Measuring Legislators' Ideological Position in Large Chambers using Pairwise-Comparisons

Abstract

Our understanding of politics often relies on the ideological placement of political actors – ranging from scaling of legislative roll-call voting in the United States to text-based classifications of political parties in Europe. A particularly thorny problem remains estimating individual positions in legislatures with strong partisan discipline. We provide a novel approach for estimating legislators' ideological positions: an expert survey in which respondents compare pairs of representatives on a left-right dimension. Our approach is innovative for four reasons. First, we rely on political youth leaders who are insightful and easy to recruit. Second, the rating task does not involve numeric scaling and consists of simple pairwise comparisons. Third, we efficiently and automatically detect informative comparisons to reduce the cost and length of the survey without compromising our estimates. Fourth, we use a Bayesian Davidson model with random effects in order to generate an ideological position for each legislator. As an empirical illustration, we estimate the placement of the 709 members of the 19th German Bundestag. Several validity tests show that our model captures variation within and across political parties. Our estimates offer a thorough benchmark to validate alternative measurement strategies. The presented measurement strategy is flexible and easily extendable to diverse political settings because it is able to capture comparisons among political actors across time and space.

Word Count: 6053
Keywords: ideology, legislature, measurement, pairwise comparisons, Germany
Competing interests: The authors declare none.

Policy preferences of members of parliaments (MPs) are a central concept in comparative research. They are core to the study of MP's individual behavior, are helpful in understanding the relationships between representatives and their electorate, and are also an essential factor driving public policies. Starting with the scaling of legislative roll-calls (K. T. Poole 2005), several methodological advancements have improved our understanding of the ideological placement of political actors. However, a particularly thorny problem remains in estimating individual positions in legislatures with strong partisan discipline and rare opportunities of unconstrained voting. In particular, a roll-call vote is more likely to express government-opposition dynamics than policy positions (Spirling and McLean 2007; Dewan and Spirling 2011; Hix and Noury 2016). Alternative measurement strategies, therefore, relied on data generated outside of parliament, for example, based on campaign finance (Bonica 2014) or social media (Barberá 2015).

This paper follows these advances and proposes a novel and low-cost approach for estimating an individual MP's ideological position in large legislatures. Our design consists of four features: First, we survey the leadership of parties' youth organizations because they possess detailed knowledge about elected representatives' ideological stances. Second, these experts compare pairs of legislators along a left-right dimension in a simple online survey. Third, we designed an algorithm to efficiently select informative pairs, which minimizes the cost of the survey without compromising the results. Finally, we utilize these comparisons for estimating a Davidson model that generates an ideological position and its accompanying uncertainty for each legislator. After discussing the costs and benefits of the design, we illustrate our design with an estimation of the ideological position of the 709 members across six parties of the 19th German Bundestag.

The German Bundestag is not only one of the world's largest parliaments, but it also consists of homogeneous and disciplined parties. It hence constitutes a challenging case

for measuring the individual ideology of its members. For the German case, 24 participants produced over 10000 comparisons. The resulting estimates of individual ideological positions display evident variation across and within political parties. These estimated positions follow expected partisan differences. We demonstrate our estimates' face and convergent validity, which mirror well-known differences between party wings and correspond with legislators' self-placement. In conclusion, we explore the feasibility and flexibility of our design. Our survey technique and subsequent estimation are simple to implement. They can be extended easily - by using common anchors, such as heads of governments - across jurisdictions and over time.

1 Measuring Individual Ideological Positions

Since the seminal work of Downs (1957) , political preferences are conceptualized as spatial models. The number of dimensions on which these positions are evaluated is typically small and they often correspond to a single left-right ideological dimension (K. T. Poole 2005). Spatial models are especially compelling when preferences are linked to institutional rules (Plott 1991), e.g. the electoral system, in order to explain collective outputs, the formation of a government, or the adoption of a policy. For instance, Proksch and Slapin (2012) investigate the determinants of floor access in legislatures and show that in mixed-member proportional electoral systems, party leaders prevent MPs with extreme positions from accessing the floor. In a different setting, Hix (2004) shows that the voting behavior of members of the European Parliament is determined by the distance between the positions of the member, their European parliamentary group, and their national party. Such models populate the field of legislative politics, where MPs' individual positions play a crucial role in the analysis of legislative behaviors and outcomes. Testing these mod-

els requires an accurate measure of individual MPs' ideological positions. Substantive research that relies on MPs' individual positions goes beyond legislative and coalition politics. These works extend to constituency preferences and their representation (for a review, see Canes-Wrone (2015)). Ultimately, they all inquire about how democratic politics works.

Separating preferences from behaviors sheds light on electoral and legislative politics. On the aggregate level, the intra-partisan distribution of core preferences affects a party's ability to adopt policies, negotiate coalition agreements or represent its electorates. On the individual level, MPs' level of sincerity reflects their capacity to represent their constituent and is likely to vary across contexts and considered behaviors. In parliamentary systems characterized by strong partisan discipline, speeches and social media posts are, for instance, more likely to be more sincere than roll-call votes. A central challenge of this research constitutes the distinction between MPs' individual preferences and their "revealed preferences" as legislative behavior (for the larger debate, see Knox, Lucas, and Cho (2022)). Simply, an MP's roll-call vote, parliamentary speech or public communication entails preferences and (cor-)responds to contextual and other strategic factors (Hix and Noury 2016). Behavior can vary despite stable core preferences, yet the two concepts must be kept apart theoretically and empirically. Our goal, therefore, is to determine MPs' preferences independent of their behavior.

In short, three reasons for measuring individual preferences of legislators exist: (1) elected representatives are the basic unit of political action, (2) much of politics organizes around a single left-right ideological dimension and (3) measures of MPs' preferences reveal important aspects of individual behaviors and collective choice.

Measuring legislators' ideological positions is a formidable task. Laver (2014) and Carmines and D'Amico (2015) summarize state of the art. Laver (2014) identifies several

challenges when measuring ideology. First, it's necessary to choose between discovering the substantive meaning of the ideological dimension inductively or developing it deductively. Second, ideology is a latent concept whose empirical reality varies across space and time. For example, holding a left-leaning position amounts to different policy preferences in Germany and France. A deductive approach makes the measure more adaptive to the context but less comparable across contexts. On the other hand, proceeding inductively helps to measure a comparable phenomenon across contexts but raises the risk of locally applying a wrong definition of ideology. Finally, the numeric scale onto actors' preferences is projected needs to stay stable across actors.

Research pertaining to the measure of individual positions can be classified into two groups: survey and behavioral approaches. We introduce these and briefly discuss their respective advantages and drawbacks. Concentrating on survey approaches, it is worth mentioning that the usual large number of representatives renders classic expert surveys inconceivable. Assessing the left-right position of an actor requires in-depth knowledge of this actor. A small handful of experts enjoy this sort of knowledge on all legislators. MPs themselves can offer such depth and breadth; thus, they can be asked to place themselves on an ideological scale. Directly surveying legislators has been carried out in the American states (Maestas, Neeley, and Richardson Jr 2003, e.g.) and across different democracies. The Comparative Candidate Survey (CCS) is an international effort that asks many legislative candidates to place themselves on an eleven-point left-right scale. This basic form of self-placement follows typical voter surveys and assumes that candidates know their own position and conceive the left to right space in the same way.

Individual self-placement possesses three drawbacks. First, its scope is limited. Despite the impressive and substantial size of their sample, their compliance rate is, as already noticed in previous elite surveys (Bailer 2014), low. In Germany in 2017, 803 of the 4828

(about 17%) candidates took the survey. 186 were eventually elected, which represents only a quarter of all German legislators. Second, self-placement questions expect the respondent to perceive the numeric scale and its association with the underlying dimension in the same way (Lesschaeve 2017). If two candidates respond with the same position, it is tempting to conclude they hold the same view. Yet, they just might possess a different perception of how their position translates on the numeric scale. Finally, it is impossible to prevent strategic answers. This misrepresentation might partly be occurring because the population is small and anonymity hard to uphold. Fearing potential backlash, respondents might consequently take over the position of their constituency or leadership.

A recent study by Hopkins and Noel (2022) offers a valuable advancement of survey-based measures. In order to construct the ideological positions of US Senators, the authors ask politically engaged citizens to compare pairs of legislators on an ideological scale. Then, they leverage these comparisons and estimate the underlying ideological positions. Their design overcomes most issues related to scaling actors' positions. It still possesses two potential disadvantages that we remedy here. First, it assumes that politically active citizens know all considered political actors well enough. As chamber size increases, familiarity with legislators is likely to decline and obtaining a complete ideological picture of a chamber becomes complicated. To solve this issue, our approach relies on sets of respondents, whose daily work brings in proximity to legislators, such as members of the parties' executives, political journalists or parliamentary staff. Second, as the number of actors increases, the comparison space becomes huge. Random exploration of that comparison space, as in Hopkins and Noel (2022), also becomes very inefficient[1]. One does not

[1]To be more precise, the authors are aware of this and exploit the bipartisan characteristic of American politics to reduce the comparison space and create two pools, that are explored randomly. This split reduces the applicability of their methods to countries where parties' ideological positions have few overlaps and where good ideological proxies such as Nominate scores exist.

learn much from a pairing of a very conservative with a very progressive member of the same party. Again, their design works well for small and active legislatures, such as the US Senate and the 10,000 possible comparisons, but is of limited use in larger chambers.

The second group of strategies consists of behavioral measures. These measures posit that ideological positions can be derived from observable behavioral patterns. The more similar the behaviors of actors, the closer their ideological positions are. In the context of legislators, two types of behaviors have been scrutinized extensively: roll-call vote (K. T. Poole and Rosenthal 1985; Carroll and K. Poole 2014) and speeches (Proksch and Slapin 2010; Lauderdale and Herzog 2016; Rheault and Cochrane 2020). At first sight, it seems reasonable to expect MPs sharing an ideological view to vote together and to deliver similar speeches. But, both measurement approaches have been extensively scrutinized in the last 20 years, and estimating positions with behaviors has proven to be more complicated than expected.

Roll-calls are often discretionary (Ainsley et al. 2020) and suffer from three limitations. First, roll-call specifically are triggered when MP have incentives not to vote sincerely (Carrubba, Gabel, and Hug 2008). Second, roll-call analyses cluster together *fringe* representatives without necessarily distinguishing between different ideological orientations (Spirling and McLean 2007). Finally, roll-calls have been developed in the American context, where partisan constraints on voting are lower than in most other legislatures (Hix and Noury 2016). In other cases, especially in western European democracies, partisan discipline is high and unconstrained voting is rare (Spirling and McLean 2006; Dewan and Spirling 2011; Hix and Noury 2016). After carefully designing roll-call models to estimate the ideal points of German MPs, Bräuninger, Müller, and Stecker (2016) conclude that non-spatial factors irregularly but extensively influence roll-call votes. "Off-the-shelf estimates may be biased in various ways, and we should instead turn to more complex

behavioral models to arrive at valid point estimates" (p. 191).

Measures based on speeches, such as wordscore (Laver, Benoit, and Garry 2003), wordfish (Proksch and Slapin 2008) and wordshoal (Lauderdale and Herzog 2016), address some of these drawbacks. Floor speeches are less affected by partisan discipline and more likely to reflect individual preferences. Similarly, speeches contain more information than discrete roll-calls. The speech of a *fringe left-leaning* MP is unlikely to be confused with the speech of a *fringe right-leaning* MP, even if they both oppose the same law. Yet, extracting position from speeches is not as straightforward as it seems. Text-scaling methods aim to project high-dimensional data - word frequencies - onto a few dimensions. A transcribed speech entails precise information about ideology, but it also contains much non-ideological content. In this context, systematically linking word patterns with ideological positions is challenging. Even once speeches are located along one dimension, it is necessary to validate that the obtained dimension corresponds to the desired latent left-right ideology. This validation is complicated without an actual gold standard accurately measuring left-right positions. Lauderdale and Herzog (2016) shows, for instance, that the first dimension structuring debates in the Irish Dáil amounts to the divide between the government and the opposition and, hence, does not match the left-right dimension as previous studies suggested.

With these two approaches of measurement and their accompanying trade-offs in mind, the following section presents the design of an expert survey that overcomes the limitations of existing surveys. In a nutshell, we ask national experts to repeatedly compare pairs of MPs. In doing so, we overcome issues related to the subjective and potentially varying interpretations of a numeric scale. We are able to provide point estimates and uncertainties for all legislators within a parliament, independent of behavior.

2 Measurement Strategy

Our measurement strategy consists of a simple expert survey using pairwise comparisons and a Davidson model to fit those responses. We propose to run the expert survey with politically active party members that work for or are closely associated with legislators. In the illustration below, we rely on leaders of the youth wings of political parties in Germany[2]. Instead of the classical scale-placement question, we take advantage of simple pairwise comparisons (Carlson and Montgomery 2017). In our case, respondents compared 500 pairs of MPs according to an ideological criterion. This number seems low compared to the 500,000 possible pairs - the Bundestag has over 700 members-. Our approach uses one respondent's previous answers to identify the most informative pairs of legislators, hence compressing as much information as possible in these 500 pairs. This approach allows an efficient exploration of the comparison space.

2.1 National experts

Measuring the position of individual legislators with an expert survey requires participants who are able to distinguish between as many legislators as possible.[3] Importantly, only a limited number of potential participants know backbenchers relatively well. We believe that leaders of the German youth parties are a good source of expertise. First, these organizations are highly institutionalized and work hand in hand with their mother organizations. They are part of the daily routine of the party: they hold executive positions at the local level, they commonly work as parliamentary assistants, they participate

[2]If young leaders worked well in Germany, as they were easy to recruit and remarkably close to legislators, it might not be the case in all countries. Alternatively, parliamentary journalists or parliamentary staffers would constitute excellent candidates for the survey.

[3]In a pre-test with political scientists, we noticed that they knew many legislators from different parties, but they all knew the same set of prominent politicians.

in grass-root activities such as campaigning or rallying, they also have a voice in the national executive of the parties and even get some of their members elected as representatives. Their daily contact with political parties makes them ideal subjects for estimating the positions of both prominent and inconspicuous legislators. Beyond youth leaders, we considered using parliamentary staffers, parliamentary journalists, or MP themselves. We focused on youth leaders for practical reasons, too: youth leaders are very accessible, and their participation was easier to incentivize. As shown in Appendix A, compliance was indeed very high.

A crucial aspect of this inductive measurement strategy regards the absence of a clear definition of left and right. Respondents are presented with two MPs and must identify which MP holds the most left-leaning position. We did not provide any further explanation on how *left-leaning* should be understood. All respondents perceived the task as straightforward and did not ask for more details. In doing so, we rely on their subjective interpretation of left and right, which is relatively homogeneous within a given country at a given point of time (Huber 1989). This is particularly true among politically active respondents, who are unlikely to misconceive left and right when comparing two legislators. Respondents were actually very consistent in their answers (inter-coder reliability of .91), supporting the hypothesis of a shared left-right definition.

The potential drawback of subjectivity is low compared to its advantages. The absence of a fixed definition of left and right-leaning improves the flexibility of the resulting measure, which can be applied across contexts. If, in a given context, left-leaning positions are about defending state intervention in the economy, respondents will be aware of it and compare MPs accordingly. If such positions are instead related to a decentralization debate, respondents will instead compare MPs with regard to decentralization. Furthermore, it relaxes all behavioral assumptions. There is no need to link an ideological position with

a specific behavior (voting with or against the party; holding a specific speech). When comparing MPs, respondents rely on their personal knowledge of the legislator and his work, voting record, network, agenda, etc. The resulting estimates are consequently not tied to one type of behavior. Instead, they reflect how respondents perceive the global behavior of an MP.

Assuming the respondents know the meanings of left and right, two types of biases still threaten our measure's validity: in-group bias and collective non-ideological heuristics. First, respondents may assess differently MPs from their own party than MPs from other parties. For instance, they might project their own position on MPs of the same party to prevent cognitive dissonance. For example, the German Young Greens are known to be much more left-leaning than their older counterparts. In-group bias might encourage Young Greens to systematically label Green legislators as more left-leaning, resulting in misstated position. To limit the potential effect of in-group bias, we implemented three safeguards. First, we recruited members from each major German party, so that our sample of respondents is representative of the German political landscape. Second, each participant had to classify members from all parties and not only from their own. Third, our models took into account respondent heterogeneity and modeled it explicitly.

The second type of bias happens when respondents mobilize external cues instead of their personal knowledge to estimate the ideology of an MP. There is a trade-off between providing respondents with enough information on the MPs for identification and cueing their answers by providing too many or particular pieces of information. We settled on offering two pieces of information: a name and an official portrait taken from the parliamentary website. We removed the party of each MP and explicitly asked the respondent not to look for more information, such as the Wikipedia page. In addition, respondents were encouraged to declare an MP as unknown when they were not clearly remembering

a particular MP. Providing a picture is debatable because visual cues can be very influential: gender, race, facial expression, background, etc (). In a pre-test, we only showed names and respondents complained about the difficulty of identifying an MP on the mere basis of the name.[4] By providing the respondents with the name and picture of an MP and offering them to declare an MP as unknown, we tried, as much as possible, to minimize the use of external cues.

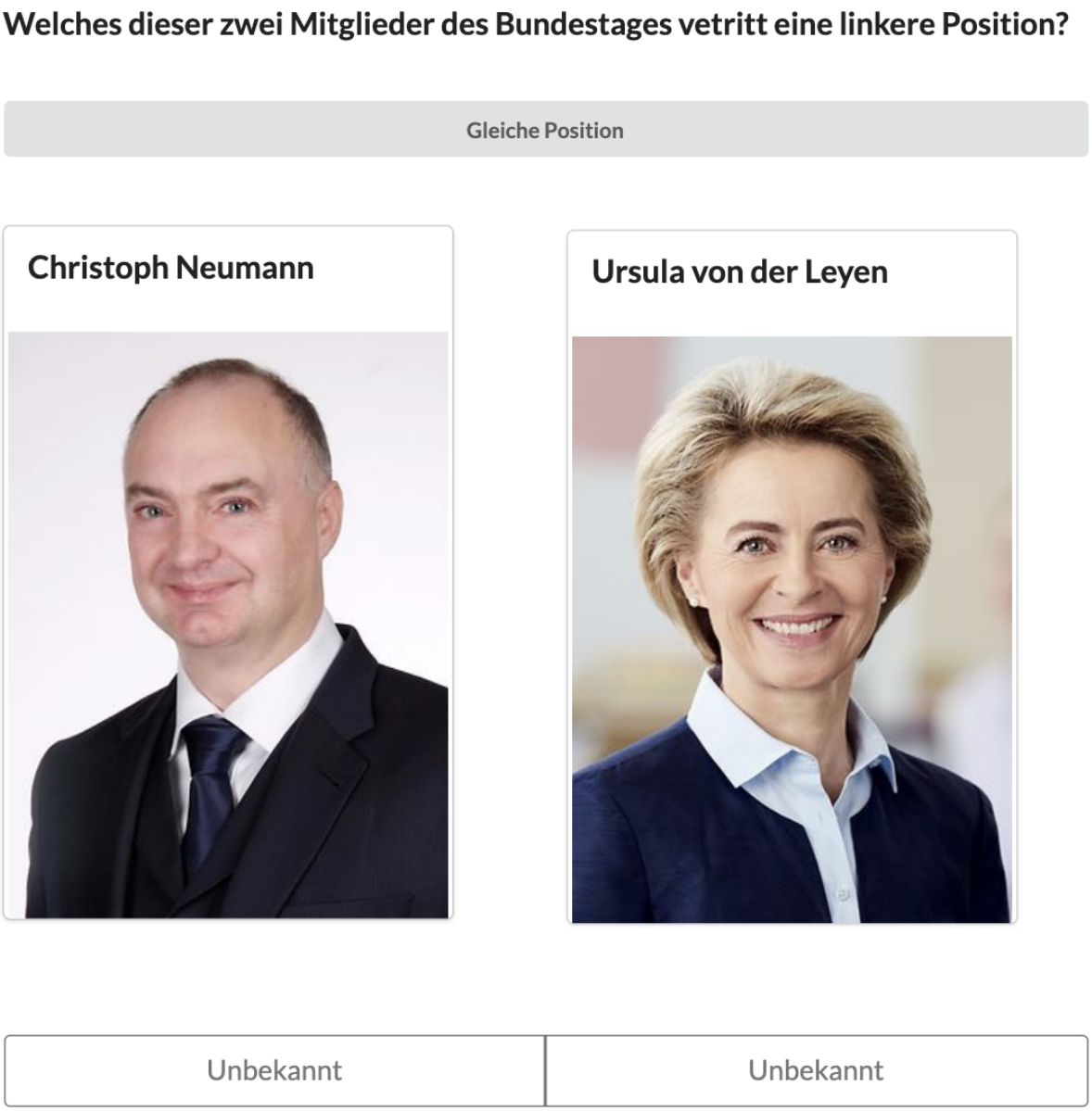
2.2 Pairwise comparisons

Pairwise comparisons constitute a simple and valuable tool to measure latent trait (Benoit, Munger, and Spirling 2019). Classical scaling approaches ask respondents to place MPs on an absolute scale. But, "completing such tasks requires workers to continuously maintain in their memory how previous [MPs] were coded and remember detailed rules dictating how stimuli are placed into categories." (Carlson and Montgomery 2017) Respondents have to come up with a rule system differentiating numerical values for each step on a scale, e.g. a "4" from a "5". These rule systems are likely to vary across time and respondents, as "individuals understand the 'same' question in vastly different ways" (Brady 1985). Instead, pairwise comparisons compress an MPs' ideology in relative terms. It does not matter whether an MP is moderate or radical in the absolute; only their relative position to each other matters. The task is consequently more reliable across coders and easier to perform because of the binary nature of the decisions.

In our illustration, respondents were asked to compare 500 pairs with the following

[4]During pre-test, respondents acknowledge they often had to search for the name on Google to make sure they associated the right person with the name. This is not surprising, given that there are a dozen "Müllers" in the Bundestag. Accordingly, we decided to include MP's picture. To reduce the heterogeneity of the pictures, we used standardized official parliamentary portraits. They all have a similar size, a similar arrangement and a similar neutral background.

Figure 1: Screenshot of the survey application. The question is: "Which of these two Members of the Bundestag holds the more leftist position?" Respondents also have the options that two MPs hold the same position (*Gleiche Position*) or declare them as unknown (*Unbekannt*).



task description: "Which of these two MPs holds a more leftist position?". They could choose between three answers: "A is more leftist than B", "B is more leftist than A" or

"A and B defend a similar position" [5]. On average, respondents needed 103 minutes (13 seconds/comparison) to complete the survey and were rewarded with 75€.

The 709 German MPs generate over 500,000 possible comparisons. Drawing randomly from this set would be inefficient, as it would include many uninformative comparisons. There is no need to compare far-right members with far-left members. A random draw would accordingly increase the number of comparisons required to estimate accurate and precise positions. To limit uninformative comparisons, we assume respondents' ranking to be transitive. After each comparison, we automatically use the new information to detect informative pairs. If a respondent declared MP_a to be more leftist than MP_b and MP_b to MP_c, it would be redundant to compare MP_a and MP_c as MP_a is much more likely to be more leftist than MP_c. Instead, the algorithm would focus on introducing another MP MP_d. Again, if MP_d was rated more leftist than MP_b, it would be uninformative to compare it with MP_c. Instead, the next pair would either compare MP_d and MP_a (as both are more leftist than MP_b) or introduce a fifth MP. This focus on the most informative pairs enables to explore the comparison space efficiently. It reduces the number of comparisons required to obtain stable estimates and the direct costs of the survey.

The transitivity assumption strictly relates to the exploration of the comparison space and does not affect the further estimation of the ideological positions. Let us hypothetically consider a mistaking respondent who rated Alice Weidel (AfD/Far-Right) as more leftist than Annalena Baerbock (Greens, Left). Transitivity implies future Weidel - comparisons to necessarily feature MPs, who were (1) previously rated as more leftist than Baerbock or (2) not rated yet. Despite being inefficient - only limited information can be extracted from comparing Weidel to leftist MPs -, these future comparisons will still be

[5] In case they did not know one of the MPs, respondents could click on "Unknown". Once an MP was declared as unknown, he would not be proposed anymore to the respondent for the rest of the survey

valid. There are reasons to believe that such mistakes constitute a trivial threat. First, final ideological scores rely on all respondents' comparisons. Idiosyncratic mistakes committed by one respondent are corrected by other respondents. Second, both the probability of such mistakes and their negative influence on efficiency decrease as the ideological distance between the two MPs increases. If such mistakes happen, they are likely to be with limited consequences. Third, the implications of those mistakes diminish as respondents move through the survey: in the early stage of the survey, we implemented a mechanism to help respondents familiarize themselves with the task (see next paragraph). Finally, the survey is initiated with prominent[6] MPs to help the respondent get familiar with the task. Indeed, we assume prominent MPs would be easier to compare. As respondents advanced in the survey, the assumption was relaxed and comparisons featuring less important MPs were also collected.

In order to safeguard our procedure, we use two tests for estimating the potential impact of mistaken ratings. First, we estimate the agreement between coders who rated the same pairs. The high inter-coder reliability (.94) suggests that such mistakes were, at most, marginally committed. Second, we used a Jack-Knife sampling scheme and re-estimated the model after removing all ratings from each respondent. The results remain extremely stable [7]. Unless the same mistake was committed by several respondents, we are confident to rule out the hypothesis that mistakes have been amplified by the transitivity assumption.

[6]Following Munzert (2018), we use the yearly traffic on the Wikipedia page of a given MP as a proxy measure for prominence.

[7]Across the 24 models fitted after removing each of the 24 respondents, the individual scores have an average standard deviation of .07 for a scale going from -3 to 3

2.3 Estimating latent positions from pairwise comparisons

We use a Davidson model to model the comparison and estimate MPs' ideological positions. Our model accounts for the nested structure of the data, with respondent-specific random effects and standard errors clustered at the level of the respondent. We describe three aspects of our model - incorporating ties in ratings, accounting for multiple comparisons by each rater, and estimating the model in a Bayesian framework.

Statistical models for pairwise comparison have been widespread, especially in psychology, since the 1920s. A well-known variant with applications in political science is the Bradley-Terry model (Bradley and Terry 1952; Agresti 2013; Loewen, Rubenson, and Spirling 2012). The goal of the Bradley-Terry model is to provide an ordering of objectives based on simple pairwise comparisons. In our empirical example, we compare two politicians to identify all ideological positions within a legislative chamber.

Typical models for pairwise comparisons and their estimation are well-established (e.g. Cattelan (2012) for a review) and can, for our purpose, be summarized as follows. Y_{sij} is a random variable containing the ratings of legislators pairs $(i, j)_s$ comparing legislators i and j made by the raters $s = 1, ..., S$. In the model, we denote $\boldsymbol{\lambda} = \lambda_1, ..., \lambda_n$ as the vector of individual ideological positions for a set of n MPs. Following conventions, $\lambda_i > \lambda_j$ is equivalent to λ_i is "more right" when compared to λ_j. Consequently, higher and positive scores mean right-leaning positions, while lower and negative scores mean left-leaning positions. For each pair of MPs (i, j), there is a probability $\pi_{i,j}$ that respondents rate i as more right than j. This probability is linked to the ideological scores λ_i and λ_j with a logistic function:

$$\pi_{ij} = \frac{e^{\lambda_j}}{e^{\lambda_i} + e^{\lambda_j}}$$

Our first extension considers that legislators might be rated to hold similar ideological positions. Classical Bradley-Terry only allows for strict comparisons and forces observers to discard pairs of objects judged to be similar. We explicitly allow respondents to judge two legislators as similar and wish to incorporate this information in the model (about 17% of the pairs of MPs were rated as similar)[8]. According to Davidson (1970), we can incorporate these ties by adding a parameter $\nu \in \mathbb{R}$. Adding this information to the model, we obtain the following parametrization for probabilities (1) $\pi_{ij|i\neq j}$ that respondents rate i as more right than j given that i and j are not rated as holding similar positions and (2) $\pi_{i=j}$ that respondents rate i and j as having similar positions.

$$\begin{cases} \pi_{ij|i\neq j} = \dfrac{e^{\lambda_j}}{e^{\lambda_i} + e^{\lambda_j} + e^{\nu + \frac{\lambda_i + \lambda_j}{2}}} \\ \pi_{i=j} = \dfrac{e^{\nu + \frac{\lambda_i + \lambda_j}{2}}}{e^{\lambda_i} + e^{\lambda_j} + e^{\nu + \frac{\lambda_i + \lambda_j}{2}}} \end{cases}$$

Here, ν can be interpreted as the degree to which the probability of $i = j$ is affected by the relative difference in ideological scores of i and j. Notably, when $\nu \longrightarrow -\inf$, i and j never have the same position, but when $\nu \longrightarrow +\inf$ i and j are systematically rated as holding similar positions.

The second extension to a simple Bradley-Terry model acknowledges that our observations, i.e. ratings by each youth group leader, are not independent of each other. Each rater s makes multiple comparisons. In order to account for multiple judgments, we use an extension of the Davidson model, proposed by Böckenholt (2001), and decompose the prediction into a fixed and a random component. The fixed effect component estimates each legislator's average (log) position, while the random component accounts for respondent-

[8]For robustness purposes, we estimated a Bradley-Terry model in parallel to the Davidson model and found very similar, but less precise results, (For more details, see Appendix C)

specific effects. Given a set of S subjects, then $\lambda_{is} = \lambda_i + U_{is}$, where $U_{i,s}$ refers to the random effect on the ideological score of MP i, when rated by $s \in \{S\}$. This extension can be incorporated in the parametrization above.

The third consideration pertains to the Bayesian estimation of the outlined Davidson model. Instead of detailing identification and estimation (Cattelan 2012), we concentrate on two aspects. First, the full identification of the model requires the constraint $\sum_i^n \lambda_i = 1$. Second, different estimation methods have been proposed for approximating the resulting likelihood. We have many raters who provided a lot of comparisons for many legislators, so we face a computationally intensive exercise. A Bayesian approach, therefore, is a sensible strategy. For our implementation, we used weakly informative priors for λ_i, ν and $U_{i,s}$ (normally distributed, centered around 0 and with variance 3.0). Our model is estimated in R using bpcs (Mattos and Ramos 2021), which uses stan and its No-U-Turn (NUTS) Hamiltonian Monte Carlo sampler (Hoffman, Gelman, et al. 2014) for estimating the parameters λ, υ and U. We present these estimates and the accompanying credible intervals visually.

3 Individual Positions in the 19th Bundestag in Germany

We illustrate this new research design using data from the 19th Bundestag in Germany. With 709 members across 6 different parties, the German parliament is a challenging environment for measuring MPs' individual positions because of its size and the large number of backbenchers. The ideological space populated by German parties is reasonably narrow, especially among governing parties, and essentially structured along a single left-right dimension. A mixed-member electoral system and a strong second chamber set

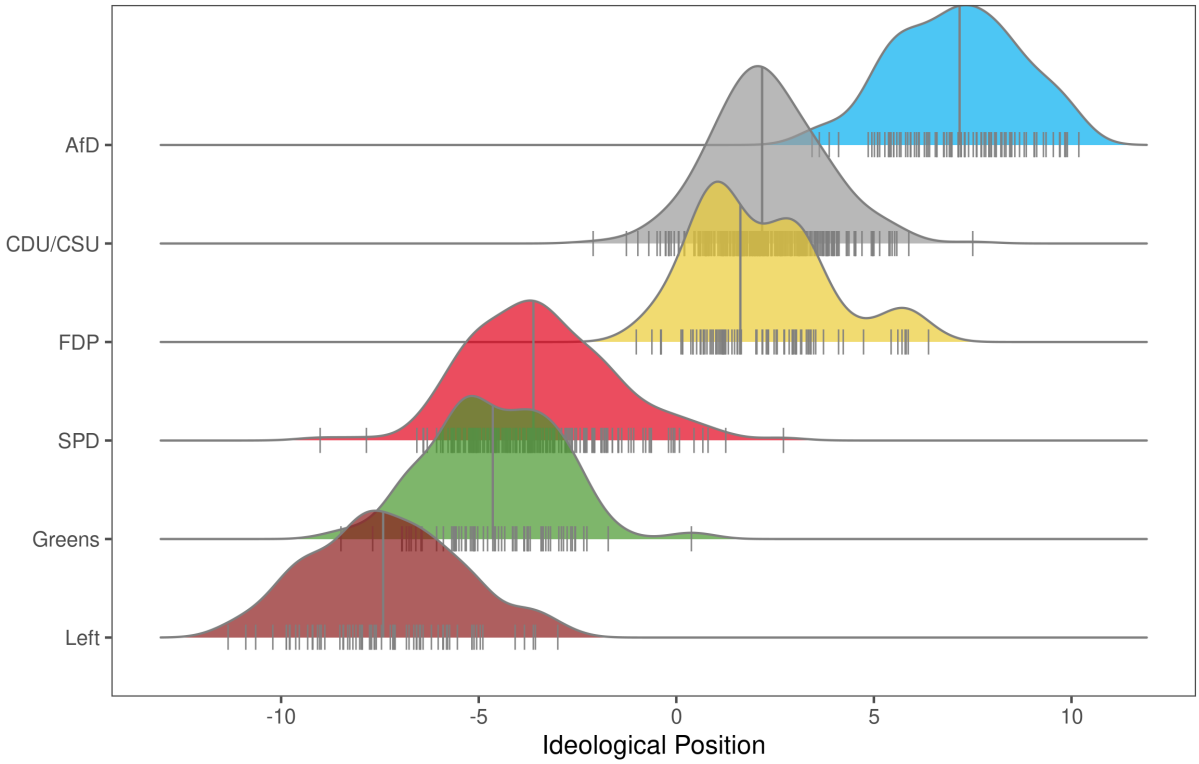
incentives for German political actors to cultivate a personal vote and pursue consensual positions.

For the survey, we recruited members of executive committees among six German youth party organizations: Junge Alternative für Deutschland (AFD), Junge Union (CDU); Junge Liberale (FDP); Jusos (SPD); Grüne Jugend (Bündnis 90/Die Grünen); and Linksjugend Solid (Die Linke). We emailed each member of these executives, asking them to contact us if they were interested in taking the survey, and selected participants on a first-come-first-serve principle. Participants who completed 500 comparisons were rewarded with 75€. Considering our budget constraint, we could afford up to five participants for each organization. The survey was taken by 24 participants between March, 30th 2020 and June, 15th 2020. All participants were asked to complete 500 comparisons.

The survey produced about 11,900 comparisons for the 709 members of the Bundestag. The built-in sorting of choice sets enabled us to avoid asking respondents to classify 178,053 uninformative pairs. The most prominent MP, Angela Merkel, was compared 2,500 times. Representatives were, on average, compared 31 times. Using those data, we fit a Davidson model and present the estimation results visually.

As a first step, Figure 2 present the ideological distribution of the MPs' point estimates for each party. When aggregated at the partisan level, the ideological positions of the six parties correspond to their well-known positions. Within each party, centrists are more common than extremists, as the close-to-Normal distribution of MPs in each party attests. Going from the left to the right, we observe the Left (Die Linke), the Greens (Bündnis 90/Die Grünen), the SPD, the FDP, the CDU/CSU and finally the AfD. As one would expect, the AfD is more distant from the CDU/CSU than the FDP. MPs from the CDU/CSU and FDP are, in aggregate, ideologically very similar to each other. On the left, the Greens have an average ideological score very similar to the SPD, but have a less

Figure 2: Ideological distribution of MPs in the 19th Bundestag by political party



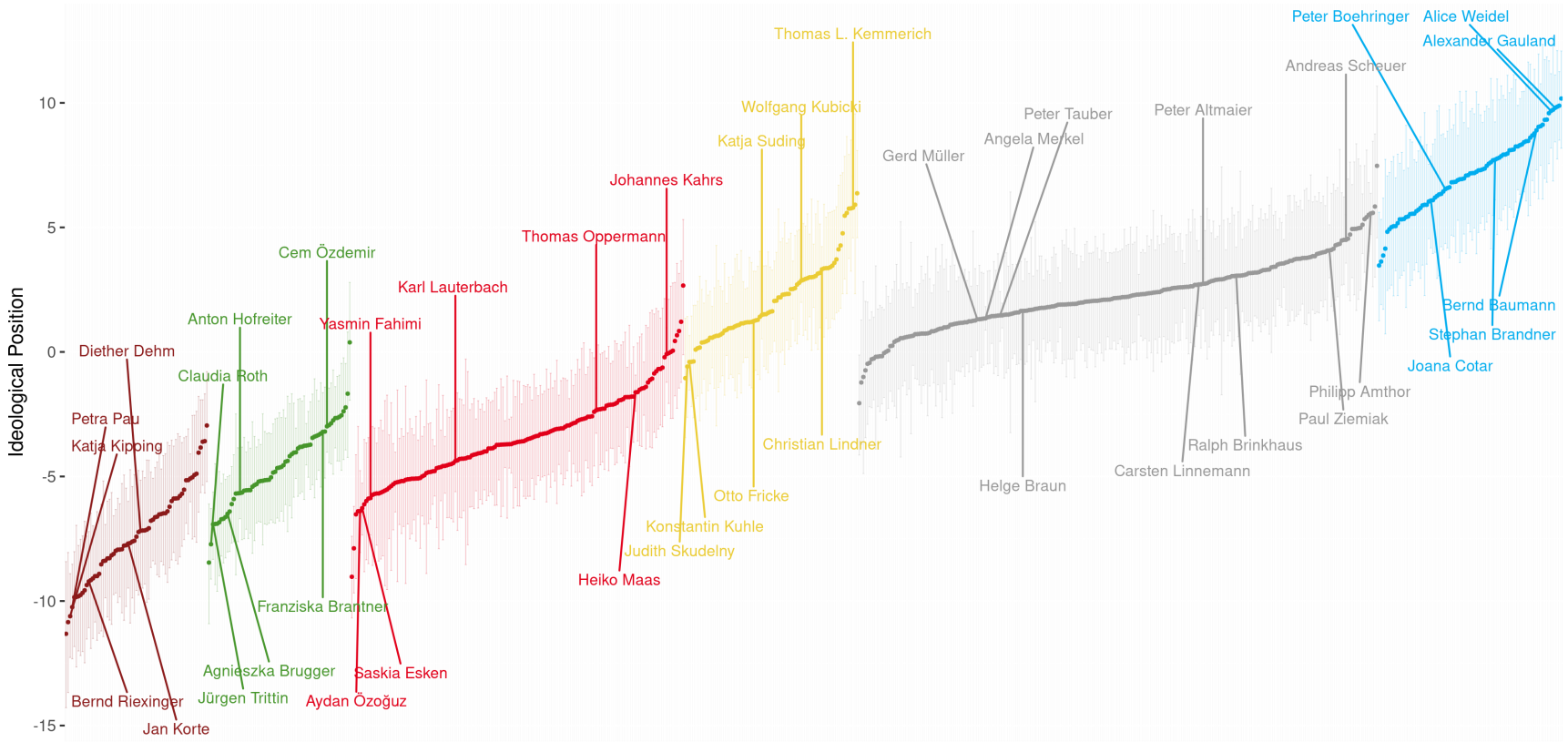
Notes: Lower score indicates a more leftist position. Vertical lines show the median position of the party. Vertical bars correspond to individual MPs

pronounced right tail. Most members of the Left are more leftist than the median member of SPD and Greens.

Taking a closer look at the ideological estimate for each legislator, Fig 3 presents these point estimates (symbolized by a dot) and their uncertainties (as a shade) for each representative. We group these estimates by party and move from ideologically left to right. Based on the rankings by our national experts, the ideologically most left and most right members of the 19th Bundestag are Tobias Pflüger (Die Linke) and Frank Magnitz (AfD). For each political party, the graphic labels some of the most prominent members of each party. For the Left, Jan Korte appears to be close to the median member of his party. Some prominent legislators circumscribe the ideological range of the Green party: Clau-

dia Roth on the left and Cem Özedemir on the right. Karl Lauterbach, a professor of health economics and an illustrious voice during the Corona crisis, is slightly to the left of the median within the Social Democratic Party. The FDP is estimated to have a relatively wide ideological range and its party leader, Christian Lindner, is placed slightly right to the party's median MP. Angela Merkel, the chancellor at the time, is among the most centrist legislators in the 19th Bundestag. This position places her among the more leftist members of her party, the CDU. Philipp Amthor, who expressed a strong disagreement with his own party's progressive immigration policy, is an example on the right of his party. According to our estimates, Alexander Gauland is one of the most right parliamentarians of the AfD. All in all, this illustration provides a realistic and detailed picture of the ideological composition of the German Bundestag.

Figure 3: Individual positions for the members of the 19th Bundestag



Notes: Each dot represents the point estimate for a legislator. The shade is the 95% uncertainty band. Lower ideological scores mean a more leftist position. Prominent members of each party are labeled by name.

3.1 Establishing the estimates' validity

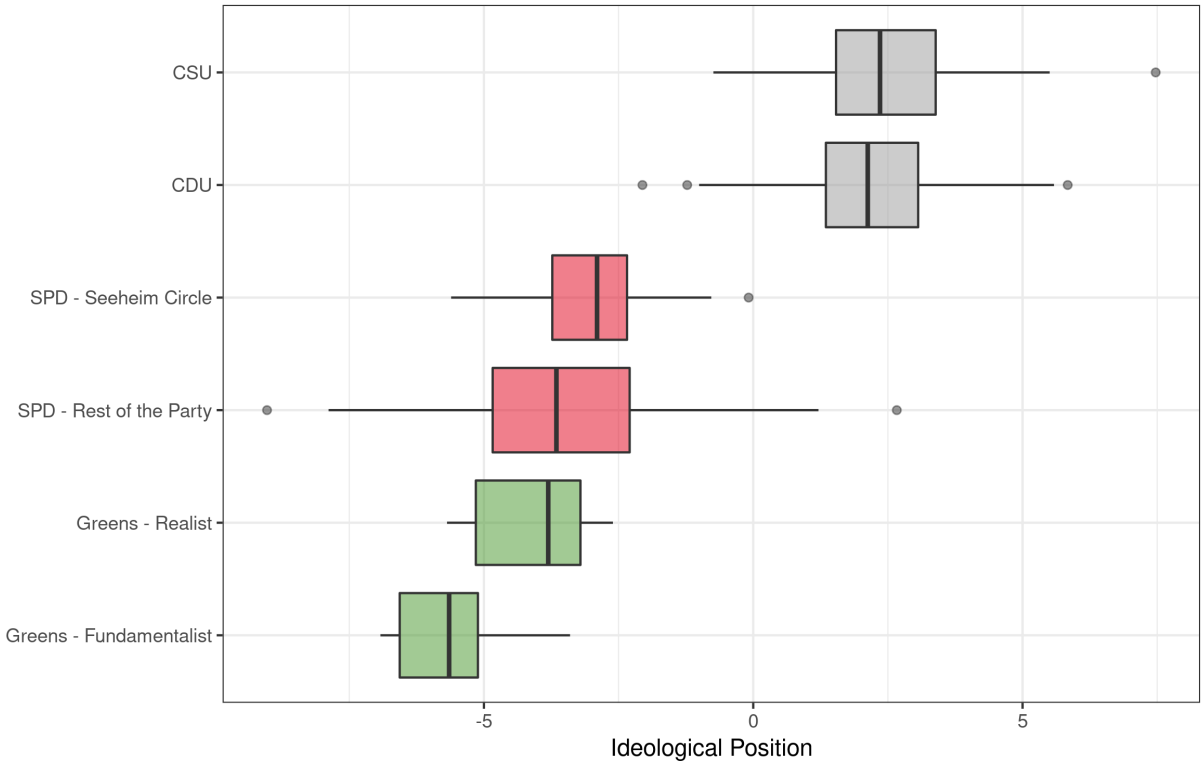
The illustration above provides basic face validity to our estimates. Political parties and some well-known legislators are placed appropriately on the ideological scale. Two additional benchmarks add further validity: our estimates behave as expected when compared with MP's membership in ideologically distinct party wings and with MPs' own ideological placement. The comparison to this external measuring instrument is particularly valuable. Appendix D summarizes further analyses.

German parties organize around wings and factions (for a review, see Sältzer (2020)). Because these wings hold different and homogeneous ideological positions, we can verify our estimates by matching members of party wings to our estimates. For example, Jürgen Trittin, Claudio Roth and Anton Hofreiter are known to belong to the "fundamentalist" - FUNDI- faction of the Greens, which can be distinguished from the "realist" - REALO- like Franziska Brantner or Cem Özdemir. Likewise, for the SPD, ideological differences prevail among party wings. The co-party leader - Saskia Esken - was a vocal critic of the decision to enter a coalition with the CDU and is recognized as more left-leaning than the rest of her party. Contrastingly, representatives members of the "Seeheimer Kreis" like Heiko Maas, Thomas Oppermann or Johannes Kahrs belong to the right segment of their parties. Given these differences, we expect the CSU (Bavarian conservative), the Seeheimer Circle (economically liberal democrats) and the Realists (economically liberal greens) to be to the right of their respective parties.

While no official listing of the faction membership exists, Sältzer (2020) compiled data on MP's group association. Mapping faction membership with our ideology scores, we can assess whether our estimates accurately placed factions beyond the few prominent individuals mentioned above for two parties in the Bundestag. These comparisons consist of 442 legislators. Figure 4 groups legislators of the SPD and the Greens according to their

faction. As expected, REALOs are more right and clearly distinguishable from the FUNDIs among the Greens. For the SPD, legislators belonging to the "Seeheimer Kreis" are placed to the right of the SPD. In a similar vein, we can affirm that conservative members from Bavaria (CSU) still are, on average, more right-leaning than other conservatives (CDU).

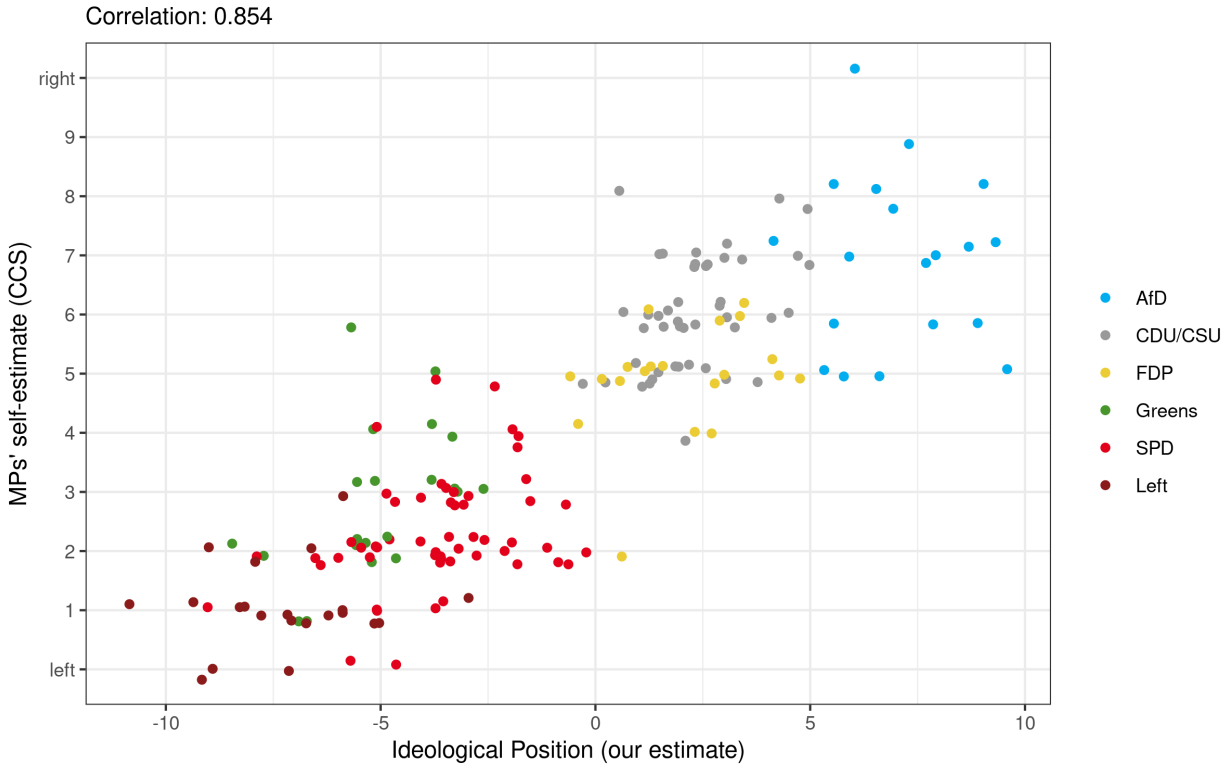
Figure 4: Comparison of legislators belonging to different partisan factions and between sister parties (CDU and CSU).



In the absence of a gold standard capturing the individual position of representatives, an appropriate alternative external measure for comparison and validation comes from non-behavioral data. The Comparative Candidate Survey (CCS) is, to our knowledge, the only possible measure here. Before a legislative election, the CCS asks candidates to estimate their own ideological position on an eleven-point scale. For the 2017 German election, 803 candidates took the survey and 186 were eventually elected. We use this available data to investigate the convergent validity of our estimate. As seen in Figure 5,

our expert-based estimates correlate highly with the self-placement from the CCS (r = .86). There seems to be a slight mismatch among MPs of the AfD who place themselves as centrists but are estimated to be among the far-right legislators. Overall, these figures and statistics deliver face validity for individual estimates and partisan aggregates, as well as convergent validity when compared to self-placement.

Figure 5: Comparison of expert-based estimates with self-placement.



Notes: Each point represents an MP, which took part in the Comparative Candidate Survey. The x-axis measures our ideological estimates, and the y-axis represents the MP's self-placement on an eleven-point ideological scale. For the sake of readability and considering the eleven-point scale used by the CCS, points are jittered. N = 186.

Finally, we found no evidence for systematic respondent bias based on $U_{i,s}$. Only 2.5% of all posterior distributions did not include 0 in their 95%-credible interval. Moreover, 95% of the estimated respondent-specific random effects fall within the interval [-.99, .92], which is tiny considering the width of the overall scale [-10;10]. o

4 Conclusion

Measuring individual positions of political actors is a fundamental task for political science. Especially in settings with high partisan discipline like the German parliament (Sieberer et al. 2020), obtaining ideological placements of legislators is challenging. In this paper, we introduce a simple research design for measuring the ideological positions of legislators. The measurement strategy comprises four components. First, we recruit members of the executive committee of the youth wings of parties. These experts are easy to reach and familiar with both MPs in leadership positions and backbenchers. Second, we ask these respondents to compare the ideological positions of many pairs of legislators. These relative comparisons are quick and reliable. This avoids differences in interpretations of numerical scales among respondents. Third, the transitivity of comparisons helps us to efficiently explore the vast comparison space and concentrates respondents on making informative comparisons. Fourth, we estimate a Davidson model based on all pairwise comparisons.

Naturally, the resulting estimates can be used substantially to improve our understanding of parliamentary processes. Additionally, our study enables new validation strategies for behavioral measures. For instance, our estimates can be used to understand what factors influence the validity of speech-based estimates so that the latter can be, when suited, systematically deployed. Our survey-based measure greatly contributes to making behavioral measures of ideology more robust.

In addition to the methodological aims, our article offers a substantive contribution by identifying the ideological positions of individual legislators in the 19th German Bundestag. The German parliament is a large national assembly with strong partisan discipline. Employing our design, we estimate the ideological position of its 709 members. These estimates coincide with the common perceptions of prominent MPs and with ide-

ological demarcations within party wings. We show that our estimates are valid and insensitive to potential biases among surveyed respondents. Overall, the proposed design is easy to implement and delivers accurate estimates and associated measures of uncertainty for legislators' ideological stance.

Our method is applicable to any political system. Pair-wise comparisons are a simple and robust psychometric tool for scaling preferences. The preferences can be easily retrieved using a straightforward design and estimation strategy. The biggest challenge lies in the identification and recruitment of knowledgeable experts. Young partisan leaders might not always be the ideal choice. In other countries, parliamentary journalists, parliamentary staffers, or even political actors themselves might provide informative comparisons. A common strategy relies on contacting country experts and academics by the researchers themselves, but several other options exist. For the American case, Hopkins and Noel (2022) used a polling company to recruit political activists and screen them before the survey. The central concern for recruitment centers on identifying highly knowledgeable political observers who are willing to participate in an acceptable rate.

We also believe that the design is flexible and can be extended easily. Two avenues seem particularly worthwhile to explore. First, one can expand the substantive scope of inquiry. In our application, we focus on a single left-right ideological dimension. Instead, a survey might ask a different set of questions altogether or let respondents decide whether they would like to rate legislators on more than one dimension (e.g., post-materialist values). Second, a common problem with spatial estimates is that they are based on a latent scale, making comparisons across political units and time difficult. Our design offers a simple solution to this problem. One might use a common anchor, such as a head of government, or even a fictional anchoring vignette in order to project individuals from different units, such as different branches of government, jurisdictions or even countries, into a common

scale.

References

Agresti, Alan (2013). *Categorical Data Analysis*. 3rd ed. Hoboken: Wiley & Sons.

Ainsley, Caitlin et al. (2020). "Roll-call vote selection: Implications for the study of legislative politics". In: *American Political Science Review* 114.3, pp. 691–706.

Bailer, Stefanie (2014). "Interviews and surveys in legislative research". In: *The Oxford handbook of legislative studies*, pp. 167–193.

Barberá, Pablo (2015). "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data". In: *Political analysis* 23.1, pp. 76–91.

Benoit, Kenneth, Kevin Munger, and Arthur Spirling (2019). "Measuring and explaining political sophistication through textual complexity". In: *American Journal of Political Science* 63.2, pp. 491–508.

Böckenholt, Ulf (2001). "Hierarchical modeling of paired comparison data." In: *Psychological Methods* 6.1, p. 49.

Bonica, Adam (2014). "Mapping the Ideological Marketplace". In: *American Journal of Political Science* 58.2, pp. 367–386. ISSN: 1540-5907.

Bradley, Ralph Allan and Milton E Terry (1952). "Rank analysis of incomplete block designs: I. The method of paired comparisons". In: *Biometrika* 39.3/4, pp. 324–345.

Brady, Henry E (1985). "The perils of survey research: Inter-personally incomparable responses". In: *Political methodology*, pp. 269–291.

Bräuninger, Thomas, Jochen Müller, and Christian Stecker (2016). "Modeling preferences using roll call votes in parliamentary systems". In: *Political Analysis* 24.2, pp. 189–210.

Canes-Wrone, Brandice (2015). "From mass preferences to policy". In: *Annual Review of Political Science* 18, pp. 147–165.

Carlson, David and Jacob M Montgomery (2017). "A pairwise comparison framework for fast, flexible, and reliable human coding of political texts". In: *The American Political Science Review* 111.4, p. 835.

Carmines, Edward G and Nicholas J D'Amico (2015). "The new look in political ideology research". In: *Annual Review of Political Science* 18, pp. 205–216.

Carroll, Royce and Keith Poole (2014). "Roll call analysis and the study of legislatures". In: *The Oxford handbook of legislative studies*, pp. 103–125.

Carrubba, Clifford, Matthew Gabel, and Simon Hug (2008). "Legislative voting behavior, seen and unseen: A theory of roll-call vote selection". In: *Legislative Studies Quarterly* 33.4, pp. 543–572.

Cattelan, Manuela (2012). "Models for paired comparison data: A review with emphasis on dependent data". In: *Statistical Science* 27.3, pp. 412–433.

Davidson, Roger R (1970). "On extending the Bradley-Terry model to accommodate ties in paired comparison experiments". In: *Journal of the American Statistical Association* 65.329, pp. 317–328.

Dewan, Torun and Arthur Spirling (2011). "Strategic Opposition and Government Cohesion in Westminster Democracies". In: *American Political Science Review* 105.2, pp. 337–358. ISSN: 1537-5943, 0003-0554.

Downs, Anthony (1957). *An economic theory of democracy*. Harper New York.

Hix, Simon (2004). "Electoral institutions and legislative behavior: Explaining voting defection in the European Parliament". In: *World politics* 56.2, pp. 194–223.

Hix, Simon and Abdul Noury (May 2016). "Government-Opposition or Left-Right? The Institutional Determinants of Voting in Legislatures*". In: *Political Science Research and Methods* 4.2, pp. 249–273. ISSN: 2049-8470, 2049-8489.

Hoffman, Matthew D, Andrew Gelman, et al. (2014). "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *J. Mach. Learn. Res.* 15.1, pp. 1593–1623.

Hopkins, Daniel J and Hans Noel (2022). "Trump and the shifting meaning of "conservative": Using activists' pairwise comparisons to measure politicians' perceived ideologies". In: *American Political Science Review*, pp. 1–8.

Huber, John D (1989). "Values and partisanship in left-right orientations: Measuring ideology". In: *European Journal of Political Research* 17.5, pp. 599–621.

Knox, Dean, Christopher Lucas, and Wendy K Tam Cho (2022). "Testing causal theories with learned proxies". In: *Annual Review of Political Science* 25, pp. 419–441.

Lauderdale, Benjamin E and Alexander Herzog (2016). "Measuring political positions from legislative speech". In: *Political Analysis*, pp. 374–394.

Laver, Michael (2014). "Measuring policy positions in political space". In: *Annual Review of Political Science* 17, pp. 207–223.

Laver, Michael, Kenneth Benoit, and John Garry (2003). "Extracting policy positions from political texts using words as data". In: *American political science review*, pp. 311–331.

Lesschaeve, Christophe (2017). "The predictive power of the left-right self-placement scale for the policy positions of voters and parties". In: *West European Politics* 40.2, pp. 357–377.

Loewen, Peter John, Daniel Rubenson, and Arthur Spirling (2012). "Testing the power of arguments in referendums: A Bradley–Terry approach". In: *Electoral Studies* 31.1, pp. 212–221.

Maestas, Cherie, Grant W Neeley, and Lilliard E Richardson Jr (2003). "The state of surveying legislators: Dilemmas and suggestions". In: *State Politics & Policy Quarterly* 3.1, pp. 90–108.

Mattos, David Issa and Érika Martins Silva Ramos (2021). "Bayesian Paired-Comparison with the bpcs Package". In: *arXiv preprint arXiv:2101.11227*.

Munzert, Simon (2018). *Measuring the Importance of Political Elites*.

Plott, Charles R (1991). "Will economics become an experimental science?" In: *Southern Economic Journal*, pp. 901–919.

Poole, Keith T (2005). *Spatial models of parliamentary voting*. Cambridge University Press.

Poole, Keith T and Howard Rosenthal (1985). "A spatial model for legislative roll call analysis". In: *American Journal of Political Science*, pp. 357–384.

Proksch, Sven-Oliver and Jonathan B Slapin (2008). "WORDFISH: Scaling software for estimating political positions from texts". In: *Version* 1, pp. 323–344.

– (2010). "Position taking in European Parliament speeches". In: *British Journal of Political Science* 40.3, pp. 587–611.

– (2012). "Institutional foundations of legislative speech". In: *American Journal of Political Science* 56.3, pp. 520–537.

Rheault, Ludovic and Christopher Cochrane (2020). "Word embeddings for the analysis of ideological placement in parliamentary corpora". In: *Political Analysis* 28.1, pp. 112–133.

Sältzer, Marius (2020). "Finding the bird's wings: Dimensions of factional conflict on Twitter". In: *Party Politics*, p. 1354068820957960.

Sieberer, Ulrich et al. (2020). "Roll-call votes in the German Bundestag: A new dataset, 1949–2013". In: *British journal of political science* 50.3, pp. 1137–1145.

Spirling, Arthur and Iain McLean (2006). "The rights and wrongs of roll calls". In: *Government and Opposition* 41.4, pp. 581–588.

– (2007). "UK OC OK? Interpreting optimal classification scores for the UK House of Commons". In: *Political Analysis*, pp. 85–96.