_0 = -\infty$, $\tau_7 = \infty$.

---

6 Similar approaches can also be found in the literatures on item response theory (i.e., unidimensional unfolding; see Andrich 1988, van Schuur & Kiers 1994) and community ecology (i.e., Gaussian ordination; see ter Braak 1985, ter Braak & Šmilauer 2015).

7 In Online Appendix A we elaborate on this connection by showing that Eq. (1) and the model studied in Lowe (2008) are special cases of the same generalized linear model.
\( \tau_{k-1} < \tau_k \) for all \( k \).

Our ordered logit model in Eq. (2) is essentially a one-item version of the graded response model from item response theory (Samejima 1969). In this model, the probabilities of tag assignment are single-peaked, except for the left- and rightmost tags. The model parametrizes these outcome probabilities via \( \Pr(z_i \leq k) = F(\tau_k - \gamma x_i) \), the cumulative probability of observing tag \( k \) or lower (i.e., the probability that party \( i \) is tagged as \( k \) or further to the left). These cumulative probabilities define a set of binary logit models, each with its own intercept \( \tau_k \) and a common slope parameter \( \gamma \). The slope parameter \( \gamma \) measures how rapidly the probability of tag assignment changes in response to party position on the latent dimension. Values of \( \gamma \) further from zero imply steeper cumulative response curves and more peaked outcome response curves.

The intercepts \( \tau_k \) define a set of adjacent intervals on the latent dimension, corresponding to the outcome values. The boundaries of these intervals can be interpreted as the points at which the cumulative probabilities are tied, and they are given by \( \tau_k / \gamma \), for \( k = 1, 2, \ldots, 6 \). For example, a party located at \( x = \tau_2 / \gamma \) has \( F(\tau_2 - \gamma x) = F(0) = 0.5 \), and thus a 50:50 chance of being tagged as left-wing or something further to the left, as opposed to being tagged as centre-left or something further to the right. By construction, the interval boundaries are defined with respect to the cumulative responses, not the observed responses. However, the boundaries are also related to the observed responses in that the probability of outcome \( k \) peaks in the middle of the interval associated with it.

As explained in the previous section, for some parties we observe more than one lr-position tag. In these instances we treat each outcome value as an independent realization of \( z \), conditional on \( x \), and model their joint probability. Formally, this means that instead of \( z_i \) in Eq. (2), we model the outcome variable \( z_{il} \), where \( l = 1, 2, \ldots, 7 \) indexes a party’s first, second, etc. observed lr-position tag.

In what follows, we pursue two approaches to recovering party positions from Wikipedia articles:

1. Estimating Eq. (1) with data on ideology tag assignment.

2. Jointly estimating Eqs. (1) and (2) with data on ideology and lr-position tag assignment.

The first approach recovers the positions of parties and tags solely from their co-occurrences without any prior assumptions about their locations on the latent dimension. This approach has applicability beyond Wikipedia. It can be used whenever political entities are indexed with a set
of keywords, following a logic of ‘pick any(-thing that applies)’ (Levine 1979). When applied to Wikipedia tags, a downside of the approach is that it discards the additional information contained in lr-position tags. Since we know the ordering of these tags, their relative locations on the latent dimension do not need to be estimated. This allows us to constrain the estimation problem and recover the locations of tags and parties with greater precision. The key assumption behind this second approach is that ideology and lr-position tags can be placed on the same latent dimension. As a simple check of this assumption, we compare the placement of ideology tags under both estimation approaches. Strong differences in results would be an indication that ideology and lr-position tags might not form a common scale.

4 Estimation

We employ Bayesian Markov Chain Monte Carlo (MCMC) methods to obtain a posterior distribution for all model parameters (Albert & Chib 1993). MCMC simulations are performed with JAGS, Version 4.3.0 (Plummer 2017). JAGS code used to estimate of both models is provided in Online Appendix C.\(^8\)

4.1 Identification and parametrization

Like all latent variable models, Eqs. (1) and (2) are not identified without some restrictions on the parameters. Three constraints are necessary to identify all parameters in a unidimensional model (Rivers 2003). To resolve invariance to addition, we center the underlying scale at \(\bar{x}\); to resolve invariance to multiplication, we standardize the scale to units of \(SD(x)\); and to resolve invariance to reflection, we impose the constraint \(\bar{x} < o_j\), where \(j\) is the index value for the tag conservatism. The identifying restrictions are imposed on each posterior draw, with offsetting transformations on the other parameters to keep outcome probabilities unchanged (see Online Appendix B for further details).

To facilitate convergence to the target distribution, we estimate a re-parametrized version of Eq. (1). As shown in (Lowe 2008, 366), Eq. (1) can be equivalently stated as

\[
Pr(y_{ij} = 1) = F(\delta_j + \lambda_j x_i - \beta_j x_i^2),
\]

\(^8\)Code and data to reproduce all results will be made available upon publication.
where \( \delta_j = \alpha_j - \beta_j o_j^2 \) and \( \lambda_j = 2\beta_j o_j \), and where the negative sign on \( \beta \) follows from the assumption of concavity (i.e., the response function must be single peaked, not single dipped). Eq. (3) has the same number of parameters as Eq. (1) but is quadratic only in \( x \), while Eq. (1) is quadratic in \( x \) and \( o \). In practice, we find that parametrizing the model as in Eq. (3) yields faster convergence to the target distribution both in terms of iterations and runtime.\(^9\) We therefore use Eq. (3) as our estimation equation. To obtain estimates of the parameters in Eq. (1), we transform posterior draws of \( \lambda \), \( \delta \), and \( \beta \) into posterior draws of \( \alpha \), \( \beta \), and \( o \), using the re-parametrization relations stated above, and subsequently apply the identifying restrictions.

### 4.2 Priors

We assign standard normal priors to \( x \) (Albert & Johnson 1999), and normal priors with mean zero and variance 5 to \( \delta \) and \( \lambda \). To enforce the concavity constraint in Eq. (3), we assign log-normal priors with log-mean zero and log-variance parameter 2 to \( \beta \), which implies a prior variance on \( \beta \) of about 47. To see what these priors mean, consider the prior variation of \( o \) relative to \( x \) (cf. Clinton & Jackman 2009, 601-2): Monte Carlo simulation shows that the prior variances on \( \lambda \) and \( \beta \) imply a prior 95\% credibility interval for \( o \) of about \([-9.6, 9.6]\). Party positions are thus a priori interior to ideologies with the prior variance of ideologies being large relative to that of party positions. Since the locations of ideologies and parties are only identified relative to each other, because the scale of the latent dimension is unknown, the wide range of ideologies relative to party positions suggests that the priors are permissive enough to allow the data to inform the estimation result. Using wider priors on ideologies leaves results unchanged (but slows down convergence). Likewise, using somewhat tighter priors also yields similar results.

For \( \gamma \) in the ordered logit model, we use a normal prior with mean zero and variance 25. For the category cut points we choose normal priors with mean zero and variance 25, subject to the order constraint \( \tau_1 < \tau_2 < \ldots < \tau_6 \).

### 4.3 Starting values

We employ correspondence analysis (CA) to generate starting values for party positions. CA is a deterministic dimension-reduction technique, similar to principal components analysis (Greenacre

\(^9\)Following Bafumi et al. (2005, 176), we use potential scale reduction factors to compare convergence under both parametrizations for a given number of iterations.
Our use of it is motivated by a result in ter Braak (1985) proving that first-dimension coordinates from a CA of the binary data matrix yield approximate maximum likelihood (ML) estimates of \( x \) and \( o \) if the data-generating process adheres to Eq. (1).\(^{10}\) This property of CA has made it particularly popular in scaling applications involving sparse data matrices (i.e., matrices with many more zeroes than ones), which are common in other fields (see, e.g., Smith & Neiman 2007, ter Braak & Šmilauer 2015) and which we also encounter in our application. ML estimators tend to be numerically unstable in these situations, while CA is guaranteed to yield a solution regardless of how large or sparse the data matrix is (ter Braak & Šmilauer 2015). Moreover, CA tends to approximate ML more closely when the data are sparse (ter Braak 1985, 863).\(^{11}\) In our MCMC estimation approach, numerical instability is not an issue; however, convergence to the target distribution can be slow with sparse data. To facilitate convergence, we follow the proposal in ter Braak (1985) and use CA estimates of \( x \) and \( o \) as starting values. Online Appendix G compares our initial CA estimates to the final Bayesian estimates.

To generate starting values for the remaining parameters, we first estimate logistic regression models—one for each ideology—of the form given in Eq. (1) using CA estimates of \( o \) and \( x \) as inputs. This gives us estimates of \( \alpha \) and \( \beta \). We then transform those estimates, using the re-parametrization relations given in section 4.1, to obtain starting values for \( \delta \) and \( \lambda \). To obtain starting values for \( \gamma \) and \( \tau \), we estimate the ordered logistic regression model in Eq. (2) using CA estimates of \( x \) as inputs.

### 4.4 Convergence

We set up four parallel chains and run each model for 22,000 iterations, discarding the first 2,000 draws as burn-in. We thin the result, keeping every 10th draw, to obtain 8,000 samples from the posterior distribution. Assessment of posterior draws via traceplots and potential scale reduction factors suggest convergence of all parameters to their target distribution (see Online Appendix D for details).

\(^{10}\)The result of ter Braak (1985) also applies if the observed data are frequency counts rather than presences or absences. Lowe (2008) introduces correspondence analysis to the literature on political text scaling by showing its similarity to the Wordscores method, as well as discussing its properties as an estimator of word and document scores.

\(^{11}\)That said, CA estimates can never be equal to ML estimates and they show some well-known biases. In one dimension, the CA solution compresses the ends of the scale, pulling estimates of \( x \) and \( o \) that lie at the boundaries toward the middle of the scale (ter Braak 1985, 864). This defect does not take away the usefulness of CA in identifying the underlying dimension, but it highlights the added value from estimating the full parametric model in Eq. (1) (on this point, see Lowe 2008, 369).
5 Selection of parties and tags

Collecting information from Wikipedia requires a list of parties that defines our target population. To maximize coverage, we draw on the largest list of parties that is currently available in political science: the Party Facts database (version 2019a; Döring & Regel 2019, Bederke et al. 2019).

Party Facts is a collaborative project that aims to solve the problem of delineating the universe of political parties. It offers an authoritative reference list of relevant parties in the world, based on the parties included in the CLEA, ParlGov, Manifesto Project, PolCon and a number of smaller datasets. Party Facts includes nearly all parties that won at least 5% seat share in a national election, as well as parties with at least 1% vote share for some countries. Of all political science datasets it currently provides the broadest coverage of political parties worldwide, with over 4,500 parties and links to the Wikipedia pages of about 3,900 of these (see Table 1).

We collect all ideology and lr-position tags for parties in Party Facts that have an infobox on Wikipedia. Data collection took place on April 27, 2019. We consider only tags that include a link to an associated Wikipedia article and code a party as tagged with the ideology or lr-position to which the link refers (see Online Appendix E for further details).

To get some insight into the variety of tags and their usage, Figures 2 and 3 give a breakdown of the raw numbers. The first thing to note is that tags are used sparingly. As Figure 2 shows, Wikipedia authors typically use one to four tags to describe party ideology; only a small fraction of parties show more than seven ideology tags. For lr-position tags, the modal frequency is one but a considerable number of parties also receive two such tags. A closer inspection of these latter cases reveals that authors almost always assign adjacent lr-position tags to parties (e.g., centre-right and right-wing).

In addition to being used sparingly, tags are also used unequally: the 20 most often-used
ideology tags, which make up less than 5 percent of all observed tags, account for over 50 percent of total tag usage. Among this core of widely applied tags, we find well-known, major ideologies such as liberalism, conservatism, and social democracy (see Figure 3).

To reduce the potential for bias and inaccuracy in our data, we focus on tags that are widely used on Wikipedia, as such tags should get more exposure and, as a result, be better known and more easily scrutinized than tags referring to rare or obscure ideologies. To achieve this, we implement a threshold of 50 tag occurrences. Tags observed in fewer parties are omitted from the analysis. This restriction allows us to rule out narrow ideologies that depend on national context (e.g., Basque nationalism), as well as peculiar ones (e.g., eurocommunism) about which only a small number of authors probably know enough to be able to apply them adequately, and spot and correct mistakes. For parties, we choose an inclusive threshold of at least two tags. In sum, this gives us scalable information for more than 2,100 parties (see Table 1).

---

12We employ the 50-tag threshold to be conservative. Our results remain substantively and quantitatively the same if we include tags that are used fewer than 50 times.
Figure 3: Frequency of usage for tags that are used at least 20 times
6 The resulting scale

We begin by inspecting the scale that we obtain for its face validity. Figures 4 and 5 summarize the estimation result via the estimated response curves and party positions for Models 1 and 2. Each response curve indicates the probability that a particular tag is assigned to a party with a given position on the underlying dimension. For example, a party at position 0 in Figure 5 has an estimated probability of about 60 percent of being tagged with liberalism, about 30 percent probability of being tagged with social liberalism, about 20 percent probability of being tagged with pro Europeanism, etc. The use of politically informative keywords thus lends a substantive interpretation to areas of the scale based on the ideologies that are particularly prominent there. In turn, this is reflected in party positions.

Figures 4 and 5 generally support the idea of a global left-right dimension underlying Wikipedia classifications of party ideology. Ideologies that are traditionally associated with the labels ‘left’ and ‘right’ (e.g., communism, socialism, social democracy, liberalism, conservatism, national conservatism) line up in the familiar order. These ideologies’ response curves are also steep in the sense that their probability of being assigned to a party rises and falls quickly along the underlying spectrum. The ideologies thus help discriminate parties on the underlying dimension. By contrast, ideologies such as agrarianism or regionalism have flat response curves and thus contribute less information on where to place parties on the underlying dimension. The fact that major ideologies line up in the familiar order suggests that party position estimates can be interpreted in terms of left vs. right. In other words, the scale has face validity.

We emphasize that we obtain substantively the same result, regardless of whether we include lr-position tags or not (compare Figures 4 and 5). Ideology tags alone are thus sufficient to obtain a meaningful left-right scale. However, the fact that the assignment of ideology tags follows a familiar left-right reasoning suggests that Wikipedia’s lr-position tags can be used to tighten our inferences about the underlying scale (besides allowing us to include even more parties in the estimation) by construing them as ordered indicators of adjacent intervals on the underlying dimension—the assumption that gives rise to Model 2.

As Figure 5 shows, the inclusion of lr-position tags leaves the ordering of ideologies largely unchanged but helps distinguish some ideologies more clearly from one another. For example, social democracy and democratic socialism are now separated more clearly, with the former falling

---

13 Point estimates for the locations of all ideologies (i.e., the locations where the probability curves peak) along with credible intervals are given in Online Appendix F.
Figure 4: Response curves and estimated party positions (indicated by tick marks) from a scaling of ideology tags only; 1,367 parties and 27 tags
Figure 5: Response curves for ideology tags, estimated intervals for lr-position tags, and estimated party positions (indicated by tick marks) from a joint scaling of ideology and lr-position tags; 2,147 parties and 35 tags
firmly into the centre-left bracket and the latter into the left-wing bracket. Likewise, Christian
democracy now clearly peaks left of conservatism and within the centre-right bracket. Lastly,
the inclusion of lr-position tags reduces the probability of tag assignment for most ideology tags,
especially for some on the far-right end (i.e., right-wing populism, national conservatism, etc.).
This is due to the higher frequency with which lr-position tags are observed compared to most
ideology tags, as well as their patterns of co-occurrence. In particular, the rightmost ideology tags
often occur in conjunction with the far-right tag as well as the right-wing tag. This leads to more
equal probabilities of far-right and right-wing parties being tagged with one of these ideologies.

7 Validation with expert surveys

Having established our estimates’ face validity, we now consider their convergent and discriminant
validity (cf. Adcock & Collier 2001). Accordingly, a valid measure should correlate highly with
other measures of the same construct, while showing lower correlations with measures of related
but different constructs.

To test this, we draw on expert ratings of party left-right position as well as party position on
other dimensions from the two largest existing expert surveys, the Democratic Accountability and
Linkages Project (Kitschelt 2014) and the Chapel Hill Expert Survey (CHES) (Polk et al. 2017).
The DALP covers 506 parties from 88 electoral democracies worldwide. Ratings were collected
between 2008 and 2009 and may not be entirely accurate with respect to parties that have recently
altered their position significantly. Nevertheless, no other expert survey provides broader country
coverage. From the CHES we use the most recent set of ratings, which were collected in 2014
and cover 268 parties from 31 European democracies.

Regarding convergent validity, Figure 6 shows that our estimates correlate well with expert rat-
ings of party left-right position. Estimates from Model 2 generally fit expert ratings more closely
than those of Model 1, suggesting that the inclusion of lr-position tags adds useful information
above and beyond ideology tags. Table 2 further shows how the correlation between our scores
and expert ratings varies over countries. Since expert surveys ask for placements of parties only
within a given polity (i.e., each expert rates parties from his or her country), pooling of ratings
across countries may introduce measurement error due to differential item functioning (Hare et al.
2014, Struthers et al. 2019).14 Comparing estimates on a country-by-country basis, the median

14For the CHES, Bakker et al. (2014) show that differential item functioning in experts’ ratings of parties’ eco-
Figure 6: Comparison of party position estimates to expert ratings from the Democratic Accountability and Linkages Project (top row) and from the 2014 Chapel Hill Expert Survey (bottom row); Model 1: N = 320 and N = 203; Model 2: N = 434 and N = 247
Table 2: Country-wise correlations with expert ratings (means and percentiles)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>10%</th>
<th>20%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>DALP</td>
<td>Model 1</td>
<td>0.72</td>
<td>0.26</td>
<td>0.69</td>
<td>0.77</td>
<td>0.91</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>0.74</td>
<td>0.42</td>
<td>0.73</td>
<td>0.76</td>
<td>0.90</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>CHES</td>
<td>Model 1</td>
<td>0.83</td>
<td>0.66</td>
<td>0.78</td>
<td>0.79</td>
<td>0.90</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>0.90</td>
<td>0.75</td>
<td>0.86</td>
<td>0.90</td>
<td>0.94</td>
<td>0.97</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Correlation is above 0.9 and the 25th percentile is above 0.75 in all four comparisons. Thus, for three-quarters of the countries covered by either the DALP or the CHES, our estimates of party positions accord well—and for half the countries covered, they accord very well—with the judgment of country experts.

Regarding discriminant validity, we compare our estimates to expert ratings of party position on economic policy, redistribution, GAL-TAN, and identity politics. As Table 3 (first column) shows, experts’ perceptions party left-right position are substantially correlated with their perceptions of party stances on all other measured dimensions. This suggests that each of the more specific dimensions reflects some aspect of left vs. right. Reassuringly, our estimates also correlate with every other dimension measured in expert surveys. However, the correlations remain weaker than those with left-right, suggesting that our estimates most likely represent general left-right position, as opposed to more specific policy stances.

A more detailed inspection of Table 3 reveals some further nuances. While experts’ judgments of party left-right position relate most strongly to their perceptions of parties’ stances on economic or redistributive policy, Wikipedia-based estimates are more strongly associated with expert ratings of parties’ positions on GAL-TAN or identity politics than on economic policy. With the inclusion of lr-position tags, correlations increase somewhat between our estimates and ratings of parties’ economic and redistributive stances (compare Model 1 and Model 2) but, overall, our scores remain slightly more associated with social and identity politics. Although small, these differences suggest that Wikipedia authors’ understanding of left vs. right is driven more by value considerations than by economic considerations, whereas experts tend to view left vs. right somewhat more in terms of economic policy and redistribution than social policy.

Economic left-right position is small.
Table 3: Correlations with expert ratings on other dimensions

<table>
<thead>
<tr>
<th></th>
<th>Expert L-R</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DALP Left-Right</td>
<td>1.00</td>
<td>0.75</td>
<td>0.79</td>
</tr>
<tr>
<td>Economy</td>
<td>0.72</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>Redistribution</td>
<td>0.67</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td>GAL-TAN</td>
<td>0.54</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>Identity</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>CHES Left-Right</td>
<td>1.00</td>
<td>0.84</td>
<td>0.90</td>
</tr>
<tr>
<td>Economy</td>
<td>0.81</td>
<td>0.53</td>
<td>0.63</td>
</tr>
<tr>
<td>Spend-Tax</td>
<td>0.77</td>
<td>0.51</td>
<td>0.61</td>
</tr>
<tr>
<td>GAL-TAN</td>
<td>0.70</td>
<td>0.77</td>
<td>0.75</td>
</tr>
</tbody>
</table>

8 Reliability

As content on Wikipedia is open to instant updating and revision by anyone, the raw data is constantly evolving. Some edits may be temporary, others may be lasting. Thus, one might ask: How reliable are estimates of party position derived from Wikipedia classifications at a single point in time? To gauge this uncertainty, we estimate party positions using ideology and lr-position tags observed at different points in time. Specifically, we compare our current scores to scores obtained in the same way using data collected some 15 weeks earlier (on January 11, 2019). Given Wikipedia’s 500 million monthly visitors, there should be ample opportunity for observing change in party classifications over this period.

Our results confirm the presence of editing behavior. A total of 111 and 279 of the parties included in Models 1 and 2, respectively, experience a change in their tags. However, we find that such edits lead to only small differences in scores. As Figure 7 shows, very few parties exhibit a substantial change in their score: out of the 1,161 and 1,958 parties included in each model comparison, only 10 and 13, respectively, show a change greater than 0.5. To put this in perspective, a difference of 0.5 roughly corresponds to a move from one left-right bracket to the next (see Figure 5). For over 99 percent of all parties in our sample, estimated scores differ by a smaller amount. This suggests that edits are mostly refinements rather than sweeping revisions of parties’ infobox classifications.\(^\text{15}\) Party scores thus exhibit test-retest reliability.

While the time frame for our comparison is limited to 15 weeks, the strong correlations be-

\(^{15}\)For example, adding the tag liberal conservatism to a party tagged with liberalism and conservatism would be a refinement that does not alter the party’s estimated position to any significant degree.
Figure 7: Reliability of party position estimates. The x-axis shows party position estimates obtained from Wikipedia classifications collected 15 weeks prior to our current estimates, which are shown on the y-axis. A 90-degree line is superimposed. Model 1: N = 1,161; Model 2: N = 1,958.

The raw information about party position. We stipulated an ideal point model that allows us to evaluate the reliability of our position estimates. The graphs show the correlation between party scores 15 weeks prior and 15 weeks later. The correlation coefficient for both models is r = 0.995.

9 Conclusion and outlook

Party positions are central to explanations of politics. In this paper we entertained the hypothesis that semi-standardized classifications of party ideology provided on Wikipedia carry valid and reliable information about party position. We stipulated an ideal point model that allows us to exploit variation in ideological keywords across parties and obtained scores of party position on a latent dimension whose substantive interpretation fits common intuitions of left vs. right. In line with our hypothesis, we found party scores to correlate well with independent expert judgments of party left-right position and we found them to be reliable in a test-retest scenario. This demonstrates that our approach yields a novel measure of party left-right position.

As a proof of concept, our results hold great promise for future research. Party ideology and the notion of ideological differences between parties are key causal factors in explanations of the
political process. They feature prominently in theories of political change and stability, legislative decision making, coalition politics, public policy, economic growth, inequality and redistribution, as well as in accounts of democratization, state building, and violent conflict. Conversely, parties’ ideological positions are the object of study in many research areas, from electoral competition to the quality of representation to political fragmentation and polarization. With its potentially unlimited coverage, our measure of party position opens up new possibilities for researchers to study these and other topics on a much larger universe of cases. This includes many parties for which we currently lack ideological placements, as well as new countries not covered by extant data sources.

Compared to expert surveys, our measure provides additional longitudinal coverage of parties, including parties that no longer exist as well as parties founded only recently. Apart from the CHES, no expert survey currently provides such longitudinal coverage. Compared to manifesto-based approaches (which do provide longitudinal coverage), our measure does not depend on the existence of an electoral manifesto. This allows for the inclusion of parties that do not produce such documents, including nascent parties, non-democratic parties, or parties in regimes with restricted electoral competition.
References


Clinton, J. D. & Jackman, S. (2009), ‘To simulate or NOMINATE?’, *Legislative Studies Quarterly* 34(4), 593–621.


**URL:** https://web.duke.edu/democracy/index.html


**URL:** http://mcmc-jags.sourceforge.net


Online Appendix

A Connections to other scaling models

Our model of ideology tag assignment is related to extant models for word counts in party manifestos through the following generalized linear model

$$\text{link}(E(y_{ij})) = a_j + b_j x_i + c_j x_i^2,$$  \hspace{1cm} (4)

where $x$ is an unobserved latent variable, $a$, $b$, $c$ are unknown parameters, and $y$ is a matrix of either

1. word ($j$) presences and absences in party infoboxes ($i$), or
2. word ($j$) counts in party manifestos ($i$).

**Case 1** yields our model of tag assignment: Since, for binary data, $E(y_{ij}) = \Pr(y_{ij} = 1)$, our model as stated in Eq. (3) is equivalent to Eq. (4) with a logit link function and the constraint $c_j < 0$.

**Case 2** yields the model in Lowe (2008) and the Wordfish approach: Applying a log link function and constraining $c_j < 0$, Eq. (4) becomes Eq. (12) in Lowe (2008, 366). As shown there, the Wordfish model results from replacing $c_j < 0$ with $c_j = 0$. 
B Identification

The identifying restrictions stated in section 4.1 imply the following offsetting transformations on the other parameters. Let the unidentified parameters be denoted with a star and the identified parameters without. Consider first Eq. (1). For identification to leave outcome probabilities unchanged, we require that

\[ \alpha_j - \beta_j (o_j - x_i)^2 = \alpha_j^* - \beta_j^* (o_j^* - x_i^*)^2. \]

Substituting the identifying restrictions

\[ x_i = \frac{(x_i^* - \bar{x}^*)}{\text{SD}(x^*)} \]
\[ o_i = \frac{(o_i^* - \bar{x}^*)}{\text{SD}(x^*)} \]

the left-hand side of the above relation becomes

\[ \alpha_j - \beta_j (o_j^* - x_i^*)^2 / \text{SD}(x^*)^2, \]

from which we deduce that \( \beta_j = \beta_j^* \text{SD}(x^*)^2 \) and \( \alpha_j = \alpha_j^* \) are the offsetting transformations required for identification.

Consider now Eq. (2). For outcome probabilities to remain unchanged, we require that

\[ \tau_k - \gamma \alpha_i = \tau_k^* - \gamma^* x_i^*, \]

which, after imposing the identifying restriction

\[ x_i = \frac{(x_i^* - \bar{x}^*)}{\text{SD}(x^*)}, \]

implies the offsetting transformations \( \gamma = \gamma^* \text{SD}(x^*) \) and \( \tau_k = \tau_k^* + \gamma^* \bar{x}^* \).
C  JAGS code

The input data for Model 1 are $y$, a binary matrix with parties $i$ in rows and ideology tags $j$ in columns, $D$, a vector containing the dimensions of $y$, and $\text{con}$, the column number of the tag conservatism.

```jags
model {
  # likelihood
  for (i in 1:D[1]) {
    for (j in 1:D[2]) {
      y[i, j] ~ dbern(pi[i, j])
    }
  }

  # priors
  for (i in 1:D[1]) {
    xstar[i] ~ dnorm(0, 1)
  }
  for (j in 1:D[2]) {
    delta[j] ~ dnorm(0, 0.2)
    lambda[j] ~ dnorm(0, 0.2)
    beta[j] ~ dlnorm(0, 0.5)
  }

  # transformation and identification of posterior draws
  xbar <- mean(xstar)
  sdx <- sd(xstar)
  ocon <- lambda[con] / (2 * beta[con])
  polarity <- ifelse(xbar < ocon, 1, -1)
  for (i in 1:D[1]) {
    x[i] <- polarity * (xstar[i] - xbar) / sdx
  }
  for (j in 1:D[2]) {
    b[j] <- beta[j] * sdx^2
    o[j] <- polarity * ((lambda[j] / (2 * beta[j])) - xbar) / sdx
  }
}
```

The primary input data for Model 2 are $y_\text{ideo}$, a binary data matrix with parties $i$ in rows and ideology tags $j$ in columns, and $y_\text{lr}$, a matrix with two columns. Column $y_\text{lr}[, 2]$ is a
stacked vector of lr-position tags (coded 1 through 7) that are observed for each party. The vector is stacked to allow for multiple observations of lr-position tags within parties. A corresponding index vector $y_{lr[ , 1]}$ identifies each party. Auxilliary inputs $D$ and $con$ are as stated in Model 1.

```r
model {

  # likelihood
  for (i in 1:D[1]) {
    for (j in 1:D[2]) {
      y_ideo[i, j] ~ dbern(pi[i, j])
    }
  }
  for (n in 1:N) {
    mu[n] <- gamma * xstar[y_lr[n, 1]]
    y_lr[n, 2] ~ dordered.logit(mu[n], tstar[1:6])
  }

  # priors
  for (i in 1:D[1]) {
    xstar[i] ~ dnorm(0, 1)
  }
  for (j in 1:D[2]) {
    delta[j] ~ dnorm(0, 0.2)
    lambda[j] ~ dnorm(0, 0.2)
    beta[j] ~ dlnorm(0, 0.5)
  }
  for (k in 1:6) {
    tau[k] ~ dnorm(0, 0.04)
  }
  tstar[1:6] <- sort(tau)
  gamma ~ dnorm(0, 0.04)

  # transformation and identification of posterior draws
  xbar <- mean(xstar)
  sdx <- sd(xstar)
  ocon <- lambda[con] / (2 * beta[con])
  polarity <- ifelse(xbar < ocon, 1, -1)
  for (i in 1:D[1]) {
    x[i] <- polarity * (xstar[i] - xbar) / sdx
  }
  for (j in 1:D[2]) {
    
```
b[j] <- beta[j] * sdx^2
o[j] <- polarity * ((lambda[j] / (2 * beta[j])) - xbar) / sdx
}
for (k in 1:6) {
t[k] <- polarity * (tau[k] - gamma * xbar)
}
g <- polarity * gamma * sdx
D Assessing parameter convergence

We ran four chains for 20,000 iterations each, after a burn-in of 2,000 iterations. Keeping every 10th posterior draw from each chain, we obtain 8,000 posterior samples for each parameter. Given the large number of parameters, we assess potential non-convergence via scale reduction factors (see below), with additional inspection of traceplots for tag parameters (not shown).

Tables D.1 and D.2 give a summary of the scale reduction factors $\hat{R}$ for all parameters in Models 1 and 2. By inspection, there are no signs of non-convergence. Scale reduction factors are generally close to one and no value exceeds 1.1. Tables D.3 and D.4 give an indication of the efficiency of the MCMC samplers for each model. Effective samples sizes suggest that 8,000 posterior samples yield sufficient information to quantify parameter uncertainty at conventional levels, in particular for parameters $x$ and $o$.

Table D.1: Scale reduction factors $\hat{R}$ for parameters of Model 1

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
<th>Npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
<td>27</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
<td>1.02</td>
<td>27</td>
</tr>
<tr>
<td>$o$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
<td>27</td>
</tr>
<tr>
<td>$x$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1,367</td>
</tr>
</tbody>
</table>

Table D.2: Scale reduction factors $\hat{R}$ for parameters of Model 2

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
<th>Npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>6</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.00</td>
<td>1.01</td>
<td>1.07</td>
<td>1.00</td>
<td>1.01</td>
<td>1.03</td>
<td>1.04</td>
<td>1.07</td>
<td>28</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.03</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>28</td>
</tr>
<tr>
<td>$o$</td>
<td>1.00</td>
<td>1.01</td>
<td>1.08</td>
<td>1.00</td>
<td>1.02</td>
<td>1.03</td>
<td>1.04</td>
<td>1.07</td>
<td>28</td>
</tr>
<tr>
<td>$x$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>2,147</td>
</tr>
</tbody>
</table>
Table D.3: Effective sample sizes for parameters of Model 1

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>Npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>333</td>
<td>3,583</td>
<td>8,000</td>
<td>339</td>
<td>359</td>
<td>443</td>
<td>978</td>
<td>2,440</td>
<td>27</td>
</tr>
<tr>
<td>$\beta$</td>
<td>196</td>
<td>2,120</td>
<td>8,000</td>
<td>206</td>
<td>243</td>
<td>267</td>
<td>501</td>
<td>1,638</td>
<td>27</td>
</tr>
<tr>
<td>$\phi$</td>
<td>795</td>
<td>4,524</td>
<td>8,000</td>
<td>822</td>
<td>923</td>
<td>1,141</td>
<td>2,164</td>
<td>3,729</td>
<td>27</td>
</tr>
<tr>
<td>$x$</td>
<td>890</td>
<td>6,366</td>
<td>8,000</td>
<td>1,546</td>
<td>2,242</td>
<td>2,903</td>
<td>4,668</td>
<td>8,000</td>
<td>1,367</td>
</tr>
</tbody>
</table>

Table D.4: Effective sample sizes for parameters of Model 2

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>Npar</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>5,321</td>
<td>5,321</td>
<td>5,321</td>
<td>5,321</td>
<td>5,321</td>
<td>5,321</td>
<td>5,321</td>
<td>5,321</td>
<td>1</td>
</tr>
<tr>
<td>$\tau$</td>
<td>2,856</td>
<td>5,350</td>
<td>8,000</td>
<td>2,875</td>
<td>2,950</td>
<td>3,044</td>
<td>3,670</td>
<td>5,007</td>
<td>6</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>288</td>
<td>5,159</td>
<td>8,000</td>
<td>293</td>
<td>532</td>
<td>963</td>
<td>3,093</td>
<td>5,298</td>
<td>28</td>
</tr>
<tr>
<td>$\beta$</td>
<td>101</td>
<td>4,319</td>
<td>8,000</td>
<td>117</td>
<td>164</td>
<td>175</td>
<td>1,181</td>
<td>3,734</td>
<td>28</td>
</tr>
<tr>
<td>$\phi$</td>
<td>414</td>
<td>5,691</td>
<td>8,000</td>
<td>447</td>
<td>608</td>
<td>1,168</td>
<td>3,757</td>
<td>7,454</td>
<td>28</td>
</tr>
<tr>
<td>$x$</td>
<td>1,016</td>
<td>6,675</td>
<td>8,000</td>
<td>1,914</td>
<td>2,779</td>
<td>3,472</td>
<td>5,355</td>
<td>8,000</td>
<td>2,147</td>
</tr>
</tbody>
</table>
E Data collection and coding of tags

As stated in the main text (section 5), we consider only tags that include a link to an associated Wikipedia article and code a party as tagged with the ideology or lr-position to which the link refers.

Technically, a tag offers two pieces of information: a link and an associated label that is presented in the infobox. Our coding of tags focuses on the targets of links rather than their labels. On their own, labels are merely words that authors can choose freely in order to describe a party’s ideology. This lack of constraint in labels can make it difficult or even impossible for other users to verify their meaning and accuracy. Links resolve this difficulty by requiring authors to commit to already existing content on Wikipedia. Their use ensures that the veracity of tags in a party’s infobox can be assessed and verified by other Wikipedia users. To reduce measurement error, we therefore use link targets instead of labels.

Links sometimes refer to Wikipedia urls that redirect to another Wikipedia article. For example, the urls for ‘liberalist’ and ‘liberal politics’ both redirect to the Wikipedia article Liberalism. In such cases, the final destination article is what we code as a party’s tag.

We use Python and R scripts to extract the desired information from Wikipedia and to create a dataset for analysis. Our data collection proceeds in five steps:

1. We download the entire infobox content (all categories) from parties’ Wikipedia articles.
2. We create a dataset that stores the downloaded information and the raw Wikitext markup code for all infoboxes.
3. We select all entries for the infobox categories ideology and position and extract their links.
4. We download the titles of the Wikipedia articles (including redirect targets) to which the links refer.
5. We create a stacked dataset with the following variables: party, category (ideology vs. position), link, tag (i.e., article title) from which we construct all input data used in the estimation of Models 1 and 2.

To distinguish ideology tags from lr-position tags, we employ the following definition: tags referring to the Wikipedia articles Far-left politics, Left-wing politics, Centre-left politics, Centrism, Centre-right politics, Right-wing politics, or Far-right politics are lr-position tags. All other
tags are ideology tags. For ease of exposition, we refer to lr-position tags as far-left, left-wing, centre-left, centre, etc. (i.e., we refer to their labels), rather than using their full article titles in Figures 3 and 5.

Our above definition is consistent with the modeling assumption that lr-position tags encode successive intervals on an underlying continuum (see section 3). It is also consistent with where we find tags in infoboxes. Table E.5 lists all the tags that we observe in the infobox category on political position. It shows their respective frequency in that category, as well as their frequency in the ideology category. We find that lr-position tags, as defined above, are used almost exclusively in the position category of an infobox (with the exception of centrism). Conversely, ideology tags, as defined above, are used almost exclusively in the ideology category, as the table lists only seven such tags.

An alternative way of defining ideology and lr-position tags would be solely with reference to their infobox category. As Table E.5 shows, such a definition would lead to inconsistencies. Tags used across both categories would have to be treated in the estimation as ideology as well as lr-position tags, depending on their category. The tag centrism in particular is used in both the ideology and position category in a total of 36 parties. Furthermore, tags in the lower half of Table E.5 do not refer to successive intervals on the underlying dimension, which makes them incompatible with our assumptions about the ordering inherent in lr-position tags (see section 3).
Treating these tags as ideology tags allows us to estimate their locations on the underlying scale freely.
Figure F.1: Point estimates and 95% credibility intervals for the locations (i.e., optima) of ideology and position tags. Panel A shows tags whose 95% credibility intervals lie within the range of party positions (−2.5 to 2.5), as indicated by the dashed lines; panel B shows tags whose estimated 95% credibility intervals lie beyond the range of party positions.
G Comparison to CA estimates

Figure G.2 compares our Bayesian estimates to estimates obtained through correspondence analysis of the data matrix. For parties from left to center both approaches yield very similar results, and CA apparently gets better at distinguishing center-right parties from more right-wing parties, once lr-position tags are included (notice the longer linear slope in the right-hand panel). Yet, CA does not satisfactorily distinguish right-wing and far-right parties, ‘squeezing’ them together relative to left-of-center parties. Figure G.3 shows that CA nevertheless yields useful estimates of party positions that successfully distinguish left from right. The overall correlation between CA estimates and expert ratings is not much worse than for our Bayesian estimates, especially after including lr-position tags. Yet, the correlation does not hold up for both sides of the political spectrum. In particular for right-of-center parties with an expert score of 6 or higher, the correlation disappears.

Figure G.2: Comparison of party position estimates to their starting values, i.e., estimates obtained via correspondence analysis
Figure G.3: Comparison of CA estimates to expert ratings of party positions: left-hand panels show results based on ideology tags only (i.e., the data used to estimate Model 1); right-hand panels show results based on ideology and Ir-position tags (i.e., the data used to estimate Model 2)
H  Software statement

Python 3.7
  - https://github.com/siznax/wptools
  - https://github.com/5j9/wikitextparser
  - https://pandas.pydata.org/

R 3.5
  - https://www.tidyverse.org/

JAGS 4.3.0
  - http://mcmc-jags.sourceforge.net/