Can (and do) Voters Use Simple Mental Models to Forecast Cabinet Composition? Experimental Evidence from Ten Parliamentary Systems

David Fortunato
School of Global Policy and Strategy
University of California, San Diego
San Diego, CA, USA
dfortunato@ucsd.edu

Seonghui Lee
Department of Government
University of Essex
Colchester, UK
s.lee@essex.ac.uk

Randolph T. Stevenson
Department of Political Science
Rice University
Houston, TX, USA
randystevenson@rice.edu

Abstract

Elections in parliamentary democracies rarely yield single-party majorities, requiring voters to form expectations about potential governing coalitions to cast outcome-oriented votes—a task often dismissed by scholars as too difficult. This paper proposes that voters overcome this challenge using an ecologically rational mental model that relies on simple cues derived from party sizes and ideologies. Using original survey experiments across ten countries and more than 20,000 subjects, we demonstrate that voters expect coalitions that include the largest party as prime minister, are of "winning" size, and are ideologically compatible. Our theoretical model deterministically predicts survey respondents' expectations in up to 41% of cases. Furthermore, the weights voters assign to these cues vary contextually, mirroring the actual empirical regularities of cabinet formation in their respective countries. Ultimately, we provide robust experimental evidence that many voters can, and do, generate sensible coalition expectations using only a few simple cues.

How Voters Predict Coalition Bargaining Outcomes

Abstract

Coalition-directed voting in multiparty democracies requires expectations about how likely various coalition governments are to form following an election. We present a heuristic theory of how voters form coalition expectations and derive global and contextual empirical expectations from the model. To test these predictions, we design an experimental platform that allows us to learn which cues voters use to infer which parties will coalesce, and which roles the parties will play in government. Experimental evidence from ten parliamentary systems shows that voters are able to generate quite reasonable coalition expectations without relying on elite messages or coalition histories (the prevailing explanations in extant literature). The data further suggest that many voters possess an abstract mental model of coalition formation that is consistent with our theoretical arguments and that the cues they use to generate coalition expectations vary within-country according to the local empirical regularities of government formation.

> *"...each voter's ballot does not support the policies of any one party. Instead it supports the whole coalition that party joins. Thus the meaning of a vote for any party depends upon what coalitions it is likely to enter..."*
>
> Downs (1957, 162-163)

Elections in parliamentary democracies rarely produce single-party majorities, creating a problem for voters who want to use their vote to influence government composition: in order to cast such "coalition-directed" ballots, these voters must first have expectations of how likely different potential cabinets are to form following an election. Without such expectations, citizens cannot determine how to use their vote to increase the chances their preferred cabinets come to power or calculate the policy implications of their vote.[1] That said, forming expectations about the chances potential cabinets form (or, inferring "coalition probabilities") appears, on its face, to be difficult. The process of coalition formation is often complex and opaque, and the number of potential cabinets routinely runs into the thousands, seemingly making the mere enumeration of possibilities, let alone inferring their likelihoods, extraordinarily difficult.

For decades the ostensible difficulty of this task led researchers to dismiss the possibility that voters could accomplish it. Downs famously wrote that, instead of forecasting coalition outcomes, "...each voter either abstains, votes after cutting off his deliberation at some unpredictable point, or decides it is easier just to vote for his favorite party" (1957, 163). Decades later, Powell (2000) clarified the troubling normative significance—if voters are unable to infer which governments are and are not likely to form, their ability to issue prospective mandates or to coordinate on viable alternatives to underperforming incumbents is severely hindered, degrading both democratic responsiveness *and* accountability.

We propose that voters can solve this problem by employing an abstract mental model of coalition formation. The model, which is built upon interdisciplinary research on the ecological rationality of heuristics (e.g., Gigerenzer and Sturm 2012; Todd and Gigerenzer 2012), including recent applications in political science (e.g., Fortunato and Stevenson 2019, 2021; Fortunato et al. 2021), posits that voters can infer coalition probabilities using a set of "cues" (i.e., observable characteristics of potential coalitions) that depend on only two kinds of information about parties: their relative sizes and ideological positions. Further, the model posits that voters aggregate these cues into inferences about coalition probabilities using three simple heuristic rules: 1) the largest party will provide the prime minister; 2) the coalition will be of "winning" size;

[1] We write "coalition-oriented" rather than "policy-oriented" to emphasize that coalition expectations are required not just for voters who are voting to achieve certain policy outcomes, but also for voters who simply prefer to see some set of parties in cabinet for any reason at all.

and 3) to get to winning size, the largest party will (if necessary) coalesce with "ideologically compatible" partners.

Given the ubiquity of these three factors in explanations of real-world government formation in multiparty democracies, as well as the widespread availability of information on parties' relative sizes and positions, we expect this general model to apply across many different contexts. That said, ecologically rational models of heuristic inference depend on a set of empirical regularities that link potentially useful cues (characteristics of potential cabinets) to the inferential target (cabinet formation probabilities) *in a given context*. Thus, we can also use our theoretical model to derive hypotheses about how the weights voters place on different cues in generating their coalition expectations in a given context will vary in response to differing empirical regularities of coalition formation.

To test our arguments, we conducted two series of original survey experiments in which we presented subjects with hypothetical parliamentary compositions (randomized party sizes and ideological positions) and then elicited their expectations about which cabinet would form. The first is a two-trial experiment with five randomized parties, included on a multi-purpose, nine-country survey. The second is a 12-15 trial experiment with five, six, seven, eight, and nine randomized parties, conducted in three countries. Between these sets of experiments, we are able to study 80,000 coalition expectations, from over 20,000 subjects representing ten countries: Denmark, France, Germany, Hungary, Italy, the Netherlands, Poland, Spain, Sweden, and the United Kingdom. The results of these experiments provide, for the first time, unambiguous evidence that many voters possess an abstract mental model of coalition bargaining that is able to generate sensible coalition expectations using only information about parties' sizes and positions and without relying on elite messages or governance histories, the prevailing explanations in the literature. Indeed, as we will show, as many as 40% of our subjects seem to employ our proposed model.

In the rest of this manuscript, we briefly review the literature on coalition-directed voting, explain why such voting requires voters to possess coalition expectations, and examine how other scholars have addressed the concept. The review concludes with an explanation of why a clean causal assessment of whether individuals can forecast coalition provabilities with a simple size- and ideology-based mental model is the appropriate next step in this literature. We then propose our ecologically rational heuristic model of coalition expectations, derive testable hypotheses from it, and explain of our research design. Finally, we present the results, conclusions, and discuss the implications of our findings.

Coalition-directed voting and coalition expectations

Coalition-directed voting is the act of casting a ballot for an individual party with the intention of increasing the likelihood that one's preferred coalitions come to power.[2] To do this optimally, one must consider all the coalitions that each party could join, their associated formation probabilities, and how these probabilities are likely to change if one votes for each party (or not). For example, Duch and Stevenson (2008) propose a theoretical model of coalition-directed economic voting that formalizes the influence of coalition expectations and finds that if voters expect a given party to be excluded from *all* viable cabinet coalitions, a coalition-directed economic vote for that party will be "wasted" and therefore uncast. Likewise, if a voter thinks a particular party is *certain* to get into power in some coalition, maximizing the coalitional impact of her vote will require voting for some other party. Like Duch and Stevenson (2008), every model of coalition-directed voting of which we are aware assumes that voters know or can infer coalition probabilities (e.g., Gschwend 2001, 2007; Kedar 2005; Duch, May and Armstrong 2010; Indridason 2011; Meffert et al. 2011).

These models represent a clear departure from the empirical literature's previous consensus—either convinced by Downs (1957) or otherwise doubtful of voters' competence—that voters could not calculate or intuit the coalition expectations necessary for coalition-directed voting.[3] The change began in the early 2000's (e.g., Gschwend 2001, 2007; Kedar 2005), as some researchers developed a new approach to studying coalition-directed voting empirically: begin by building a theoretical model of coalition-directed voting that assumes voters have well-formed coalition expectations, produce estimates of those expectations from a stipulated model, assign these estimated expectations to respondents in an election survey, predict their votes under the model, and examine if actual voting behavior is consistent with the model's predictions.

One prominent example of this approach is Duch, May and Armstrong (2010) who assumed that voters employ a mental model of coalition formation that relates the observable characteristics of potential coalitions to coalition probabilities in the same way as political scientific models of coalition formation. Thus, they estimated the coalition expectations each of their respondents should have from the Martin and Stevenson (2001) empirical model of cabinet formation and then used those estimates to calculate how a rational voter should have voted given those stipulated coalition expectations and a Gamsonsian policy-making model. Influential work by Kedar (2005) and Duch and Stevenson (2008) takes a similar approach. We refer to such

[2]Many theories of coalition-directed voting are policy motivated (e.g., Kedar 2005; Duch, May and Armstrong 2010), but not all (e.g., Gschwend 2001, 2007; Duch and Stevenson 2008). Because coalition expectations are required for both policy-oriented and non-policy oriented coalition-directed voting, our discussion will focus on the more general idea that outcome-oriented voting in multiparty democracies in general requires voters to form coalition expectations.

[3]Many theorists were undeterred by Downs' skepticism, however. See Austen-Smith and Banks (1988).

models of voters' coalition expectations as "Voter as Political Scientist" (VaPS) models.

Thus, by the early 2000's we see a switch from almost universal skepticism that voters can form sensible coalition expectations to a willingness to write down complex models of coalition-directed voting that assume form coalition expectations as well or better (e.g., Kedar 2005) than political scientists. But this change in the literature occurred without first providing evidence that the previous consensus was wrong. Instead, these scholars advanced an "as-if" justification for their theoretical models that tolerated potentially tenuous assumptions about coalition expectations as long as they produced empirically verified down-stream predictions about voting behavior (and they did).

The importance of the "as-if" justification behind these models is apparent—and a continued underlying skepticism of voter capabilities is revealed—when we examine subsequent efforts to explore voters' coalition expectations directly. Rather than design studies to test if voters actually use the kinds of rational mental models of coalition formation that their theoretical (VaPS) models assumed, scholars instead focused on two alternative ways voters might form expectations that, while unreflective of their theoretical assumptions, require much less of voters: adaptive expectations derived from historical patterns of coalition formation (Gschwend 2001, 2007; Armstrong and Duch 2010; Bowler, McElroy and Müller 2022) and direct (or mediated) message-based expectations derived from elite signals and media reports of coalition probabilities (Bargsted and Kedar 2009; Meffert et al. 2011; Irwin and Van Holsteyn 2012; Eberl and Plescia 2018; Bowler, McElroy and Müller 2022).

Given this, one goal of the manuscript is to evaluate, for the first time, if voters can infer coalition probabilities that conform to political scientists' empirical models of cabinet formation without relying on coalition histories or elite messages. If the answer is "yes," then we can advise students of coalition-directed voting, contra Downs' intuition and the skepticism evident in much of the literature, to be bold in their assumptions about voters' abilities to infer coalition probabilities. That said, the heuristic model of coalition expectations we develop below is not the VaPS model that previous researchers have assumed. Instead, it can be thought of as a simplified version of the VaPS model that focuses on the most powerful and universal predictors of coalition formation: those based on the sizes and ideological profiles of potential coalitions. To preview, our analysis allows us to reject a "null model," that would predict, in the absence of coalition histories or elite messages, that coalition expectations are simply noise, as Downs and other skeptics may suggest. But the data are not perfectly consistent with a full VaPS model either. Instead, the data suggest that coalition expectations are the output of a mental model that, while not perfect, leverages the most predictive, non-redundant, universally applicable, and easily obtainable information available to produce inferences that are likely to be "good enough" to allow coalition-directed voters to cast a satisfying vote, or

perhaps simply talk about multi-party politics in a satisfying way.

A heuristic model of coalition expectations

Almost all explanations of expectation formation take one of three approaches: 1) adaptive expectations, which rely on historical trends; 2) direct and mediated message-based expectations, which outsource expectation formation to elite signals, experts, media, etc.; and 3) rational or model-based expectations, which utilize some mental model of how the world works to transform a set of informational inputs into expectations. Unfortunately, however, empirical testing has not kept pace with theorists' models. For example, no study (observational or otherwise) has ever tested whether voters actually condition their coalition expectations on previous histories of cabinet composition.[4] Likewise, while several observational studies have included measures of media and elite messages in empirical models of coalition expectations, these have seldom been paired with measures of coalition characteristics (like their sizes and ideologies)—likely for the reasons discussed in the previous section). Further, the X attempts to experimentally examine the direct impact of specific elite messages on coalition expectations have produced mixed results, with the largest effects in the moderate range and many null effects. Indeed, the message of one of the most studies often cited in support of the idea that elite messages can drive expectations, actually shows that those messages seem to only be received by party partisans. [FN: One challenge in accessing the empirical literature on coalition expectations is that such expectations are often not the direct focus of the studies. For example, while Gerswend... is certainly relevant, this work is really about testing theories of strategic, coalition directed voting, with discussions and empirical results around expectations playing a supporting role. As such researchers in this area have tended to devote limited recourses to ... [e.g., relying on preexisting data which tends to severely circumscribe the set of potential coalitions about which expectations are elicited) the This was one of the motivations for this study, which seeks to directly assess]

As discussed above, while scholars have often proposed theoretical models with model-based coalition expectations, these have not been directly tested. Below we fill this gap by designing a survey experiment in which we hold constant, by design, information about coalition histories and elite/media messages in order to explore whether our subjects can form sensible coalition expectations in their absence by using a mental

[4]This includes Armstrong and Duch (2010), which is often misread as providing evidence that history is an important source of coalition expectations. In fact, as the authors make quite clear, the article is simply a historical review of coalition combinations across countries. The authors do not measure coalition expectations and do not provide any empirical evidence for the hypothesis that coalition histories inform voters' coalition expectations, or, for that matter, that voter even know those histories.

model of the coalition formation process. That said, we do not expect these mental models to be perfectly informed rational models. Instead, we propose and test a model-based approach to expectation formation based on limited inputs, building on the large interdisciplinary literature on heuristic inference.

Heuristics are effort-reducing strategies that people use to make inferences or decisions based on limited information. Such strategies depend on a set of cues, in this case observable characteristics of potential coalitions, cue-weights, and aggregation rules that map the cues and cue-weights into an inference about some target, in this case coalition probabilities. Further, we can call a heuristic, and the person using it, *ecologically rational* if the heuristic meets four criteria in its particular context. First, it uses simple rules to aggregate cues and their weights, for example a simple linear additive function or a lexicographic "take-the-best" rule. Second, it produces inferences that are sufficiently accurate relative to alternative inferential strategies that they will be satisfying. This requires cues that are empirically associated with the inferential target in the real world, though these associations need not be causal. Third, it discounts cues, ceteris paribus, that are more costly to obtain, calculate, or infer than other potential cues. Fourth, it avoids the use of redundant cues.

Keeping these criteria in mind, we first build a core model of coalition expectations built upon cues that are readily available and strongly associated with government formation in all contexts. The theory of ecologically rational heuristics then produces the testable hypothesis that these cues should impact heuristic inferences in proportion to the size of their empirical associations, controlling for differences across cues in the cost to obtain, calculate, or infer their values.[5]

To conserve space, we do not review in detail the rather large literature on cabinet formation here—interested readers can find such reviews in Martin and Stevenson (2001, 2010), Müller and Strøm (2003), and De Marchi and Laver (2020, 2023). Instead, we can quickly summarize the characteristics of potential cabinets that political scientists have found to be predictive of coalition formation by relying on the recent comprehensive review and reanalysis of the data for 31 parliamentary democracies by Kayser, Orlowski and Rehmert (2023).[6] These variables broadly fall into one of six classes:

1. size characteristics (e.g., minority/minimal-winning/surplus majority or number of parties)

[5]See the definition of "validity" in Gigerenzer (2019, 3559) or the derivation of the "ecological rationality hypothesis" in Fortunato and Stevenson (2019, 313-315).

[6]Kayser, Orlowski and Rehmert (2023) do not account for some very recent (agent-based) models of coalition formation that allow the exploration models of coalition formation that are considerably more complex than previous efforts (De Marchi and Laver 2020, 2023). While these models improve our understanding of coalition formation and offer substantial promise for future revelation, it is presently unclear if they reveal new empirical associations between observable cues and coalition formation that voters may find useful to their expectation formation process. We discuss these issues in more detail in the appendix.

2. ideological compatibility (e.g., divisions within coalition or opposition)

3. bargaining power (e.g., plurality and median status of member parties)

4. electoral performance (e.g., changes in seat share)

5. coalition history (e.g., incumbency and record of co-governance)

6. elite signals (e.g., pre-electoral pacts)

Categories 1-3, size characteristics, ideological compatibility, and the inclusion of parties with strong bargaining power, are universal to all coalition formation episodes and have been found to be robustly associated with coalition formation (Martin and Stevenson 2001; Kayser, Orlowski and Rehmert 2023) and prime ministerial selection (Glasgow, Golder and Golder 2011). To be more specific, the most predictive characteristics of *coalitions* are their overall compatibility (i.e., minimizing ideological dissimilarity within the coalition), being minimal-winning rather than minority or surplus majority, and inclusion of both the largest party, who enjoys bargaining power in excess of its absolute size from formateur/informateur privileges, and the median party, whose bargaining leverage is self-evident. For *prime ministerial* selection, whether a party is the largest or the median are the only two variables that are consistently predictive across contexts. Critically, all variables from categories 1-3 are either directly observable or are a calculable function of parties' sizes and positions, two characteristics that research shows voters know quite well (e.g., Fortunato, Stevenson and Vonnahme 2016; Lee, Haime and Stevenson 2019, respectively). These are the variables our heuristic model is built upon.

In contrast, we do not integrate covariates from categories 4-6 into our heuristic model. We eliminate 4, electoral performance (average change in seat share), from consideration because it is uncorrelated with coalition formation (Kayser, Orlowski and Rehmert 2023). Categories 5 and 6, coalition history and elite signals, are the covariates representing the literature's prevailing view on how voters can come to possess coalition expectations, without a rational mental model, via either adaptive or direct message-based expectations, and therefore the characteristics we want to hold constant in our research design.[7][8]

[7]We include the well-documented negative association between pariah parties and coalition formation in these categories as the association is derived from history (they have not governed or rarely governed) and concurrent anti-pacts.

[8]It is of course possible that factors such as incumbency or pre-electoral pacts may be used to supplement a mental model "in the wild" (see Eberl and Plescia 2018; Bowler, McElroy and Müller 2022).

Core model and expectations

Our heuristic model generates expectations for which coalition will form and which party will lead it as prime minister. We begin our discussion with prime ministerial selection as most real-world government formation episodes are led by a formateur party that nearly always goes on to win the premiership.

PM selection research has distilled the universe of potentially significant, context-free predictors of which party will provide the PM to essentially two: size and ideological centrality—both reflections of bargaing power. The literature is unanimous that the best predictor, by far, of which party will be Prime Minister is the one that holds a plurality of legislative seats (Bäck and Dumont 2008; Glasgow, Golder and Golder 2011). Indeed, looking over all post-election government formations in the ParlGov database (Döring and Manow 2021), the largest party wins the premiership 76% of the time. Considering that an average of eight parties competed for the premiership over these 815 elections, predictive success of over 76% is very efficient.

Controlling the median legislator is also predictive, but less so than plurality, and there are at least three reasons why it may not factor heavily into voters' PM expectations. Most importantly, plurality status is much more predictive than median status in the real world. Second, median status is a more complex cue, requiring a rather complex aggregation of information on both size and position, substantially increasing its relative cost. Third, it is unclear whether many voters understand median status and its relevance to bargaining. Though the median voter theorem is well understood by political scientists, it is hardly a universal concept and rarely discussed in popular media. These factors led Fortunato et al. (2021) to argue that median status should *not* be used by voters to heuristically infer policy influence, and their analysis confirms this. Contrast this with plurality status, which *is* discussed in the media, is much less costly to observe, and is very intuitive (see, e.g., Duch, Przepiorka and Stevenson 2015). Thus, while our design will certainly allow estimation of the weight respondents place on median status in their expectations, our central empirical expectation for PM selection is that plurality status will be most the salient cue in prime ministerial prediction:

H_1: voters will expect the largest party to provide the PM

Empirical models of coalition formation tend to either identify which constellation of parties will coalesce without distinguishing their role in the government, or, identify which parties will join a PM (or formateur party) as partners in government (Martin and Stevenson 2001). Regardless of which approach is taken, however, there is clear consensus on the most important cues. Potential coalitions that are ideologically cohesive are much more likely to form than those that are ideologically divisive and,[9] majority coalitions,

[9]Nearly all empirical models of cabinet formation measure compatibility in a single dimension, but recent contributions by

specifically minimum-wining majorities, are more likely to form than minority coalitions (though this varies in interesting ways, see Strøm 1990). In addition to being the most predictive, these cues are also cheap, intuitive, and we believe often sufficient to generate "good enough" expectations. As such, we predict respondents will focus on overall ideological compatibility (or perhaps specifically compatibility with the PM) and identifying a coalition of a winning size. This yields the two following testable hypotheses.

H_2: voters will expect coalition parties to be ideologically compatible

H_3: voters will expect the coalition to be of a winning size

The reader may notice that we have not yet defined the terms "compatible" and "winning" in a rigorous way. This is intentional on our part because we lack guidance in the existing research on the specific way(s) that voters understand idealogical compatibility and winning status. That is, there are many ways voters may apply these two criteria. They may think of ideological (in)compatibility among coalition partners as the total variance their positions, the distance between the two most dissimilar parties, the average distance of all parties from the PM, etc. Of course, political scientists have not agreed upon a single concept of what it means for a coalition to be ideologically compatible—measurements of compatibility vary substantially across and even within studies. Likewise, voters may think of coalitions as "winning" if they control a majority of seats, are minimal-winning, are minimum-safe, etc.[10] Indeed, as voters are not political scientists, they may even use more exotic definitions of winning than are usually found in the scholarly literature. For example, they may intuit that a larger majority is safer or more stable than a minimal-winning majority, or, that a minority seat share that is nearly, but not quite, 50% is "close enough." Rather than impose, ex ante, strict definitions of compatibility and winning, we evaluate our model's predictions using a variety of specific (plausible) definitions for ideological compatibility and winning and then interrogate which seem to fit the data the best (including the possibility that none do).

There are two other covariates from the categories of size, compatibility, and bargaining power that have been consistently associated with the chances a potential coalition will form: whether it includes the largest party or the median party. Importantly, however, the direction and size of these associations may depend on whether the focal party is included in the potential coalition as PM or as a junior partner. While including the largest party as PM has a large positive impact on formation probabilities, including the largest party

De Marchi and Laver (2020, 2023) and Martin and Vanberg (2025) take a multidimensional approach. We discuss multidimensionality in the detail in the appendix, but note here that for simplicity—and the tractability of our experiments—we do not derive hypotheses or tests for multidimensional approaches to compatibility.

[10] A "minimum-safe majority" is a majority that is large enough to tolerate expected ad hoc defection (Provins, Monroe and Fortunato 2022, 1701).

as a junior partner has the opposite effect.[11] In contrast, the impact of median status is usually found to be positive (if modest in size) for both cabinet roles. That said, we are skeptical that voters will rely on median party status as a cue in identifying partner parties for the same reasons noted above regarding use of median status in identifying the PM: the cue is less predictive, more costly, and less intuitive than other available cues (see Fortunato et al. 2021). Of course, our empirical model will allow us to estimate the weight that our subjects place on these cues in inferring junior partners, so we can let the data "speak for themselves."

Finally, there are other indicators of bargaining power apart from plurality and median status, like minimum integer weights, which have been shown to influence an individual party's probability of cabinet participation and share of ministerial portfolios (Laver, De Marchi and Mutlu 2011; Cutler et al. 2016). For the same reasons that we expect voters to discount median status in their ecologically rational coalition expectations, we should be very skeptical that they somehow calculate or intuit complex bargaining weights—even if these were known to be strongly and consistently associated with formation outcomes. Similar to median status, Fortunato et al. (2021) find no evidence that such cues are understood or utilized by voters when inferring policy influence.

In sum, our ecologically rational heuristic model is composed of three cues—plurality, compatibility, and winning size. These cues are then aggregated using a simple rule. For example, ecologically rational voters may use a linear combination of cues and cue weights, where the cue weights are proportional to the cues' real world associations with coalition outcomes (discounted by cue availably and complexity, e.g., Fortunato et al. 2021). Alternatively, voters may use a "take-the-best" aggregation rule, in which alternatives are first evaluated on the basis of the most predictive cue (discounted as above) and are only evaluated on subsequent cues (ordered by their discounted predictive power) if previous cues fail to discriminate to a single alternative (e.g., Gigerenzer and Goldstein 1996, 1999).[12] In practice, expectations based on weighted linear cue aggregation and take-the-best should produce similar empirical patterns in a between-subjects experiment like ours, which gives us only a single response per formation opportunity (the coalition a subject thinks is most likely to form).

[11] To date, no study has modeled coalition formation and prime minister selection simultaneously as we do below, so these effects must be backed out by comparing across models of coalition formation and models of partner selection to a stipulated prime minister in extant research (e.g., Martin and Stevenson 2001). We confirm these associations with our data below.

[12] One process consistent with take-the-best would be to 1) identify the plurality party as PM and then, iff the PM alone is not winning, 2) add the nearest ideological party as partner and evaluate whether the coalition is winning. If yes, stop; if no, 3) add the next nearest neighbor, reevaluate the winning status, and so on.

Contextual correspondence

We expect the core heuristic model to be widely applied across contexts; however, the relative weights voters place on cues should vary contextually in relation to how accurately the cues predict real-world coalition formation outcomes. Consider an illustrative example from the following party system: $\{A_1^{36}, B_3^{12}, C_7^8, D_8^{12}, E_{10}^{32}\}$, where letters identify parties, subscripts indicate left-right positions, and superscripts indicate seatshares—so, party B is located at position 3 and has 12% of the seats. Suppose a respondent chooses party A (plurality) as their expected PM. At this point, she may then choose to add only party B to the coalition, which falls just shy of a majority, but keeps the cabinet ideologically compact. Alternatively, she may add parties B and C, building a majority coalition but also tripling its ideological spread.

The theory of ecological rationality suggests that the weight that voters put on majority vs compactness in forming their expectations depends on local context because the intuited weights are a function of the long-term correlations between these different cues and real coalition outcomes in their local context. For example, if our respondent lives in Germany, which has only been governed by majority cabinets in the post-war period, she may prioritize majority status over ideological compactness. In contrast, if she lives in Denmark or Sweden, which have been routinely governed by minority cabinets, she may prioritize ideological compactness over majority status. More generally, we expect a strong positive correlation between the relative weights placed on different cues in the formation of voters' expectations and the predictiveness of those cues in the formation of real-world governments within a particular country.

These contextual variations are part and parcel of an ecologically rational model in which people gravitate toward *contextually* predictive cues to make inferences by "selecting building blocks that are adapted to [a] particular environmental structure" (Todd and Gigerenzer 2007, 169). Thus, we predict that, within country, the relative weights placed on cabinet characteristics in voters' coalition expectations should positively correlate to the relative weights placed on cabinet characteristics in real-world government formation.

> H_4: cue-weights in respondents' coalition expectations will be positively correlated to cue predictiveness in real coalition formation in their country

Research design

Previous studies of coalition expectations or preferences have typically taken an observational approach to the question, eliciting expectations or preferences (or assessing their effects) over the formation of real governments in pre-electoral surveys (e.g., Eberl and Plescia 2018; Bowler, McElroy and Müller 2022; Harsgor, Itzkovitch-Malka and Tuttnauer 2023; Welz 2023). There are advantages to this approach. One is its

high external validity—the researcher is asking respondents to engage the very problem required to cast an outcome-oriented vote. Another is that assessing whether or not a respondent's expected coalition is the coalition that actually formed is a natural way to evaluate the accuracy of expectations.

There are also drawbacks to this approach. First, there are many possible coalitions, but survey infrastructure may only allow researchers to make a small subset of possibilities available to respondents. Second, relying on a single (real) party system limits the information available to estimate the weights respondents place on characteristics of potential coalitions—all subjects interact with the same array of party positions, the same plurality party, the same nearest neighbor to the plurality party, etc. Third, subjects bring varying levels of information to the design. Some respondents may have a detailed memory of which coalitions have formed in the past or know which parties have struck a pre-electoral pact, while other respondents may not. Further, these observational designs make it difficult to separate respondents' usage of a general mental model, based on observable characteristics of parties, from external signals about coalition viability from political elites and other sources.

We constructed our experimental design in light of these issues. First, we allowed participants an unconstrained choice set over included parties and their roles rather than providing a subset of alternatives. Second, we presented subjects with a diverse array of party systems. This means not only sweeping through multiple iterations of bargaining environments as defined by the distribution of parties' relative voting weights (Laver and Benoit 2015), but also mixing these bargaining environments with a diverse distribution of ideological arrangements, such that plurality parties are allowed to manifest at left, right, and center, pivotal members of minimum-wining coalitions may present as moderate or extreme, etc. Third, we controlled the *internal* information environment, providing relevant party characteristics to the subjects, so that we need not worry about "contamination" from subjects' differing levels of political information. Fourth, we controlled the *external* information environment, eliminating potential outside sources of coalition expectations, including elite messages and parties' histories of co-governance.[13]

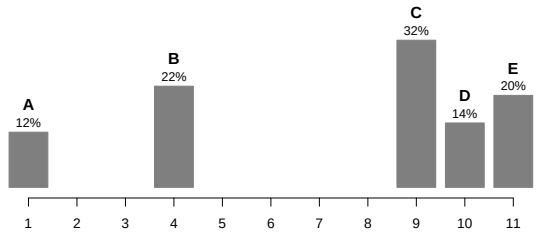
The experimental instrument presented subjects with a hypothetical legislative composition (parties' positions and seat shares) and then asked them to form a government by assigning each party to one of the following roles: prime minister, partner in cabinet, or opposition using a single response matrix. The only constraints on our subjects in this process are that they identify one and only one prime minister, and, that

[13]In an unpublished study that is superficially similar to ours, Duch and Tyran (2012) present subjects with eight coalition vignettes, asking them to identify the most likely coalition. This study does not tend to the four points above as it does not allow unconstrained choice over parties and roles, does not present a diverse array of party systems, and specifically instructs respondents to choose a *majority* coalition.

all parties must be assigned to one and only one role.[14] To be clear, we do not elicit coalition *probabilities* directly from our subjects. Rather, we have them identify the most likely (singular) coalition. We can then estimate the correlates of these expectations to test our hypotheses, and, assess our heuristic model holistically by calculating lexicographic correct classification and prediction efficiency. Importantly, this experimental format takes the election result as *given*, but evidence suggests that voters absorb information from polling such that they are able to infer post-electoral party sizes quite accurately before the election (Lee, Haime and Stevenson 2019). Thus, any conclusions drawn from this design should generalize to pre-election periods when polls are being regularly reported.

There are several benefits to this response structuring, including the freedom to identify any cabinet, with any array of parties serving in any role (save the PM constraint). The text of the prompt (with a sample electoral result) reads:[15]

"Imagine that a legislative election produced a legislature composed of the following five parties."



"What role do you think each party above is most likely to play in the post-election government?"

We identify the parties with single letters that are not utilized as party abbreviations in the respondents' respective countries. The image of the electoral result portrays the parties' share of legislative seats and ideological positions with an unlabeled x-axis, which allows the respondents to interpret that dimension in any way they see fit. This constrains dissimilarity to a single dimension in the instrument for tractability

[14]The experiment also allowed subjects to assign parties to be "outside supporter of the cabinet," but we collapse that role with opposition for our analysis here for the simple reason that there is insufficient information in the extant research about how voters think about external support parties, or even if they understand what such a party is, to treat it as a distinct category. This is evident in the small amount of research that we do have on voter's treatment of external support parties, which reveals substantial cross-national variation, and perhaps raises more questions than it answers (Tromborg, Stevenson and Fortunato 2019; Fortunato et al. 2021; Hjermitslev 2024). Nonetheless, we show in the appendix that our central results are unaffected by changing the designation of support parties from opposition to cabinet partners—the only substantive difference (in addition to much-degraded model fit) is a redistribution of minority cabinet probability to surplus majority, as one would expect.

[15]To be sure our subjects understood the graphic, we asked them to identify a correct description of one of the parties (e.g., "party A is the leftmost party and has 12% of the seats") from four alternatives.

purposes, but we make no claims that voters' broader understanding of politics is unidimensional, and, in the appendix, we discuss recent multidimensional models of cabinet formation by De Marchi and Laver (2020, 2023) and Martin and Vanberg (2025) and their relevance to our purposes here.

The array of sizes and positions were constructed by first defining all possible (unordered) divisions of legislative seats for a party system of n parties with the following constraints:[16] seats must sum to 100% across all parties, seats are allocated in 2% intervals, no party may have less than 2%, and no party may have more than 52%. Next, we defined all unordered distributions of ideological positions for an n-party system where positions are allocated between 0 and 10 in whole numbers and only one party may occupy any position. For our smallest party system of five, one can make over 100m unique electoral results, far too many to actually have our subjects evaluate. Instead, we take a random sample of 1,000 seat distributions and ideological arrays, and match one to the other (randomly assigning the ideological positions to the seat shares) to create the electoral result.

It is important to be clear about what we can and cannot learn from this experiment. All subjects are given the seat shares and ideological positions of all parties, allowing us to detect application of any cue derived from size, ideology, or both, including plurality and median statuses, ideological compatibility, cohesion of the opposition, minimal-winning status, and so on. Factors deriving from history and elite messaging, like incumbency, pre-electoral pacts, coalition forecasts by media, etc. are held constant by design. This means that we cannot estimate the weight voters would give such information in forming their expectations, or, whether or how voters may utilize such information "in the wild." As such, we cannot conclude that voters use our model and *not* history or elite messages—the prevailing explanations in the literature. We may only draw conclusions about the degree to which expectations are 1) consistent with our general model and 2) covary across contexts in response to differing empirical regularities as discussed in reference to H4.

We gathered data in two ways. First, we administered the experiment to about 1,000 subjects each in Denmark, Germany, and the Netherlands and presented each subject with 12-15 government formation episodes, each with a different "electoral result" (i.e., a different array of party sizes and positions).[17] These subjects interacted with party systems of five, six, seven, eight, or nine parties in similar proportions. Second, we embedded our experiment in a large, institutional survey of about 2,000 respondents each in France, Germany, Hungary, Italy, the Netherlands, Poland, Spain, Sweden, and the United Kingdom,

[16] By "unordered" we mean that we care only about the package of values, not the order in which the values may appear such that set $\{24, 24, 24, 26, 2\}$ and $\{26, 2, 24, 24, 24\}$ count as just *one* distribution of seat shares.

[17] We contracted the surveys from Qualtrics.

where each respondent was presented with two government formation episodes from two different electoral results for a five-party system.[18] Additional information on the surveys and experimental administration, including compliance with APSA's Principles and Guidance for Human Subjects Research is given in the appendix. These data allow us to compare how the cues voters may rely on to generate coalition expectations may change as the party-system grows, increasing the number of potential alternatives and therefore the complexity of the task. We can also compare the typical weights our respondents place on each cue in generating the expectations across a large number of countries.

The analysis proceeds in three steps. First, we estimate conditional logit models of our subjects' coalition expectations that identify the constellation of parties that will coalesce and the party will lead as PM. We present these estimates with a parallel model of real-world coalition formation utilizing the same covariates. These models allows us to describe the correlates of our subjects' expectations, assess a version of the VaPS model of expectations by comparing the correlates of our subjects' expectations to the real-world model results, and test H1-3. Second, we appraise the overall fit of our core model by calculating how many of our subjects' expectations are consistent with the model, evaluating different definitions of compatibility and winning status. Finally, we test H4 by assessing the extent to which the weight voters place on cues is correlated with the real-world empirical association between those cues cabinet and formation outcomes in each specific context. Together, these analyses show that respondents' expectations are reasonable in that they comport with findings from political scientific research, but not perfectly so; the application of our general heuristic model is likely widespread; and that we can soundly reject the nulls of H1, H2, and H4 in every country, and H3 in all but one country.

Analysis

Assessing a VaPS model

As discussed above, political scientists studying coalition-directed voting have proposed (or implicitly assumed) that voters can generate coalition expectations by employing a mental model that incorporates all the same factors that political scientists have found to be predictive of coalition formation (e.g., Duch and Stevenson 2008; Duch, May and Armstrong 2010; Indridason 2011). We call this the Voter as Political Scientist (VaPS) model of coalition expectations, which has, to date, never been tested directly. Instead, political scientists—perhaps convinced that the VaPS model was implausible in practice—proposed alternative explanations for how voters could come to generate or otherwise obtain coalition expectations, largely

[18] This survey was administered by [author's former institution].

focusing on signals from party elites or media (Bargsted and Kedar 2009; Irwin and Van Holsteyn 2012; Eberl and Plescia 2018; Bowler, McElroy and Müller 2022), or the the history of co-governance (Gschwend 2001, 2007; Armstrong and Duch 2010). Indeed, even Duch, May and Armstrong (2010), who impute coalition probabilities derived from conditional logit models of cabinet formation to voters in their data, discount the possibility that voters actually apply a VaPS model, and instead write that, "voters are informed about [coalition probabilities] because of, among other factors, the stability in coalition formations, public opinion polling that indicates which coalitions are likely to form, and party efforts that signal which coalitions are more or less likely to occur after the election" (706).

One goal in this initial description of our data is to actually evaluate the VaPS model and we do this in four ways. First, we specify conditional logit models following Martin and Stevenson (2001) and estimate them our experimental data and data on real-world cabinet formation from the same 10 countries. This allows us to coarsely assess whether or not the predictors of our subjects' experimental coalition expectations are broadly similar to the real-world predictors of coalition formation in those countries. Second, we compare the fully specified VaPS model to a stripped down specification that includes just the three covariates required for our heuristic model and show that the predictive accuracy of these two models is virtually identical. Third, we estimate separate conditional logit models across experimental data gathered from episodes involving 5, 6, 7, 8, and 9 parties to assess whether the stability (or predictability) of our respondents' expectations is sensitive to the complexity of the bargaining environment. It does not appear to be. And fourth, we estimate separate models for each of our 10 countries to analyze patterns of similarity and difference in expectation predictors across context.

Our VaPS model specification is based Martin and Stevenson (2001), specifically Model 7, which is what Duch, May and Armstrong (2010) use to estimate the coalition probabilities they would impute to respondents in their survey data. Our model departs from Martin and Stevenson (2001) in three ways. First, because we analyze experimental data, which only includes information on party sizes and positions, we can only include the characteristics of potential coalitions derived from party sizes and positions. These are: inclusion of the largest party, inclusion of the median party, the potential coalition's minority and minimal-winning status (relative to a surplus majority baseline), ideological divisions in the coalition, ideological divisions in the opposition (interacted with the coalition's minority status), and the (logged) number of parties in the coalition.

Second, we model the identification of the coalition and the party of the prime minister *simultaneously*, which has not been done before. This requires formatting our data such that there is one row for each potential PM in each potential coalition. For example, for coalition AB, there are two lines of data cor-

responding to two potential outcomes, one in which A is PM and one in which B is PM. It also requires differentiating covariates that describe characteristics of individual parties—inclusion of the plurality party and median party—by whether they refer to the prime ministerial party or one of its partners.

Third, we employ a different measure of the coalition's "ideological compatibility." There are many different ways that compatibility can be defined and measured—e.g., ideological span of partners, average distance from mean of partners, average distance from PM, and so on—but there is no specific guidance in the literature on how voters may understand the similarity or difference of more than two cardinal values. In the next section we compare different definitions of compatibility when assessing our heuristic model, but here, we use the measure that we find provides the best fit to the data, which is the sum of all partners' distances to the PM. For example, if the PM coalesces with one party that is 1 unit away and one party that is 3 units away, the measure takes on a value of 4. We use this measure of compatibility throughout our VaPS analyses, but our substantive conclusions are not sensitive to substituting alternative measures.

We estimate the same specification for the real-world model and experimental data VaPS model, while our limited specification includes only the covariates central to our heuristic model: largest party included as PM, total distance from the PM, and minimal-winning size. Before estimating, all continuous covariates (total distance from PM, opposition span, and logged number of parties) are standardized (mean=0; SD=1) to ease interpretation. We show the regression results from our experimental data side-by-side with results obtained from analyzing real government formation data from those same ten countries from 1970-2020 in Table 1.[19] [20]

There are important similarities and differences between the real-world and experimental correlations. Immediately clear is that the *signs* on all coefficients are identical across the models. It is also clear that the inclusion of the largest party as PM is critical in both data. Indeed, 81% of real cabinets in this sample (and 78% across all 36 post-war countries in the Parlgov archive) include the largest party as PM and 80% of our subjects' expected coalitions do as well.[21] Ideological compactness is similarly very important in both models, making it clear that voters have homed in on two of the most salient predictors of coalition formation in these systems.

[19] We use ParlGov data (Döring and Manow 2021) to model real cabinet formation.

[20] The real-world model has more correct predictions, given by $E[\hat{y}|y]$, than the experimental models. Much of this is driven by the relatively small number of single party majorities in the experimental data ($\frac{6971}{80873} = 0.086$) relative to the real-world data ($\frac{47}{209} = 0.225$).

[21] In complex party systems, say, those of more than five parties where the plurality party is not a majority (which corresponds to 562 of the 1,031 governments in the ParlGov archive) the coalition that ultimately forms contains the largest party as PM 69% of the time.

Table 1: Conditional logit regressions of cabinet formation and prime ministerial selection using real-world and experimental data. Experimental data only include 5-party systems.

	Real-world cabinet formation	Subjects' coalition expectations VaPS	Limited
Largest party is PM	2.792	2.358	2.795
	(0.237)	(0.012)	(0.009)
Largest party is a junior partner	−0.372	−0.524	
	(0.392)	(0.018)	
Median party is PM	1.503	0.393	
	(0.273)	(0.014)	
Median party is a junior partner	1.194	0.568	
	(0.238)	(0.012)	
Minimal-winning coalition	0.214	0.660	1.048
	(0.233)	(0.010)	(0.007)
Minority	−1.389	−0.730	
	(0.303)	(0.015)	
Total distance from PM	−5.647	−1.629	−1.706
	(0.503)	(0.011)	(0.007)
Opposition span	−0.021	−0.264	
	(0.120)	(0.006)	
Opposition spread × minority coalition	−0.161	−0.427	
	(0.161)	(0.010)	
Number of parties in coalition (logged)	−0.470	−0.438	
	(0.084)	(0.006)	
Episodes	209	80,873	80,873
N	6,617,804	35,918,432	35,918,432
Log Likelihood	−763	−296,326	−302,188
$E[\hat{y}\|y]$	0.354	0.205	0.211

Model construction based on Martin and Stevenson (2001), Duch, May and Armstrong (2010), and Kayser, Orlowski and Rehmert (2023).

There are also differences, though overall they do not strike us as stark. Perhaps most striking is that inclusion of median as PM is very important to the formation of real cabinets (third most salient predictor), but substantially less so in our respondents' expectations (ninth most salient predictor). This is consistent with our discussion of the complexity of the median status cue. Interestingly, the estimates on size features are also fairly different. In both models, the estimate on minority (relative to surplus majority) is negative, large, and efficient. However, while the experimental estimate on minimal-winning is positive, large, and efficient, it is smaller in the real-world data and does not clear traditional significance tests. These size estimates imply that our subjects' expectations are more similar to formal theoretical predictions than are the cabinets that have actually formed in these countries.

We estimate one more model, shown in the third column, to demonstrate which features are really doing the "work" in expectation formation. This model includes just the three covariates that are focal to our heuristic model: largest party is PM, ideological compatibility, and an indicator for minimum-wining size. The overall fit to the data, given by the log likelihood estimates, is virtually identical between the two models. Further, the plurality prediction from the limited model is significantly more accurate than the fully specified VaPS model: $E[\hat{y}|y]_{\text{limited}} = 0.211 > E[\hat{y}|y]_{\text{VaPS}} = 0.205; t = 3.242$. This is consistent with an ecologically rational model, built upon just a few cues. The other factors that we exclude from the limited model *are* predictive, but the data suggest that they are primarily influencing expectations at the margins, perhaps when the core cues fail to discriminate to a single alternative.

In sum, our conclusion from the estimates in Table 1 is that our subjects possess a mental model of coalition expectations that is broadly consistent with the true patterns of government formation in these countries and also consistent with a VaPS understanding of how competently voters can generate coalition expectations. However, the limited model analysis also reveals that voters' "true" mental models are likely much simpler than the VaPS conception and consistent with our ecologically rational expectations theory in that they appear to rely heavily on a small set of easily accessible cues that robustly and consistently predict coalition formation.

Before disaggregating the experimental data to compare across party-system size and countries, we use the estimates in Table 1 to test H1-3. Recall that H1 predicts subjects will expect the plurality party to provide the PM (positive estimate); H2 predicts subjects will expect coalition parties to be ideologically compatible (negative estimate on total distance from PM); and H3 predicts subjects will expect the coalition to be winning (positive estimate). Both the VaPS model and the limited model clearly allow us to reject the null of no association for largest party is PM, total distance from PM, and minimal-winning size.

We now ask whether the contours of our subjects' coalition expectations are consistent as the number of parties in the legislatures grows from five to nine parties. We pool data from these three samples, estimate the VaPS model for each party system size, and record the results in Table 2. The correlations and their relative magnitudes are remarkably (perhaps surprisingly) consistent across party system size, even though complexity increases rapidly with the number of potential alternatives growing from 80 to 2,304. Indeed, the estimates on our core cues—largest party as PM, total distance from PM, and minimal winning—secularly increase in size and efficiency as the party system grows, but the estimates on most ancillary cues do not. The estimates reassure us that tests of H1, H2, and H3 are not somehow an artifact of the smaller, simpler bargaining environment and also assuage concern about the external validity of our conclusions from the larger sample being limited by the five-party system.

Table 2: Comparing cue-weights across party system size in Denmark, Germany, and the Netherlands.

	Party system size				
	Five	Six	Seven	Eight	Nine
Largest party is PM	2.106	2.180	2.317	2.402	2.541
	(0.040)	(0.039)	(0.038)	(0.037)	(0.036)
Largest party is a junior partner	−0.730	−0.838	−0.723	−0.825	−0.808
	(0.060)	(0.060)	(0.058)	(0.059)	(0.059)
Median party is PM	0.642	0.681	0.676	0.796	0.704
	(0.045)	(0.045)	(0.045)	(0.046)	(0.046)
Median party is a junior partner	0.723	0.714	0.743	0.861	0.801
	(0.040)	(0.039)	(0.037)	(0.037)	(0.037)
Minimal-winning coalition	0.515	0.577	0.648	0.726	0.825
	(0.032)	(0.031)	(0.030)	(0.030)	(0.030)
Minority	−0.787	−1.006	−1.150	−1.161	−1.282
	(0.062)	(0.058)	(0.055)	(0.052)	(0.051)
Total distance from PM	−1.240	−1.353	−1.537	−1.664	−1.739
	(0.029)	(0.029)	(0.030)	(0.030)	(0.031)
Opposition span	−0.186	−0.315	−0.414	−0.489	−0.556
	(0.019)	(0.019)	(0.020)	(0.020)	(0.021)
Opposition spread × minority coalition	−0.232	−0.203	−0.221	−0.381	−0.303
	(0.040)	(0.042)	(0.044)	(0.044)	(0.045)
Number of parties in coalition (logged)	−0.300	−0.465	−0.559	−0.643	−0.704
	(0.023)	(0.020)	(0.018)	(0.016)	(0.015)
Alternatives per episode	80	192	448	1,024	2,304
Episodes	8,084	8,052	8,063	8,087	8,069
N	646,720	1,545,984	3,612,224	8,281,088	18,590,976
Log Likelihood	−24,025	−28,324	−32,338	−36,148	−39,865
$E[\hat{y}\|y]$	0.279	0.229	0.202	0.170	0.160

In table 3 we compare our subjects expectations across contexts by estimating the VaPS specification for each country individually. We limit the data to 5-party systems to maintain comparability. In each model, the largest positive estimate is on the largest party providing the PM and the largest negative estimate is on total distance from PM, allowing for conclusive rejection of the null of H1 and H2 in all countries. Estimates on minimal-winning are always large and positive, but are less stable across contexts. In Poland and Spain, the minimal-winning estimate is just a bit larger than the estimate on minority size, though still significantly larger ($p = 0.009$ and $p = 0.002$, respectively), and, in France, the estimate on minimal-winning and minority are indistinguishable to traditional levels of significance ($p = 0.698$). Thus, we cannot reject the null of H3 in France using minimal-winning as our criterion for winning, but we can reject the null

Table 3: Conditional logit regressions of cabinet formation and prime ministerial selection for individual countries.

	Country of sample									
	Denmark	France	Germany	Hungary	Italy	Netherlands	Poland	Spain	Sweden	UK
Largest party is PM	1.963	2.202	2.136	2.553	2.370	2.397	2.356	2.425	2.088	2.881
	(0.068)	(0.045)	(0.046)	(0.054)	(0.050)	(0.043)	(0.048)	(0.048)	(0.049)	(0.060)
Largest party is a junior partner	−0.472	0.134	−0.838	−0.086	−0.120	−0.493	−0.253	−0.502	−0.437	−0.147
	(0.100)	(0.064)	(0.068)	(0.075)	(0.070)	(0.061)	(0.071)	(0.080)	(0.067)	(0.089)
Median party is PM	0.710	0.032	0.510	−0.135	−0.123	0.408	0.024	0.257	0.393	0.015
	(0.077)	(0.051)	(0.051)	(0.059)	(0.056)	(0.046)	(0.054)	(0.055)	(0.055)	(0.066)
Median party is a junior partner	0.601	0.252	0.697	0.208	0.241	0.590	0.351	0.380	0.535	0.417
	(0.075)	(0.047)	(0.042)	(0.047)	(0.046)	(0.037)	(0.046)	(0.048)	(0.049)	(0.053)
Minimal-winning coalition	0.293	0.711	0.822	0.788	0.798	0.460	0.773	0.817	0.518	0.949
	(0.059)	(0.050)	(0.035)	(0.047)	(0.047)	(0.033)	(0.048)	(0.049)	(0.041)	(0.048)
Minority	−0.915	0.732	−0.893	0.169	0.423	−0.520	0.630	0.639	−0.510	0.383
	(0.111)	(0.066)	(0.073)	(0.072)	(0.069)	(0.060)	(0.067)	(0.069)	(0.073)	(0.077)
Total distance from PM	−1.508	−0.826	−0.940	−1.125	−1.026	−0.766	−0.822	−1.160	−1.099	−0.864
	(0.058)	(0.036)	(0.031)	(0.037)	(0.035)	(0.026)	(0.034)	(0.037)	(0.034)	(0.040)
Opposition span	−0.212	0.083	−0.069	−0.060	−0.088	−0.147	−0.076	−0.125	−0.300	0.016
	(0.035)	(0.027)	(0.020)	(0.025)	(0.025)	(0.018)	(0.025)	(0.026)	(0.023)	(0.027)
Opposition spread × minority coalition	−0.056	−0.475	−0.417	−0.707	−0.570	−0.451	−0.500	−0.442	−0.301	−0.471
	(0.068)	(0.041)	(0.048)	(0.047)	(0.044)	(0.041)	(0.043)	(0.043)	(0.048)	(0.048)
Number of parties in coalition (logged)	−0.310	−0.160	−0.447	−0.080	−0.053	−0.301	−0.094	0.011	−0.269	−0.349
	(0.041)	(0.027)	(0.026)	(0.030)	(0.029)	(0.024)	(0.028)	(0.028)	(0.029)	(0.031)
Alternatives per episode	80	80	80	80	80	80	80	80	80	80
Episodes	2,375	4,432	7,013	4,346	4,432	7,542	4,480	4,422	5,206	4,354
N	190,000	354,560	561,040	347,680	354,560	603,360	358,400	353,760	416,480	348,320
Log Likelihood	−7,160	−15,433	−19,763	−13,989	−14,971	−24,482	−15,392	−14,383	−16,694	−12,470
$E[\hat{y}\|y]$	0.265	0.139	0.322	0.214	0.178	0.210	0.168	0.191	0.236	0.237

of H3 in every other country, and, majority coalitions (combining minimal-winning and surplus) are more commonly expected than minorities in every country. We also note that there are clear differences in some of the estimates on the cues that are not focal to our heuristic model, mostly in relative magnitude, but also occasionally in sign. Some of these differences reveal departures the correlates of coalitions expectations from what the VaPS perspective would suggest (e.g., median is PM, minority). We will explore this variation in more detail in our tests of H4, and simply note for now that estimates on our focal cues are remarkably consistent across contexts, that we can reject the null of H1 and H2 in all countries, and reject the null of H3 in all but France.

Heuristic model fit

In the preceding section we used standard statistical models to explore the plausibility of a VaPS model of coalition expectations and found that, on average, expectations seem largely consistent with that perspective. However, comparison to a limited specification inspired by our heuristic model suggests that a substantial amount of the predictive power is driven by just three size and positions characteristics: largest party is

PM, ideological compatibility, and minimum-winning status. In this section, we assess the heuristic model directly by calculating its logical (tor lexicographic) fit to the data (e.g., Atsusaka 2021)—for each iteration of the experiment, we identify which coalition our heuristic model would predict, identify the coalition our respondent actually chose, and calculate the overlap between the two.

Identifying the heuristic model prediction requires specific definitions of "ideological compatibility" and "winning." Lacking guidance in the literature, we compare across eight different definitions of compatibility and four definitions of winning. The compatibility measures are:

1. *Connected*: parties in coalition are contiguous
2. *Coalition span*: total range of the coalition
3. *Average mean distance*: average of coalition parties' distances to unweighted coalition mean
4. *Average weighted mean distance*: average of coalition parties' distances to seat-weighted coalition mean
5. *Average PM distance*: average of partners' distances to PM
6. *Average weighted PM distance*: average of partners' seat-weighted distances to PM
7. *Max PM distance*: the maximum partner distance to PM
8. *Total PM distance*: the sum of partner distances to PM

All of these definitions are correlated, but vary in implied formation process. For example, minimizing distance to the prime minister implies the PM is central to government formation, whereas minimizing total dissimilarity implies that no one party is more important than others. We see definitions 1-3 as more holistic, while 4 is ambiguous, and 5-8 are clearly PM-driven. If PM-centered definitions fit the data better than holistic measures, we may conclude that subjects think of cabinet formation as driven by the PM.

There are two more important ways these measure differ. First, some measures are more complex, or, difficult to calculate or intuit. For example, *average weighted PM distance* is substantially more difficult to calculate than *total PM distance*. Second, some measures discriminate, or identify a unique prediction, more often than others. For example, there will be fewer "ties" when evaluating *average weighted PM distance* than *total PM distance*, as the seat weights induce variation discriminating between close sets. As such, we will want to consider how many predictions our heuristic model makes under different ideological compatibility measures when we compare them below.

Our definitions of winning from least to most restrictive, are:

1. Majority inclusive of ties
2. Majority
3. Minimal-winning inclusive of ties
4. Minimal-winning

"Majority" refers to any majority, minimal-winning or surplus. By "ties" we mean coalitions that have precisely 50% of seats rather than a proper majority. As with compatibility measures, there is little guidance in the literature as to how voters think about winning or majority status, whether or not they understand that 50% is not a majority, etc., so we let the data speak for themselves.[22]

With these definitions in hand we assess how well the heuristic model classifies or predicts our subjects' choices. For each episode, we identify all coalitions that could form, differentiating the PM party from its partners (if any). For episodes including 5, 6, 7, 8, and 9 parties, there are a total of 80, 192, 448, 1,024, and 2,304 potential coalitions, respectively. We first identify the subset that include the plurality party as PM and then the subset that meet the winning criterion. Of these winning coalitions with the plurality party as PM, the heuristic model prediction ($\hat{y}$) is the coalition that meets (minimizes) the compatibility criterion. All sets of definitions make at least prediction per episode, with the exception of the *connected* compatibility definition under minimal-winning inclusive of ties and strict minimal-winning size criteria. It is impossible to form a connected coalition around the plurality as PM that is minimal-winning or minimal-winning inclusive of ties in about 18% and 12% of episodes, respectively.

Recall that in our full sample there is an average of 444 potential coalitions per episode, so the baseline, naïve probability of any one alternative the respondent's expected coalition is $E[y] = 0.002$. In Table 4 we describe the fit of the heuristic model to the data under combinations of compatibility and winning by calculating three simple statistics. First, we calculate the average number of predictions per episode. Second, we calculate the rate of correct predictions, which is the number of times the respondent's expected coalition was also the heuristic model prediction, divided by the total number of respondent expectations: $E[\hat{y}|y]$. Third, we calculate predictive efficiency, which is the number of times the heuristic model prediction was also the respondent's expected coalition, divided by the total number of heuristic model predictions: $E[y|\hat{y}]$.

[22]There is some evidence that some voters may not differentiate between a majority and a large plurality (Duch, Przepiorka and Stevenson 2015), but drawing a "close enough" threshold would be an arbitrary exercise, relying on the presumption that voters are making small mistakes (or employing a "close enough" decision rule) and not intentionally creating a minority coalition, so we do not attempt that. Considering oversized definitions of winning, such as "minimum safe" (Provins, Monroe and Fortunato 2022), also requires an arbitrary threshold and only degrades model fit.

Table 4: Describing heuristic model fit across definitions of ideological compatibility and winning. In each column, the best performing measure is **bolded** and worst is *italicized*. Correct classification is $E[\hat{y}|y]$; predictive efficiency is $E[y|\hat{y}]$. There is an average of 444 potential coalitions per episode.

	Majority + ties			Majority		
Definition	Average predictions	Correct classification	Predictive efficiency	Average predictions	Correct classification	Predictive efficiency
Connected	*8.508*	**0.361**	*0.042*	*8.216*	**0.333**	*0.041*
Coalition span	1.578	0.246	0.156	1.658	0.251	0.151
Average mean distance	1.236	0.206	0.166	1.254	0.204	0.163
Average weighted mean distance	1.135	0.202	0.178	1.140	0.201	0.176
Average PM distance	1.299	0.205	0.158	1.300	0.200	0.154
Average weighted PM distance	**1.063**	*0.173*	0.162	**1.067**	*0.175*	0.164
Max PM distance	2.022	0.254	0.125	2.151	0.255	0.119
Total PM distance	1.201	0.231	**0.192**	1.208	0.235	**0.194**
Aggregation	2.110	0.405	0.192	1.885	0.383	0.203

	Minimal-winning + ties			Minimal-winning		
	Average predictions	Correct classification	Predictive efficiency	Average predictions	Correct classification	Predictive efficiency
Connected	*1.758*	**0.248**	*0.141*	*1.466*	0.220	*0.150*
Coalition span	1.326	0.236	0.178	1.298	**0.237**	0.182
Average mean distance	1.213	0.222	0.183	1.218	0.225	0.184
Average weighted mean distance	1.133	0.218	**0.193**	1.136	0.221	**0.194**
Average PM distance	1.183	0.216	0.183	1.171	0.217	0.185
Average weighted PM distance	**1.059**	*0.175*	0.165	**1.060**	*0.179*	0.168
Max PM distance	1.340	0.230	0.172	1.263	0.227	0.180
Total PM distance	1.201	0.231	0.192	1.208	0.235	**0.194**
Aggregation	1.142	0.290	0.254	1.097	0.268	0.244

High rates of classification are desirable, but so too is discrimination or specificity. The efficiency statistic discrimination by penalizing multiple predictions per episode. In each column of the table we **bold** the most desirable statistic and *italicize* the least desirable. Finally, we also calculate the aggregated performance of the heuristic model across all compatibility definitions and winning criteria. For each episode under each winning criterion, we identify the compatibility measure that does best—the measure that correctly classifies the respondent's choice by making the smallest number of predictions. We then calculate the summary statistics by aggregating the predictions and classification of the best performing compatibility measure across all choices. Our notion here is that different respondents may gravitate toward different definitions of compatibility and the aggregation represents the heuristic model's best fit while letting each respondent employ their own preferred compatibility definition. We reiterate that these are logical classifications, not

predictions from a statistical model. As such, there is no uncertainty or confidence interval to report.

The aggregation shows that as many as 41% of all expectations are consistent with our heuristic model. Given the very large number of alternatives, we interpret this as excellent model fit.[23] Comparing across measures, it is clear that measures that make more predictions produce more classifications, but also discriminate poorly. *Connected* is the best example, which makes by far the most predictions on average across all winning criteria and also makes the most classifications for all but minimal-winning. *Average weighted PM distance* discriminates best (makes the fewest predictions per episode), but also classifies the worst, under all winning criteria. *Total PM distance* is the most efficient compatibility definition under three of four winning criteria and produces the highest efficiency score across all compatibility and winning definition pairs (0.194). This definition has two interesting and appealing properties. The first is that, because it is additive in parties, all winning coalitions that minimize *total PM distance* are minimal-winning—hence identical fit statistics across majority and minimal-winning definitions (both inclusive and exclusive of ties). The second is that this measure captures an intuitive process for identifying an expected cabinet: identify the plurality party as PM and, if it is not a single-party majority, add its nearest neighbor. If this produces a majority then stop. If not, then add the next nearest neighbor, and so on.

We use *total PM distance*, the most efficient compatibility definition, with minimal-winning to assess heuristic model to explore our heuristic model a bit further. We contend that correct classification of 0.235 and predictive efficiency of 0.194 is very impressive in light of the very large number of alternatives (444 on average) in each choice. But the model is perhaps even more predictive than this efficiency measure implies as there may potentially be many "near misses," expectations that are very similar to the model prediction, but just miss classification. For example, consider set $\{A_1^{40}, B_3^4, C_5^{12}, D_7^8, E_9^{36}\}$, where a respondent identifies A as PM, adds nearest neighbor B, assesses total size (44%, not a majority), and then adds the next party C to create a majority (56%). This coalition is in keeping with the spirit of our heuristic model, but would not be classified by the model, because B is superfluous to the majority and increases the sum of distances.

[23] We can contextualize our model's fit with a measures designed to compare across choice sets of varying size by Selten (1991). This definition of accuracy is, r, its "hit rate" (rate of correct classification), minus a, its predictive "area"—the proportion of available alternatives the model predicts. In the case of the aggregation, where $r = 0.405$ and $a = \frac{2.111}{444} = 0.005$, this is: $r - a = 0.405 - 0.005 = 0.4$. In order for theoretical model predicting binary outcomes to achieve a score this high, it would have to produce a hit rate of 0.9. The partisan model of US senate voting (ie., that senators vote their party line), for example, produces such a hit rate in the modern Congress (Fortunato and Stevenson 2019). Recent research on voting in the United States shows that a partisan model of vote choice (i.e., that an individual selects the candidate representing the party that the individual identifies with)—possibly the most well-documented choice model in political science—would produce score of $r - a = 0.8 - 0.5 = 0.3$ for presidential elections and $r - a = 0.79 - 0.5 = 0.29$ for Congressional elections (Bankert 2021).

To get a sense of how many such near misses there may be, we calculate a measure of deviance from the model prediction: how similar are all potential cabinets in *implied ideology*, or overall position, to the heuristic model prediction. We assess this by calculating the absolute difference in seat-weighted left-right position between the model prediction and all alternatives. Examining the 83% of cases in which the heuristic model discriminates to a single alternative (in order to avoid multiple comparisons between potential alternatives and the heuristic model prediction), we find that respondents' expected coalitions are 0.8 units closer to the model prediction than we would expect by chance (coalitions can be up to ten units away from the model prediction).

Table 5: Heuristic model strict classification and near misses across party system sizes. Compatibility definition is *total PM distance* and winning criterion is minimal-winning. Only episodes in which heuristic model makes a unique prediction are included (83% of episodes). Episodes with 6 or more parties are only available from Denmark, Germany, and Netherlands, identified as "DGN" below.

Parties	Alternatives	Baseline hit rate	Strict classification	Near misses	Sample
Five	80	0.0125	0.2340	0.5723	all countries
Six	192	0.0052	0.2578	0.6043	DGN
Seven	448	0.0022	0.2322	0.6016	DGN
Eight	1024	0.0010	0.2048	0.6038	DGN
Nine	2304	0.0004	0.1898	0.5816	DGN
Total (average)	444	0.0023	0.2295	0.5822	

In Table 5 we explore near misses in more detail. These data only include episodes where the heuristic model makes a unique prediction. Given that only one prediction is made, we only have one prediction from which to calculate "near misses," and, the classification rate is equivalent to predictive efficiency ($E[\hat{y}|y] = E[y|\hat{y}]$). The table disaggregates the data according to the party system size and records the number of potential alternatives, the baseline hit rate (the probability of selecting any one of those alternatives by chance), the heuristic model's strict classification rate, and the classification rate inclusive of near misses. Note that we only have data on party systems of six or more from Denmark, Germany, and the Netherlands (DGN), so the increase in classification rate when we move from five to six parties primarily reflects differences in model fit across countries (which we explore below). All told, the heuristic model's strict classification and near miss performance is fairly stable across party system, which is remarkable given the significant growth in the choice set as the system grows. In the most complex party system of 9, the heuristic model inclusive of near misses classifies at an impressive rate of 0.58 and has a strict classification rate of 0.19, or, about 437× better than the baseline hit rate. Across the whole sample, the heuristic

model inclusive of near misses still has a classification rate of 0.58 and a strict classification rate of 0.23, or, over 100× better than the baseline hit rate. We conclude that this is very strong evidence for widespread application of an abstract mental model of coalition expectations that is consistent with our theoretical argument.

Table 6: Heuristic model fit across country samples. Compatibility definition is *total PM distance* and winning criterion is minimal-winning. Only five-party episodes.

Sample	Strict classification	Near misses
Germany	0.328	0.581
Denmark	0.260	0.602
United Kingdom	0.251	0.593
Hungary	0.240	0.562
Netherlands	0.234	0.545
Sweden	0.221	0.560
Italy	0.211	0.566
Spain	0.210	0.570
Poland	0.196	0.555
France	0.178	0.536
France	0.178	0.536

Finally, using five-party episodes to ease comparison, we describe differences in heuristic model fit across countries. We again compare model fit using *total PM distance* and minimal-winning as our compatibility and winning criteria. Table 6 records heuristic model fit across countries using both strict classification and classification inclusive of near misses.[24] The Table reveals substantial variability in fit across contexts. Indeed, the model classifies subject expectations in Germany at 180% the rate at which it classifies French expectations. There are several potential reasons for this variation, including the fact that the heuristic model is a better fit to the real empirical regularities of coalition formation in some contexts than others. For example, at the time our experiments were administered, every post-election cabinet formed in Germany in the post-war era had been a minimal-winning coalition that minimized partner distance from the plurality PM—a precise fit to the heuristic model. But the empirical regularities of French—and Polish, Italian, Spanish, etc.—cabinets frequently diverge from the heuristic model. Of course, the cross-sectional variability in the empirical regularities goes beyond the cues that our model is built upon, as one may expect, but this level of variation in core model application is still interesting. We analyze and discuss this variation in more

[24]We adjust for the specific array of party sizes in each choice, because model fit varies over arrays and the distribution of arrays varies over countries. To adjust, we simply regress the classification of each episode $\{0, 1\}$ on country while absorbing variation across arrays with fixed effects. Each classification estimate has a standard error of 0.01

detail in the next section, but first a final note. Even in France, where the model's strict classification is the poorest, its relative classification is still quite high—over 14× the base hit rate (0.0125 for five-party episodes). Taken together with the classification rate of near misses, the data suggest that a substantial number of voters possess a mental model of coalition formation that is broadly consistent with our proposed heuristic model, from which they can derive reasonable coalition expectations.

A note on external validity

The preceding analyses clearly illustrate that many voters possess a mental model of coalition formation that they may call upon to generate reasonable coalition expectations. Our experiments allow clean identification of a causal impact of various size and position characteristics on our subjects' expectations. The results are robust and internally valid. However, this positive result does not allow us to draw strong conclusions about how voters generate coalition in the real world. As our anonymous reviewers point out, our external validity is limited by at least two factors.

First, most of our subjects interacted with a small number of parties, which may makes the task easier than in the real world, particularly countries, like the Netherlands, with quite a large number of parties in parliament. We believe that comparison of cue-weights derived from statistical models (Table 2) and heuristic model fit across party system sizes (Table 5) should assuage this external validity concern.

Second, it is possible that calling upon a mental model in the real world is simply unnecessary, or more effortful than alternative means of generating expectations, in particular gleaning expectations from elite and media messages or calling upon parties histories of co-governance. In the appendix we report that results of two experiments administered in Denmark and Germany, that speak to these concerns. The experiments present subjects with the same task of determining which government will form given an array of parties' sizes and positions, but this time the parties are named after the real parties composing the Danish and German parliaments at that time. Further, we made only quite small changes to sizes and positions of the parties from their true values at the time. Of course, by giving the parties real, familiar names, we invite our subjects to bring in all the various competing information and associations that they have regarding those parties that we specifically designed our primary experiments to eliminate. The survey that housed the experiments also asked respondents for their perception of how often each party-pair had co-governed over the last few decades and the degree to which they believed each party was committed to democracy, which allows us to account anti-system party inclusion and proxy for media messages signaling negative coalition probability for pariah parties (like the Danish People's Party or the Alternative for Germany). For the German data, we were further able measure the media narrative regarding the likelihood of various

coalitions forming.

Analysis of these ancillary experiments reveals that our subjects do, in fact, place the weight on these factors in the expected directions. They are more likely to expect coalitions to form if they have (real or) perceived longer histories of co-governance, if the parties are perceived as committed democracy, and if media portray the parties as likely to coalesce. The analysis also reveals that subjects are still quite responsive size and position characteristics, and, indeed, that the relative weights placed on inclusion of the plurality party as PM, ideological compatibility, and winning status are substantially larger than the weights placed on history, anti-system sentiment, and media narrative. In sum, this is evidence for the application of a mental model of coalition formation in a context that is very close to the real world.

A very important corollary, is that the results of the central experiments *could have been negative*. If they had been—showing that even in the rather pristine setting of the lab, with other influences carefully controlled, few respondents use size and ideology to predict cabinet compositions—it would have cast considerable doubt on the proposition that voters use them in the real world. In other words a null result in this laboratory setting would almost certainly imply a real absence of a mental model and therefore an impossibility of model usage in the real world. Of course, our central experiments found robust evidence for a sensible mental model that these ancillary experiments imply is utilized in real world situations as well.

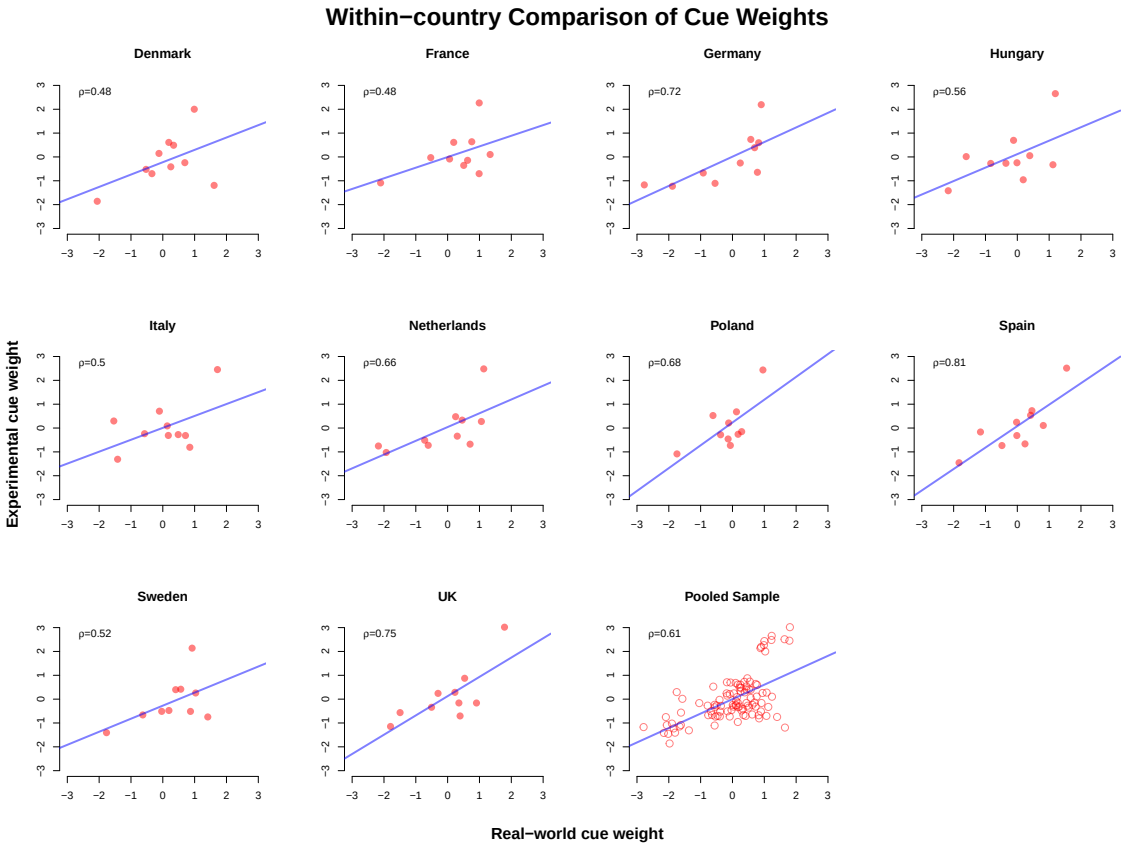
Assessing contextual correspondence

In this section we test H4 by comparing the extent to which the characteristics of potential coalitions that predict real-world cabinets in a given context are the same characteristics (cues) that predict our subjects' experimental coalition expectations. In keeping with the theory of ecologically rational heuristics (Fortunato and Stevenson 2019; Fortunato et al. 2021), we expect a strong, positive correspondence of experimental and real-world cue-weights *within-countries*. Recovering such correlations would support H4, demonstrating that voters' mental models of coalition formation are ecologically rational, or, well-adapted to their specific context.

To estimate the country-specific cue-weights, we return to the analyses of real-world cabinet formation and our subjects' coalition expectations presented in Table 1 and make two small changes to the models. First, we replace the compatibility measure, *coalition span*, with *total PM distance* as *total PM distance* performed best in the data fitting comparisons above. Second, for the expectations data, we interact each covariate with country in order to derive country-specific weights for each cue. For the real-world data, we follow Fortunato, Stevenson and Vonnahme (2016) and estimate country-level random coefficients for each cue. Random effects are necessary here due to separation issues; e.g., all German cabinets are minimal-

winning and include the plurality party as PM. Random coefficients estimators are able to produce reasonably efficient, accurate, and comparable coefficient estimates under separation without placing strict ex-ante constraints on their values (Crisman-Cox 2021). With these country-level estimates of cue weights from our experiments and real-world cabinet formation, we can assess the extent to which the relative predictive power of characteristics in the real-world influences the relative weight these cues bear in structuring our subjects' coalition expectations across contexts. We report complete estimates from these models in the appendix and only compare the country-level cue weights that we derive from them here.

Figure 1: Comparison of relative weights of cues within countries.



In Figure 1 we plot the correspondence of the cue-weights derived from the real-world and experimental data for each country. The estimates are standardized to ease comparison. Each point represents a cue (e.g., inclusion of plurality party as PM) where the x-axis value conveys the predictive power of the cue in the formation of real-world coalitions in a given country and y-axis value conveys the predictive power of the cue in our subjects' coalition expectations in that country. In each and every country, there is a clear, strong, positive relationship between the two, with a pooled correlation of $\rho = 0.61; p < 0.01$. The data are quite

conclusive that the weights placed on observable characteristics of coalitions in respondents' expectation formation in a given country are strongly, positively correlated to the predictiveness of those characteristics for real-world government formation in that country. In other words, the coalitions that our subjects expect to form tend to look like the coalitions that have formed in their country over the last several decades. This strong correlation is evidence of ecologically rational coalition expectations.

We also note that each plot reveals a very strong correspondence for a few cues, but a weaker correspondence for the rest. This is clearest in the pooled plot, which shows a cluster of quite positive and quite negative cues and then a larger group of cues clustered about the experimental (y-axis) zero value, but much more variable in their real-world (x-axis) values. What these data show is that, even though the overall within-country correspondence is quite strong, it is disproportionately driven by a small number of cues. Below we estimate full- and restricted-sample regressions to probe this further.

Table 7: Relationship between real-world and experimental cue-weights in coalition expectations. All countries, 5-party systems only.

Covariate	Pooled	Within country	Within cue	Within country, only 3 focal cues	Within country, omit 3 focal cues
Real-world weight	0.242	0.246	0.019	0.345	0.070
	(0.026)	(0.027)	(0.028)	(0.022)	(0.033)
Country FEs		✓		✓	✓
Cue FEs			✓		
Observations	100	100	100	30	70
R^2	0.461	0.480	0.906	0.931	0.112

Focal cues are: largest party is PM, ideological compatibility, and minimal-winning status.

In Table 7, we regress the experimental cue-weights on their real-world counterparts, presenting results from a pooled model, models including country or cue fixed effects, and regressions restricted to only the three focal cues from our heuristic model (largest party is PM, ideological compatibility, and minimal-winning status), or including all *but* those three focal cues. The within-country model results, which basically summarize the information in Figure 1, allow us to conclusively reject the null of H4. Models examining the within-country correspondence of just our three focal cues, or omitting the three focal cues, lend further credibility to the argument that coalition expectations are ecologically rational. These models are telling us that the core heuristic model appears to be wide-spread, *and*, the weights applied to the cues vary in relation to the salience of those factors in the formation of real-world cabinets in a particular country.

Further, comparing the correspondence of the focal cues to the correspondence of the non-focal cues shows that, even though both sets are positively and significantly associated, the association for the focal cues is much, much greater in terms of both substance and statistical significance. Again, this is in keeping with an ecologically rational model, where just a few cues can be leveraged into efficient inferences. The association of the non-focal cue weights to their real-world explanatory power *is* significant, but the overall correspondence between the two is much weaker relative to the correspondence for the focal cues. This suggests that subjects applying the heuristic model seem to do so in a manner calibrated to their local context, and, that when our subjects' expectations deviate from the strict predictions of the heuristic model, they tend to do so in a way that is consistent with the empirical regularities of real-world coalition formation in their country, however, the explanatory power of the deviations is quite modest compared to the explanatory power of the focal cues in the core heuristic model.

Finally, the model examining within-cue correspondence does not reveal a significant association. This null-correlation is driven by just two outlying deviations—the estimates on the minority cue in Denmark and Sweden. By and large, our Danish and Swedish subjects expect majority coalitions to form (refer to table Table 3), even though their countries have exclusively (Denmark) or routinely (Sweden) formed minority governments in the post-war period. Indeed, both Danes and Swedes place the second largest negative weight on the minority cue, despite that cue bearing largest (Denmark) or second largest (Sweden) *positive* weight in real-world coalition formation in those countries. Omitting these two observations and re-estimating with the remaining 98 observations reveals within-cue correspondence of $\beta = 0.120$ (0.028). Speculating about why these deviations manifest, one potential explanation is that Danes and Swedes simply understand that policymaking requires majorities, and, to the extent that they have internalized the empirical regularity of minority governance in their countries, they are willing to divorce it from their interaction with the experiment.

It is also worth noting that, on average, there is substantially less within-cue variability in expectations ($\bar{sd} = 0.251$) than there is in real-world coalition formation ($\bar{sd} = 0.827$). Part of this may be due to the compression of potential alternatives in these five-party systems (more on this below), but this is also in keeping with an efficient heuristic model in which the majority of the predictive power in expectations is derived from just a few focal cues.

Before concluding, we use the data from our directed surveys of Denmark, Germany, and the Netherlands to assess how the correspondence between real-world and experimental cue weights may be influenced by party-system size. To generate the country-specific cue weights, we estimate fully interacted conditional logit models of expectation formation for each party system size. Table 8 regresses these estimates on the

Table 8: Comparing relationship between real-world and experimental cue-weights in coalition expectations across party system size in Denmark, Germany, and the Netherlands.

	Party system size				
Covariate	Five	Six	Seven	Eight	Nine
Real-world weight	0.261	0.286	0.314	0.331	0.342
	(0.044)	(0.046)	(0.050)	(0.053)	(0.056)
Country FEs	✓	✓	✓	✓	✓
Observations	30	30	30	30	30
R^2	0.577	0.595	0.605	0.604	0.592

real-world cue weights for Denmark, Germany, and the Netherlands to recover the within-country association for party systems of five, six, seven, eight, and nine. The comparison is revealing as it shows closer correspondence as the party system grows. Why? One reason is that the expansion of the party system increases the number and variability of potential alternatives which, in turn, allows our subjects more latitude to build coalitions in keeping with the empirical regularities of real-world cabinets in their country. This implies that the more constrained choice set utilized in our most of experimental instruments perhaps subdued our ability to recover evidence for our empirical predictions in general and H4 in particular, due to the compression of alternatives.

Discussion

Can voters infer coalition likelihoods without relying on coalition histories or elite messages? The argument and analyses above suggest that the answer is "yes" and that 1) many voters possess a general mental model of coalition formation that needs only information on party sizes and positions; 2) these models can generate reasonably accurate coalition expectations; and 3) many employ a version of our ecologically rational heuristic model which is composed of the following rules: the largest party will provide the prime minister, the coalition will be of "winning" size, and the largest party will (if necessary) coalesce with "ideologically compatible" partners. Indeed, analyzing coalition expectations from experiments administered in ten parliamentary democracies, we find as many as 40% of expectations are consistent with our stipulated heuristic model when aggregating different definitions of compatibility with a permissive winning definition, and, that 23% of expectations are consistent with our most efficient, singular definitions of compatibility and winning. Given an average choice set of 444 potential outcomes, we believe this a very good fit to the

data.

Analyzing expectation patterns within and across countries, we also find that, in keeping with an ecologically rational heuristic model, expectations comport with the empirical regularities of cabinet formation in that particular context—cue weights are applied in expectation formation proportionally to their real-world associations with cabinet formation. Though our subjects' expectations are not a perfect match to the Voter as Political Scientist expectations model, there is still a very close correspondence between voters' expectations and both the political scientific understanding of what their expectations *should* be and the structure of real-world governing coalitions in their country. This correspondence to within-country cue weights implies opportunity for interesting individual-level variability that we hope to investigate in the future. Recent research documenting individual- and group-level variation in "ideological awareness" (Fortunato, Stevenson and Vonnahme 2016; Martin and Vanberg 2025) suggests that these within-country correspondences may be stronger for more interested or sophisticate voters than their less interested counterparts and preliminary analysis suggests this is the case in our data (see Appendix). This is but one of many potential individual-level extensions of this study.

In general, our conclusions depart from past arguments, including "as-if" models that leave the actual process of expectation formation a black box (Austen-Smith and Banks 1988; Duch, May and Armstrong 2010), and, the presumption that voters, lacking the ability to (or interest in) forecast(ing) coalition outcomes based on the characteristics of potential cabinets, instead rely exclusively on history (Gschwend 2001, 2007; Armstrong and Duch 2010; Bowler, McElroy and Müller 2022) or elite signals to form expectations (Bargsted and Kedar 2009; Irwin and Van Holsteyn 2012; Eberl and Plescia 2018; Bowler, McElroy and Müller 2022). To be clear, we are not arguing that these approaches are incredible, simply that they are distinct. What we have discovered here is strong evidence that many voters possess an abstract mental model of a process that political scientists previously assumed was far beyond their comprehension (Downs 1957). This gives credence to the simple idea that for some specific, common, but complex tasks—like figuring our how different people (or parties) form coalitions to get what they want—voters possess *general* mental models that they can readily adapt and apply. Indeed, one could reasonably argue that the general model our subjects employ to generate coalition expectations is a close relative to others that have been used to explain responsibility responsibility attributions for multiparty policymaking (Duch, Przepiorka and Stevenson 2015; Bowler, Gschwend and Indridason 2020; Fortunato et al. 2021).

Zooming out and thinking more broadly about these tasks, this is not too surprising. Though coalition formation and policymaking can certainly be very complex processes, they are simply special cases of a general process of preference aggregation of which we think most people have an intuitive conception—we

have all allocated project tasks with co-workers, divided household responsibilities among family members, or chosen restaurants with friends. These processes typically involve a de facto agenda-setter (the boss), voting weights (parents are more influential than their children), and even veto players (the friend with dietary restrictions). In our experiment, we have asked our subjects to engage a special case of a general process for which many of us have an innate understanding. Given this, it is entirely possible that people with no experience with coalition governance (like Americans), or even those who grew up in a non-democratic country, may approach this problem in similar ways. As such, we think a fascinating route for future research would be to investigate how people engage with *unfamiliar* (weighted) majority rule problems.

This discussion also points toward another general recommendation for future research: scholars should not rule out (without evidence) the possibility that the mass electorate has the ability to solve difficult problems. For decades, empirical voting scholars looked past the idea of coalition-directed voting, in part due to skepticism in voters' ability to solve the expectations problem, until Gschwend (2001, 2007), Kedar (2005), Duch and Stevenson (2008), and others began investigate how observable behavior conformed to models in which we presume that voters *can* solve this problem. What other phenomena may we be able to theorize and explain empirically if we simply do not begin by dismissing the typical person's ability to solve seemingly difficult political economic problems?

References

Armstrong, David A and Raymond M Duch. 2010. "Why can voters anticipate post-election coalition formation likelihoods?" *Electoral Studies* 29(3):308–315.

Atsusaka, Yuki. 2021. "A logical model for Predicting Minority representation: Application to redistricting and Voting Rights cases." *American Political Science Review* 115(4):1210–1225.

Austen-Smith, David and Jeffrey Banks. 1988. "Elections, Coalitions, and Legislative Outcomes." *American Political Science Review* 82(2):405–22.

Bäck, Hanna and Patrick Dumont. 2008. "Making the First Move: A Two-Stage Analysis of the Role of Formateurs in Parliamentary Government Formation." *Public Choice* 135(3-4):353–73.

Bankert, Alexa. 2021. "Negative and positive partisanship in the 2016 US presidential elections." *Political Behavior* 43(4):1467–1485.

Bargsted, Matias A and Orit Kedar. 2009. "Coalition-targeted Duvergerian voting: how expectations affect voter choice under proportional representation." *American Journal of Political Science* 53(2):307–323.

Bowler, Shaun, Gail McElroy and Stefan Müller. 2022. "Voter expectations of government formation in coalition systems: The importance of the information context." *European Journal of Political Research* 61(1):111–133.

Bowler, Shaun, Thomas Gschwend and Indridi H Indridason. 2020. "Coalition policy perceptions." *The Journal of Politics* 82(4):1458–1473.

Crisman-Cox, Casey. 2021. "Estimating substantive effects in binary outcome panel models: a comparison." *The Journal of Politics* 83(2):532–546.

Cutler, Josh, Scott De Marchi, Max Gallop, Florian M Hollenbach, Michael Laver and Matthias Orlowski. 2016. "Cabinet formation and portfolio distribution in European multiparty systems." *British Journal of Political Science* 46(1):31–43.

De Marchi, Scott and Michael Laver. 2020. "Government formation as logrolling in high-dimensional issue spaces." *The Journal of Politics* 82(2):543–558.

De Marchi, Scott and Michael Laver. 2023. *The governance cycle in parliamentary democracies*. Cambridge University Press.

Döring, Holger and Philip Manow. 2021. "Parliaments and governments database (ParlGov)." *An infrastructure for empirical information on parties, elections and governments in modern democracies. Version* 14.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. Harper and Row, New York.

Duch, Raymond M and Jean-Robert Tyran. 2012. Coalition context, voter heuristics and the coalition directed vote. In *Prepared for the Annual meeting of the European Political Science Association, Berlin, Germany*.

Duch, Raymond M, Jeff May and David A Armstrong. 2010. "Coalition-directed voting in multiparty democracies." *American Political Science Review* 104(04):698–719.

Duch, Raymond M. and Randolph T. Stevenson. 2008. *The Economic Vote: How Political and Economic Institutions Condition Election Results*. Cambridge: Cambridge University Press.

Duch, Raymond, Wojtek Przepiorka and Randolph Stevenson. 2015. "Responsibility attribution for collective decision makers." *American Journal of Political Science* 59(2):372–389.

Eberl, Jakob-Moritz and Carolina Plescia. 2018. "Coalitions in the news: How saliency and tone in news coverage influence voters' preferences and expectations about coalitions." *Electoral Studies* 55:30–39.

Fortunato, David, Nick CN Lin, Randolph T Stevenson and Mathias Wessel Tromborg. 2021. "Attributing policy influence under coalition governance." *American Political Science Review* 115(1):252–268.

Fortunato, David and Randolph T Stevenson. 2019. "Heuristics in context." *Political Science Research and Methods* 7(2):311–330.

Fortunato, David and Randolph T Stevenson. 2021. "Party government and political information." *Legislative Studies Quarterly* 46(2):251–295.

Fortunato, David, Randolph T Stevenson and Greg Vonnahme. 2016. "Context and political knowledge: Explaining cross-national variation in partisan left-right knowledge." *The Journal of Politics* 78(4):1211–1228.

Gigerenzer, Gerd. 2019. "Axiomatic rationality and ecological rationality." *Synthese* pp. 1–18.

Gigerenzer, Gerd and Daniel G Goldstein. 1996. "Reasoning the fast and frugal way: models of bounded rationality." *Psychological review* 103(4):650.

Gigerenzer, Gerd and Daniel G Goldstein. 1999. Betting on one good reason: The take the best heuristic. In *Simple heuristics that make us smart*. Oxford University Press pp. 75–95.

Gigerenzer, Gerd and Thomas Sturm. 2012. "How (far) can rationality be naturalized?" *Synthese* 187(1):243–268.

Glasgow, Garrett, Matt Golder and Sona N Golder. 2011. "Who 'wins'? Determining the party of the prime minister." *American Journal of Political Science* 55(4):937–954.

Gschwend, Thomas. 2001. *Strategic voting in mixed electoral systems*. State University of New York at Stony Brook.

Gschwend, Thomas. 2007. "Ticket-splitting and strategic voting under mixed electoral rules: Evidence from Germany." *European Journal of Political Research* 46(1):1–23.

Harsgor, Liran, Reut Itzkovitch-Malka and Or Tuttnauer. 2023. "Do coalition and formateur expectations affect vote switching?" *European Political Science Review* 15(1):96–115.

Hjermitslev, Ida B. 2024. "Punishing the pseudo-opposition: Accountability under a minority government." *European Journal of Political Research* 63(2):774–786.

Indridason, Indridi H. 2011. "Proportional representation, majoritarian legislatures, and coalitional voting." *American Journal of Political Science* 55(4):955–971.

Irwin, Galen A and Joop JM Van Holsteyn. 2012. "Strategic electoral considerations under proportional representation." *Electoral Studies* 31(1):184–191.

Kayser, Mark A, Matthias Orlowski and Jochen Rehmert. 2023. "Coalition inclusion probabilities: a party-strategic measure for predicting policy and politics." *Political Science Research and Methods* 11(2):328–346.

Kedar, Orit. 2005. "When moderate voters prefer extreme parties: Policy balancingin parliamentary elections." *American Political Science Review* 99(2):185–199.

Laver, Michael and Kenneth Benoit. 2015. "The basic arithmetic of legislative decisions." *American Journal of Political Science* 59(2):275–291.

Laver, Michael, Scott De Marchi and Hande Mutlu. 2011. "Negotiation in legislatures over government formation." *Public Choice* 147(3):285–304.

Lee, Seonghui, Agustina Haime and Randolph Stevenson. 2019. "Measuring knowledge of parties' legislative seat shares." *Political Science Research and Methods* 7(1):185–196.

Martin, Lanny W and Georg Vanberg. 2025. *Voter Sophistication and Coalition Politics: How Ideological Awareness Shapes Government Formation in Multiparty Democracies.* Oxford University Press.

Martin, Lanny W. and Randolph T. Stevenson. 2001. "Government Formation in Parliamentary Democracies." *American Journal of Political Science* 45(1):33–50.

Martin, Lanny W. and Randolph T. Stevenson. 2010. "The Conditional Impact of Incumbency on Government Formation in Parliamentary Democracies." Unpublished Manuscript, Rice University.

Meffert, Michael F, Sascha Huber, Thomas Gschwend and Franz Urban Pappi. 2011. "More than wishful thinking: Causes and consequences of voters' electoral expectations about parties and coalitions." *Electoral Studies* 30(4):804–815.

Müller, Wolfgang C and Kaare Strøm. 2003. *Coalition governments in western Europe.* Oxford University Press.

Powell, G. Bingham. 2000. *Elections as Instruments of Democracy: Majoritarian and Proportional Views.* New Haven: Yale University Press.

Provins, Tessa, Nathan W Monroe and David Fortunato. 2022. "Allocating costly influence in legislatures." *The Journal of Politics* 84(3):1697–1713.

Selten, Reinhard. 1991. "Properties of a measure of predictive success." *Mathematical social sciences* 21(2):153–167.

Strøm, Kaare. 1990. *Minority Government and Majority Rule.* Cambridge: Cambridge University Press.

Todd, Peter M and Gerd Ed Gigerenzer. 2012. *Ecological rationality: Intelligence in the world.* Oxford University Press.

Todd, Peter M and Gerd Gigerenzer. 2007. "Environments that make us smart: Ecological rationality." *Current directions in psychological science* 16(3):167–171.

Tromborg, Mathias Wessel, Randolph T Stevenson and David Fortunato. 2019. "Voters, responsibility attribution and support parties in parliamentary democracies." *British Journal of Political Science* 49(4):1591–1601.

Welz, Robert. 2023. "At least agree on the important things: the impact of issue distance, intracoalition heterogeneity, and salience on voters' coalition preferences." *Politische Vierteljahresschrift* 64(1):19–49.

APPENDIX: "Can (and do) Voters Use Simple Mental Models to Forecast Cabinet Composition?"

Contents

1 Survey information

The directed surveys were fielded in Denmark (March-April 2020), Germany (January 2020), and the Netherlands (July-August 2020) with participants recruited from Qualtrics' online panel, matching population demographics on age and gender (n=999 in DK, 1120 in DE, and 1004 in NL). Upon completing initial screening questions, participants were randomly assigned to a party system condition—which featured between five and nine parties—and presented with a series of hypothetical election outcomes characterized by party sizes and legislative positions. The hypothetical parties were described as Parties A, B, C, D, E, F, G, H, and I in Germany, and Parties G, L, Q, T, U, W, X, Y, and Z (letters that are not already used as real-world party abbreviations in Denmark, where parties have single-letter abbreviations) in the Danish and Dutch surveys. Respondents were asked to answer which cabinets they thought would form by indicating which party would play the role as party of prime minister, coalition partner, supporting party to the government, and opposition party. The default choice was set at opposition. Each respondent repeated the task for 12 different hypothetical scenarios (within the assigned party system) in DK and DE and 15 in NL. The study was approved by [author A's institution].

The institutional survey was administered by Respondi, who was contracted by [author B's former institution]. The survey was fielded in January and February 2020 from Respondi's online panel matching population demographics on age and gender in France, Germany, Hungary, Italy, Netherlands, Poland, Spain, Sweden, and the UK (n=2216, 2196, 2173, 2216, 2249, 2240, 2211, 2604, 2186, respectively). Upon completing initial screening questions, participants were presented with 2 sets of five party hypothetical election outcomes characterized by party sizes and legislative positions. The hypothetical parties were described as Parties A, B, C, D, and E. Respondents were asked to answer which cabinets they think will form by indicating which party would play a role as party of prime minister, coalition partner, supporting party to the government, and opposition party. The default choice was set at opposition. The survey falls under [author B's former institutional IRB] exemption.

Compliance with APSA Principles and Guidance for Human Subjects Research

Subjects retain their anonymity, consent to participate at their will, and are compensated for their time (by the research firm). They are informed of the name of the university administering the survey and there are no risks of harm or trauma imposed by the survey and no broader societal impacts of survey participation. No participants meet reasonable criteria for "vulnerable" or "powerful"—they are all normal residents of wealthy, developed, Central and Western European states. There is no deception in the survey. There are

no potential risks of noncompliance with state law or institutional code of conduct.

2 Review of literature on coalition expectations

In this section, we review the small empirical literature that tried to model voters coalition expectations. Our purpose in doing so is to both identify the main drivers of coalition expectations explored in previous studies and to explain how this literature motivates, in two different ways, the need for the kind of experimental studies we have done in this paper.

First, previous work, relying survey data on voters' expectations about real cabinet formations, has provided evidence that size and ideology, elite and media messages, and historical patterns of coalition formation are associated with coalition expectations, though the scope of the relevant evidence is fairly limited both geographically and temporally.[1] Further, though there is correlational evidence for each of these potential drivers of expectations, the largest estimated associations have been for size and ideological variables.

Second, while there is correlational evidence for all three of these potential drivers, the evidence for the influence of history is limited to unpublished work. Likewise, the evidence for the influence of media messages, while strong in correlational studies, has not proved robust in experimental work. Finally, there has been no previous experimental work on potential drivers of expectations other than media/elite messages has been attempted. This leaves a significant gap in our understanding of coalition expectations that, in our view, motivates the need for the kind of direct experimental tests of whether voters can (and do) use simple heuristic models of size and ideology to form their expectations that we provide in this paper.

The earliest empirical work of which we are aware that directly explored coalition expectations is reported in Duch and Stevenson (2008), who report the results of several original surveys of coalition expectations using observational survey data and found that beliefs about which party would secure the prime ministry and which parties would join the cabinet were correlated with variables such as party seat shares, ideological distance between the partners, and past incumbency. However, since the focus of the work was not on expectations, the evidence they provide, mostly from bivariate correlations, should be considered illustrative.

Similarly, Armstrong and Duch (2010) is frequently cited by theorists to justify the assumption that voters can use the historical record of coalition formation to assess coalition probabilities (i.e., they have

[1] While the literature has also focused on how individual-level variables like respondents' partisan attachments and political sophistication impact expectations, we are less concerned with these individual-level variables (though we do include them in some of our empirical models) and so focus instead on what the literature says about the kinds of cues (history, size, ideology, media messages) upon which voters might condition their expectations.

adaptive expectations). However, the study did not actually analyze voter expectations; it merely pointed out that historical patterns of formation exist and so could hypothetically be used by voters, leaving the empirical question of whether voters actually use history (or even know it) entirely open.[2]

Indeed, we are aware of no published work that has ever tested the proposition that voters' expectations about which cabinets are likely to form are correlated with (let alone caused by) historical patterns of cabinet formation—though, as we discuss below, an unpublished 2014 paper does so.

The first (published) full-fledged attempt to model coalition expectations is from Eberl and Plescia (2018). This study is an ambitious observational analysis of cabinet formation following the 2013 elections in German and Austria and is based on the German Longitudinal Study (GLES) and the Austrian National Election Study (AUTNES), both of which asked respondents about the perceived likelihood that a given coalition would form after the election, ranging from "certainly not" to "very certain", using an 11-point scale in Germany and a 4-point scale in Austria. Unfortunately, however, the surveys only solicited respondents' expectations about the likelihood of formation of three potential German cabinets and six Austrian ones and so the ability of the authors to examine how variation in the characteristics of potential cabinets might impact expectations is limited. Indeed, even variation in the media message about the likelihood of these different potential cabinets (the main focus of the paper) is obviously limited by the focus on only 9 potential cabinets. To solve this problem, the authors shift the focus from variation in potential coalitions to individual-level variation in media exposure to messages about the likelihood of each of the nine potential cabinets. They did this by measuring the media message about each cabinet from a variety of news sources, surveying respondents about which news sources they used, and creating an exposure-weighted measure of the media message about the likelihood of each of the 3 and 6 cabinets (respectively for German and Austria) "received" by each individual. They show that this measure has a significant impact on expectations. The

[2] Meffert and Gschwend (2010) are sometimes cited as an early study of coalition expectations, but this study is mainly concerned with identifying the extent to which Germans cast strategic votes of various kinds in the 2006 German Federal election. While in one small part of the study, the authors do measure survey respondents' expectations about coalitions (by asking respondents about the likelihood that a given coalition would have a majority to form a government after the election, using a 4-point scale ranging from "certainly not" over "rather uncertain" and "certain" to "very certain"), they did not attempt to model the sources of these expectations. That said, the they did show that in a variety of settings, voters' levels of certainty about whether a given coalition would reach a majority impacted their willingness to cast a strategic vote—thus providing evidence that voters' expectations about whether the coalition would win a majority were consequential for strategic voting. Another commonly cited study in this literature, but that is not really about expectation formation is Meffert et al. (2011). Instead, this paper is about the causes and consequences of voters' expectations about the share of votes/seats that parties will get after the election and whether various combinations of parties will win a majority. A central finding is that voters overestimate the majority chances of coalitions they personally prefer (wishful thinking).

size of this effect, however, is very small in Austria and minuscule in Germany. In Austria, an increase in the tone of media mentions about a potential coalition covering the full observed range of variation in media messages (i.e., from its lowest observed value to its highest) is only 0.23 points on a 4-point scale. For Germany this same impact (over the full range of variation in message tone) is only 0.065 points on an 11-point scale. These are tiny effects and are also illustrated by the fact that the change in the Adjusted R-squared between the regression model that includes the media message variables and those that do not was 0.001 in Germany and 0.054 in Austria.[3]

In addition to the media message, the authors try to account for expected coalition size in the model by including the average of the polls for each coalition. Of course, with so few potential coalitions included in the study, the only way they could do this was to interact this variable with an individual level variable attention to the campaign—so that it is really this individual variation in attention that is driving their results for this variable. This can be seen in their results for the interaction which show a large and significant impact for the main effect of campaign attention but minuscule estimates for the relevant interactions. Realistically, one simply cannot expect to estimate the impact of the expected size (poll standing) of potential coalition on coalition expectations when one only has variation over three (or six) cases.

Unfortunately for our work, which seeks to understand how coalition characteristics impact voters' expectations, it is difficult to know what lesson to take from the empirical work reported in this article. One might see the results, as the authors do, as evidence that the media message and expected coalition size matters to coalition expectations, but the small effect sizes could also be interpreted as evidence against the importance of those drivers. In our view, however, the fact that almost all the variation in the predictors in this article come from differences in voter attention and media use, make the estimated effects incomparable to the effects of differences in coalition characteristics or the distribution of the media message about coalitions estimated in other correlational studies in this literature (e.g., Bowler, McElroy and Müller 2022)

[3]The only indicators of either the governing history of each potential coalition or their other characteristics (like size and ideology) is a dummy variable for the incumbent potential cabinet and a measure of each potential cabinet's expected size. However, the fact that they had only 3 German and 6 Austrian potential cabinets to work with (analyzed separately by country) make it difficult to draw strong conclusions. First, the "incumbent dummy" is really just a marker for one of the potential cabinets in these two small sets of potential coalitions. As such, it captures everything about this one potential cabinet that could shift respondents' expectations—of which incumbency is only one of the possibilities. For example, the incumbent German cabinet in this study was composed of the CDU/CSU and FDP—two parties that certainly differed ideologically from the other two German potential cabinets included in the study (SPD-Greens and the CDU/CSU-SPD), but these ideological differences are absorbed into the incumbency dummy. Thus, the best way to interpret the estimated impact of this variable is that coalition characteristics (of some kind) matter in this study.

so not much should be made of their sizes.[4]

The two papers most relevant to our study are [reference removed, 2014] and Bowler, McElroy and Müller (2022), who both use conditional logit models to estimate the association between various measures of each potential cabinet's size and ideology, the media message about its formation chances, and (in the case of [reference removed 2014]) its historical familiarity.

Bowler, McElroy and Müller (2022) analyzed regional and federal German elections in conditional logit models that included measures of the media message about the likelihood various coalitions would form as well as two of their most important size and ideological characteristics (majority status and ideological spread). Their analyses of these data revealed that a coalition's expected size and ideological compatibility were both strongly associated with voters' expectation that it would form. Further, a coalition's size and ideological traits had a considerably larger combined impact on expectations than media volume and tone. For example, they find that a one standard deviation increase in the media message about a coalition's likelihood of forming is associated with a 3% increase in the chance an average respondent saying the coalition is the most likely to form.[5] Contrast that with majority status and ideological spread of the potential coalition, which they also included in their model. The impact of these variables, under the same scenario as the last paragraph, change the average respondent's expected probability of that coalition by +10% and -8% respectively.

Additional correlation evidence comes from a 2014 conference paper [reference removed] that examined prime-ministerial and partner selection in New Zealand (2008), Germany (2009), Norway (2009), and the Netherlands (2012). We include the study here because it is the only one, of which we are aware, that includes measures of coalition history, size and ideological variables, and the media message in the same model. These conditional and mixed logit models focused on which coalition of parties voters expected to form, given their choice of PM (which was separately modeled and found to depend almost entirely on which party the respondent thought would have the most seats). Their models included variables capturing each potential coalition's true history, size and ideological characteristics (and in some cases respondent perceptions of these variables) and for two of their countries, a measure of the media message (hand coded newspaper articles). Their results revealed large correlations, consistent across countries, between voters'

[4] One set of drivers, which is not our focus of our study, but for which this design is powerful are individual-level characteristics. Like other studies reviewed here, this study shows that measures of voter affinity for the parties in a potential coalition (measured here as their partisan predispositions) impact their expectations about those coalitions positively (i.e., evidence of "wishful thinking").

[5] Holding all values at their standardized means, dummies at 0, and accounting for the nonstandard manner in which they standardized the reported coefficient for this variable—i.e., dividing by 2 SDs instead of 1—see their Appendix page A6

expectations that a given cabinet would form and its historical familiarity, size and ideology, and media message favorability.

Putting this together with Bowler, McElroy and Müller (2022), and discounting the relevance of the small effect sizes reported in the Eberl and Plescia (2018), we can conclude that there is correlational evidence—from 7 elections in 5 counties, as well as 19 regional elections—that the size and ideological characteristics of potential cabinets, their historical familiarity, and the media message about their chance of formation all matter to coalition expectations—along with individual variables like affinity for a given party or cabinet.

These correlations are, of course, relevant to accessing the external validity of the experimental results in this paper. Specifically, unless one thinks these correlations are entirely spurious, voters in the real-world appear to use size and ideology in building their coalition expectations and may well put more weight on these variables than on history or the media message. But should we believe these correlations?

Is it possible that in the real world in which voters form their expectations, the large correlations between the size and ideological characteristics of coalitions and voters' coalition expectations is spurious? Logically, it is certainly possible. Perhaps, for example, voters know historical patterns of coalition formation that are themselves determined (in part) by the real size and ideological characteristics of potential coalitions (as the literature on real world coalition formation tells us they are). If they use those historical patterns to condition their expectations, then they will appear to choose winning, ideologically compact coalitions—even if they do not even know those characteristics, let alone use them in a mental model to forecast coalition probabilities. Thus, it is entirely possible that the observed associations between size and ideological characteristics of coalitions are non-causal—-and so, at this stage in the development of this literature, it is important to investigate that possibility.

Usefully, exactly the kind of effort that is needed to sort out the causal status of size and ideological variables on expectations has already begun for another potential driver of expectations: media and elite messages. As described above, the correlational evidence has found a robust positive association between media/elite messages about the probability a given coalition will form and voters' coalition expectations. However, there are obvious ways this association could be spurious. If the media tends to emphasize the likelihood of historically frequent combinations and/or ones with favorable size and ideological characteristics, then voters who use the kind of mental model we describe (or who rely on history) will tend to expect exactly those coalitions that the media labels as probable—even if those voters do not monitor the political news closely or otherwise discount media messages in their expectations.

Given this causal indeterminacy, scholars interested in elite/media signals have turned to exactly the kind of survey experiments that we report in this paper (specifically, our ancillary instrument that uses real

parties).

The most probative of these studies is by Bahnsen, Gschwend and Stoetzer (2020), a survey experiment conducted during Sweden's 2018 general election, which randomizes respondents to receive very direct signals about which coalitions party elites and the media thought would form—for example a typical treatment stated, "Several political observers agree that, given the statements and signals sent by [signaling party], there is a high probability of this party joining a coalition government with [signaled coalition partners] after the election." The authors used well powered ($\sim$ 2000 respondents) causal mediation analysis to determine if the effect of the manipulated coalition signals on vote choice was indeed "mediated" by a change in voters' expectations. Most importantly for us, as a part of the mediation analysis, they estimated the direct impact of their coalition signal treatment on coalition expectations.

Of the eight hypothesized effects from the four treatments (the treatments were the statement above with four different coalition descriptions filled in) only two were significant at the 95% level (despite the well powered treatments arms) and, more importantly, the effect sizes were miniscule, especially for such a direct treatment message. On their 1-7 scale of coalition likelihood ratings (the DV), with empirical standard deviations for the DVs ranging from 1.5 to 1.7, we get effects strictly smaller than 0.3, with an average closer to 0.1—which is approximately 1/15 of the sample standard deviation of the DV.

In their conclusion the authors speculate why they failed to find strong support for the impact of elite/media messages, even with these very direct treatment messages, arguing that if a coalition signal is not at least somewhat consistent with the realities of coalition histories and/or the size and ideological combinations that tend to form, then "realistically it will not push citizens perception around" and as such they advise that parties should only use such signals "to activate existing positive prior attitudes about [the likelihood of] coalitions of parties..." In other words, such treatments are only likely to "work" when they reinforce the conclusions a voter can draw from cabinet characteristics and histories.

Another experimental study is Falcó-Gimeno and Muñoz (2017), which was conducted in the run up to a 2015 regional election and cabinet formation in Valencia, Spain. While the main focus of this paper was not expectations, in an appendix to the paper the authors present evidence about how their treatments impact coalition expectations. The treatments, like in Bahnsen, Gschwend and Stoetzer (2020), were very direct: While some respondents received a placebo, others were directly told that party leaders and informed political observers believed a specific cabinet would form, they were then asked them asked which cabinet they thought would form.

The results of the experiment were generally positive (with an average effect of treatment on the share of respondents choosing the treated coalition 13%) but varied a lot depending on which specific party coalitions

were treated. Specifically, they found large effects for messages about the new liberal party, Ciudadanos, forming a right-leaning coalition (+24%), but more modest effects for announcements that it would likely form a left-leaning coalition (+6%). Likewise, views of which coalition the more established PSOE party would likely form were about +11% and +9% for right and left-leaning coalitions respectively.

In our view, we should view these results as generally supportive of the impact of coalition signals (though given the directness of the treatment, they seem modest) and so they provide a contrast to the Bahnsen, Gschwend and Stoetzer (2020) results. However, the fact that they are not larger and the unevenness of the effects across parties, also support speculation by Bahnsen, Gschwend and Stoetzer (2020) that respondents appear to resist even somewhat heavy-handed coalition signals when those signals do not comport with their pre-existing knowledge of the parties involved (and their relationships).[6] Indeed, in discussing the pattern of results, the authors speculate that the disparity in the impact of these coalition signals on voter expectations for different messages is likely due to the fact that voters hold stronger priors regarding the commitments of older, established parties, making their coalition expectations slightly more resistant to campaign signals that run counter to those beliefs, than those of new parties with less established track records.

Overall, we think a fair assessment of the experimental evidence on the causal impact of elite/media messages on coalition expectations suggests this impact is varied, often modest, and it is certainly smaller than the correlational evidence would suggest.

For the purpose of thinking about the importance of the experiments in our paper, there is, of course, a direct analogy to these experiments on elite/media signals and several important lessons we can take from them.

First, the mostly negative results in Bahnsen, Gschwend and Stoetzer (2020) remind us that the results of the experiments reported in our paper *could have been negative!* And, if they had been—showing that even in the rather pristine setting of the lab, with other influences carefully controlled, few respondents use size and ideology to predict cabinet compositions—then it would have cast considerable doubt on the proposition that voters use them in the real world. Indeed, had that been our finding, then we would be

[6]The bigger effects Falcó-Gimeno and Muñoz (2017) found for the regional cabinet formation they studied compared to the largely negative results for Swedish national cabinet formation from Bahnsen, Gschwend and Stoetzer (2020) may result from differences in the availability of alternative sources of information in these very different contexts. If respondent knowledge about the sizes and ideologies and histories of potential cabinets was lower in the regional context (as seems likely), then delivering a very clear coalition signal (and the signals in this experiment were, like those in Bahnsen, Gschwend and Stoetzer 2020, very direct) into a low-information environment, will surely show bigger effects than in situations in which the respondent has more of the other kinds of information (e.g., relevant histories and characteristics of potential cabinets) that might compete with those signals.

forced to accept the possibility that the large observed empirical associations between size and ideology and expectations found in the literature were not due to the kind of mental model we have proposed but instead reflect spurious correlations with the media message or (more likely given the experimental results above) coalition histories.

Second, the negative and mixed results of these elite/media message experiments demonstrate the need for the kinds of experiments on coalition characteristics that we administer in this paper. The reason is that—as both Falcó-Gimeno and Muñoz (2017) and Bahnsen, Gschwend and Stoetzer (2020) speculate—the negative and mixed results of these experiments may well be due to respondents resisting the implications of those messages *because* they find them incompatible with the sizes and ideologies of the parties involved. That is unlikely to be the correct interpretation of the negative results for these coalition signaling experiments, if experiments like ours had come up negative.

Of course, the results of our experiments are not negative, or even mixed to any great degree. That is, when we isolate the impacts of size and ideology from history and elite signals, those impacts, *unlike* those of media and elite signals, remain strong. Further, the substantial impact of our size and ideological manipulations are robust across national contexts, survey environments (e.g., a large cross-national omnibus and more targeted in-depth country specific surveys), numbers of parties included in the experiments (from 5 to 9), different model specifications, different concepts/measures of what voters might mean by size and ideology, and whether we use real or fictional parties in the experiments.

Thus, our experimental results, taken together with the experiments on manipulated media messages, should increase our confidence that the uncontrolled correlations between media messages and expectations were partially driven by the fact that, in the real world, such messages are at least partially determined by coalition characteristics like size and ideology (and perhaps history, though we have no experimental verification for that).

As such, our experiments, understood in the context of both the correlational evidence and the experimental evidence on elite signals, paint a picture of a literature that is slowly zeroing-in on the causal structure driving coalition expectations (for some reasonable share of the population) that gives more weight to coalition characteristics than elite signals (with the impact of history not yet clarified).[7] [8]

[7] Barnfield et al. (2025) is the most recent attempt to model coalition expectations (using data from the 2021 German election). However, the authors only examine individual-level characteristics (affective attachments, age, gender) without examining any historical, media, or coalition-level factors.

[8] Another great illustration of importance of this kind of dialog between experimental and observational studies in this subject area comes from Gschwend, Meffert and Stoetzer (2017). While the paper only estimates the impact of coalition signals on the vote and not on expectations, is nevertheless useful for us because it parallels quite closely what we are trying to do in this

A final takeaway from the existing literature concerns why we think it is important for experiments manipulating size and ideology to do so for fictional parties as well as real ones. One thing the authors of the media message experiments were keen to point out was the necessity, when using real parties, of constructing messages that were contextually plausible so that respondents would not dismiss them out of hand.

From a causal inference perspective, this just means that if we want to isolate treatment in an experiment with real world parties, we have to assume that respondents do not disbelieve our treatment because they have other information about the value of that treatment.[9] If they do, our treatment is no longer exogenous and we lose the main advantage of the experimental design. As such, in real world experiments, we need to choose plausible treatments so this assumption will hold. Applied to our experiments, this means that when we use real world parties, we cannot create any party system we like. We have to make only relatively small, plausible changes from true sizes and positions of those parties. That is why in constructing our real-world experiments, we made only relatively small changes. This, however, comes at a price.

If we make only plausible, relatively small, changes to the party systems as they existed at the time of our survey, our treatment configurations of seats and ideological positions will necessarily look a lot like those in the real world.[10] The price we pay for this is that it necessarily increases the chances that the coalitions our theory predicts voters will expect (based on size and ideology) in the various treatment conditions will be similar to the ones one would predict if they were instead using real world media messages or coalition histories. Clever choices about the treatments can mitigate this problem,[11] but the fact that

paper. In their case, they observed that the observational literature on coalition signals left open the possibility that previous correlations between coalition signals and the vote were spurious and so experimental testing of that relationship was needed even if the experimental manipulations required the use of hypothetical scenarios that only mimic the real world. What we do that they did not (because they only used real parties in their experiments) was to definitively block all backdoor paths in the experiments (i.e., they asked respondents how they would vote in a randomly assigned hypothetical scenario in which, for example, the "the Greens would clearly reject a coalition with the ÖVP and announce the intention to form a coalition with the SPÖ?") thus allowing the possibility that respondents who understand that the Green-SPÖ combination is likely on ideological grounds vote for it for that reason and not due to the signal.

[9] It important not to confuse this with the point made above about subjects resisting a treatment because it contradicts other possible influences on expectations that they weight more heavily. In that case, they can accept the treatment (an elite signal about a coalition likelihood) at face value but because they know things about other aspects of the real-world situation, they down weight that treatment in favor of those other things. That is not what we are talking about here. In this case, we mean that they disbelieve the treatment itself. In the signaling experiments, this is akin to the respondent saying to themselves "well, they would never do that!"

[10] Indeed, we included a baseline array of parties' seats and sizes which depicted the status-quo, real-world party system at the time of our survey and then created out other treatments by making modest changes to that.

[11] For example, in one of our treatments in the German experiment using real parties, we change the size of the AfD vs its

using real world parties both allows respondents to bring other influences (like coalition histories) into the experimental setting while, at the same time, limiting our ability to separate controlled variation in our treatments from uncontrolled variation in these other influences is an important reason to do both kinds of experiments.

Only with evidence from abstract experiments can we be confident that voters are responding to our treatments and not external information that happens to lead to the same behavior.

3 On multidimensionality in expectation formation

Both our ecologically rational heuristic model of expectations formation and our experimental tests of its implications adopt a unidimensional conception of parties' left-right relationships. This is in keeping with a great deal of theoretical and empirical literature on coalition formation, but not the most recent work (De Marchi and Laver 2020, 2023), which argues that real-world bargaining over the partisan composition of cabinets takes place in high-dimensional spaces by politicians that have no need to reduce this space to low dimensional summaries. Further, that work shows that party positions in this multi-dimensional policy space are consequential for which cabinets form. Given this, why did we build a theoretical model of voters' coalition expectations, as well as design our experiments, using a unidimensional representation of parties' relationships to one another?

First, compared to a multi-dimensional conception of parties' relative ideological positions, the unidimensional conception is clearly consistent with the overall idea that voters, with limited time and attention, utilize heuristics that are "good enough" to provide satisfying inferences about the political world. Further, even a cursory reading of the vast literature in comparative political behavior leaves little doubt that the "left-right" heuristic (an inference about which party is "left" or "right" of another—and by how much—in some abstract, unidimensional space) is, along with partisan labels, the most common and important heuristic voters use to make sense of the political world.[12]

status quo size in ways that would make a CDU/CSU-AfD coalition more expected under our theoretical model. Since it is unlikely that either the real-world media messages available to our respondents or histories of coalition formation would also predict this coalition, we can be confident that our finding that respondents do increase the likelihood of this coalition (by about 10 percentage points), it is due to our manipulation.

[12]For example, one recent demonstration of the importance of this kind of ideological shortcut comes from Lee, Stevenson and Cozza (2025), who argue (and show) that systematic variation in average levels of political interest across countries is driven by corresponding differences in the availability and usefulness of partisan and (unidimensional left-right) ideological cues. Put simply, voters in countries where ideological and partisan cues are more helpful in understanding politics are able to better cope with the political world, which causes the emotion of interest to "fire" more often in response to political messages, ultimately

Second, and most importantly, our choice to rely on a unidimensional ("left-right") space to represent how voters think about the relationships between parties reflects the modern understanding of what the left-right actually means to voters (at least in the Western democracies we study). This modern understanding rejects the traditional view that most voters understand and use the relative left-right positions of parties exclusively (or even primarily) as an aggregate policy dimension—that is, as a reduction or summary of a multi-dimensional policy space. See Fortunato, Stevenson and Vonnahme (2016) for an explicit text of the hypothesis that voters gather left-right information on parties as a summary device for their multidimensional policy stands (they find no evidence that this is the case).

Instead, modern work exploring what the left-right means to voters shows that it is not used exclusively as a policy aggregator, but rather as a more general way to keep track of a variety of different, but strongly overlapping, relationships between parties.[13] For example, in a set of seven different survey experiments across 4 countries, Lee, Santoso and Stevenson (In press) asked respondents to place a sequence of hypothetical parties (described by a wide variety of policy positions, leadership characteristics, patterns of cooperation with other parties, patterns of social group support, and more) on an 11-point "left-right" scale. What they found was that party policy positions, while certainly important to voters' left-right images of parties, were not their most important drivers for any voters. Instead, for the least sophisticated 2/3 of voters, their left-right images of parties responded most strongly to previous patterns of partisan cooperation and conflict, followed by parties' very broad values, patterns of social group support for the parties, and characteristics of leaders (e.g., their gender).[14] For more sophisticated voters (the top 1/3) there is a larger impact of policy but, even for these, other non-policy variables are as important. Critical to this paper is the finding that voters' left-right images of parties are strongly influenced by patterns of partisan cooperation and conflict—that is, voters think that parties who frequently cooperate are close on the left-right.[15] This result is consistent with a mountain of recent work that has explored how voters update their left-right images of parties in response to parties governing together (Fortunato and Stevenson 2013;

causing individuals in such contexts to develop life-long political interest and knowledge more readily than in countries in which these cues are less helpful.

[13]Our focus is on parties, but voters can clearly place other political entities (other voters, groups, or themselves) in this space.

[14]Note also that even the terminology adopted in this literature (i.e., voters' "left-right images of parties" rather than "left-right positions" reflects this broader conception of how that voters understand and use the unidimensional partisan space called "left-right."

[15]The ability to randomize policy and cooperation cues experimentally (as Lee, Santoso and Stevenson In press do) is essential in establishing this result since in an observational design, party cooperation and policy agreement between the parties would be hopelessly conflated.

Fortunato 2019), making public statements about each other Weschle (2018); Adams, Weschle and Wlezien (2021), or otherwise cooperating (Santoso, Stevenson and Weschle 2024).

Other work that has identified similar, non-policy, influences on the left-right images of parties includes O'Brien (2019), Fernandez-Vazquez and Somer-Topcu (2019), Fortunato and Adams (2015), Zechmeister (2006), and Zechmeister and Corral (2013).[16] Indeed, while much of this literature developed in response to findings by Adams, Ezrow and Somer-Topcu (2011) that voters' left-right images of parties do not track party policy positions, one can find echoes of this message much earlier in (mostly comparative work) on political behavior. Indeed, perhaps the most forceful statement (and data-rich demonstration) of this perspective is the Arian and Shamir (1983) conclusion that the left-right is not only about partisan policy differences, but is used more generally by voters, "to label and to identify the good or the bad, the right and the wrong, the desirable and the despicable." (142) As Fortunato, Stevenson and Vonnahme (2016), summarize this earlier work: "quite apart from any specific policy content, when a party is on the "left" or "right," this tells the voter which other parties should be considered allies and which enemies" (1213).

Third, much of the work cited above thinks of the voter's left-right images of parties—which are by definition images arranged in a unidimensional space—as a heuristic inference that depends on both policy and non-policy cues.[17] Thus, the heuristic rules that map cues to inferences necessarily aggregate these cues into a unidimensional summary. For example, Lee, Santoso and Stevenson (In press) posit that these images are linear functions of the cues (and the cue weights) they identify. The weights, of course, are not the output of some data reduction procedure like the weights from a factor analysis, but instead reflect the long-term empirical correlations between each cue and the target of heuristic inference (i.e., the true left-right positions of the parties—see Lee, Santoso and Stevenson (In press) for a discussion of the concept of a parties "true" left-right position).

Thus, in our view, the theory of ecologically rational heuristics, as applied to the voters' left-right images of parties, aggregates different policies, along with non-policy information, into a single "left-right" dimension. Importantly, the voter does not construct such an image on the fly when we ask them about

[16] The empirical literature on the drivers of left-right self-placements has consistently found that factors other than policy play a prominent role in these placements. For example, see Inglehart and Klingemann (1976); Huber (1989); Knutsen (1995, 1997); Freire (2006*b*,*a*); Fortunato, Stevenson and Vonnahme (2016); Vegetti and Širinić (2019).

[17] It's worth emphasizing that when people talk about a party being left or right of another, this is a unidimensional comparison. One can correctly argue that bargaining over cabinets occurs over multi-dimensional policy spaces, but that does not make the voter's left-right images of parties multi-dimensional—these are inherently, conceptually, and grammatically unidimensional. The relevant question is whether these unidimensional party images depend only on policy (so can be reasonably thought of as arising as some kind of reduction or aggregation of a multi-dimensional policy space) or on other relevant kinds of (non-policy) relationships between parties.

which coalitions will form. Instead, the left-right images of parties are such an important aid in navigating politics that voters likely keep track of these images as they do other useful concepts like the state of the economy, the popularity of parties, and how much they like the current PM. Thus, voters carry these images with them—ready to use to make sense of a political messages, to have a political conversation, or to use as an input to other heuristics (like the one we propose for coalition expectations).[18]

Given this, voters likely do ultimately use multidimensional policy information in forming their coalition expectations, the impact of these policies runs through voters' left-right images of the parties, which pre-package a lot of relational information about the parties, including how they relate on many different policies, into a form that busy voters can quite easily apply to aid in their understanding of complex political reality (Benoit and Laver 2012).

Applying this understanding to the research design in this paper, we present our respondents with a description of a party system that includes the seat shares of each party and their positions on a "left-right" dimension which is otherwise unlabelled. We do not call this a "policy" dimension. Instead, we invite respondents to bring to the exercise whatever they think it means for one party to be on the left or right of another. Given the research cited above, these differences likely to reflect more than just policy for almost all voters—including their beliefs about the extent to which parties are allies or enemies—which, on its face, must be relevant for their expectations about which of these parties will coalesce in the future.

4 Bargaining power

Many political economic models of aggregate choice suggest (or show formally) that "bargaining power" is a critical determinant of parties' ability to get what they want. This influence over negotiations, or collective choice, stems from one of two features of the bargaining environment: parties' centrality (in one or more dimensions, though we assume just one here following the discussion above) or voting weight. Substantial theoretical research implies that holding the positional median, or being more generally central, should deliver outsized influence to a party, allowing it coax outcomes closer to its ideal point than it would otherwise be able without this positional advantage (e.g., Baron 1991; Krehbiel 1998; Morelli 1999). The depth and rigor of the theoretical research on centrality is sufficient to lead Laver and Schofield (1990, 111) to famously conclude that,

[18]This understanding dovetails nicely with work in political psychology (e.g., Lodge and Taber 2013) that suggest individuals maintain on-line tallies that "store" their opinions about various concepts they need in daily life (like how much they like chocolate or the PM).

> "It makes no difference if [the median party] goes off on holiday to Bermuda and sits on the beach getting a suntan... its policies should be enacted whatever it does."

The empirical research, however, is less clear. Martin and Vanberg (2014) present a recent summary of this research and also find in their own analysis that the median legislator in the assembly is less influential over policy outcomes than the preceding theoretical research predicts. This weaker-than-expeccted influence was also subsequently found in a large, multi-country study of parties' ability to deliver on their manifesto pledges by Naurin, Royed and Thomson (2019), who wrote that,

> "[o]n balance, we conclude that proximity to the median legislator has a small effect on the likelihood of pledge fulfillment at most and that this effect is mediated for the most part by other variables: the likelihood of holding executive office and party size. Being the median legislative party appears to increase the likelihood of fulfillment for opposition parties' pledges but not for governing parties' pledges. Moreover, holding the median legislator is less important than holding executive office for pledge fulfillment" (72).

Importantly, the focus of this research is policymaking, not government formation per se, which is of course our focus. While there are there are model specifications in which median inclusion has only weak or uncertain power in predicting while cabinets form (e.g., Martin and Stevenson 2001, 2010), empirical research on government formation generally finds that coalitions including the party of the median legislator are significantly more likely to form (Kayser, Orlowski and Rehmert 2023). As such, the empirical regularity that potential coalitions including the median party or more central parties are more likely (or simply more commonly observed, neglecting the selection phase), may indeed inform an ecologically rational model of coalition expectations, so long as the association is understood by voters, and, offers low-cost predictive power in excess of what is provided by cues that may be some combination of more easily understood, more predictive, or less costly.

Previous research casts doubt on the usefulness of centrality in a related task, policy influence attribution Fortunato et al. (2021), on the basis that it is 1) costly and 2) less useful (predictive) than other available cues for that task. We agree that median status is more costly and less predictive than other available cues (especially for the specific task of inferring which will lead the coalition as prime minister), but do not derive explicit hypotheses about centrality. While we believe that an efficient model of coalition expectations can be formed without understanding or consideration of median status, it may be the case that this empirical regularity is subconsciously absorbed such that cabinet near the middle of the ideological space may simply look more "right" or more familiar than cabinets to the extreme or leapfrogging the median. In this event,

when the respondent's model presents the opportunity to choose between a coalition that looks more familiar or less familiar—meaning that the model does not discriminate to a single, unique alternative, but instead discriminates to more than one alternative requiring the respondent to choose amongst alternatives across which the model is indifferent—a general tendency toward centrality could yield a positive and significant correlation between median inclusion and coalition expectations, even without median or centrality being part of our respondents' core expectations model.

What about voting weights? Most of the literature here ignores the location of parties and instead derives their bargaining power by examining the proportion of potential coalitions to which they can lend pivotal support. That is, how many minimal winning coalitions is the focal party a part of and how many minimal winning coalitions are there in total? Minimal integer weights, the Banzhaf power index, and the Shapely Shubik index are variants of this calculation, and bargaining power has recently been shown to matter in coalition bargaining—parties with larger minimum integer weights, holding their raw seat share constant, are much more likely to enter government and perhaps marginally more likely to obtain more ministerial portfolios, given their cabinet entrance (Cutler et al. 2016). Given this apparent association, it may be the case that voters would find such a cue to be predictive of coalition formation. However, we find it implausible that voters would actually utilize such a cue for the simple reason that, even if the underlying concept of bargaining power, as understood by formal theorists was widely understood by voters (and there is no evidence that this is the case), the calculation of such weights is almost certainly beyond the capacity of nearly all voters. As Cutler et al. (2016) write, "this is because the problem of coalition enumeration is NP-hard, which of course generates problems for real politicians as well as for political scientists" and of course therefore voters as well. Because the costs are exorbitant, and because Fortunato et al. (2021) find no evidence that voting weight cues are understood by voters or applied in their heuristic models of responsibility attribution, we do not consider the application of such voting weight cues in our analysis.

5 Comparing across compatibility and winning definitions

Having described the correlates of our subjects' coalition expectations, we can conclude that some significant portion of subjects are employing a mental model to generate their expectations and that the models most commonly employed are likely simpler than the VaPS model. In this section we assess the extent to which expectations are consistent with the specific model we have proposed above. We define consistent behavior as subjects expecting the coalition that our heuristic model would predict. That coalition must include the largest party as PM, which is unambiguous, but must also include ideologically compatible partners and be winning—two characteristics that can be measured in many different ways. Unfortunately, to our knowledge, there is no research on how voters think about what it means for two or more parties to be more or less ideologically compatible or for coalitions to be winning. As such, we evaluate eight different measures of compatibility and four measures of winning and let the data speak for themselves. Our compatibility measures are:

1. *Connected*: parties in coalition are contiguous
2. *Coalition span*: total range of the coalition
3. *Average mean distance*: average of coalition parties' distances to unweighted coalition mean
4. *Average weighted mean distance*: average of coalition parties' distances to seat-weighted coalition mean
5. *Average PM distance*: average of partners' distances to PM
6. *Average weighted PM distance*: average of partners' seat-weighted distances to PM
7. *Max PM distance*: the maximum partner distance to PM
8. *Total PM distance*: the sum of partner distances to PM

All of these definitions are correlated, but may also imply different formation processes. For example, minimizing distance to the prime minister implies the PM is central to government formation, whereas minimizing total dissimilarity does not imply that any one party is necessarily more important to the formation process. We see definitions 1-3 as more holistic, while 4 is ambiguous, and 5-8 are clearly PM-driven. If PM-centered definitions fit the data better than holistic measures, we may conclude that subjects think of cabinet formation as driven by the PM.

There are two more important ways these measure differ. The first is complexity. Some measures are more difficult to calculate (or intuit) than others. For example, *average weighted PM distance* is substantially

more difficult to calculate than *total PM distance*. The second is variation which enables discrimination. There will be fewer "ties" when evaluating *average weighted PM distance* than *total PM distance*, as the seat weights induce variation discriminating between close sets. As such, we will want to consider how many predictions our heuristic model makes under different ideological compatibility measures when we compare them below.

What does it mean for a coalition to be "winning?" Political scientists categorize coalition types as minority, minimal-winning, or surplus majority in disciplined legislatures (Laver and Shepsle 1990), as well as minimum-safe in undisciplined legislatures (Provins, Monroe and Fortunato 2022). It is not clear that these concepts are understood by voters. To the extent that we can glean hints from research that indirectly touches upon this question, the signals are mixed. On the one hand, Proksch et al. (2024) show that voters prefer majority to minority cabinets, all else equal, when they are *specifically told* one alternative has a majority and the other does not. On the other hand, experimental research on responsibility attribution suggests that voters do not necessarily differentiate between a majority and a large plurality (Duch, Przepiorka and Stevenson 2015). As for surplus majorities, the closest analogue we could find is research on preferences for consensus decision rules relative to simple majority, which imply people tend to prefer the former (Hare 1980), though this does not allow us to infer whether and how voters may understand the difference between minimal-winning and surplus majorities, particularly in cases where the aggregate seat shares of some minimal-winning majorities may be *larger* than some surplus majorities. We evaluate the following measures of winning:

1. Majority inclusive of ties
2. Majority
3. Minimal-winning inclusive of ties
4. Minimal-winning

Analyzing all four of these definitions allows us to evaluate model fit using the winning criterion that theoretical research tells us *should* be the desired outcome in government formation—*minimal-winning*—against less strict criteria that may be more intuitive for voters or simply used more frequently. Namely, we believe *majority*—inclusive of surplus majorities—is the simplest and most intuitive definition for our subjects. However, it is also possible that many subjects do not understand, or have not intuited, that precisely 50% is not a majority, so we also want to evaluate winning criteria that are inclusive of ties (50%). Interestingly, model fit is less sensitive to changes in the winning criterion under some ideological

compatibility measures than others, as some compatibility criteria narrow many heuristic model predictions to minimum-winning.

We fit the data by first identifying all potential party constellations that may form, differentiating among cabinets that have the same party membership but a different party designated as PM. This yields a total of 80, 192, 448, 1,024, 2,304 potential outcomes per episode for systems of five, six, seven, eight, and ninie parties, respectively. Second, we identify all proto-coalitions satisfying the target winning criterion in which the plurality party serves as PM. Of these alternatives, those that *minimize* the dissimilarity criterion—i.e., are the most compatible—are our heuristic model predictions. With the more permissive winning criterion (*majority inclusive of ties*), only *connectedness* routinely (79% of the time) produces surplus majority predictions. For all other ideological compatibility measures, coalitions that minimize the dissimilarity criterion tend toward minimal-winning. *Average mean distance* and *max PM distance* are the most inclusive of surplus majority predictions after *connectedness* and make them 26% and 34% of the time, respectively. Conversely, over 95% of predictions made by the *average weighted PM distance* measure are minimum-winning (or exactly 50% if we define winning as inclusive of ties) and all *total PM distance* predictions are minimum-winning (or exactly 50%).

To assess how well the heuristic model fits our subjects' coalition expectations, we calculate two fit statistics. First, we calculate the portion of respondent expectations that are consistent with a heuristic model prediction under each ideological compatibility and winning definition. We call this *correct classification* and it captures gross fit, or, the number of times the respondent expectation (y) was also a heuristic model prediction ($\hat{y}$), divided by the total number of respondent expectations ($E[\hat{y}|y]$). Second, we calculate the portion of heuristic model predictions that are consistent with respondent expectations under all measures. Martin and Stevenson (2001) call this *predictive efficiency* and we calculate it here as the number of times the respondent expectation was also a heuristic model prediction, divided by the total number of heuristic model predictions ($E[y|\hat{y}]$). As such, *correct classification* "rewards" a measure for being inexact, or making many predictions per episode, while *predictive efficiency* "punishes" a measure for being inexact, or making many predictions. We give these fit statistics, along with the average number of heuristic predictions made by each compatibility measure per episode in Table **??**, using majority inclusive of ties as the winning definition, in order to emphasize fit differences across the compatibility measures (we plot the estimates for all winning criteria below). The table groups the compatibility criteria by measure focus (PM-centric or holistic), and rank-orders them by *predictive efficiency*. The figures at the point of the braces represent the aggregation of all compatibility measures within the braces (*average mean weighted distance* is aggregated in both PM focal and holistic groups).

We begin with number of predictions. Of all ideological compatibility criteria, *connectedness* makes by far the most predictions per episode (8.508), resulting in very high correct classification, but very low efficiency. *Average weighted PM distance* (1.063), *average weighted mean distance* (1.135), and *total PM distance* (1.201) make the fewest predictions and discriminate to a single alternative in well over 80% of episodes. *Connectedness*, by contrast, *never* discriminates to a single alternative under this permissive winning criterion. In general, the compatibility measures that make fewer predictions are in one way or another sensitive to the addition of more parties, or, increase variance through a weighting scheme. For the former, compare *connectedness*, which does not change if the coalition is two parties or three parties (they are either connected or not), to *total PM distance*, which is definitionally increasing with the addition of member parties. For the latter, compare *average PM distance* to *average weighted PM distance*, which are constructed from the same distances, but of course one is aggregated according to the seat shares associated with those distances and is therefore much more variable.

Connectedness classifies more subject expectations than any other measure (0.361), and, in general, measures that make many predictions classify better. Aggregating across measures of ideological compatibility, our heuristic model classifies 40% of subject expectations. But even focusing on the more exacting measures, like *total PM distance* (0.231) and *average weighted mean distance* (0.202), the heuristic model classifies more than 20% of subject expectations, which we think is very impressive given the massive choice set (444 alternatives per choice averaging across the sample). For example, *total PM distance* classifies at over 2750% the rate of chance (0.008).

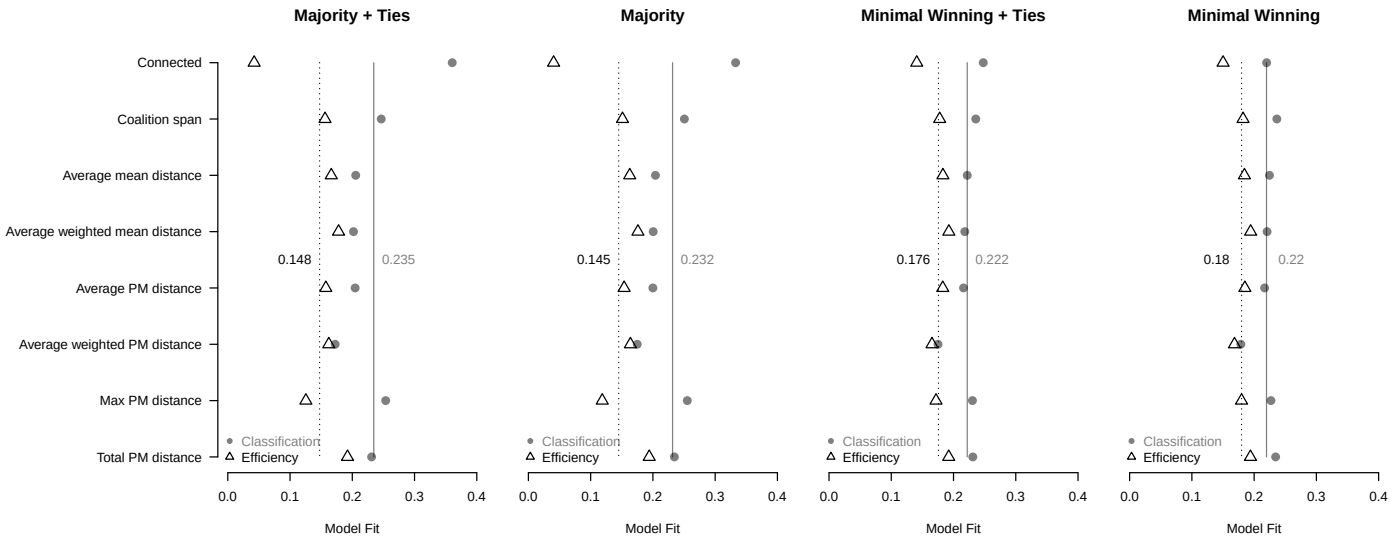
Finally, consider *predictive efficiency*. Of course, *connectedness* is by far the least efficient measure. *Total PM distance* is the most efficient measure (0.192), followed by *average weighted mean distance* (0.178), and *average mean distance* (0.166). *Total PM distance* clearly makes the PM focal to the choice and *average weighted mean distance* gives the PM primacy by assigning more influence to larger parties where, definitionally, the PM, as plurality, must bear the most weight, but *average mean distance* is a fairly holistic measure. As a result, the data are a bit ambiguous as to whether our subjects see the PM as really driving the formation process under this very permissive winning criterion.[19]

Recall that all Table **??** values were calculated using *majority inclusive of ties* as the winning criterion. Figure 1 compares the fit of our heuristic model across all eight compatibility criteria and all four winning criteria to show how more restrictive winning criteria influence the overall fit of the heuristic model. The gray

[19]It is worth noting, however, that the most efficient measure, *total PM distance*, perfectly captures the iterative process recorded in Footnote 11, in which respondents first select the plurality party as PM and then add nearest neighbor parties to the coalition until it reaches a winning seat share.

circles show correct classification and the hollow black triangles show predictive efficiency, while the vertical lines show the average of each across compatibility criteria. Under most winning criteria, *connectedness* has the greatest rate of correct classification. *Total PM distance* has the greatest predictive efficiency under the two more permissive winning criteria and *total PM distance* and *average weighted mean distance* are about equally efficient under the two more restrictive criteria. In all cases, rate of correct classification is greater than predictive efficiency and, on average, efficiency is increasing with the restrictiveness of the winning criterion. Further, the average efficiency gains for increasing the restrictiveness of the winning criterion tend to outpace classification losses. This means that there are fewer expectations being missed by the more restrictive winning criterion than there are excess predictions being eliminated. Changes in the performance of *connectedness* best showcase this. Under most winning criteria, *connectedness*, the most permissive compatibility measure, has both the highest correct classification and also the lowest predictive efficiency. The gap between those two is reduced with each increase in the restrictiveness of the winning criterion. This means that the additional predictions being generated by the more permissive winning criteria are less accurate than the predictions generated by the *minimal-winning* criterion. This increase in the heuristic model's predictive efficiency under the minimal-winning criterion tells us that the typical respondent using our heuristic model (or a close variant of it) seems to intuit and value the difference between minimal-winning and a surplus majority.

Figure 1: Comparison of heuristic model fit across ideological compatibility (row) and winning (column) criteria.



In addition to intuiting the difference between minimal-winning and surplus majority, the data also suggest that many respondents seem to intuit that precisely 50% is *not* a majority. The best example of this comes from our most efficient measure, *total PM distance*, which, due to its additive nature, will always discriminate to a minimal-winning seat share, whether defined as a true minimal-winning majority or a tie. When a tie is considered winning, *total PM distance* classifies at a rate of 0.231 and has a predictive efficiency of 0.192. When we employ a proper majority for the criterion, correct classification increases to 0.235 and efficiency increases to 0.194. How could even classification be improving with the more restrictive winning criterion? Consider the set $\{A_1^{24}, B_3^{26}, C_5^{10}, D_7^{20}, E_9^{20}\}$, where our model says B, as plurality, should be PM. When ties are considered winning, then AB is the winning coalition that minimizes *total PM distance* and ABC is a surplus majority that definitionally increases *total PM distance*. However, when ties are considered to be shy of a majority, ABC is the winning coalition that minimizes *total PM distance*. The increase in classification success when we shift from *majority inclusive of ties* to (true) *majority* as the size criterion tells us that more of our respondents would choose ABC than AB when faced with this party set—more respondents construct a true minimal-winning coalition than a (more compact) coalition that produces a seat share of precisely 50%

We contend that correct classification of 0.235 and predictive efficiency of 0.194—looking at the entire sample and using *total PM distance* and minimal-winning as the compatibility and winning definitions—is very impressive, given the very large number of alternatives in each choice. But the model is perhaps even more predictive than this efficiency measure implies as there may potentially be many "near misses," expectations that are very similar to the model prediction, but just miss classification. For example, consider set $\{A_1^{40}, B_3^{4}, C_5^{12}, D_7^{8}, E_9^{36}\}$, where a respondent identifies A as PM, adds nearest neighbor B, assesses total size (44%, not a majority), and then adds the next party C to create a majority (56%). This coalition is in keeping with the spirit of our heuristic model, but would not be classified by the model, because B is superfluous to the majority and increases the sum of distances.

To get a sense of how many such near misses there may be, we calculate two measures of deviance from the model prediction: 1) how similar a potential cabinet is in *size* to the heuristic model prediction by calculating the absolute difference in seats between the model prediction and all alternatives; and 2) how similar a potential cabinet is in *ideology*, or overall position, to the heuristic model prediction by calculating the absolute difference in seat-weighted left-right position between the model prediction and all alternatives. Examining the 82% of cases in which the heuristic model discriminates to a single alternative (in order to avoid multiple comparisons between potential alternatives and the heuristic model prediction), we find that respondents' expected coalitions are 5.3 seats closer to the model-predicted seat count than we would expect

by chance. Further, where identified coalitions can be up to ten units away from the model prediction, expectations are 0.8 units closer to the model prediction than we would expect by chance. Together, these deviance measures suggest that there may be many "near misses" among our respondents' expectations.

As above, we can take a closer look at the data by analyzing potential differences stemming from party system size using our dedicated surveys from Denmark, Germany, and the Netherlands. We compare model fit using *total PM distance* and minimal-winning as our compatibility and winning criteria and we adjust for bargaining context—whether the party system has single-party majority, a strongly dominant party, or is a top-two, top-three, or open system following Laver and Benoit (2015)—using fixed effects regression.[20] More specifically, we regress our outcome, which is correct classification—an indicator that the subject's expected coalition was also a heuristic model prediction—on the size of the party system, while holding bargaining context constant. We then calculate expected classification by system size and calculate the improvement of expected classification over chance and record them in Table **??**.[21] As one would expect, the heuristic model's classification in absolute terms diminishes as party system size increases. However, the drop off in absolute classification pales in comparison to the increases in the number of alternatives (from 80 to 2,304) such that the relative predictive power of the model is increasing substantially. The results suggest we can dismiss the concern that the heuristic model's applicability may be limited to only very simple and perhaps unrealistic contexts and that a significant number of our subjects are able to solve an abstract and complex political problem with relative ease by employing an ecologically rational heuristic model.

[20] We adjust because the distribution of bargaining contexts varies quite a bit across the system sizes.

[21] Each classification estimate has a standard error of 0.006.

6 Reclassifying external support parties

Finally, we analyze the all coalition expectations under two treatments of outside supporting parties, using just the five-party system data. The left column treats support parties are opposition, as we did in the main text, and as nearly all empirical studies of cabinet formation (and portfolio allocation) have in the past. The right column treats them as cabinet members. The results are quite similar with the obvious exceptions of differences in the estimates on size covariates. Treating support parties as cabinet members, rather than opposition, results in a large (and common sensical) shift in the probability density away from minority and minimal-winning and toward surplus majority. This is manifest in the large reduction in the size of the estimates on minority and minimal-winning status indicators.

Table 1: Analysis of coalition expectations under different treatment of outside supporting parties. The left column assigns them to "opposition" status, while the right column treats them as part of the cabinet.

	Opposition	Cabinet
Largest party is PM	2.366	2.432
	(0.016)	(0.020)
Largest party is partner	−0.309	−0.223
	(0.023)	(0.023)
Median party is PM	0.218	0.558
	(0.017)	(0.020)
Median party is partner	0.437	0.792
	(0.015)	(0.016)
Minimal-winning coalition	0.703	0.341
	(0.014)	(0.015)
Minority coalition	0.045	−0.527
	(0.022)	(0.029)
Partners' distance from PM	−0.974	−0.502
	(0.011)	(0.009)
Opposition spread	−0.100	−0.448
	(0.008)	(0.008)
Opposition spread × minority coalition	−0.471	−0.154
	(0.014)	(0.024)
Number of parties (logged)	−0.212	0.277
	(0.009)	(0.013)
Episodes	42,893	42,893
N	3,431,440	3,431,440
Log Likelihood	−139,913.900	−148,144.100

7 Models to derive country-specific cue weights

We use a mixed logit model with random coefficients allowed at the country-level to derive empirical Bayes predictions of the country level parameter estimates—how much predictive weight does each cue exert in the formation of real-world cabinets in each country. We use random coefficients, rather than a fully interacted conditional logit model, due to separation issues (perfect prediction, or, a lack of variation in within-unit covariate values when outcome is equal to 1). These estimates are used for the analyses in Figure 2 and Tables 7 and 8 in the main text. Results are given in Table 2 below.

Table 2: Mixed logit model of real-world cabinet formation.

	Real-world	
Parameter	Mean	SD
Largest party is PM	3.226	0.595
	(0.345)	(0.291)
Largest party is partner	-0.700	2.751
	(0.508)	(1.723)
Median party is PM	1.501	0.429
	(0.335)	(0.282)
Median party is partner	1.008	1.696
	(0.273)	(0.413)
Minimal-winning coalition	0.189	1.314
	(0.299)	(0.237)
Minority	0.073	2.379
	(0.454)	(0.401)
Partners' distance from PM	-5.743	0.424
	(0.634)	(0.687)
Opposition spread	0.090	1.256
	(0.139)	(0.516)
Opposition spread × Majority coalition	0.777	0.366
	(0.507)	(0.480)
Number of parties (logged)	-0.485	0.424
	(0.136)	(0.118)
Episodes	209	
N	6,617,804	
Log Likelihood	−678.397	

To estimate the county-level cue weights in our subject experimental coalition expectations, we estimate a fully-interacted conditional logit model, where every covariate is interacted with country. This model utilizes only the fiver-party system data. These estimates are used for the analyses described with Figure 2 and Table 7 in the main text. Results are given in Table 3 below.

Table 3: Fully interacted conditional logit model of experimental coalition expectations. 5-party systems only.

	Largest is PM	Largest is Partner	Median is PM	Median is Partner	Minimal-winning coalition	Minority	Partners' distance from PM	Opposition spread	Opposition spread × minority	Number of parties (logged)
Baseline estimate (Germany)	2.136	−0.838	0.510	0.697	0.822	−0.896	−0.947	−0.070	−0.421	−0.447
	(0.046)	(0.068)	(0.051)	(0.042)	(0.035)	(0.072)	(0.031)	(0.020)	(0.048)	(0.026)
Country-level interactions:										
Denmark	−0.173	0.366	0.200	−0.095	−0.529	−0.018	−0.569	−0.144	0.364	0.137
	(0.082)	(0.121)	(0.093)	(0.086)	(0.068)	(0.133)	(0.066)	(0.041)	(0.084)	(0.049)
Spain	0.289	0.336	−0.253	−0.317	−0.006	1.538	−0.205	−0.054	−0.019	0.458
	(0.067)	(0.105)	(0.075)	(0.064)	(0.060)	(0.100)	(0.048)	(0.033)	(0.064)	(0.039)
France	0.066	0.972	−0.478	−0.445	−0.111	1.631	0.125	0.152	−0.052	0.287
	(0.064)	(0.093)	(0.072)	(0.063)	(0.061)	(0.098)	(0.048)	(0.034)	(0.063)	(0.038)
Hungary	0.417	0.752	−0.644	−0.489	−0.034	1.069	−0.171	0.010	−0.282	0.367
	(0.071)	(0.102)	(0.078)	(0.063)	(0.058)	(0.102)	(0.048)	(0.032)	(0.067)	(0.040)
Italy	0.234	0.718	−0.632	−0.456	−0.024	1.323	−0.073	−0.018	−0.147	0.394
	(0.068)	(0.097)	(0.075)	(0.063)	(0.059)	(0.100)	(0.047)	(0.032)	(0.065)	(0.039)
Netherlands	0.261	0.345	−0.102	−0.107	−0.362	0.377	0.183	−0.077	−0.029	0.146
	(0.063)	(0.092)	(0.069)	(0.056)	(0.048)	(0.094)	(0.040)	(0.027)	(0.063)	(0.036)
Poland	0.220	0.585	−0.486	−0.346	−0.049	1.530	0.130	−0.006	−0.077	0.352
	(0.066)	(0.098)	(0.075)	(0.063)	(0.059)	(0.099)	(0.046)	(0.032)	(0.064)	(0.039)
Sweden	−0.048	0.400	−0.117	−0.161	−0.304	0.384	−0.160	−0.232	0.117	0.178
	(0.067)	(0.096)	(0.075)	(0.065)	(0.053)	(0.102)	(0.046)	(0.031)	(0.068)	(0.039)
UnitedKingdom	0.745	0.690	−0.495	−0.280	0.127	1.275	0.074	0.086	−0.055	0.098
	(0.075)	(0.112)	(0.083)	(0.068)	(0.059)	(0.105)	(0.051)	(0.034)	(0.068)	(0.040)
Episodes						48,602				
N						3,888,160				
Log Likelihood						−154,736				

Finally, to derive country-level cue weights for our subjects experimental coalition expectations for each party-system size, we estimate fully-interacted conditional logit models of expectations for each party-system size utilizing our directed surveys of Germany, Denmark, and the Netherlands. These estimates are used for the analyses described in Table 8 in the main text. Results are given in Table 4 below.

Table 4: Interacted conditional logit models of experimental coalition expectations. 5-9 party systems. Baseline country is Germany in all models.

	Five	Six	Seven	Eight	Nine
Largest party is PM	1.972	1.996	2.035	2.305	2.360
	(0.072)	(0.068)	(0.066)	(0.067)	(0.065)
Denmark × Largest party is PM	−0.009	0.022	0.217	−0.034	0.090
	(0.099)	(0.094)	(0.093)	(0.092)	(0.089)
Netherlands × Largest party is PM	0.381	0.509	0.613	0.302	0.441
	(0.100)	(0.096)	(0.094)	(0.092)	(0.092)
Largest party is partner	−0.857	−1.194	−1.154	−1.047	−1.089
	(0.104)	(0.107)	(0.102)	(0.104)	(0.104)
Denmark × Largest party is partner	0.384	0.670	0.917	0.565	0.657
	(0.145)	(0.145)	(0.138)	(0.142)	(0.141)
Netherlands × Largest party is partner	−0.012	0.395	0.357	0.108	0.181
	(0.151)	(0.151)	(0.148)	(0.149)	(0.151)
Median party is PM	0.723	0.718	0.737	0.726	0.724
	(0.082)	(0.081)	(0.080)	(0.083)	(0.082)
Denmark × Median party is PM	−0.013	0.120	0.137	0.225	0.128
	(0.113)	(0.111)	(0.110)	(0.113)	(0.113)
Netherlands × Median party is PM	−0.216	−0.221	−0.338	−0.007	−0.180
	(0.113)	(0.113)	(0.113)	(0.115)	(0.116)
Median party is partner	0.934	0.837	0.899	0.935	0.837
	(0.071)	(0.068)	(0.065)	(0.065)	(0.065)
Denmark × Median party is partner	−0.332	−0.120	−0.172	−0.037	0.081
	(0.103)	(0.098)	(0.095)	(0.095)	(0.093)
Netherlands × Median party is partner	−0.294	−0.230	−0.283	−0.174	−0.164
	(0.096)	(0.092)	(0.089)	(0.089)	(0.089)
Minimal-winning coalition	0.673	0.825	0.781	0.943	1.081
	(0.055)	(0.053)	(0.051)	(0.051)	(0.051)
Denmark × minimal-winning coalition	−0.379	−0.424	−0.255	−0.384	−0.503
	(0.081)	(0.079)	(0.078)	(0.079)	(0.079)
Netherlands × minimal-winning coalition	−0.128	−0.351	−0.185	−0.318	−0.351
	(0.075)	(0.073)	(0.071)	(0.071)	(0.071)
Minority	−0.917	−1.090	−1.596	−1.369	−1.469
	(0.117)	(0.106)	(0.106)	(0.095)	(0.092)
Denmark × minority	0.002	0.160	0.776	0.456	0.417
	(0.161)	(0.148)	(0.143)	(0.132)	(0.129)
Netherlands × minority	0.329	0.103	0.495	0.172	0.143
	(0.154)	(0.142)	(0.138)	(0.126)	(0.123)
Partners' distance from PM	−1.053	−1.216	−1.340	−1.555	−1.653
	(0.049)	(0.049)	(0.049)	(0.051)	(0.052)
Denmark × partners' distance from PM	−0.457	−0.273	−0.555	−0.213	−0.145
	(0.076)	(0.075)	(0.078)	(0.077)	(0.078)
Netherlands × partners' distance from PM	−0.178	−0.183	−0.137	−0.140	−0.127
	(0.068)	(0.069)	(0.069)	(0.071)	(0.072)
Opposition spread	−0.150	−0.336	−0.449	−0.539	−0.594
	(0.032)	(0.033)	(0.034)	(0.034)	(0.035)
Denmark × opposition spread	−0.063	−0.014	0.083	0.036	0.083
	(0.047)	(0.049)	(0.051)	(0.052)	(0.054)
Netherlands × opposition spread	−0.040	0.074	0.030	0.101	0.038
	(0.044)	(0.046)	(0.047)	(0.048)	(0.049)
Opposition spread × minority	−0.359	−0.347	−0.118	−0.451	−0.438
	(0.078)	(0.080)	(0.085)	(0.082)	(0.082)
Denmark × opposition spread × minority	0.303	0.285	−0.054	0.272	0.351
	(0.104)	(0.109)	(0.114)	(0.113)	(0.115)
Netherlands × opposition spread × minority	0.046	0.076	−0.239	−0.080	0.007
	(0.102)	(0.106)	(0.112)	(0.109)	(0.111)
Number of parties (logged)	−0.352	−0.467	−0.660	−0.698	−0.753
	(0.042)	(0.037)	(0.032)	(0.030)	(0.027)
Denmark × number of parties (logged)	0.042	−0.021	0.229	0.094	0.074
	(0.059)	(0.051)	(0.045)	(0.041)	(0.037)
Netherlands × number of parties (logged)	0.129	0.040	0.090	0.065	0.068
	(0.057)	(0.050)	(0.044)	(0.040)	(0.036)
Episodes	8,084	8,052	8,063	8,087	8,069
N	646,720	1,545,984	3,612,224	8,281,088	18,590,976
Log Likelihood	−23,882	−28,167	−32,156	−35,994	−39,683

8 Full results for H4 regressions

The full model results for main text Table 7, including all fixed effects estimates, are below.

Table 5: Relationship between real-world and experimental cue weights in coalition expectations. Full model results. Baseline categories are Germany for country and largest party is PM for cue.

Covariate	Pooled	Within country	Within cue	Only 3 focal cues	Omit 3 focal cues
Real-world weight	0.242	0.246	0.019	0.345	0.070
	(0.026)	(0.027)	(0.028)	(0.022)	(0.033)
Denmark		−0.165		−0.230	0.051
		(0.307)		(0.373)	(0.226)
Spain		0.129		0.041	0.220
		(0.307)		(0.373)	(0.224)
France		0.174		0.242	0.260
		(0.307)		(0.373)	(0.224)
Hungary		0.201		0.349	0.129
		(0.307)		(0.373)	(0.224)
Italy		0.114		0.270	0.142
		(0.307)		(0.373)	(0.224)
Netherlands		0.147		0.103	0.106
		(0.307)		(0.373)	(0.224)
Poland		0.187		0.150	0.218
		(0.307)		(0.373)	(0.224)
Sweden		−0.084		−0.046	0.034
		(0.307)		(0.373)	(0.225)
United Kingdom		0.204		0.205	0.189
		(0.307)		(0.373)	(0.224)
Largest party is partner			−2.563		
			(0.191)		
Median party is PM			−2.097		
			(0.138)		
Median party is partner			−1.866		
			(0.145)		
Minimal-winning			−1.592		
			(0.151)		
Partners' distance from PM			−3.176		
			(0.287)		
Opposition spread			−2.380		
			(0.154)		
Minority			−2.246		
			(0.173)		
Opposition spread × minority coalition			−2.727		
			(0.148)		
Number of parties in coalition (logged)			−2.471		
			(0.167)		
Constant	0.204	0.114	2.276	0.817	−0.198
	(0.067)	(0.217)	(0.129)	(0.264)	(0.158)
Observations	100	100	100	30	70
R^2	0.461	0.480	0.906	0.931	0.112
Country FEs		✓		✓	✓
Cue FEs			✓		
Observations	90	90	90	90	
R^2	0.448	0.456	0.984	0.985	

The full model results for main text Table 8, including all fixed effects estimates, are below.

Table 6: Comparing relationship between real-world and experimental cue-weights in coalition expectations across party system size in Denmark, Germany, and the Netherlands. Country baseline is Germany.

	Party system size				
Covariate	Five	Six	Seven	Eight	Nine
Real-world weight	0.261	0.286	0.314	0.331	0.342
	(0.044)	(0.046)	(0.050)	(0.053)	(0.056)
Denmark	−0.178	−0.097	−0.019	−0.061	−0.042
	(0.293)	(0.309)	(0.333)	(0.351)	(0.373)
Netherlands	0.090	0.119	0.147	0.115	0.122
	(0.292)	(0.309)	(0.332)	(0.351)	(0.372)
Constant	0.125	0.042	−0.010	0.005	−0.016
	(0.207)	(0.218)	(0.235)	(0.248)	(0.263)
Country FEs	✓	✓	✓	✓	✓
Observations	30	30	30	30	30
R^2	0.577	0.595	0.605	0.604	0.592

9 The role of political interest

This analysis replicates the H4 tests in the main text of the manuscript using two sets of country-level cue weights: one derived from high interest subjects and one derived from low interest subjects, where we sort into those groups according to whether their level of reported political interest is above the sample median value or otherwise. The analysis reveal that the correspondence between the experimental and real-world cue-weights is higher for high interest subjects than for low interest subjects. This implies that higher interest subjects have better observed, or otherwise internalized, the empirical regularities of coalition formation in their country and generate coalition expectations that closer approximate these empirical regularities than their lower interest counterparts. 2in

Table 7: Contextual correspondence for high and low interest subjects

	Pooled		Within-state	
	High interest	Low interest	High interest	Low interest
Real-world weight	0.257	0.228	0.262	0.230
	(0.027)	(0.027)	(0.027)	(0.028)
Country FEs			✓	✓
Observations	100	100	100	100
R^2	0.487	0.427	0.519	0.440

10 Experiment using a real party system

10.1 Germany

In our directed survey of German voters we included a second instrument that was designed to learn about tradeoffs German voters may make in their coalition expectations under very realistic scenarios. We were particularly interested in how voters would react to changes in the size and position of the Alternative f̀ur Deutschland (AfD)—a far-right nationalist party that had been successful in winning representation at the state level and had recently won representation in the Bundestag for the first time, securing about 13% of the seats in the chamber in the 2017 election, with which all other parties in Bundestag had publicly committed to not coalesce.

Using real-world parties in the experiment allows respondents to condition their expectations about which coalitions will form in our experimental setting on any of their preexisting beliefs about the parties and their relationships to each other or to the respondent. This might include the parties' histories of co-governance, their recent statements and interactions, the media's portrayal of their compatibility, our respondents' affective orientation toward them and so on; covariates that we measure and include in the analysis. Nonetheless, we can still estimate a causal effect of size and position characteristics on respondents' coalition expectations, as long as our manipulation of those characteristics is sufficiently credible that our respondents do not reject them. To this end, we made only modest changes to the size and positions of the parties from their "true" values at that time. To preview, the analysis reveals that respondents' coalition expectations remain consistent with our heuristic model, despite the real-world context of the experiment.

Further, the historical, media, and other new variables we included in the analysis all work as expected, suggesting that our respondents really are responding, in the context of our experiment, to the relevant historical context and media narrative. Indeed, when we examine the substantiative details of these expectations for a set of potential coalitions—those involving the AfD—for which we have a strong intuition about how the real world context should impact coalition expectations, we find exactly what we expect: voters are much less likely to include the AfD in coalitions that we would otherwise expect them to based on size and idealogical considerations. At the same time, however, such effects do not crowd out the impact of size and ideology. For example, holding all other size and position characteristics constant, we find that a small redistribution of seats from smaller parties to the AfD, granting the AfD enough seats to complete a majority with the CDU, leads to an increase of 8 percentage points in the proportion of respondents that expect an CDU-AfD coalition to form (from 0.5% $\rightarrow$ 8.9%) and an increase of 15 percentage points in the proportion of respondents that expect any coalition including both the CDU and AfD to form (from

6.9% $\rightarrow$ 21.8%). Even if this increase is significantly muted by all of the signals associated with the AfD, we can still conclude that when respondents bring in this extra information suggesting a coalition *should not* form—information that *should* drive them to derive expectations contra to what the size and position characteristics alone would imply—these size and position characteristics *still matter*.

The instrument, which was administered in January of 2020, presents each respondent with 7 different compositions of the Bundestag. The first composition showed the parties approximately as they were polling at the time (superscript=seatshare; subscript=LR position):

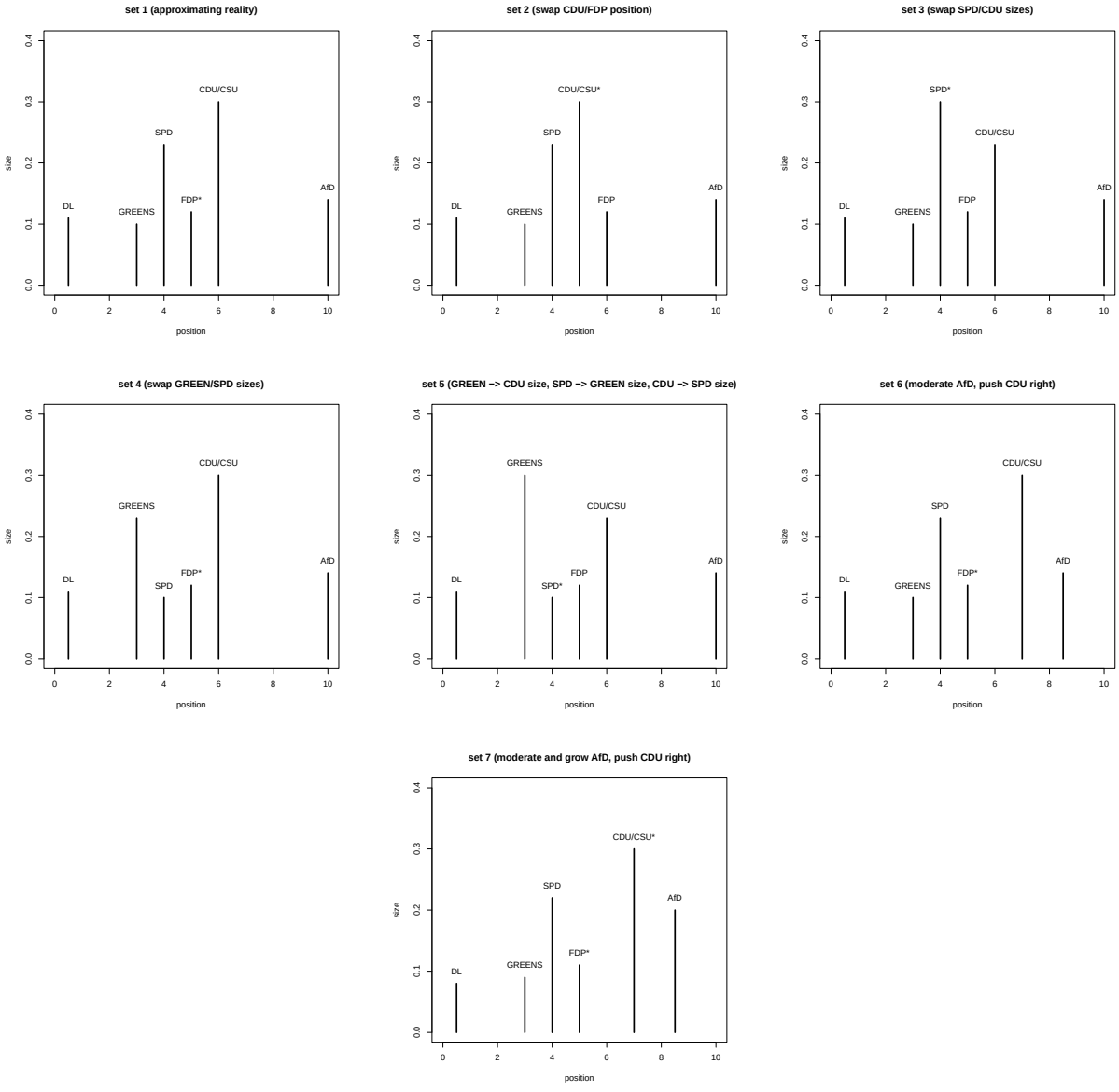
$$\{\text{Die Linke}_{0.5}^{0.11}, \quad \text{Green}_{3}^{0.10}, \quad \text{SPD}_{4}^{0.23}, \quad \text{FDP}_{5}^{0.12}, \quad \text{CDU}_{6}^{0.30}, \quad \text{AfD}_{10}^{0.14}\}$$

All compositions held the position (on a left-right scale of 0-10) of Die Linke (0.5), the Greens (3), and SPD (4) constant. In most potential outcomes, the FDP, CDU/CSU, and AfD were positioned at 5, 6, and 10, respectively. We show all party systems in Figure 2, where the title records the change from the baseline setting. In each picture we denote the party controlling the median legislator with an asterisk (in Figure 7 there is no true, singular median: FDP has the 50^{th} legislator and CDU has the 51^{st}). Note that the differences between outcomes are subtle (to maintain credibility). For each party system, we asked our respondents to identify the role each party would most likely play in the government to follow: prime minister, partner to the prime minister in cabinet, or opposition.

Because we varied aspects of the party system experimentally and asked our respondents for their perceptions of certain party characteristics, we can answer some lingering questions readers may have. First, we can measure the true history of cabinet participation for each party dyad, which we will call "familiarity" following Martin and Stevenson (2010), allowing us to assess the effect of coalition histories in voters' coalition expectations. Following Martin and Stevenson (2010), we measure this as the percentage of days a given party pair has been in government together over a number of preceding years (we measure this from 1980, 1990, and 2000 and assess the predictive power of each below) and take the average of all party dyads included in a given coalition. Second, we can measure the distribution of media messages regarding the likelihood of various coalitions forming, allowing us to assess the effect of media messages on coalition expectations. We call the media's description of the likelihood that each coalition will form the "media likelihood."

Our survey also asked respondents for their perceptions of three important characteristics of parties. First, we asked respondents for their perceptions of how often each unique pair of parties had governed together in the past, allowing us to evaluate the potential role of parties' *perceived* history of co-governance

Figure 2: German parliamentary composition vignettes



in coalition expectations, apart from the *true history*.[22] We also asked respondents for their perceptions of all parties' commitment to democracy, allowing us to evaluate the potential role of parties' "anti-system" nature in coalition expectations.[23] Finally, we asked respondents to tell us how likely they were to vote (at any time in the future) for all of the parties, allowing us to measure our respondents' overall affinity for all coalitions.[24]

These data allow us to estimate relationships that our central study does not. First, we can assess whether true histories of co-governance and media messages regarding coalition likelihood positively predict coalition expectations in a real-world setting. Second, we can estimate whether or not voters use their perceptions of parties' histories of co-governance, perceptions parties' anti-system characteristics, and their overall affinity for the parties when generating their coalition expectations. Third, we can assess whether our subjects are responsive to *small* changes in the size and position characteristics of real parties with which they are familiar. Most importantly, we can estimate these correlations simultaneously to assess their relative weight in shaping coalition expectations. In other words, we can each of the prevailing models of coalition expectations in the extant literature—history (Armstrong and Duch 2010), media (Eberl and Plescia 2018; Bowler, McElroy and Müller 2022), and rationalist models like our own—in the same model, which has not yet been done in the published literature.

For our analyses, we calculate all the same size and position covariates as in the main text, using the values shown in the survey experiment to our respondents. These characteristics are constant across all respondents—in each vignette, the size and position correlates for all potential cabinets are identical for all respondents. We then calculate each coalition's true familiarity, which is the average percentage of time since 2000 (or 1990 or 1980) spent co-governing for each dyad in the coalition. For example, the SPD and Greens co-governed Germany for 27.93% of the time between 2000 and our survey's administration. So the familiarity score for an SPD-Green coalition is 27.93. The familiarity score for an SPD-Green-Die Linke coalition, however, would be $9.31 = \frac{fam_{SDP-Green}+fam_{SDP-Linke}+fam_{Green-Linke}}{3}$, or, the total co-governance

[22]The question proposed 2 parties, then asked: "How many times have the following parties together formed the government in the past 20 years? If you don't know the exact answer, please guess." Respondents answered for every party pair. The available responses were: Never (1), About 25% of the time (2), About 50% of the time (3), About 75% of the time (4), Always (5).

[23]The question was, "For each party below, how committed do you think the party is to maintaining and strengthening democracy in Germany?" Available responses were: Completely committed (1), Somewhat committed (2), Neither committed nor uncommitted (3), Somewhat uncommitted (4), Completely uncommitted (5).

[24]Question wording: There are a number of parties in Germany, each of which would like to get your vote. How likely is that you would *ever* vote for each of the following parties in a future German federal election? Please answer on a scale from 0 to 10, where 0 means "very unlikely" and 10 "means very likely."

history for all dyads, divided by the number of dyads in the coalition (recall Die Linke has never served in government, so its familiarity with both SPD and Greens is 0). All single-party governments enter the data with a familiarity score of 0. In Table 8 we regress familiarity measured from 1980, 1990, and 2000 on coalition expectations, alone, in order to learn which value predicts most efficiently. We also model all three simultaneously and the average of all three, which has the effect of weighting more recent cabinets more heavily than more distant cabinets. Comparing likelihood estimates, the data show that familiarity since 2000 is the best singular predictor, and, including the other two measures (as in model 4) does not significantly improve model fit. In the models to follow, we use familiarity since 2000.

Table 8: Comparing different measures of true familiarity.

Covariate	Model 1	Model 2	Model 3	Model 4	Model 5
Familiarity from 1980	0.745			0.093	
	(0.006)			(0.067)	
Familiarity from 1990		0.736		−0.177	
		(0.005)		(0.121)	
Familiarity from 2000			0.637	0.723	
			(0.004)	(0.061)	
Average of all three					0.715
					(0.005)
Episodes	7,840	7,840	7,840	7,840	7,840
Observations	1,505,280	1,505,280	1,505,280	1,505,280	1,505,280
Log Likelihood	−35,903.280	−34,680.980	−34,402.130	−34,401.060	−34,565.160

We derived the media likelihood from 65,173 news articles published by 371 German language outlets in the four months leading to the administration of our experiment (articles were retrieved using the EventRegistry API). We describe this process in detail in the final section of the appendix, from the initial gathering of articles, our inclusion rules for them, to the coding, etc, but summarize here. The articles were coded with assistance from the Gemini 2.5 LLM which we evaluated against itself and other available LLMs. Each time a potential coalition (or a portion of a potential coalition) was mentioned in one of the articles, the model used the article text to score how likely or unlikely the article was communicating that the coalition was to form on a 0 to 10 scale in which a 10 indicated that the article communicated that the cabinet was very

likely to form and 0 that it was very unlikely to form. Once all the coalition likelihood scores are coded and validated, we simply take the mean of all individual likelihood score for each coalition as the final "media likelihood."

These finals score are intuitive face validity. There is a 16-way tie for the least likely coalitions, all containing four or more parties, and all large surplus majorities. The least likely coalitions of three parties or less are CDU-Green-AfD; CDU-SPD-AfD; and CDU-FDP-Linke. The coalitions coded as most likely are CDU-FDP and CDU-SPD, which was also the most commonly mentioned potential coalition. The least discussed potential cabinet is Die Linke governing alone, with only 8 mentions (all quite negative). These figures read as plausible on their face and, as we will see, predict our subjects' expectations in reasonable ways.

Finally, we calculate respondent-coalition level covariates. The first, "affinity," uses responses to the survey question, "How likely is that you would *ever* vote for [party]?" to calculate each respondent's average propensity to ever support members of each potential coalition. We calculate "average anti-system" by taking the mean of each respondent's perception of each party's lack of support for democracy for all parties included in each potential coalition (higher values indicate less perceived commitment to democracy). We calculate "perceived familiarity" by taking the average of the respondent's perceived history of co-governance for each party-dyad included in the coalition. As with true familiarity, all single-party coalitions enter the data with a 0-value on this covariate (about 9% of expectations are a single-party cabinet and about half of those are the CDU governing alone). As in the main text, all continuous variables are rescaled to be standard-normal (mean = 0, SD = 1) before estimation, which allows comparison across coefficient estimates to infer relative effect magnitudes within model.

Before the statistical analyses, we simply describe the difference in expectations across Vignettes 6 and 7. In both 6 and 7, the CDU is the plurality and the AfD is its nearest neighbor. Indeed, all party positions are identical in Vignettes 6 and 7, however, in Vignette 6 the CDU and AfD combined do not control a majority, but they do in Vignette 7. How do expectations respond to this subtle change? We summarize the results in Table 9. For Vignette 6, only 6 respondents predict that the CDU and AfD will coalesce on their own, 77 predict that the CDU and AfD will govern with one or more other parties, and a total of 110 expect the AfD to enter government with any constellation. In Vignette 7, 99 respondents predict that the CDU and AfD will coalesce on their own, 244 predict that the CDU and AfD will govern with one or more parties, and a total of 277 expect the AfD to enter government with any constellation. These differences in expectations are solely attributable to the small change in the AfD's seat share relative to the other parties. All party locations are identical between the two Vignettes and the sizes of the plurality and first runner

up (CDU and SPD, respectively) are also identical. The media narrative and histories of co-governance are unchanged. All parties' universal commitment to ostracizing the AfD from cabinet remains unchanged. But still expectations of the AfD's participation in cabinet—and specifically coalescing with the CDU—surged due to just a small redistribution of seats. Clearly, our subject are responsive to changes in size and position characteristics, even in presence of known associations, elite signals, etc.

Table 9: Summary of differences in AfD-included coalitions between Vignettes 6 and 7 from Figure 2.

Vignette	Respondents choosing CDU/CSU-AfD alone	Respondents choosing CDU/CSU-AfD with others
Vignette 6	6	77
(no CDU/CSU-AfD majority)	0.5%	6.9%
Vignette 7	99	244
(CDU/CSU-AfD majority)	8.8%	21.8%

The first model estimates the correlation between coalition expectations and true familiarity and media likelihood. These covariates vary across coalitions, but not within coalitions across respondents (the value on CDU-SPD is identical for all respondents), role allocations (the value on CDU-SPD is identical regardless of which party is PM), or vignettes (the value on CDU-SPD is identical for all 7 vignettes). As one may expect, both reveal strong, positive correlations. Omitting observable characteristics pertaining to size and position, as well as respondent-perceived measures of familiarity, affinity, and support for democracy, our respondents are more likely to expect coalitions with longer histories of co-governance and coalitions that media report are more likely.

Model 2 regresses only respondent-perceived measures of familiarity, affinity, and anti-system characteristics on coalition expectations. These variables vary across coalitions and respondents (the value on CDU-SPD varies across respondents) but not role allocations, or vignettes. The model reveals that these variables also perform as one would likely predict. Respondents are more likely to expect coalitions composed of parties they perceive as having longer histories of co-governance, parties they like more, and parties they perceive as supportive of democracy.

Model 3 and 4 replicate the VaPS and limited models from Table 1 in the main text. The estimates on our core covariates—largest party is PM, total distance to PM, and minimal-winning—are large, efficient, and in the expected direction in both models.

Next, we add to the VaPS specification true familiarity and media likelihood (Model 5) and the respondent-

Table 10: Conditional logit model estimating correlation between objective size and position characteristics against real histories of co-governance and media narrative as well as perceptions of co-governance histories, perceptions of anti-system characteristics, and overall affinity on coalition expectations. Experiment conducted in Germany using real German parties.

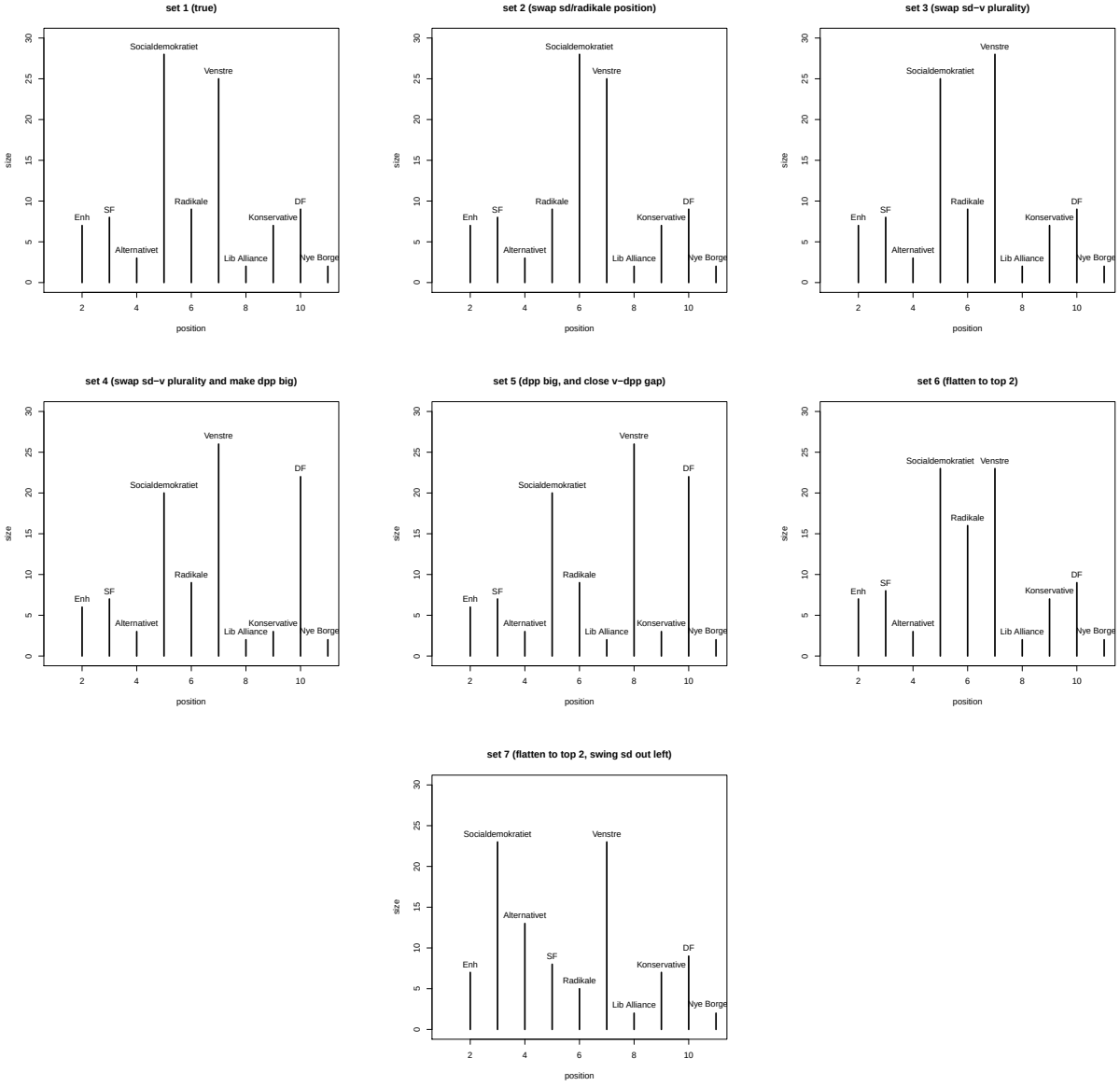
Covariate	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
True familiarity	0.273				0.286	
	(0.005)				(0.006)	
Media likelihood	0.737				0.404	
	(0.009)				(0.015)	
Perceived familiarity		0.970				0.519
		(0.011)				(0.010)
Affinity		0.242				0.148
		(0.024)				(0.020)
Perceived anti-system		−0.676				−0.284
		(0.029)				(0.026)
Largest party is PM			2.023	1.482	1.189	1.800
			(0.035)	(0.023)	(0.041)	(0.037)
Largest party is partner			1.256		0.378	0.937
			(0.042)		(0.047)	(0.043)
Median party is PM			−0.324		−0.012	−0.356
			(0.033)		(0.036)	(0.034)
Median party is partner			−0.980		−0.633	−0.984
			(0.038)		(0.040)	(0.038)
Minimal-winning coalition			0.704	0.289	0.176	0.527
			(0.036)	(0.027)	(0.041)	(0.037)
Minority			0.102		−0.260	−0.124
			(0.059)		(0.061)	(0.060)
Total distance from PM			−0.588	−1.745	−0.895	−1.205
			(0.037)	(0.021)	(0.045)	(0.042)
Opposition spread			0.662		−0.125	0.127
			(0.027)		(0.029)	(0.028)
Opposition spread × Minority coalition			0.633		0.198	0.669
			(0.037)		(0.039)	(0.038)
Episodes	7,840	7,840	7,840	7,840	7,840	7,840
N	1,505,280	1,505,280	1,505,280	1,505,280	1,505,280	1,505,280
Log Likelihood	−31,366.360	−36,873.550	−32,270.010	−33,697.230	−29,599.520	−30,578.330

perceived covariates (Model 6). The inclusion of these covariates attenuates some of the most salient estimates, but estimates on our core covariates remain in the expected direction, large, and efficient in both models. We conclude from these models that history, media, narrative, and individual affinities and perceptions of history and commitment to democracy matter for coalition expectations, but certainly do not crowd out the salience of size and position characteristics—particularly our three core covariates. Indeed, we can calculate odds ratios to compare the effect sizes from Model 5. A one-SD increase in media likelihood increases expectation odds by 50%; a one-SD increase in familiarity increases expectations odds by 33%; including the largest party as PM increases expectations odds by 228%; switching to minimal-winning (from surplus majority) increases expectations odds by 19%; switching to minority decreases expectations odds by 23%; finally a one-SD increase in ideological distance from the PM reduces expectations odds by 59%.

10.2 Denmark

Here, we report estimates from a nearly identical experiment administered in Denmark around the same time. Unfortunately, this case is far from the preceding and succeeding elections and the cabinet was in no risk of collapse, so there is no real relevant media narrative on coalition likelihoods around the time of administration. As such, we do not model objective score for media likelihood (or history) here. We do have all the same information on respondent's perceptions, however.

Figure 3: Danish parliamentary composition vignettes



As in the previous experiment, we present our respondents with potential parliamentary compositions, given in Figure 3, which are all slight variations from the status quo at that time. However, because this party system is quite a bit busier than the German case, we included a few more rank-order swaps in the left-right alignment of the parties than in the last experiment, where we only made one rank-order change. Again, we mark the most relevant substantive changes in the panel headers. Just as before, we ask our respondents identify which cabinet they believe will form from each parliamentary composition.

In the German version, we asked respondents for their perceptions of how often each unique pair of parties had governed together in the past, allowing us to evaluate the potential role of parties' *perceived* history of co-governance in coalition expectations. We calculate their perceived familiarity by taking the mean perceived familiarity of all dyads included in a given coalition. Unfortunately, the Danish party system was too large for us to ask every respondent about every party dyad (there are 45). Instead, each respondent was given a random sample of one third of the dyads. Because the missing dyads are, by design, missing entirely at random, we are able to efficiently impute the missing values using the other survey responses and respondent demographics following King et al. (2001) and Honaker et al. (2011).

Again, we asked respondents for their perceptions of all parties' commitment to democracy, allowing us to evaluate the potential role of parties' "anti-system" nature in coalition expectations. Each coalition anti-system score is the average of all included parties. Finally, we asked respondents to tell us how likely they were to vote (at any time in the future) for all of the parties, allowing us to measure our respondents' overall affinity for all coalitions. Each coalition's affinity score is the average of all included parties.[25]

As above, we arrange the data into observations identifying both the member parties and their role and calculate all size and position covariates from the vignettes to estimate the same VaPS and limited specification conditional logit models from Table 1 in the main text. Because this is a 10-party system there are 5,120 potential outcomes for each respondent-vignette observation. As in the German experiment, every respondent interacts every vignette. We report model results in Table 11. Model 1 regresses our respondent-perceived covariates on their coalition expectations. While affinity and perceived anti-system produce the expected correlation, perceived familiarity does not. We believe that this may be function of covariance with other perceptions variables combined with attenuation of the cross-respondent variance in the perceived familiarity covariate itself stemming from the imputation (model 4 estimates, which partial out some of this covariance with the size and position variables suggests this may be the case).

Models 2 and 3 estimate the VaPS and limited specifications and show the expected results, particularly for our core covariates which produce, large, efficient estimates in the predicted direction. These estimates

[25]The format of these questions and the perceived history of co-governance question was identical to the German survey.

Table 11: Conditional logit model estimating correlation between objective size and position characteristics against perceptions of co-governance histories, perceptions of anti-system characteristics, and overall affinity on coalition expectations. Experiment conducted in Denmark using real Danish parties

Covariate	Model 1	Model 2	Model 3	Model 4
Perceived familiarity	−0.049			0.016
	(0.018)			(0.010)
Affinity	0.743			0.238
	(0.019)			(0.013)
Perceived anti-system	−1.266			−0.412
	(0.027)			(0.020)
Largest party is PM		2.593	2.751	2.480
		(0.037)	(0.028)	(0.038)
Largest party is partner		−0.250		−0.353
		(0.056)		(0.057)
Median party is PM		0.416		0.497
		(0.042)		(0.043)
Median party is partner		0.288		0.405
		(0.033)		(0.033)
Minimal-winning coalition		1.315	1.087	1.301
		(0.043)	(0.032)	(0.044)
Minority		−0.341		−0.233
		(0.056)		(0.056)
Total distance from PM		−3.323	−3.235	−3.110
		(0.031)	(0.026)	(0.032)
Opposition spread		−0.721		−0.722
		(0.025)		(0.026)
Opposition spread × Minority coalition		0.785		0.751
		(0.051)		(0.051)
Episodes	6,993	6,993	6,993	6,993
N	35,804,160	35,804,160	35,804,160	35,804,160
Log Likelihood	−55,799.080	−39,619.930	−40,052.190	−38,905.070

are basically unchanged in Model 4, which adds the subject perceived variables. In this specification, all perceived covariates are estimated in the predicted direction. Our subjects are more likely to expect coalitions to form if they are perceived to have a deeper history of co-governance, are perceived to committed to democracy, and the subject is more fond of the participating parties. None of these correlations, however, crowd out the effects of size and position. The three largest coefficient estimates are for our three core covariates.

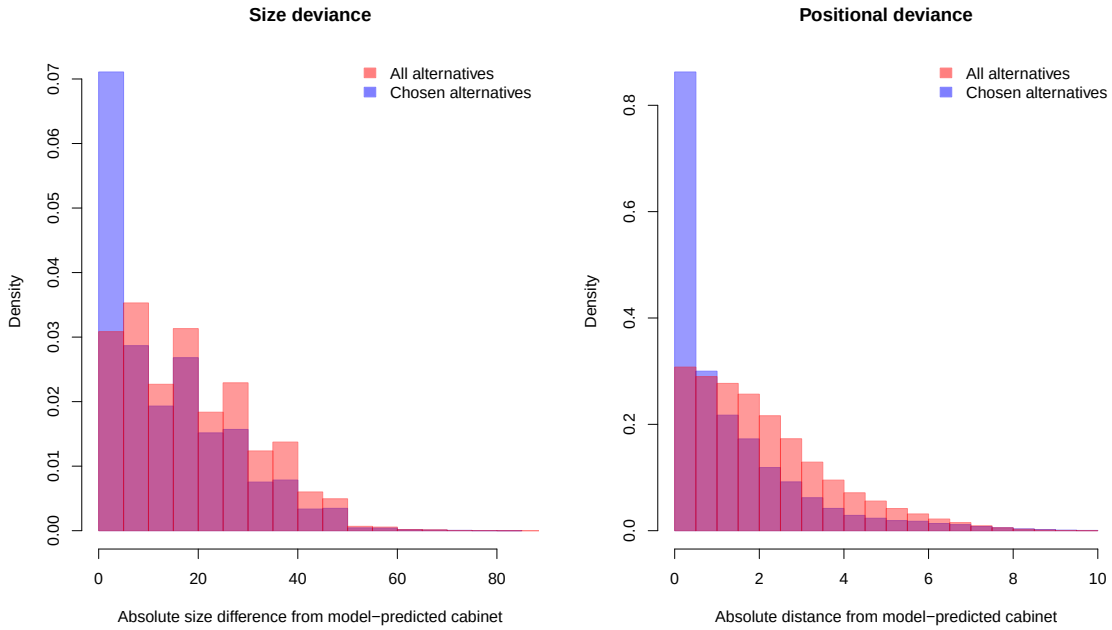
11 Near misses

It is possible that many coalition expectations that are not correctly classified by our heuristic model are nonetheless the result of attempting to employ our heuristic model or an alternative, yet very closely related mental model. Here we ask, among the responses that are not correctly classified, are there more "near misses" to the heuristic model prediction, or, more whimsical expectations?

We calculate two summary measures of deviance from our heuristic model. The first assesses how similar a potential cabinet is in *size* to the heuristic model prediction, by calculating the absolute difference in total seat share between all alternatives in an episode and the heuristic model prediction. The second assesses how similar a potential cabinet is in *ideology*, or overall position, to the heuristic model prediction by calculating the seat-weighted left-right position of all alternatives and taking the absolute difference from the seat-weighted left-right position of the heuristic prediction. The intuition here is that there may be many respondents applying our model (or something very similar) but making small mistakes. Consider the set of parties $\{A_1^{40}, B_3^4, C_5^{12}, D_7^8, E_9^{36}\}$, where a respondent identifies A as PM, adds nearest neighbor B, assesses total size (44%, not a majority), and then adds the next party C to create a majority (56%). This coalition is in keeping with the spirit of our heuristic model, but would not be classified by the model, because B is superfluous to the majority and increases the sum of distances. Nonetheless, this coalition is quite similar in size and aggregate ideological position to the heuristic model-predicted coalition of the AC. This is what we mean by "near misses."

For this analysis, we consider only the subset of episodes for which plurality as PM, *total PM distance*, and minimal-winning majority discriminate to a single alternative (83%), in order to avoid multiple comparisons between potential alternatives and heuristic model predictions. In Figure 4, we plot the distribution of absolute differences in the size and position of alternatives from our heuristic model predictions, coloring all potential cabinets in red and respondents' chosen alternatives in blue. The figure shows that respondents' expectations cluster heavily toward the heuristic model prediction in terms of both aggregate size and position—the blue density of *chosen* alternatives skews significantly left of the red density of all alternatives. Respondents' expectations are 5.3 seats closer to the model-predicted seat count than we would expect by chance. In terms of position, respondents' expectations are over 0.8 units closer to the heuristic model prediction than we would expect by chance. Further, where expectations can be up to ten units away from the model prediction, 58% of expectations are less than one unit away—what we would call a "near miss." If one were to adopt a more permissive cut point of 2 units away from the heuristic model prediction, 77% of expectations fall within this interval.

Figure 4: Absolute difference in aggregate position and size of coalition expectations from heuristic model prediction



In Table 12, we record the proportion of expectations that are either correctly classified by our heuristic model or are near misses to the heuristic model prediction using the ideological proximity criterion—one unit or less from the seat-weighted average of the parties in the heuristic model prediction—by the size of the party system. Across all system sizes, there are about twice as many "hits" when including near misses as compared to strict classification. Across all system sizes, for each strictly predicted expectation, there are about 1.58 *additional* expectations that are near misses to the heuristic model.

12 Party-level classification

In the main text we primarily evaluate the fit of our theoretical model to the experimental data lexicographically—we simply count up the number of theoretical predictions that are also realizations of the dependent variable and vice versa. That classification exercise examines the composition of the coalition *and* the roles assigned to the parties in cabinet. Here, we show how well the theoretical model classifies all parties into government or opposition roles. We do this in two ways. First we compare the theoretical prediction—the minimal winning coalition that minimizes partners' total distance from the plurality party serving as PM—to our respondents' expected coalition and measure the agreement between government and opposition roles for all

Table 12: Heuristic model fit across party system sizes inclusive of near-misses.

Parties	Alternatives	Baseline hit rate	Strict classification	Near misses	Sample
Five	80	0.0125	0.2340	0.5723	all countries
Six	192	0.0052	0.2578	0.6043	DGN
Seven	448	0.0022	0.2322	0.6016	DGN
Eight	1024	0.0010	0.2048	0.6038	DGN
Nine	2304	0.0004	0.1898	0.5816	DGN
Total (average)	444	0.0023	0.2295	0.5822	

parties. That is, if our theoretical model predicted a coalition of parties A and B, where C, D, and E are in opposition, and our respondent chose a coalition of A and B, where C, D, and E are in opposition, we would mark this as role classification of 1. If, instead, the respondent chose a coalition of B and C, where A, D, and E where in opposition, we would mark the classification success as 0.6 (B, D, and E are correctly classified, while A and C are not).

The second classification exercise compares not only the specific theoretical prediction to respondent choices, but also *near misses*—cabinets that whose seat-weighted mean position is within one unit of the theoretical prediction (where all cabinet positions are between 0 and 10). Note that both classification exercises only examine the 83% of choice settings where the theoretical model makes a *single prediction*.[26] The results are given in Table 13 and show that the specific model predictions classify over 70% of parties correctly, while classification inclusive of near misses is 93% overall.

Table 13: Theoretical model classification of parties into government and opposition roles by system size

System size	Strict predictions	Near misses
5	0.689	0.884
6	0.734	0.906
7	0.744	0.918
8	0.750	0.925
9	0.752	0.931
Total	0.710	0.897

[26]Most settings in which the theoretical model makes multiple predictions are instances in which there is a tie for plurality, or, two majority-making parties equidistant from the plurality party. These are events that are very rare in the real world.

13 Media messages

To capture discussions surrounding the composition of the next federal German cabinet (post-Merkel) in periods before the 2021 election date was formally set, we used python scripts to retrieve articles from EventRegistry API, a paid service supplying electronic access to a very large number of German language news publications.

Step 1: Filtering the October 2019-January 2020 newspaper data

We intentionally downloaded a very broad set of articles from the EventRegistry API using standard coalition keywords ("regierungsbildung", "k-frage", "neuwahlen", "jamaika", "ampel", "minderheitsregierung", "koalition", "bÃ¼ndnis).

While most of these articles are unlikely to contain relevant discussions, this was done to prevent any article selection bias at this stage. This yielded the following numbers of articles from 371 different German Language media outlets:

- 27,251 articles for January 2020
- 32,381 articles for December 2019
- 37,452 articles for November 2019
- 42,046 articles for October 2019

Step 2: The Proximity Filter

The next step used a python script to filter these articles. The script looked for the coalition key words above and examined a 50-word window around them.

The Negative Filter

: If the script found terms indicating the discussion was about a state-level or foreign event, the article was thrown out. We did a comprehensive search of the articles to identify important regional level and foreign political events likely to be included in the raw data (because they contain the key words from step 1). The key words we used to filter were: "österreich", "wien", "sebastian kurz", "ibiza-affäre", "övp", "fpö", "israel", "jerusalem", "netanjahu", "knesset", "gantz","spanien", "madrid", "pedro sánchez", "sánchez", "vox", "podemos", "gro§britannien", "london", "boris johnson", "theresa may", "brexit", "unterhaus",

“tories”, “italien”, “rom”, “matteo salvini”, “salvini”, “giuseppe conte”, “conte”, “fünf sterne”, “lega”, “irland”, “dublin”, “leo varadkar”, “sinn fein”, “fianna fail”, “slowakei”, “bratislava”, “igor matovic”, “caputová”, “schweiz”, “bern”, “svp”, “bundesrat”, “polen”, “warschau”, “pis”, “kaczynski”, “griechenland”, “athen”, “mitsotakis”, “tsipras”, “syriza”, “dänemark”, “kopenhagen”, “mette”, “frederiksen”, “finnland”, “helsinki”, “sanna marin”, “belgien”, “brüssel”, “flandern”, “usa”, “vereinigte staaten”, “washington”, “trump”, “biden”, “wei§es haus”, “kanada”, “ottawa”, “justin trudeau”, “neuseeland”, “jacinda ardern”, “australien”, “scott morrison”, “landtagswahl”, “landtag”, “bürgerschaftswahl”, “kommunalwahl”, “hamburg”, “bremen”

The Positive Anchor

: After a lot of testing, we also required that the article include (in the window) a word anchoring the discussion to German Federal politics: “bundestag”, “bundestagswahl”, “bundesregierung”, “bundeskanzler”, “bundeskanzlerin”, “kanzleramt”, “k-frage”, “kanzlerfrage”, “kanzlerkandidat”, “kanzlerkandidatin”, “post-merkel”, “merkel-nachfolge”, “groko”, “gro§e koalition”, “spd”, “cdu”, “csu”, “merkel”

Results of the filtering

:

- Jan 2020: 9,376 articles kept
- Dec 2019: 17,367 articles kept
- Nov 2019: 20,159 articles kept
- Oct 2019: 18,271 articles kept

The total number of articles that passed these filters was 65,173.
The top 10 publications contributing articles were as follows:

1. DIE WELT: 1,768 articles
2. S̀uddeutsche Zeitung: 1,688 articles
3. FinanzNachrichten.de: 1,634 articles
4. svz: 1,600 articles
5. T-online.de: 1,163 articles

6. stern.de: 949 articles
7. Augsburger Allgemeine: 904 articles
8. Der Tagesspiegel: 828 articles
9. NNN.de: 787 articles
10. GMX: 785 articles

Test for whether the LLM has discarded articles we should have kept

After applying the filters described above, we pulled a composite random sample of 160 of the discarded articles (80 from the foreign-noise discards, 80 from the missing-federal anchor discards) distributed evenly across the 4 months. We then had Gemini 3.1 pro (the most advanced LLM available to us at the time) read and evaluate whether our filters were working as intended. The results of that test are below:

The Foreign/State Noise Discards (80 Articles Sampled)

Accuracy: 100% Examples of articles discarded based on these criteria included:

- The UK Brexit elections (Boris Johnson, Corbyn)
- The Israeli Knesset coalition failure
- The Austrian T̀urkis-Gr̀un formation
- Global security (Syria, Soleimani, NATO summits)

The "Missing Federal Anchor" discards (80 Articles Sampled)

Accuracy: 100%

By dropping data that failed to anchor to a federal party or parliamentary term, we successfully removed irrelevant articles that happened to use terms like "Neuwahlen" or "Koalition." For example, this removed:

- Reports about local traffic accidents, Fahrerflucht, and police interventions (e.g., Ampel, meaning traffic light).
- Local sports divisions, elementary schools, and neighborhood organizations holding board elections.
- Incidental foreign coverage for regions we had not specifically checked in our foreign filter (e.g. Bolivian protests, Tunisian elections).

Step 3: Coding the remaining articles

Our set of filtered articles likely still includes articles that are not about the next coalition formation, but further key words filtering was likely to be counter-productive. Thus, at this point we used an LLM to read each article and evaluate its relevance before coding it (if relevant).

To decide whether each article should be ignored (because it was not discussing government formation), identify all potential cabinets that were mentioned, and code each mention for how likely or unlikely it was to form the next cabinet, we used the Gemini API to send each article to Gemini 2.5 flash. We also coded the full set of articles using Gemini 3. The key substantive coding instructions used are provided below. For each article, we asked the LLM to decide if the article was relevant to the formation of the next federal German cabinet and, if so, to code each mention of a specific party combination (or partial party combination), including "negative mentions" in which specific combinations are excluded. For each mention of a coalition, the mention was also coded for the tone of the coverage from very unlikely to form to very likely to form on a 0-10 scale. The main prompt was as follows:

> You are an expert political science coder analyzing German media from late 2019 and early 2020.
>
> Your task is to identify mentions of potential governing cabinets for the Federal German Election or cabinet formation dynamics within the Bundestag.
>
> IMPORTANT RULES
>
> 1. FEDERAL CONTEXT ONLY Only code articles regarding the Federal level German government (Bundestag).
>   - Discard: State-level elections or negotiations (e.g., Thüringen, Sachsen).
>   - Discard: Generic news where a Minister performs daily duties (e.g., speaking at a conference) without mention of coalition stability, makeup, or future formation.
> 2. SHORTHAND TRANSLATION Translate all shorthand names (e.g., 'Ampel', 'Jamaika', 'GroKo') into the exact standardized labels provided in the list below.
> 3. LIKELIHOOD LEVEL (FUTURE/STABILITY ONLY) $*$ Score (0 to 10): Use this range if the article discusses a FUTURE cabinet formation (e.g., post-election) OR whether the CURRENT cabinet will survive, reform, or collapse.

4. Output NULL (NaN): Use this if the article discusses the CURRENT cabinet performing daily tasks (e.g., passing a budget) with NO mention of its future stability or formation.
5. THE EXCLUSION RULE If a party explicitly excludes a coalition with another party, you MUST generate a separate mention for EVERY potential cabinet containing both parties.
   - Code: likelihood-level: 0.0
   - Set: is-exclusion-rule: True
6. THE "UNSPECIFIED" RULE If an article mentions that "Party A and Party B" need an unnamed "third partner," use the label that includes 'U' (Unspecified).
7. SINGLE ROW PER CABINET Synthesize multiple mentions of the same cabinet within a single article into one output row.
8. STRICT LABEL ORDERING Always output letters exactly in this alphabetical/hierarchical order: C, S, F, G, A, L, U. Example: Use 'C-S-G'; NEVER 'G-S-C'.
9. EXAMPLES
   - Article: "Linke, SPD und Grüne einigen sich auf Minderheitsregierung in ThÃ¼ringen." → IGNORE (State level).
   - Article: "Minister Scholz (SPD) spricht auf UN-Klimagipfel in Madrid." → IGNORE (Daily work; no stability discussion).
   - Article: "SPD-Parteitag stimmt Ã¼ber Verbleib in der GroKo ab. AKK (CDU) lehnt Nachverhandlungen ab." → CODE: cabinet-label='C-S', likelihood-level=5.0 (Discussion of future stability).

After coding each of our 65,173 filtered articles, the LLM concluded that an addition 26,505 did not contain any relevant discussion of the next German federal cabinet formation.

One important feature of our coding rules was the "Exclusion Rule.
This rule told the LLM how to code an article discussing a situation in which one or more parties excluded cooperation with some other party. In this case, we created negative mentions (coded 0) for all the excluded coalitions.

Verifying the quality of the Coding

Duplicated Article Test

Of the 65,173 articles in the dataset, there were only 47,956 unique title and article body combinations, because some articles were released by several different news sources. We decided to keep these in the data since such duplication of the same information across sources is a common feature of media environments and helps us capture the extent to which a specific article is widely circulated.

This also created the opportunity to use these duplicated articles to test the consistency of the coding across articles, since the article coding (even if they were duplicates) was done independently across articles.

The first thing we checked was whether the LLM found the same coalition mentions across the duplicated articles (i.e., if it identified discussion of the CDU/CSU-FDP coalition in one instance of the article, did it also do that in duplicated instances). We call this "Label agreement."

- Total Duplicate article clusters evaluated: 3,864
- 100% Perfect Label Agreement: 3,545 clusters (91.74%)

For ∼92% of the replicated articles, the LLM extracted the exact same set of potential cabinet coalitions, despite them being processed as completely independent articles in the coding.

Next, for the 3,545 clusters for which the mentioned coalition labels matched perfectly, we checked the consistency of the likelihood scores assigned to each do the 9,586 coalition mentions in these articles.

- Total unique mentions compared: 9,586
- 100% Exact match for likelihood scores (11-point scale): 8,769mentions (91.48%)

Thus, for ∼92% of the identified cabinets, the LLM gave the exact same 0-10 likelihood score across all duplicates.[27]

For the ∼8% of duplicated mentions that had a disagreement, the overall average absolute difference in likelihood scores was very small (0.13 points). The table below provides the exact differences in scores for these cases.

[27] This includes 521 cases in which likelihood score (following the rules we set out in our prompt) were independently coded null across different coding of the same article (because while the article mentioned a coalition, it gave no indication of its likelihood).

Table 14: Summary of duplicate coding

Absolute distance between ratings of duplicate mentions	# mentions	%	Cumulative %
0	8769	91.48	91.48
1	556	5.8	97.28
2	196	2.04	99.32
3	47	0.49	99.81
> 3	18	0.19	100

Full Recoding of all articles Using a different LLM

We also independently coded all 65,173 articles using Gemini 3. We then compared the coding between the two LLMs

We start by comparing the overall agreement between the average coalition likelihood levels of the media message about each of the 64 potential cabinets that could have been formed by the six German parties.

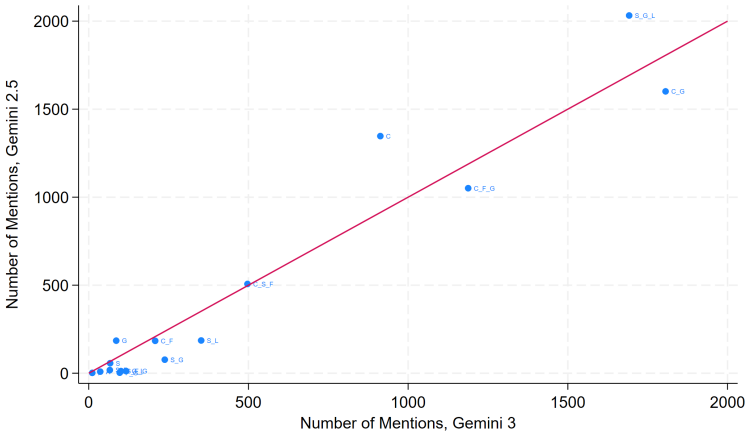
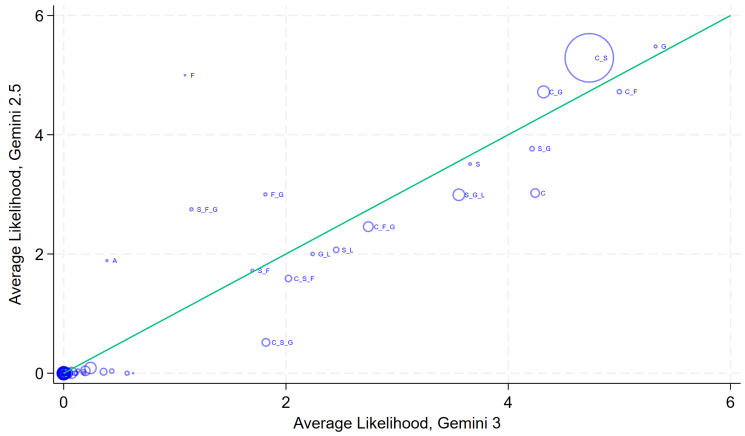


Figure 5: C: CDU/CSU, G: Greens, S: SPD, L: Linke, F: FDP, A: AfD

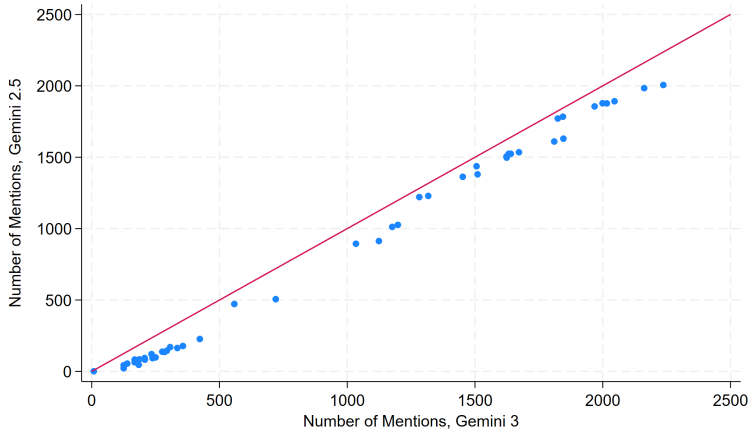
The sizes of the circles in the graph indicate the relative number of mentions of each potential coalition and we can clearly see (unsurprisingly) that a lot of discussion about coalition formation at the end of 2019 and early 2020 mentioned the incumbent CDU-SPD coalition, often commenting on its chances of

re-formation. The main point, however, is the high level of agreement in the average coalition likelihood scores using the two different LLMs.

The next graphs shows that this agreement extends to the number of mentions of each potential coalition. Because many coalitions are almost exclusively "mentioned" only because of the application of the exclusion rule described above, we examine those cases separately. First, for mentions not generated by the exclusion rule, we again get a close correspondence between models.[28]



For coalitions whose mentions are almost entirely due to the exclusion rule, the correspondence is still strong but Gemini 3 (as we discuss more below) consistently identifies fewer such exclusions.



As we discuss below, the extent to which the two models differ in this respect is inflated by the fact that a single disagreement about an exclusion (e.g., whether and article says the SDP has ruled out a coalition

[28]This graph excludes the CDU-SPD coalition because it has many more mentions than other coalitions, so makes it hard to see these patterns. Both models have similar numbers of mentions for that coalition as well.

with the AfD) generates many "coalition mentions" in the data (16 for two parties that say they will not coalesce and 31 if a party that says it will join no coalition or is ruled out by all others).

We can also look at the level of correspondence in the coding across these two LLMs at the level of individual articles. The level of agreement about which articles should be ignored (because they are not about cabinet possibilities) was 89%.

Table 15: Summary of duplicate coding

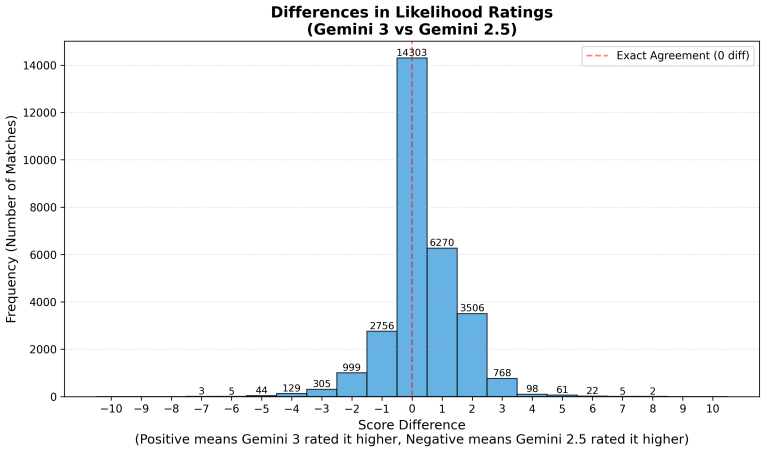
	Gemini 4 ignored	Gemini 3 ignored
Gemini 2.5 ignored	24738	1767
Gemini 2.5 coded	5555	33113
Total examined	65173	

For the 33,113 articles that both LLMs coded, we want to know if they identified the same sets of potential coalitions mentioned in the article: To do this, we can use Jaccard similarity, which quantifies how many elements two sets have in common relative to the total number of unique elements across both sets. We apply this separately to direct mentions of coalitions and "negative mentions" created by the exclusion rule.

There were 43,405 unique direct mentions of potential coalitions in the 33,113 articles that both LLMs coded (that were not generated by an application of an exclusion rule). Of these mentions, 85% matched between the two LLMs (i.e., the two LLMs identified the same party combinations in the same article) and the differences were approximately symmetric: 3327 were only identified Gemini 2.5 and 3300 were only identified by Gemini 3.

Next, for the 36,778 coalition mentions identified by both LLMs (and not generated by the exclusion rule), the coded likelihood levels (on our 11-point scale) match exactly 49% of the time. When they differ, however, they do not differ by much. The average difference on the 11-point scale was 0.79 points. The histogram of these differences is provided below.

When we examine the match between mentions created by the exclusion rule, the agreement between the two LLMs drops. Specifically, when we examine the Jaccard similarity for the 109,671 mentions generated by applying an exclusion rule in either model, we only have 62085 perfect matches (56.61%), with Gemini 3 applying the exclusion rule in only about a third of the case that Gemini 2.5 did. As we explained above, however, the size of this discrepancy is inflated because of the fact that a single disagreement between the



LLMs gets magnified, by the exclusion rule as much as 31 times.[29] [30]

[29] We did not include a comparison of the likelihood scores for mentions generated by the exclusion rule since they are, by definition, all coded as very unlikely.

[30] Because all such exclusions get coded as very unlikely, and we aggregate mentions of such excluded coalitions over all mentions, it turns out that the aggregate scores for such coalitions come up close to 0 even when the LLMs disagree somewhat about the specific set of articles that underly those scores.

References

Adams, James, Lawrence Ezrow and Zeynep Somer-Topcu. 2011. "Is anybody listening? Evidence that voters do not respond to European parties' policy statements during elections." *American Journal of Political Science* 55(2):370–382.

Adams, James, Simon Weschle and Christopher Wlezien. 2021. "Elite interactions and voters' perceptions of parties' policy positions." *American Journal of Political Science* 65(1):101–114.

Arian, Asher and Michal Shamir. 1983. "The primarily political functions of the left-right continuum." *Comparative politics* 15(2):139–158.

Armstrong, David A and Raymond M Duch. 2010. "Why can voters anticipate post-election coalition formation likelihoods?" *Electoral Studies* 29(3):308–315.

Bahnsen, Oke, Thomas Gschwend and Lukas F Stoetzer. 2020. "How do coalition signals shape voting behavior? Revealing the mediating role of coalition expectations." *Electoral Studies* 66:102166.

Bargsted, Matias A and Orit Kedar. 2009. "Coalition-targeted Duvergerian voting: how expectations affect voter choice under proportional representation." *American Journal of Political Science* 53(2):307–323.

Barnfield, Matthew, Joseph Phillips, Florian Stoeckel, Vittorio Mérola, Sabrina Stöckli, Benjamin Lyons, Jack Thompson, Paula Szewach and Jason Reifler. 2025. "Wishful thinking in response to events: Evidence from the 2021 German federal election." *Electoral Studies* 95:102940.

Baron, David P. 1991. "A Spatial Bargaining Theory of Government Formation in Parliamentary Systems." *American Political Science Review* 85(1):137–64.

Benoit, Kenneth and Michael Laver. 2012. "The dimensionality of political space: Epistemological and methodological considerations." *European Union Politics* 13(2):194–218.

Bowler, Shaun, Gail McElroy and Stefan Müller. 2022. "Voter expectations of government formation in coalition systems: The importance of the information context." *European Journal of Political Research* 61(1):111–133.

Cutler, Josh, Scott De Marchi, Max Gallop, Florian M Hollenbach, Michael Laver and Matthias Orlowski. 2016. "Cabinet formation and portfolio distribution in European multiparty systems." *British Journal of Political Science* 46(1):31–43.

De Marchi, Scott and Michael Laver. 2020. "Government formation as logrolling in high-dimensional issue spaces." *The Journal of Politics* 82(2):543–558.

De Marchi, Scott and Michael Laver. 2023. *The governance cycle in parliamentary democracies*. Cambridge University Press.

Duch, Raymond M. and Randolph T. Stevenson. 2008. *The Economic Vote: How Political and Economic Institutions Condition Election Results*. Cambridge: Cambridge University Press.

Duch, Raymond, Wojtek Przepiorka and Randolph Stevenson. 2015. "Responsibility attribution for collective decision makers." *American Journal of Political Science* 59(2):372–389.

Eberl, Jakob-Moritz and Carolina Plescia. 2018. "Coalitions in the news: How saliency and tone in news coverage influence voters' preferences and expectations about coalitions." *Electoral Studies* 55:30–39.

Falcó-Gimeno, Albert and Jordi Muñoz. 2017. "Show Me Your Friends: A Survey Experiment on the Effect of Coalition Signals." *The Journal of Politics* 79(4):1454–1459.

Fernandez-Vazquez, Pablo and Zeynep Somer-Topcu. 2019. "The informational role of party leader changes on voter perceptions of party positions." *British Journal of Political Science* pp. 1–20.

Fortunato, David. 2019. "The Electoral Implications of Coalition Policy-Making." *British Journal of Political Science* 49:59–80.

Fortunato, David and James Adams. 2015. "How voters' perceptions of junior coalition partners depend on the prime minister's position." *European Journal of Political Research* 54(3):601–621.

Fortunato, David, Nick CN Lin, Randolph T Stevenson and Mathias Wessel Tromborg. 2021. "Attributing policy influence under coalition governance." *American Political Science Review* 115(1):252–268.

Fortunato, David and Randolph T Stevenson. 2013. "Perceptions of partisan ideologies: The effect of coalition participation." *American Journal of Political Science* 57(2):459–477.

Fortunato, David, Randolph T Stevenson and Greg Vonnahme. 2016. "Context and political knowledge: Explaining cross-national variation in partisan left-right knowledge." *The Journal of Politics* 78(4):1211–1228.

Freire, André. 2006*a*. "Bringing social identities back in: The social anchors of left-right orientation in Western Europe." *International political science review* 27(4):359–378.

Freire, André. 2006*b*. "Left-right ideological identities in new democracies." *Pôle Sud* 25(2):153–173.

Gschwend, Thomas, Michael F Meffert and Lukas F Stoetzer. 2017. "Weighting parties and coalitions: How coalition signals influence voting behavior." *The Journal of Politics* 79(2):642–655.

Hare, A Paul. 1980. "Consensus versus majority vote: A laboratory experiment." *Small Group Behavior* 11(2):131–143.

Honaker, James, Gary King, Matthew Blackwell et al. 2011. "Amelia II: A program for missing data." *Journal of Statistical Software* 45(7):1–47.

Huber, John D. 1989. "Values and partisanship in left-right orientations: Measuring ideology." *European Journal of Political Research* 17(5):599–621.

Inglehart, Ronald and Hans-Dieter Klingemann. 1976. "Party identification, ideological preference and the left-right dimension among Western mass publics." *Party identification and beyond* pp. 243–273.

Kayser, Mark A, Matthias Orlowski and Jochen Rehmert. 2023. "Coalition inclusion probabilities: a party-strategic measure for predicting policy and politics." *Political Science Research and Methods* 11(2):328–346.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing incomplete political science data: An alternative algorithm for multiple imputation." *American Political Science Review* 95(1):49–70.

Knutsen, Oddbjørn. 1995. "Value orientations, political conflicts and left-right identification: A comparative study." *European journal of political research* 28(1):63–93.

Knutsen, Oddbjørn. 1997. "The partisan and the value-based component of left-right self-placement: A comparative study." *International Political Science Review* 18(2):191–225.

Krehbiel, Keith. 1998. *Pivotal politics: A theory of US lawmaking*. University of Chicago Press.

Laver, Michael and Kenneth A Shepsle. 1990. "Coalitions and cabinet government." *American Political Science Review* 84(3):873–890.

Laver, Michael and Kenneth Benoit. 2015. "The basic arithmetic of legislative decisions." *American Journal of Political Science* 59(2):275–291.

Laver, Michael and Norman Schofield. 1990. *Multiparty Government: The Politics of Coalition in Europe*. Oxford: Oxford University Press.

Lee, Seonghui, Philip Santoso and Randolph Stevenson. In press. *Heuristic Inference and the Left-Right: An Experimental Analysis of How Voters Form Their Left-Right Images of Parties.* Cambridge.

Lee, Seonghui, Randolph Stevenson and Joseph Francesco Cozza. 2025. "Explaining Cross-National Differences in Political Interest: The Role of Partisan and Ideological Cues." *Manuscript: University of Essex* .

Lodge, Milton and Charles S Taber. 2013. *The rationalizing voter.* Cambridge University Press.

Martin, Lanny W. and Georg Vanberg. 2014. "Parties and Policymaking in Multiparty Governments: The Legislative Median, Ministerial Autonomy, and the Coalition Compromise." *American Journal of Political Science* , Forthcoming.

Martin, Lanny W. and Randolph T. Stevenson. 2001. "Government Formation in Parliamentary Democracies." *American Journal of Political Science* 45(1):33–50.

Martin, Lanny W. and Randolph T. Stevenson. 2010. "The Conditional Impact of Incumbency on Government Formation in Parliamentary Democracies." Unpublished Manuscript, Rice University.

Meffert, Michael F, Sascha Huber, Thomas Gschwend and Franz Urban Pappi. 2011. "More than wishful thinking: Causes and consequences of voters' electoral expectations about parties and coalitions." *Electoral Studies* 30(4):804–815.

Meffert, Michael F and Thomas Gschwend. 2010. "Strategic coalition voting: Evidence from Austria." *Electoral Studies* 29(3):339–349.

Morelli, Massimo. 1999. "Demand Competition and Policy Compromise in Legislative Bargaining." *American Political Science Review* 93(4):809–20.

Naurin, Elin, Terry J Royed and Robert Thomson. 2019. *Party mandates and democracy: Making, breaking, and keeping election pledges in twelve countries.* New Comparative Politics.

O'Brien, Diana Z. 2019. "Female leaders and citizens' perceptions of political parties." *Journal of Elections, Public Opinion and Parties* 29(4):465–489.

Proksch, Sven-Oliver, Jens Wäckerle, Lukas Hetzer, Paula Hoffmeyer-Zlotnik, Jasmin Rath and Hauke Licht. 2024. "Rise of the Radical Right and Minority Governments: A Survey Experiment of Voters' Coalition Preferences." *Manuscript: University of Cologne* .

Provins, Tessa, Nathan W Monroe and David Fortunato. 2022. "Allocating costly influence in legislatures." *The Journal of Politics* 84(3):1697–1713.

Santoso, Lie Philip, Randolph T Stevenson and Simon Weschle. 2024. "What drives perceptions of partisan cooperation?" *Political Science Research and Methods* 12(4):888–896.

Vegetti, Federico and Daniela Širinić. 2019. "Left–right categorization and perceptions of party ideologies." *Political Behavior* 41(1):257–280.

Weschle, Simon. 2018. "Quantifying Political Relationships." *American Political Science Review* 112(4):1090–1095.

Zechmeister, Elizabeth. 2006. "What's left and who's right? A Q-method study of individual and contextual influences on the meaning of ideological labels." *Political Behavior* 28(2):151–173.

Zechmeister, Elizabeth J and Margarita Corral. 2013. "Individual and contextual constraints on ideological labels in Latin America." *Comparative Political Studies* 46(6):675–701.