A two-dimensional party-position panel from text and expert data

Armin Seimel
GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany
armin.seimel@gesis.org

Abstract

This Data Descriptor releases an election-year panel of economic left-right and cultural cosmopolitan–traditionalist party positions for 708 parties in 65 countries from 1944 to 2025. The panel supports comparative analyses connecting party-level ideological supply to elections, regions, voters, institutions, or political economy over long periods. Estimates are produced by a Bayesian dynamic item-response model that places heterogeneous text and expert evidence on a common latent scale and reports posterior uncertainty for every party-election estimate. The model combines the temporal reach of manifesto and media text data with the precision of expert placements by treating expert surveys, manifesto codings, media sentence codings, and historical expert data as noisy item-level signals of underlying party positions. Text counts enter through a two-part hurdle binomial likelihood, aggregate expert means enter through beta-binomial likelihoods, general left-right ratings inform the relationship between dimensions, and party positions evolve through random walks within party segments. Technical validation combines item-response diagnostics, random and party-blocked held-out expert prediction, convergent and discriminant checks, posterior predictive coverage, evidence-composition balance, robustness analyses, and convergence diagnostics.

Background and Summary

Expert surveys and text analysis of party documents have transformed comparative research on party competition, yet the resulting measures remain fragmented. Comparative party-position research often relies on source-specific measures, such as expert placements or manifesto-based indicators, each with distinct coverage, precision, and measurement assumptions. Manifesto and media text data provide the temporal and geographic reach needed for estimates from 1944 to 2025, while expert surveys provide direct and precise party placements at selected time points. These sources differ in scale, methodology, temporal coverage, and substantive emphasis [1–3]; prior cross validation studies show that they recover related positional signals while also containing source-specific artifacts and disagreement [4–7]. The dataset described here combines these complementary strengths in a single measurement

model: the extent of manifesto and media text coverage and the precision of expert placements.

Recent findings reinforce this existing rationale. [6] find that manifesto, expert-survey and voter-survey measures produce very similar left–right party rankings despite differences in their construction. [7] and [8] show that remaining discrepancies include systematic centrist bias in manifesto estimates and scale-linking problems when external party positions are combined with voter-preference scales. For the text sources used here, [9] find strong convergence between PolDem and Manifesto position measures on economic, cultural and European-integration domains, although issue salience differs more strongly. [10] show that text-based estimates depend on how algorithms extract and aggregate politically relevant textual variation and therefore require systematic benchmarking. Together, these findings reinforce the need for an explicit measurement model: the sources measure common underlying party positions, but their observed scores use different scales and contain source-specific biases, so they cannot simply be pooled.

This data record addresses this measurement problem by integrating heterogeneous measurements into economic left-right and cultural cosmopolitan–traditionalist party-position estimates on bounded unit scales. The economic dimension captures distributive and state-market conflict, including welfare, redistribution, market regulation, public ownership, economic liberalization, labour and related economic policy. The cultural dimension captures a cosmopolitan–traditionalist conflict: lower values indicate more cosmopolitan, socially liberal, multicultural and internationalist positions, while higher values indicate more traditionalist, nationalist, morally conservative and law-and-order-oriented positions. This terminology follows the green, alternative and libertarian versus traditional, authoritarian and nationalist (GAL–TAN) tradition in comparative party research while using a descriptive label for the broader set of text and expert indicators integrated here [11, 12]. Together, the two dimensions correspond to the distinction between economic conflict and cultural conflict in the comparative party competition literature [11, 13]. This structure allows researchers to distinguish whether observed changes in party support, party-system competition or voter-party alignment are associated with movement on the economic dimension, the cultural dimension or both. The estimates are posterior summaries from a Bayesian dynamic item-response model, with each source contributing through its own measurement relationship. In this setting, a latent party position is the unobserved economic or cultural position inferred from observed expert ratings and text-coded policy signals after estimating how each source and item relates to the two dimensions. The election-year panel is the primary research product for 708 parties in 65 countries from 1944 to 2025. The released data include posterior uncertainty and source-composition metadata for every estimate, documenting the evidence informing each estimate across countries and periods.

The statistical model is central to the data record. For each party segment and year, it estimates two latent positions, economic left-right and cultural cosmopolitan–traditionalist. Readers can think of the model as pooling many imperfect measurements of party positions while allowing each source to have its own scale, noise, and relationship to the economic and cultural dimensions. Observed expert means and text counts are linked to the latent positions through source and item-specific response functions. Text indicators contribute through a two-part hurdle binomial likelihood that separates issue mention from directional position. Expert means contribute through beta-binomial likelihoods that use scale bounds

Component	Role in dataset construction
Latent party-year positions	Two-dimensional economic and cultural positions for each party segment and year.
Text counts	Two-part hurdle binomial item observations separating issue mention from directional position.
Expert means	Beta-binomial aggregate expert observations using scale bounds and expert counts.
General left-right ratings	Cross-dimensional information about how economic and cultural positions combine in one-dimensional left-right judgments.
Random walk	Temporal propagation within party segments, with uncertainty increasing where direct evidence is sparse.
Posterior summaries	Election-year means, standard errors, and credible intervals released as the primary data product.
Source-composition fields	Metadata describing the observed text and expert inputs around each election-year estimate.

Table 1: Main components of the model that generates the released data.

and expert counts. General left-right expert placements inform how one-dimensional left-right ratings combine economic and cultural information in different countries, sources, and decades. Latent positions evolve through random walks within party segments, allowing irregular observations to inform nearby election years while widening uncertainty where evidence is sparse.

The model addresses the integration problem by estimating common latent positions and source-specific measurement relationships jointly. The raw inputs are not treated as directly interchangeable. Instead, each source variable enters as a distinct item with its own intercept, loading, likelihood, and uncertainty structure. Comparability applies to the posterior estimates on the common scale, not to the raw source scores. The model is needed precisely because the inputs differ in coverage, scale, and measurement properties. The empirical question is whether these heterogeneous indicators contain coherent directional signals once item, source, country, and temporal differences are modeled. Text sources supply observations across the 1944–2025 period and across party systems where expert surveys are absent or sparse; expert surveys supply precise direct placements that help locate and calibrate the latent dimensions. The posterior distribution combines the item-level evidence, source-specific measurement parameters, temporal dynamics, and prior structure to produce election-year estimates that are comparable on the model's common latent scale and released with uncertainty and source-composition metadata. Table 1 summarizes the main components of this model.

This measurement claim depends on the model's assumptions and can be assessed with the diagnostics reported below. The model assumes that parties have latent economic and cultural positions that generate observable expert ratings and text-coded policy signals with error. It does not assume that the raw input variables are equivalent. The empirical question

is whether, after allowing item, source, country, and temporal differences, the indicators display coherent directional response patterns. The item response diagnostics, held-out expert validation, posterior predictive checks, and evidence-composition balance checks make this measurement structure transparent.

The main data product is a modeled election-year measurement panel rather than a set of separate source-specific scales. Manifesto observations are item-level indicators in the same latent measurement model as expert observations; they are not a standalone manifesto scale. The model estimates a latent party position from all available signals; it does not treat any single raw input as the party's position. The item response diagnostics reported below assess whether the heterogeneous indicators behave as expected under this latent structure. The data record is accompanied by validation results that quantify how well the integrated measurement model recovers observed and held-out expert placements, how the two dimensions discriminate from each other, how family level positions align with theoretical expectations, and how well uncertainty is calibrated.

The dataset is intended for comparative election-year analyses requiring model-based comparable party positions across 708 parties, 65 countries, and the 1944–2025 period. The release is intended primarily for election-year analysis. The model estimates annual latent trajectories because observations can enter at different points in time, including expert surveys between elections, campaign text at elections, and historical expert evidence. The random walk lets these irregular observations inform the election-year estimates while widening uncertainty as temporal distance increases. The annual trajectory is also released as secondary model output for applications that require inter-election estimates. It provides a model-based alternative to simple linear interpolation: rather than mechanically connecting election-year endpoints, the model combines irregularly timed text and expert evidence and propagates temporal uncertainty into each annual estimate. These values remain posterior estimates from the dynamic measurement model rather than independently observed yearly party positions.

Methods

Model overview

The full model has three connected parts: a latent position process, source-specific measurement models, and posterior release construction. The latent process defines economic and cultural party positions for each party segment and year and links adjacent years through random walks. The measurement models connect heterogeneous observations to those positions: text data enter as directional issue counts, dimension-specific expert items enter as aggregate bounded ratings, and general left-right ratings enter as country, source, and decade-specific mixtures of the two latent dimensions. Posterior draws from the joint model are then transformed to the unit interval and summarized for election years, producing the released means, standard errors, credible intervals, and source-composition fields.

Figure 1 shows how these stages connect. Original source licences continue to govern access to third-party inputs; the public release contains the derived party-position products, metadata and reproducible code rather than implying redistribution of every raw source.

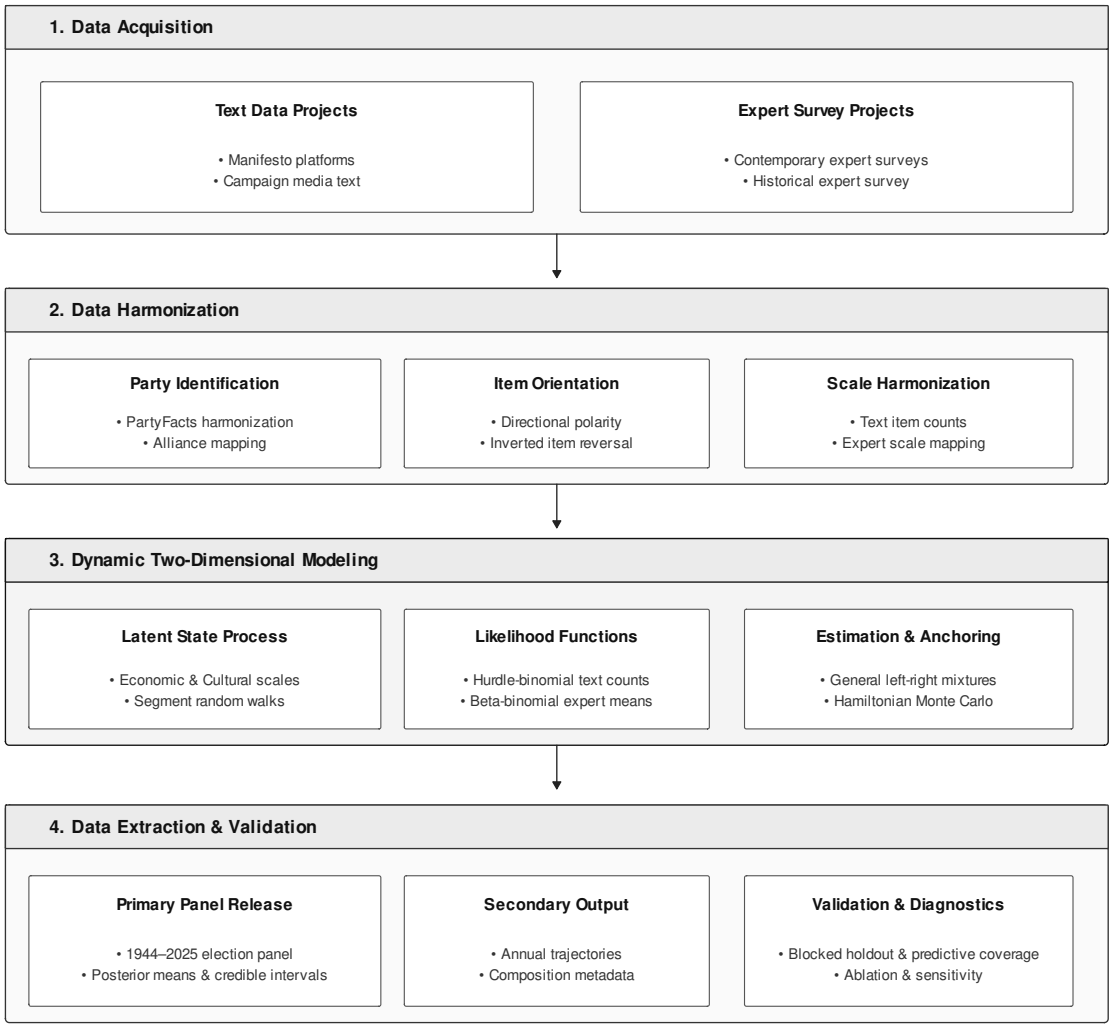


Figure 1: Research and release workflow. Heterogeneous text and expert sources pass through access checks, party and variable harmonization, model-ready construction and the dynamic measurement model. The posterior annual trajectories then generate the election-year panel, usage metadata and validation artifacts. Third-party inputs remain subject to their providers' licences.

Input data

The model integrates multiple text and expert data projects. Table 2 summarizes the sources, dimensions, party counts, year ranges, and source-specific citations. Party identifiers were harmonized through PartyFacts wherever possible [14]. Harmonization prioritized individual party identifiers rather than electoral alliances. Cases of party renaming were treated as continuous only where the historical and successor labels referred to the same continuing organization; mergers, major splits, and newly founded successor parties were kept separate. Party family information used for construct validation comes from ParlGov linked through PartyFacts [14, 15].

Source	Type	Dimensions	Parties	Years	Citation
Manifesto Project	Text	Econ, Cult	713	1920 to 2025	[1]
PolDem	Text	Econ, Cult	93	1972 to 2017	[16]
Chapel Hill Expert Survey (CHES) family	Expert	Econ, Cult, LR	390	1999 to 2024	[2, 17–20]
V-Party	Expert	Econ, Cult	589	1970 to 2019	[3]
Populism and Political Parties Expert Survey (POPPA)	Expert	Econ, Cult, LR	225	2018 to 2023	[21]
Global Party Survey (GPS)	Expert	Econ, Cult	271	2019	[22]
Morgan	Expert	LR	72	1945 to 1973	[23]

Table 2: Model-ready input data sources entering the latent measurement model. Econ denotes economic left-right, Cult denotes the cultural cosmopolitan–traditionalist dimension, and LR denotes a general left-right scale used for anchoring rather than as an output dimension. Party counts and year ranges describe input coverage before construction of the 1944–2025 election-year release. Manifesto Project input begins in 1920, but all pre-1944 observations are from the United States and therefore provide no comparative country coverage; the released comparative panel begins in 1944. The citation column gives the source-specific reference for each input dataset or source family.

Text observations come from the Manifesto Project and the PolDem national election campaign dataset [1, 16]. The Manifesto Project codes party platforms into quasi sentences assigned to policy categories and provides the longest temporal coverage, especially before contemporary expert surveys become available. All Manifesto Project observations before 1944 are from the United States and therefore do not provide comparative country coverage. The released panel starts in 1944, which defines the comparative postwar window used for the data record. PolDem applies a related coding approach to media coverage of parties and contributes an additional campaign and media based text source for a smaller set of countries and years. Both sources were transformed into party-year item counts, with a positive count and a sample count for each directional item. The model uses counts rather than pre-computed manifesto scales so that issue salience and position can be separated in the likelihood.

Expert observations come from the CHES family, V-Party, POPPA, GPS and Morgan's historical expert survey [2, 3, 17–23]. These sources differ in scale granularity, number of experts, temporal coverage, and whether they measure dimension-specific positions or gen-

eral left-right. Some sources provide direct placements for recent party systems, while others extend historical and cross-national coverage. The model treats these inputs as distinct evidence streams and records source composition in the released metadata. Published expert means were normalized to the unit interval with higher values indicating rightward economic or traditionalist cultural positions. Items whose original direction was opposite to this convention were reversed before estimation.

Morgan's historical expert data provide early general left-right evidence for party systems before contemporary expert surveys. Morgan collected expert placements on a continuous 0 to 100 scale for European and Israeli parties. Only the postwar portion enters the dataset. The continuous scale is discretized to 10 points to match CHES resolution, and the number of expert respondents is used as the expert count in the beta-binomial likelihood. Morgan's 25 percent trimmed means reduce the effective number of contributing experts, and remaining measurement differences are absorbed by the model's precision and item parameters.

Latent positions

Let $\theta_{d,r}$ denote the latent position for dimension d and segment year index r, where $d = 1$ is economic left-right and $d = 2$ is cultural cosmopolitan–traditionalist. The reported data use the inverse logit transformation

$$\theta^*_{d,r} = \text{logit}^{-1}(\theta_{d,r}) = \frac{1}{1 + \exp(-\theta_{d,r})}. \tag{1}$$

The transformed value $\theta^*_{1,r}$ is the economic position, with 0 indicating left and 1 indicating right. The transformed value $\theta^*_{2,r}$ is the cultural cosmopolitan–traditionalist position, with 0 indicating cosmopolitan and 1 indicating traditionalist. The unit interval is used for the released estimates because the observed expert means and directional text shares are bounded summaries and because the endpoints give users an interpretable common reporting scale. Estimation nevertheless proceeds on the unbounded logit scale above, and the release includes posterior standard errors and credible intervals so that downstream analyses can propagate uncertainty rather than treating the transformed posterior mean as exact. Users who model the estimates as dependent variables should account for the bounded transformation and posterior uncertainty, especially near the endpoints where the inverse-logit scale compresses differences.

In practical use, values are coordinates on a model-estimated common scale. A value near 0.20 is relatively leftward economically or cosmopolitan culturally, a value near 0.50 is near the middle of the reporting scale, and a value near 0.80 is relatively rightward or traditionalist. A change of 0.10 is ten percentage points on this reporting scale, so the same numerical contrast can be compared across parties, countries and years conditional on the model. It is not a ten-percent change in a literal policy package, nor does it guarantee equal substantive policy distance at every point: 0 and 1 are asymptotic transformed-scale endpoints and estimation occurs on a nonlinear logit scale. This makes posterior party-year positions comparable on the shared scale; it does not make the underlying raw manifesto categories and expert ratings interchangeable.

Temporal structure and segmentation

Party positions are modeled as random walks within continuous party segments. A segment is a continuous period in a party history, uninterrupted by an evidence gap longer than the preprocessing threshold, within which adjacent observed or latent years can be linked by the dynamic model. For segment s with ordered years $r_1^{(s)}, \ldots, r_{T_s}^{(s)}$, the model specifies

$$\theta_{d,r_1^{(s)}} = \mu_{d,s} + \sigma_{d,f(s)} \epsilon_{d,r_1^{(s)}}, \tag{2}$$

$$\theta_{d,r_t^{(s)}} = \theta_{d,r_{t-1}^{(s)}} + \sigma_{d,f(s)} \epsilon_{d,r_t^{(s)}}, \quad t = 2, \ldots, T_s, \tag{3}$$

where $\epsilon_{d,r} \sim \mathcal{N}(0,1)$ and $\mu_{d,s} \sim \mathcal{N}(0, \sigma_d^{\text{init}})$. The non centered parameterization improves posterior geometry for Hamiltonian Monte Carlo.

Innovation standard deviations are partially pooled through country and party family structure. The model allows different contexts and party families to exhibit different rates of positional change while sharing information across countries and families:

$$\sigma_{d,c} = \text{softplus}\left(\mu_d^{\sigma} + \tau_d^{\text{country}} \eta_{d,c}\right), \tag{4}$$

$$\sigma_{d,f} = \text{softplus}\left(\log \sigma_{d,c(f)} + \tau_d^{\text{family}} \eta_{d,f}\right). \tag{5}$$

This structure lets the data determine whether party families or countries exhibit faster or slower positional movement.

The random walk is mainly a device for integrating evidence observed at different time points and for extracting election-year posterior summaries from a continuous annual latent process. Expert surveys sometimes occur between elections, and text observations are tied to election campaigns or media coverage. The dynamic structure allows these observations to inform nearby election-year estimates and expresses the decay of that information through posterior uncertainty. It also supports inter-election interpolation for research designs that require annual positions, while retaining the interpretation of annual values as model-generated posterior estimates. Party histories are split at gaps exceeding seven years, a threshold chosen to be longer than one regular electoral cycle in most parliamentary systems while avoiding long unconstrained bridges across party histories. Applying a random walk through longer gaps would produce trajectories where no data constrain the estimates. Segmentation treats post-gap periods as new segments with independent starting positions. The sensitivity checks below re-estimate the model with five- and ten-year thresholds and show that the released election-year estimates are nearly unchanged. Thirty-nine parties in the release are split into multiple segments, and the annual model output used to derive the election-year release contains 17,585 segment-year observations for 708 parties.

For example, if a party has manifesto evidence in an election year and an expert survey observation two years later, the random walk allows both sources to inform the election-year posterior while increasing uncertainty with temporal distance and with the absence of direct observations. Conversely, if a party has no nearby expert or text observations, the posterior combines the previous latent state, the random walk variance, and the hierarchical prior structure; wider intervals record the resulting posterior uncertainty.

Text likelihood

Text observations are modeled as directional counts with a two-part hurdle binomial likelihood. For each party-year item observation n, the sample count records the number of relevant non-neutral mentions, and the positive count records the number supporting the item's positive pole. The first part models whether the item has no relevant text. Let $z_n = 1$ if the sample count is zero and $z_n = 0$ otherwise; then $z_n \sim \text{Bernoulli}(\pi_n)$ with

$$\text{logit}(\pi_n) = \alpha^{\text{zs}} + \xi^{\text{year}}_{t(n)} + \xi^{\text{seg}}_{s(n)} + \xi^{\text{ciy}}_{j(n)}. \tag{6}$$

For observations with positive sample size, the positive count follows

$$y_n \mid \text{sample}_n > 0 \sim \text{Binomial}\left(\text{sample}_n, \text{logit}^{-1}(\eta_n)\right), \tag{7}$$

with

$$\eta_n = \gamma^{\text{int}}_k + d_n \gamma^{\text{load}}_k \theta_{m(n),r(n)} + \delta_{c(n),k}. \tag{8}$$

Here γ^{int}_k is an item intercept, $\gamma^{\text{load}}_k > 0$ is the item loading, $d_n \in \{-1, +1\}$ records the substantively coded item polarity, $m(n)$ is the dimension, and $\delta_{c,k}$ captures country-item differences in the expression or coding of policy categories. Item polarity is assigned before estimation so that higher expected responses correspond to the economic right or cultural traditionalist pole; the loading then estimates how strongly the item discriminates along that oriented dimension. This specification separates issue salience from position: silence on an issue is modeled by the hurdle part rather than treated as centrism.

A manifesto or PolDem observation therefore does not enter the model as a pre-computed left-right or cultural score. Each directional category enters as an item-level count. This lets the model estimate whether, for example, pro-market, welfare expansion, nationalist, cosmopolitan, or law and order categories behave as informative signals of the latent dimensions, and how strongly they do so. Items with weak or flat estimated response functions contribute less information than items with steep and theoretically ordered response functions.

The Manifesto Project mapping uses 34 Comparative Manifesto Project (CMP) categories organized into pro-market, pro-welfare, cosmopolitan, and traditional item families following the logic of text-based policy scaling [6, 24]. Economic items cover market regulation, welfare state expansion and limitation, economic liberalization, nationalization, education, and labor groups. Cultural items cover internationalism, European integration, multiculturalism, national way of life, traditional morality, freedom and human rights, and law and order. PolDem contributes nine directional media issue categories: economic liberalization, welfare state, immigration, cultural liberalism, nationalism, European Union integration, Euro currency, defense, and security.

All text and expert items are oriented before estimation so that higher expected responses correspond to the economic right or to the cultural traditionalist pole. For text data, each source category enters a single dimension-specific item family and contributes positive and sample counts for the relevant pole; categories used as opposing poles within a bipolar item are not double-counted within that item. For expert data, items whose original direction is opposite to the release convention are reversed in preprocessing. This includes the V-Party welfare-position, immigration-policy, lesbian, gay, bisexual and transgender (LGBT)-policy,

cultural-superiority, religion-policy, and gender-policy items. The source-variable and item-code mappings are implemented in the repository scripts that generate the model input files, and the diagnostics outputs document the resulting item coding and orientation checks.

Expert likelihood

Expert surveys provide bounded ordinal or quasi continuous ratings. Published expert means are treated as aggregate summaries of individual expert judgments. The consistently available information across sources is usually the mean placement, the scale bounds, and the number of experts, rather than the full distribution of individual category choices. The beta-binomial reconstruction uses these common inputs directly and avoids introducing source-specific cutpoint structures that would be weakly identified for sparse aggregate means with varying numbers of response categories. If an observation is based on K_n experts and an item has n_k scale steps, the reconstructed sum is modeled with a beta-binomial likelihood. For dimension-specific expert ratings, the expected position is

$$\mu_{d,n}^{\text{exp}} = \text{logit}^{-1}\left(\gamma_k^{\text{exp,int}} + \gamma_k^{\text{exp,slope}}\theta_{d(n),r(n)}\right), \tag{9}$$

and the observed rounded sum follows

$$y_n^{\text{exp}} \sim \text{BetaBinomial}\left(K_n n_k, \phi^{\text{exp}} K_n \mu_{d,n}^{\text{exp}}, \phi^{\text{exp}} K_n (1-\mu_{d,n}^{\text{exp}})\right). \tag{10}$$

The likelihood uses the number of experts as the main source of observation-level precision: for a given source and item, a mean based on many experts carries more information than a mean based on few experts. Scale granularity enters only through the reconstruction of published bounded aggregate ratings, because full category counts or individual expert responses are not consistently available across sources and periods. The estimated overdispersion and precision parameters prevent either a large expert count or a fine response scale from being treated as mechanically exact, allowing for inter-expert disagreement and source-specific noise.

General left-right expert ratings are used for cross-dimensional anchoring in the identification sense. They are not an output dimension of the released dataset. Instead, they provide information about how a one-dimensional left-right placement relates to the economic and cultural dimensions in a given country, source, and decade. The model combines the two latent dimensions through an observation-specific economic weight:

$$\tilde{\theta}_r^{\text{lr}} = w_n\theta_{1,r} + (1-w_n)\theta_{2,r}, \quad w_n = \text{logit}^{-1}\left(\alpha_0 + \delta_{c(n)} + \delta_{k(n)} + \delta_{d(n)}\right). \tag{11}$$

The offsets allow the content of left-right placements to vary by country, source, and decade. In some settings a general left-right rating can be more strongly economic; in others it can contain more cultural information. This component anchors the relationship between the two dimensions without imposing a fixed economic-cultural mixture across all countries and periods.

Party unions and alliances

Some party labels correspond to electoral unions or alliances. The Manifesto Project often codes joint platforms under an alliance label, while expert surveys usually rate constituent parties individually. Discarding these platforms would lose information; treating the alliance as a single party would conflate distinct parties. The model therefore uses mean constituent averaging. Union-to-constituent mappings were constructed from Manifesto Project programme-type information, PartyFacts metadata, party-name evidence, and manual verification of country-specific cases. The released party identifiers refer to individual parties; alliance labels are used only to link shared text evidence to constituent party trajectories. For a union u with constituents $\mathcal{C}(u)$, text observations coded under the union label constrain

$$\bar{\theta}_{d,r}^{(u)} = \frac{1}{|\mathcal{C}(u)|} \sum_{p \in \mathcal{C}(u)} \theta_{d,r(p)}. \tag{12}$$

Expert ratings for individual constituent parties continue to constrain those parties directly. The final mapping contains 106 constituent mappings linking 76 union or alliance identifiers to 100 constituent party identifiers across 34 countries. The repository includes the final mapping table and diagnostics outputs used to check that released identifiers represent individual parties rather than alliance records.

Identification, priors, and computation

Identification uses substantively oriented item polarities, positive text loadings, and expert slopes to fix dimension direction, a soft anchor prior on the initial position of Germany's Christian Democratic Union (CDU), and general left-right expert placements to anchor the relationship between economic and cultural dimensions. The CDU anchor places the reference party at approximately 0.55 on both dimensions after inverse logit transformation, consistent with a center-right and moderately traditionalist position. Shifting the anchor would shift the absolute location of the common scale but not the within-country party orderings or validation correlations.

The latent positions are estimated for party segments rather than as deviations from contemporaneous national party-system means. The model does not impose country-year centering or force the distribution of parties within a country-year to have a fixed mean or variance. Entry of a new party therefore adds evidence for that party and can affect other estimates only through shared measurement parameters, temporal structure, and posterior dependence, not through mechanical rescaling of all existing parties in the same party system. Because expert judgments may nevertheless contain national party-system referents, posterior uncertainty and the recorded source composition provide context for estimates around major party-system changes.

Several quantities are learned rather than fixed. Item intercepts and loadings determine how each source variable relates to a latent dimension. Country item deviations allow text categories to have different baselines or strengths across countries. Expert precision parameters determine how tightly expert observations constrain the latent position. Random walk variances determine how much information propagates across time. The identification

restrictions define the orientation and location of the latent space; they do not force all indicators to be equally informative.

Priors are weakly informative. The global random walk variance prior corresponds to a median year-to-year innovation of approximately 0.04 on the unit scale, while the hierarchical variance structure allows larger shifts where supported by data. Initial segment positions use weakly informative normal priors on the logit scale, with the CDU anchor fixing the common orientation and location. Text and expert item loadings and slopes use positive weakly informative priors centered on informative response functions, so the data determine whether a given item discriminates strongly or weakly. Country, source, decade, and item deviations use mean-zero hierarchical priors. Expert concentration parameters receive priors reflecting higher reliability for dimension-specific expert surveys than for general left-right ratings, because the latter must be decomposed into two dimensions.

The production model was estimated with Hamiltonian Monte Carlo using CmdStan. The final production run used four chains with 1,000 warmup iterations and 2,000 sampling iterations per chain, yielding 8,000 posterior draws. The model contained 35,258 monitored parameters. No divergent transitions were observed. Split $\hat{R}$ was below 1.01 for all monitored parameters, the maximum split $\hat{R}$ was 1.009, and the minimum effective sample size was 535.

Data Records

The dataset is available at Harvard Dataverse [25]. The dataset is published under DOI 10.7910/DVN/CDNTYJ; the versioned release files are available at the public Codeberg `v1` release. The primary data record supports comparative election-year analyses across 708 parties in 65 countries from 1944 to 2025. It contains one row per party segment at an election. Each row gives the model-estimated economic left-right and cultural cosmopolitan–traditionalist position of a party at an election, together with posterior uncertainty intervals and metadata describing the nearby text and expert evidence informing that estimate. Each row is a posterior summary from the full joint model, combining source-specific observations through the measurement and temporal structure described above. The annual latent trajectory is generated by the model and released as secondary output; the election-year panel is the primary release for applied research. The core identifiers are PartyFacts party identifier, parent union party identifier where relevant, segment number, a two-letter International Organization for Standardization country code (ISO2), election identifier where available, election date or year, and calendar year. Segment numbering records whether a party history was split because of long gaps in available observations. The main segment is coded as segment 1; higher segment numbers indicate additional periods after gaps exceeding seven years.

The main variables are `economic_lr` and `galtan`. `economic_lr` is the economic left-right posterior mean on the unit interval. `galtan` is the cultural cosmopolitan–traditionalist posterior mean on the unit interval, where lower values indicate cosmopolitan or socially liberal positions and higher values indicate traditionalist or nationalist positions. The variable name `galtan` preserves the connection to the GAL–TAN measurement tradition, while the text uses the descriptive cosmopolitan–traditionalist label for interpretation. For each dimension, the data include posterior standard errors and 2.5 and 97.5 percent posterior

quantiles. These uncertainty columns summarize how tightly the joint model constrains each position. The release also includes source-composition fields that record whether text or expert observations are direct, nearby, or temporally propagated for each election-year estimate.

Interpreting posterior uncertainty

A key advantage of this data record is that every party-position estimate is accompanied by posterior uncertainty, enabling researchers to account for the strength of the underlying evidence when comparing parties or tracing change over time.

In plain language, a released 95 percent credible interval is the range of latent party positions that remains plausible after the model combines the available, sometimes sparse, noisy or disagreeing text and expert evidence with the party's temporal trajectory. A wider interval means that the supplied evidence constrains the party position less tightly; it does not mean that the party necessarily moved more. The posterior standard deviation is a one-number summary of the same latent-position uncertainty, while the 2.5 and 97.5 percent quantiles delimit the reported 95 percent credible interval.

Technically, these intervals propagate finite and noisy observations, disagreement as represented by the source-specific likelihoods, uncertainty in item intercepts and loadings, relevant hierarchical parameters and the random-walk trajectory, conditional on the model, priors, source coding, party links and supplied inputs. They do not automatically capture unknown source miscoding, omitted evidence, incorrect party matches, unmodeled structural breaks, construct invalidity or every form of systematic cross-source bias. They are Bayesian probability statements about the latent position under the fitted model, not a frequentist guarantee that an unknown true value would fall inside 95 percent of intervals across repeated datasets. A posterior predictive interval considered in Technical Validation is a different object: it adds observation-level variation and describes a possible new expert rating, not only the latent party position.

The data dictionary defines these fields mechanically. Direct support refers to observations in the same calendar year as the election-year estimate. Nearby support records the distance to the nearest text and expert observations on the annual model trajectory. The `source_support_class` field distinguishes estimates with both text and expert support, text-only support, expert-only support, and estimates for which neither source type is observed near the election year and the posterior is mainly propagated through the temporal model. The metadata report the distance threshold used for the summary class, and the continuous distance fields preserve the underlying timing information. Table 3 lists the core fields in the released panel.

The annual model output records 17,585 segment-year observations, 708 unique parties, 39 parties with multiple segments, and a year range from 1944 to 2025. The primary data product is the election-year subset derived from this annual output. Release `v1` covers Albania, Armenia, Argentina, Austria, Australia, Azerbaijan, Bosnia and Herzegovina, Belgium, Bulgaria, Bolivia, Brazil, Belarus, Canada, Switzerland, Chile, Colombia, Costa Rica, Cyprus, Czechia, Germany, Denmark, Dominican Republic, Ecuador, Estonia, Spain, Finland, France, the United Kingdom, Georgia, Greece, Croatia, Hungary, Ireland, Israel, Iceland, Italy, Japan, South Korea, Sri Lanka, Lithuania, Luxembourg, Latvia, Moldova,

Field	Description
party_id	PartyFacts identifier for the individual party.
union_party_id	PartyFacts identifier of a parent union or alliance where applicable.
segment_num	Segment number within party, after splitting at gaps exceeding seven years.
country	Two-letter International Organization for Standardization (ISO2) country code.
year	Calendar year.
election_id	Election identifier where available in the election-year release.
election_date	Election date where available in the election-year release.
economic_lr	Economic position, 0 for left and 1 for right.
galtan	Cultural cosmopolitan–traditionalist position, 0 for cosmopolitan and 1 for traditionalist.
economic_lr_se	Posterior standard deviation for economic left-right.
galtan_se	Posterior standard deviation for the cultural cosmopolitan–traditionalist dimension.
economic_lr_q025	Lower bound of the 95 percent posterior interval for economic left-right.
economic_lr_q975	Upper bound of the 95 percent posterior interval for economic left-right.
galtan_q025	Lower bound of the 95 percent posterior interval for the cultural cosmopolitan–traditionalist dimension.
galtan_q975	Upper bound of the 95 percent posterior interval for the cultural cosmopolitan–traditionalist dimension.
has_text	Indicator for direct text evidence at or near the election year.
has_expert	Indicator for direct expert evidence at or near the election year.
n_text_sources	Number of distinct text source families contributing direct or nearby evidence.
n_expert_sources	Number of distinct expert source families contributing direct or nearby evidence.
nearest_text_distance	Absolute distance in years to the nearest text observation used to inform the trajectory.
nearest_expert_distance	Absolute distance in years to the nearest expert observation used to inform the trajectory.
source_support_class	Summary source-composition class distinguishing overlapping text and expert observations, text-only observations, expert-only observations, and estimates dominated by temporal propagation.

Table 3: Core fields in the released party-position panel.

Montenegro, North Macedonia, Malta, Mexico, the Netherlands, Norway, New Zealand, Panama, Peru, Poland, Portugal, Romania, Serbia, Russia, Sweden, Slovenia, Slovakia, Türkiye, Ukraine, the United States, Uruguay and South Africa. A machine-readable, versioned country table (`metadata/country_coverage_v1.csv`) gives the ISO2 code, display name, first and last election year, number of parties and rows, and source-support counts for

every country; the release location is given in Data Availability. This explicit versioned table is the up-to-date coverage pointer, while the count of 65 reported here refers specifically to `v1`.

Interpreting positions and changes

Because latent party positions are abstract quantities, Figure 2 uses recognizable party-year examples and historical trajectories to show what locations, distances and changes on the two common scales mean in substantive terms.

Labour in 1945 and the Brazilian Workers' Party in 2022 occupy almost the same lower-left region of Panel A in Figure 2. Their economic positions are 0.180 and 0.200, and their cultural positions are 0.251 and 0.198. Values around 0.20 on both dimensions therefore describe parties that are strongly left economically and cosmopolitan culturally relative to the parties in the panel, even when the cases come from different continents and are separated by 77 years. The French Communist Party in 1958 provides a different combination: its economic position of 0.016 is the most extreme leftward landmark, but its cultural position of 0.599 lies just above the midpoint. The Japanese Liberal Democratic Party in 1969 (0.423 economic, 0.487 cultural) and the African National Congress in 1994 (0.459, 0.391) locate recognizable cases nearer the middle of the two scales.

Comparing Law and Justice in 2005 with the French National Front in 1988 shows why both coordinates are needed. Their cultural positions are almost identical at 0.792 and 0.798, placing both in the strongly traditionalist region, whereas their economic positions are separated by 0.455: Law and Justice is economically leftward at 0.332 and the National Front economically rightward at 0.787. The Free Democratic Party in 2017 occupies another distinct combination, economically rightward at 0.744 but culturally cosmopolitan at 0.334. These comparisons show how the same scales locate parties across countries and historical periods without collapsing the economic and cultural dimensions into a single ranking.

Panels B–E use the same coordinates to show within-party change. Fidesz moves from 0.796 economically and 0.430 culturally in 1990 to 0.311 and 0.850 in 2022: a 0.485 shift toward the economic left and a 0.420 shift toward cultural traditionalism, consistent with its documented transformation from a liberal party into a dominant national-conservative and illiberal party [26, 27]. On this reporting scale, a change of 0.10 is approximately one fifth of the economic movement and one quarter of the cultural movement in this particularly pronounced case. Labour illustrates non-monotonic change: it moves from 0.180 economically and 0.251 culturally in 1945 to 0.540 and 0.506 in 1997, before returning to 0.286 and 0.342 in 2024. The shaded credible bands indicate how tightly the available evidence constrains each position at each election; wider bands denote less precise latent-position estimates.

Technical Validation

The validation strategy evaluates the measurement model that generates the released estimates and makes clear which property of the model-generated data product each check addresses. Table 4 summarizes the role and limits of each diagnostic. This distinction is important because no single diagnostic can establish all aspects of measurement qual-

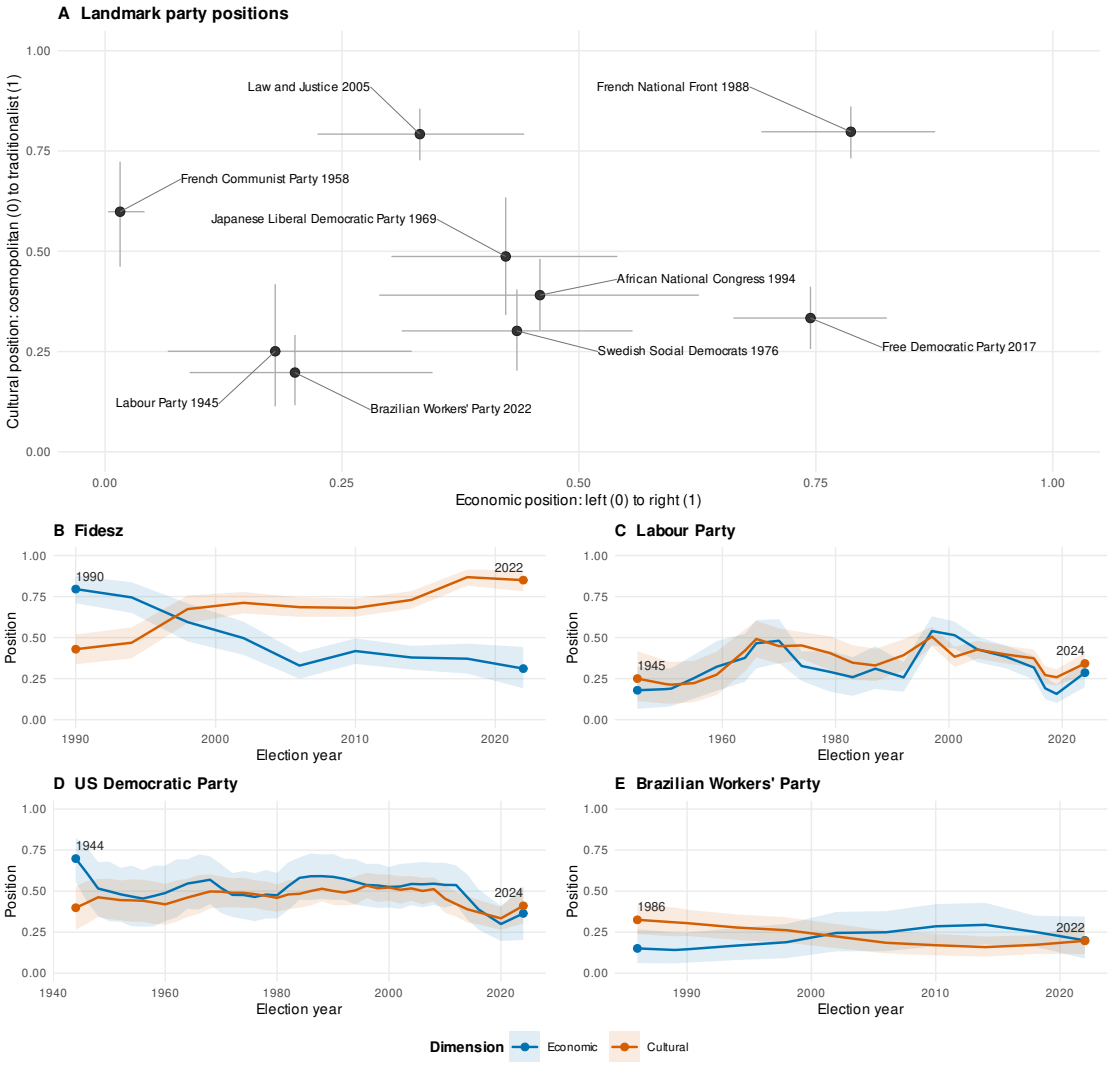


Figure 2: Landmark positions and temporal change on the common scales. Panel A shows nine labelled party-year posterior means, one from every decade between the 1940s and 2020s, jointly on the economic and cultural scales; horizontal and vertical bars indicate 95 percent latent-position credible intervals. Panels B–E show all released election-year posterior means for Fidesz (1990–2022), the Labour Party (1945–2024), the US Democratic Party (1944–2024) and the Brazilian Workers' Party (1986–2022). Economic and cultural positions are shown separately against election year, and shaded bands indicate 95 percent latent-position credible intervals, allowing historical positions and within-party changes to be read on the same scales.

ity. The evidence is cumulative: item response diagnostics test whether inputs behave as directional signals, held-out expert validation and posterior predictive coverage evaluate prediction and calibration, construct and discriminant checks assess substantive structure, evidence-composition diagnostics compare estimates informed by different source combinations, and robustness checks evaluate sensitivity to major design choices.

Check	What it tests	What it does not test
Item response diagnostics	Whether text and expert indicators show coherent, monotonic relationships with the estimated latent positions after item, source, and country differences are modeled.	They do not imply that raw source scores are directly interchangeable or that every historical estimate has equal precision.
Held-out expert validation	Whether the model can predict expert ratings withheld from estimation, using the remaining expert data, text indicators, and temporal structure.	It is focused on expert observation prediction; text-only and historically sparse estimates are assessed with complementary diagnostics.
Party-blocked validation	Whether text histories and shared model parameters recover expert placements for selected parties whose entire expert histories are excluded from estimation.	It evaluates eligible parties with both evidence types, not parties or eras for which no comparable expert criterion exists.
Convergent and discriminant checks	Whether posterior estimates align with expert observations on the same dimension and remain less associated with observations of the other dimension.	They do not replace source-specific diagnostics or guarantee that every input source is equally informative.
Posterior predictive coverage	Whether predictive intervals for expert observations are calibrated once latent uncertainty and observation-level noise are included.	It is not a substantive validation of party movement; it evaluates probabilistic calibration for observed ratings.
Evidence-composition balance	Whether posterior means are systematically shifted for election years informed by different nearby source types after accounting for country, decade, and vote share.	It is a balance diagnostic for the estimated positions; posterior intervals still summarize precision.
Robustness checks	Whether the election-year estimates are sensitive to a major source exclusion, plausible alternatives to the party history splitting threshold, or final item-screening choices.	They complement, rather than replace, source diagnostics, posterior uncertainty, and substantive checks.
Substantive face-validity	Whether party families and well-known party types fall in expected regions of the economic and cultural space.	It is a compact substantive check and cannot validate every individual party-year estimate.

Table 4: Interpretation of the technical validation checks.

Measurement structure and item response diagnostics

A central diagnostic is whether heterogeneous indicators display coherent item response behavior once modeled jointly. The model does not assume that a manifesto category, a media coding, and an expert survey rating have the same baseline, scale, or reliability. It estimates these measurement relationships through item-specific intercepts, loadings, source-specific likelihoods, country deviations, and observation-level uncertainty. The relevant empirical question is therefore whether the indicators contain a common directional signal after these

differences are modeled.

Figures 3 and 4 report estimated item response functions for representative economic and cultural indicators. These curves show how the expected probability of a text item or the expected expert rating changes across the latent position. The theoretically expected pattern is monotonic: pro-market, welfare limitation, and economically liberal indicators should increase as the economic right position increases, while welfare expansion and labor oriented indicators should decrease. Nationalist, traditional morality, immigration restriction, defense, security, and law and order indicators should increase as the cultural position moves toward the traditionalist pole, while cosmopolitan and socially liberal indicators should decrease. Expert survey items should show analogous monotonic mappings to the latent scale.

These diagnostics are important because they show that the model is not merely averaging heterogeneous input sources. The item response curves assess whether the model recovers a coherent latent structure in which substantively related indicators respond predictably to estimated party positions. Flat or weakly ordered curves identify noisy indicators, which contribute less information to the latent position. Country deviations allow the same item to have different baselines or strengths across contexts without requiring the raw source variables to be directly interchangeable. The appropriate interpretation is therefore not that the raw data sources are identical measures, but that they provide noisy, partially overlapping signals of two latent party positions under the estimated measurement structure.

The figures should be read as model diagnostics rather than substantive time series results. The black curve is the posterior mean response function for an item, and the grey curves show posterior uncertainty in that response function. Dashed curves identify items whose original scale direction was reversed before or within estimation so that higher reported values consistently correspond to rightward economic or traditionalist cultural positions. Steep monotonic curves indicate items that discriminate strongly across the latent dimension. Flat curves indicate items that contain little positional information after the model accounts for baseline frequency and source differences.

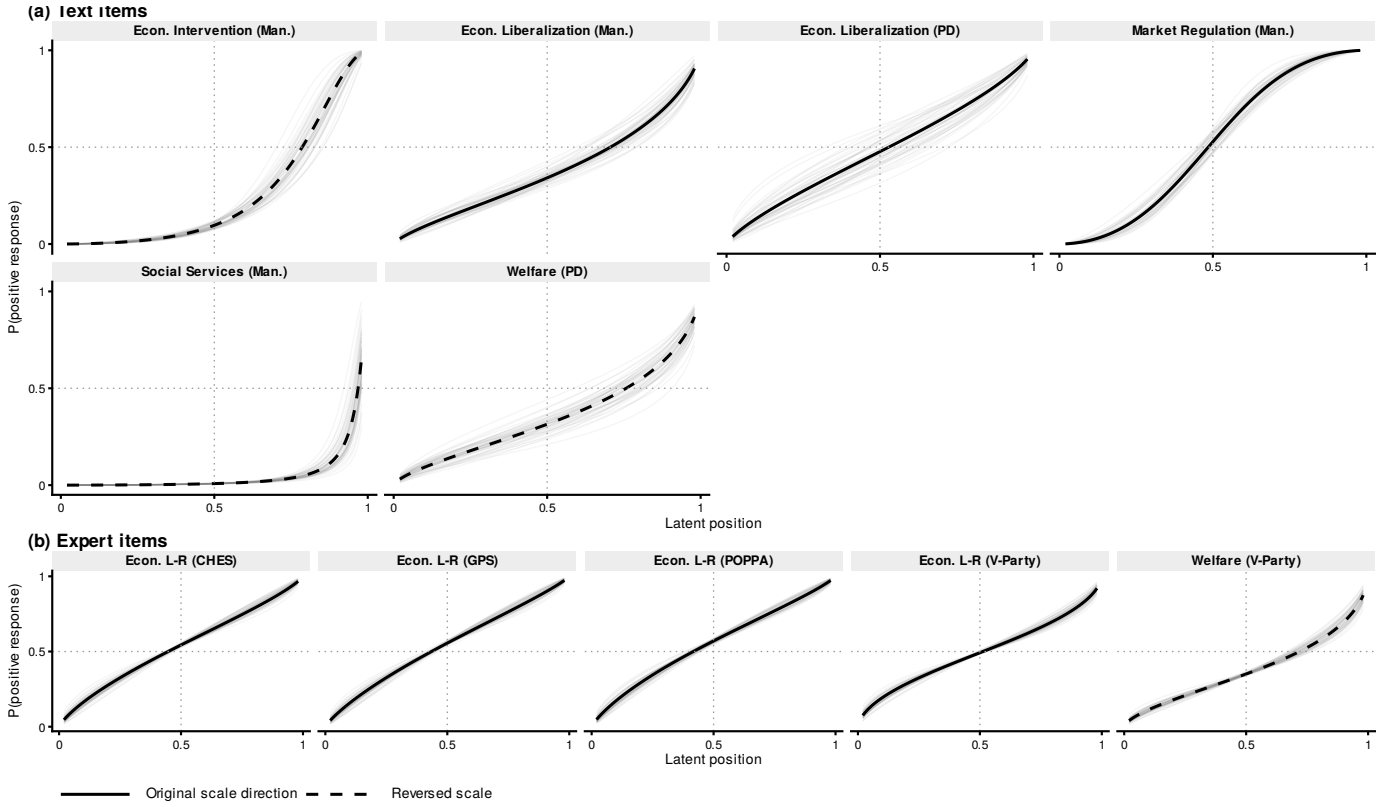


Figure 3: Item response diagnostics for the economic dimension. The panels show representative estimated response functions for economic text and expert indicators entering the latent measurement model. The curves summarize how the expected text response or expert rating changes as the latent economic left-right position changes, after estimating item-specific baselines, item loadings, source-specific likelihoods, country deviations, and observation-level uncertainty. Monotonic and substantively ordered curves indicate that heterogeneous indicators contain directional positional signal under the estimated measurement structure. Flat curves indicate weaker or noisier indicators and therefore less information for the latent position.

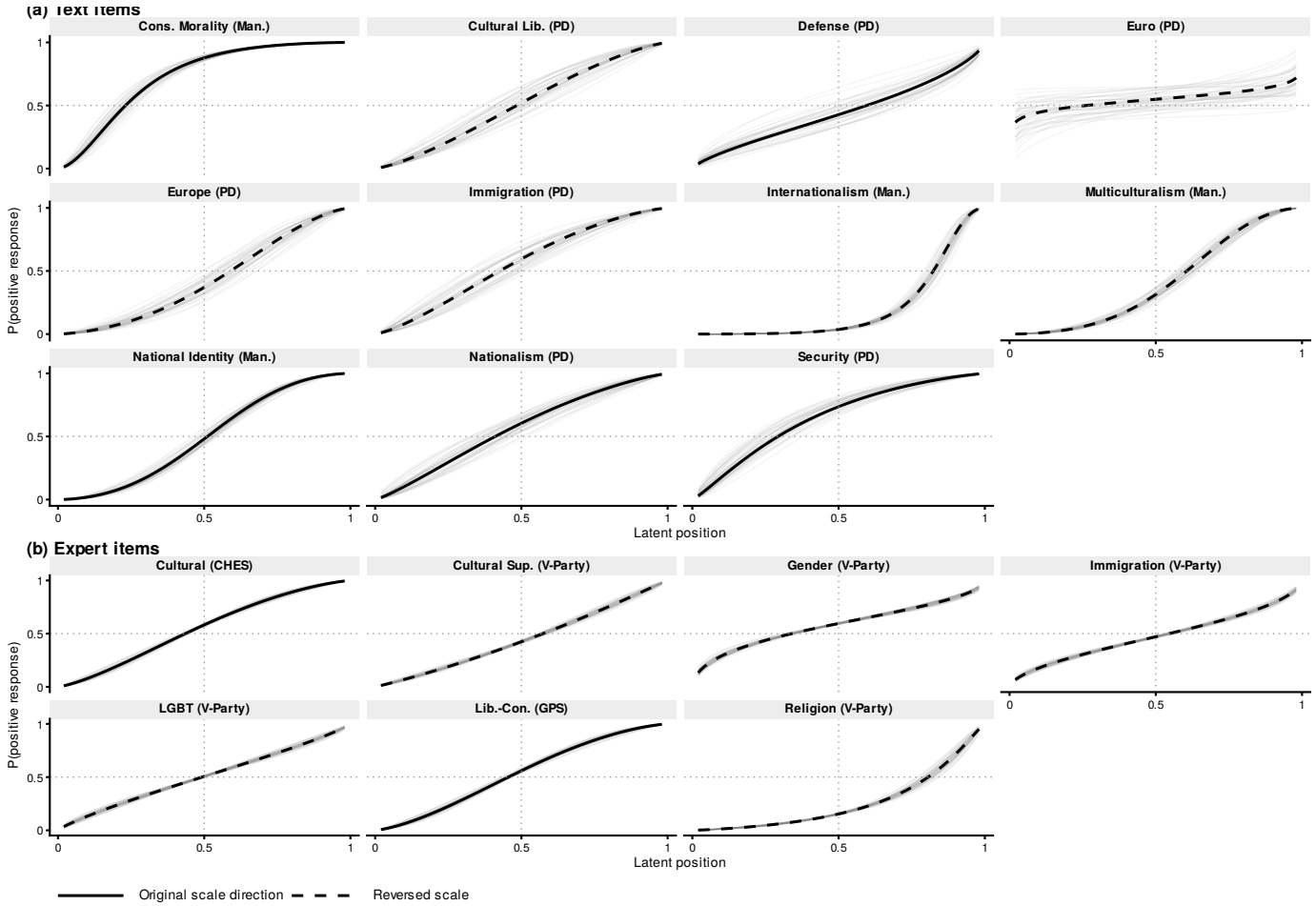


Figure 4: Item response diagnostics for the cultural dimension. The panels show representative estimated response functions for cultural text and expert indicators entering the latent measurement model. The curves summarize how the expected text response or expert rating changes as the latent cultural position changes, after estimating item-specific baselines, item loadings, source-specific likelihoods, country deviations, and observation-level uncertainty. Monotonic and substantively ordered curves indicate that heterogeneous indicators contain directional positional signal under the estimated measurement structure. Flat curves indicate weaker or noisier indicators and therefore less information for the latent position.

Diagnostic threshold	Parameters	Percentage
$\hat{R} < 1.01$	35,258	100.0
$\hat{R}$ 1.01 to 1.05	0	0.0
$\hat{R} > 1.05$	0	0.0
ESS $> 1,000$	34,853	98.9
ESS 400 to 1,000	405	1.1
ESS < 400	0	0.0

Table 5: Overall convergence diagnostics for the production model run.

Dimension	Pearson r	Spearman ρ	MAE	RMSE	n
Economic left-right	0.904	0.901	0.077	0.100	4,637
Cultural	0.960	0.967	0.078	0.098	1,425

Table 6: Convergent validity between posterior means and expert observations.

Convergence diagnostics

These diagnostics assess the reliability of posterior computation rather than construct validity; they establish that the reported posterior summaries are based on well-mixed Hamiltonian Monte Carlo draws. The production model mixed well by standard Hamiltonian Monte Carlo diagnostics. Across 35,258 monitored parameters, all parameters had split $\hat{R} < 1.01$, and no parameter had split $\hat{R} > 1.05$. Effective sample size (ESS) exceeded 1,000 for 98.9 percent of parameters and exceeded 400 for all parameters. The maximum split $\hat{R}$ was 1.009. The worst converging parameters were generally latent positions at temporal extremes of data sparse segments, where the posterior is mostly informed by the random walk prior rather than direct observations. Table 5 reports the overall convergence diagnostics.

Convergent validity against expert surveys

Convergent validity compares posterior mean estimates to expert observations measuring the same dimension. This check asks whether the latent estimates recover expert positional information when expert observations are compared to the corresponding party-year and dimension. It is not a claim that expert surveys are error free or that all text indicators are independently validated by this comparison. Economic left-right reaches Pearson $r = 0.904$ and Spearman $\rho = 0.901$ across 4,637 expert observations, with mean absolute error (MAE) 0.077 and root mean square error (RMSE) 0.100. Cultural positions reach Pearson $r = 0.960$ and Spearman $\rho = 0.967$ across 1,425 expert observations, with MAE 0.078 and RMSE 0.098. These results show that the posterior estimates recover the main same-dimension expert signal closely where comparable expert observations are available. Correlations are highest for recent expert surveys with continuous scales, and lower for coarser or more heterogeneous source items. Table 6 reports these convergent-validity results.

Held-out source and dimension	Pearson r	MAE	n
CHES economic left-right	0.973	0.056	219
POPPA economic left-right	0.980	0.066	74
V-Party economic left-right	0.860	0.091	509
V-Party welfare-position item	0.673	0.178	515
CHES cultural	0.959	0.077	218
V-Party cultural-superiority item	0.812	0.128	515
V-Party immigration-policy item	0.749	0.103	514
V-Party LGBT-policy item	0.798	0.094	519
V-Party religion-policy item	0.674	0.302	529
V-Party gender-policy item	0.571	0.169	527

Table 7: Out-of-sample validation using held-out expert observations.

Held-out expert validation

Out-of-sample validation was conducted by re-estimating the model after holding out 20 percent of expert and general left-right observations, stratified by source. The held-out observations were removed from the training likelihood before estimation; they were not merely masked after fitting the full model. Text data were retained in full so that the same parties and segment structure remained in the estimation. Expert-only parties were protected from holdout to avoid dropping parties from the training data. Held-out expert observations were then compared to posterior means from the training run.

This exercise tests whether the integrated measurement model can reproduce expert placements that were not supplied to estimation. It is a convergent prediction test for held-out expert ratings, complementing the item response, substantive, and uncertainty diagnostics used for text-only and historically sparse estimates. For CHES, held-out correlations are nearly identical to in-sample correlations: $r = 0.974$ for economic left-right and $r = 0.959$ for the cultural dimension. POPPA economic reaches $r = 0.980$. Aggregate held-out correlations are lower for retrospective and coarser expert items, especially on the cultural dimension.

This design addresses a specific question: whether the remaining expert data, text indicators, and temporal structure can predict expert placements that were not used to estimate the model. The stronger results for CHES and POPPA and the weaker results for selected V-Party items indicate that predictive performance depends on source granularity and item definition rather than being uniform across all expert evidence. The broader validation evidence for historical and text-supported election years comes from the combination of held-out prediction, item response diagnostics, source-composition balance, posterior uncertainty, and substantive validity checks. Table 7 reports the held-out expert validation results.

Party-blocked expert validation

The random holdout above retains other expert evidence for the same party. I therefore conducted a harder predeclared stress test in which the full expert history was excluded

for selected parties while their text histories remained. From 388 eligible parties with both evidence types and at least three text years, 82 parties were selected with stratification by region and first expert period. All 3,916 dimension-specific expert rows and 310 general left–right rows for those parties were removed before estimation, leaving their text observations and the evidence for unselected parties available to identify the common scale. The blocked fit produced matches for 3,523 withheld dimension-specific ratings from 76 selected parties. For the other six selected parties, the withheld expert ratings fall outside the party-year range defined by the retained text trajectory, so no same-party-year latent estimate exists for a like-for-like score.

This is a deliberately demanding test that differs from the model's usual use. Removing every expert observation for a selected party across its full history removes party-specific expert anchors, which are a central part of how the model links text and expert evidence over time through joint cross-modal anchoring and Bayesian forward and backward propagation. The test therefore asks how much party-specific expert information can be recovered from text histories and information learned from other parties when the selected party is denied its own expert anchors; it does not test the integrated production model under its usual information structure. Even under this stripped-down design, agreement remains informative ($r = 0.548$ for economic left-right and $r = 0.388$ for cultural position; Table 8). Lower agreement than in the ordinary holdout is expected, because this test removes the party-specific expert information used to calibrate its text evidence. Combining incomplete, noisy and differently timed evidence is precisely why the dynamic Bayesian measurement model uses direct expert placements to calibrate text evidence and propagate information through uneven party trajectories. Results are similar or slightly stronger when the nearest text observation is one to three years away rather than in the same year, but this contrast is descriptive and the latter group is small. Detailed decade, region, source, item and text-distance results are released with the validation artifacts. The 95 percent latent credible intervals contain 51.7 percent of withheld economic and 41.3 percent of withheld cultural expert means. This is an interval-overlap diagnostic, not posterior predictive coverage: the latter would also model the item-specific expert-observation process.

The validation fit retained 1,000 post-warmup iterations per chain, rather than 2,000 for the production release, because this deliberately demanding four-chain sensitivity analysis is computationally intensive. Diagnostics were nevertheless good overall: there were no divergent transitions or maximum-treedepth hits, mean split-$\hat{R}$ was 1.001, and 99.88% of monitored quantities were below the deliberately strict 1.01 threshold (maximum 1.025). I therefore interpret it as a sensitivity analysis, not as a replacement production release. It supports a limited conclusion: because joint expert-text anchoring and Bayesian information propagation are the model's central mechanism, removing a party's expert history disables that core feature and reduces predictive agreement, especially on the cultural dimension. It does not reflect production release accuracy where expert survey anchors are present.

Dimension	Text support	n	r	MAE	RMSE	Interval overlap
Economic	All matched ratings	1,095	0.548	0.157	0.204	51.7%
Economic	Direct text	971	0.547	0.159	0.208	50.6%
Economic	Nearby text (1–3 years)	124	0.627	0.134	0.165	60.5%
Cultural	All matched ratings	2,428	0.388	0.189	0.238	41.3%
Cultural	Direct text	2,293	0.382	0.190	0.240	40.5%
Cultural	Nearby text (1–3 years)	135	0.440	0.167	0.206	54.8%

Table 8: Party-blocked validation against withheld expert ratings. Expert evidence for selected parties was artificially removed before fitting to evaluate text-only extrapolation without expert anchors.

Discriminant and construct validity

Discriminant validity uses the multitrait multimethod logic of [28]. This check asks whether the two latent dimensions recover different positional signals rather than a single general party ideology score. The model's economic estimates correlate highly with expert economic ratings but much less with expert cultural ratings. Conversely, cultural estimates correlate highly with expert cultural ratings but much less with expert economic ratings. The cross-dimension correlations are 0.393 and 0.422, substantially below the same dimension correlations above 0.90.

Construct validity was assessed using party family orderings and a compact substantive inspection of the estimated two-dimensional space. These checks evaluate whether the estimates recover broad political patterns that should be visible if the scales have substantive content. Party families were classified using ParlGov linked through PartyFacts. Economic family means follow the expected left-to-right ordering, with Communist and Far Left parties at 0.12, Green parties at 0.27, Social Democratic parties at 0.33, Christian Democratic parties at 0.60, Radical Right parties at 0.62, Conservative parties at 0.65, and Liberal parties at 0.66. The Spearman rank correlation for a five-family core ordering excluding Liberals is 0.90. On the cultural dimension, Green parties are the most cosmopolitan at 0.25 and Radical Right parties the most traditionalist at 0.72, yielding a Spearman rank correlation of 1.00 for the six-family ordering used in the final validation.

The same pattern is visible substantively rather than only statistically. Figure 5 displays the economic and cultural dimensions jointly using decade-averaged party positions. The figure shows that the two dimensions form interpretable regional and temporal party spaces rather than a single undifferentiated left-right score. It also shows that the association between economic and cultural positions varies across regions and decades, which supports the decision to estimate the two dimensions separately. Green and social democratic families occupy the left and culturally cosmopolitan region; communist and far left parties are the most economically left; Christian democratic, conservative, and liberal families are on the economic right; and radical right parties are the most culturally traditionalist. These descriptive regularities are not used as input to the measurement model. They therefore provide a concise face-validity and substantive robustness check: the model recovers familiar cross-party structure while still allowing individual parties and party-years to deviate from

Party family	Parties	Economic mean	Cultural mean
Communist/Far Left	49	0.12	0.38
Green/Ecological	30	0.27	0.25
Social Democratic	86	0.33	0.38
Christian Democratic	41	0.60	0.53
Radical Right	50	0.62	0.72
Conservative	82	0.65	0.54
Liberal	81	0.66	0.37

Table 9: Substantive face-validity by party family.

family averages. Because country coverage varies across region-decade cells, Figure 5 should be read as a compact face-validity display of the estimated party space rather than as evidence about regional temporal change. Table 9 reports the party-family means underlying this check.

Posterior predictive coverage

Posterior predictive coverage checks whether observed expert ratings fall inside predictive intervals for new observations at the corresponding party-year and dimension. Unlike a credible interval for the latent position alone, this interval adds observation-level measurement variation. Economic left-right achieves 89.8 percent coverage for the nominal 95 percent posterior predictive interval, with 6,698 of 7,455 observations covered. Cultural estimates achieve 84.8 percent coverage, with 13,170 of 15,539 observations covered. Calibration curves show the same shortfall across interval levels: nominal 50, 80, 90 and 95 percent intervals cover 48.9, 75.9, 84.9 and 89.8 percent of economic ratings and 42.4, 67.9, 77.9 and 84.8 percent of cultural ratings.

The misses are concentrated rather than uniform. V-Party accounts for 2,985 of 3,126 misses (95.5 percent), reflecting its large number of observations and several coarse items. Within the cultural dimension, the V-Party gender-policy item has 75.6 percent coverage and contributes 683 misses, the religion-policy item has 80.0 percent and 566 misses, the cultural-superiority item has 84.4 percent and 441 misses, the immigration-policy item has 87.9 percent and 341 misses, and the lesbian, gay, bisexual and transgender (LGBT)-policy item has 91.2 percent and 249 misses. The V-Party welfare-position item is the main economic exception, with 80.2 percent coverage and 557 misses. In contrast, the Chapel Hill Expert Survey (CHES) covers 97.3 percent of economic and 94.3 percent of cultural ratings; the Populism and Political Parties Expert Survey (POPPA) covers 98.7 percent of economic ratings. Cultural coverage is 82.1 percent in the high-volume 2010s and 84.3 percent in the 2000s, compared with 97.7 percent for the smaller 2020s sample. Denmark and Sweden contribute the largest country counts of cultural misses (116 and 105), but these descriptive cells also differ in item composition and are not identified country effects. Binning by fitted value, expert count and interval width shows no single endpoint or interval-width pattern that explains the shortfall as coherently as source and item. Figure 6 displays the full item comparison, and the repository provides country, decade, source, item and residual-pattern

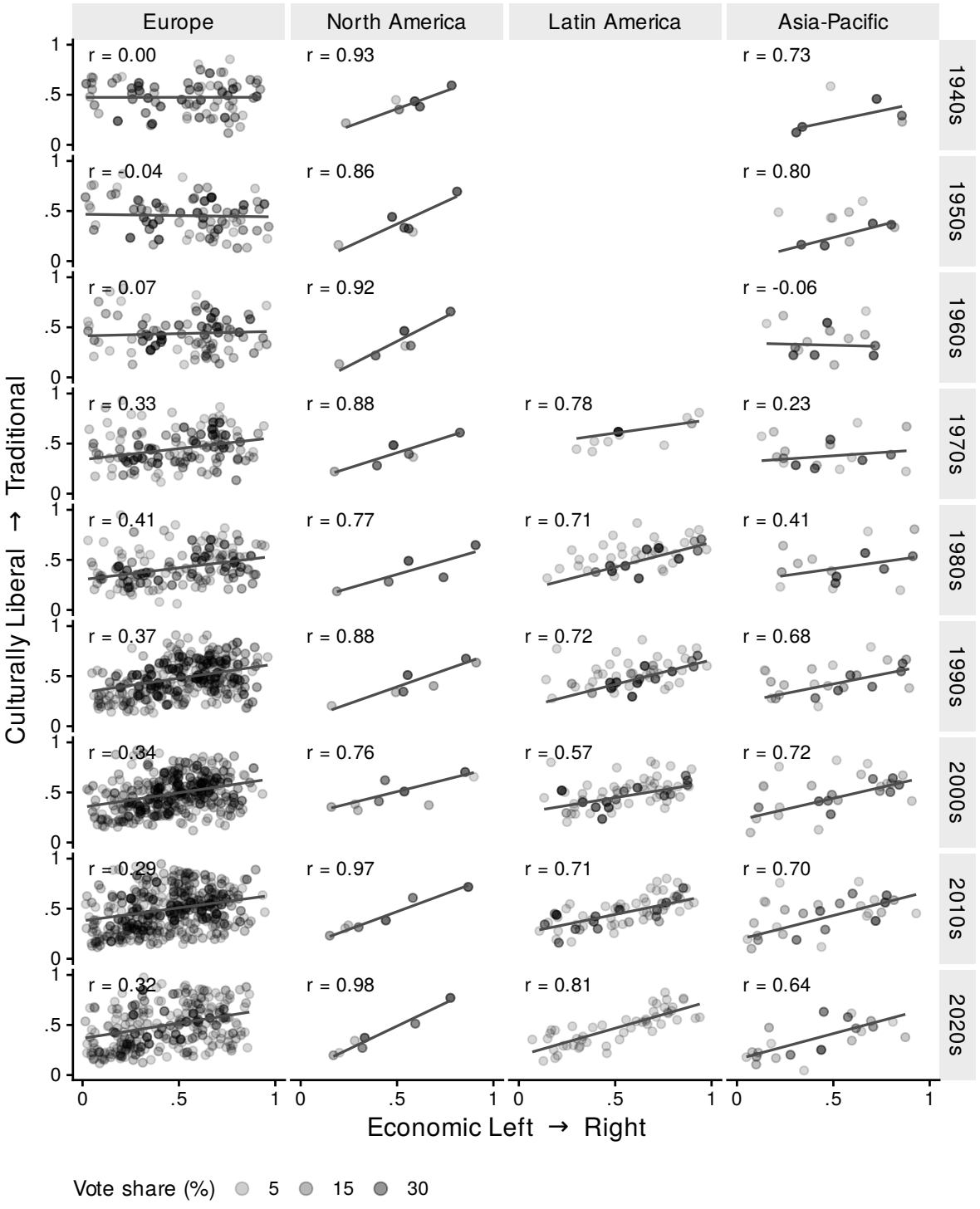


Figure 5: Face-validity check of the two-dimensional party-position space. Points show decade-averaged party positions by region, with opacity reflecting mean vote share. The figure illustrates that the economic and cultural dimensions recover substantively interpretable party space structure and that the relationship between the two dimensions varies across regions and periods.

Dimension	Coverage percent	Covered	n
Economic left-right	89.8	6,698	7,455
Cultural	84.8	13,170	15,539

Table 10: Posterior predictive coverage for expert observations.

tables with Wilson intervals, predictive-interval widths and each cell's share of all misses.

The concentration of misses in these items is consistent with a difference between contemporaneous and retrospective expert evaluation. Whereas real-time expert surveys (such as CHES or POPPA) evaluate party positioning contemporaneously during election cycles, historical expert projects (such as V-Party) rely on retrospective expert recall across multi-decade coding windows to achieve impressive temporal depth back to 1970. Retrospective coding over long historical windows can introduce additional observation-level variation and rounding in coarse items. The concentration by source and item points primarily to the expert-observation component rather than to a uniform failure of latent-position uncertainty. The model allows for source- and item-specific measurement relationships, expert precision, latent-position uncertainty and temporal dependence, but predictive coverage shows that the remaining mismatch is concentrated in a small set of coarse retrospective V-Party items. Treating that mismatch as a generic interval-width problem would impose an unvalidated dispersion correction and risk systematically biasing the common latent scale. Changing this calibration would require independently justified item- and source-specific observation-error assumptions, followed by a complete production re-estimation. I therefore retain the production estimates, report the item evidence prominently, and advise users not to treat predictive intervals for all expert items as having exact nominal coverage. This finding is distinct from the released latent-position credible intervals, which remain model-conditional summaries whose limitations are stated in Data Records and Usage Notes. Table 10 reports the overall posterior predictive coverage results.

Evidence-composition balance

The model is designed to place estimates informed by different combinations of text and expert data on the same latent scale. A direct diagnostic is therefore whether the posterior means are systematically displaced when the nearby observations are mainly text based, mainly expert based, or temporally propagated. I estimated linear balance regressions on the election-year panel, separately for the economic and cultural dimensions, with indicators for source-composition class and fixed effects for country and decade, plus log vote share. Standard errors are clustered by country. The reference category is election years with both text and expert evidence observed directly or nearby; each model uses 4,918 election-year rows.

Table 11 reports adjusted differences in posterior means on the unit scale with 95 percent confidence intervals. Text-only election years differ from the overlapping category by 0.015 economically and 0.018 culturally; both intervals include zero. The expert-only cultural contrast is 0.040 (0.005–0.074), while the remaining pooled intervals include zero. These are descriptive balance diagnostics for posterior means, not causal effects of adding a source.

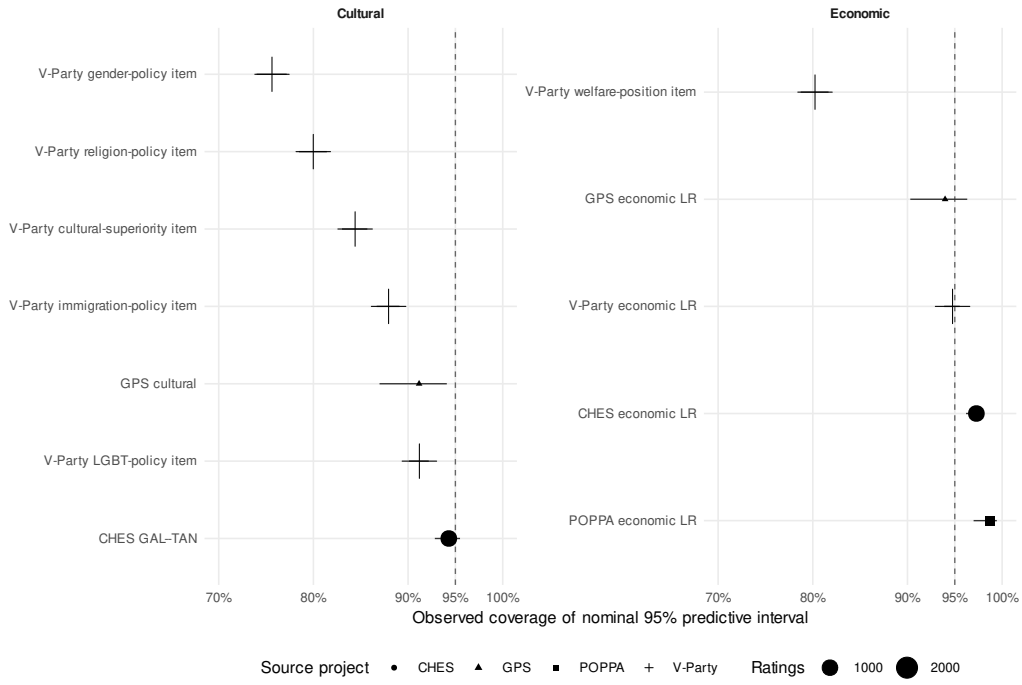


Figure 6: Observed item-level coverage of nominal 95 percent posterior predictive intervals for expert ratings. Horizontal bars are 95 percent Wilson intervals and point size represents the number of ratings. The dashed line marks 95 percent. The shortfall is concentrated in coarse V-Party items; recent multi-expert surveys are closer to nominal coverage.

To test whether pooled coefficients conceal cancellation, I also estimated a mixed model with a country intercept and partially pooled text-only contrasts by region and decade, retaining log vote share and pooled expert-only and temporal-propagation indicators. The latter categories have only 237 and 84 rows and do not support stable subgroup slopes. Figure 7 shows that the economic text-only contrast varies more by decade (partially pooled estimates from -0.061 to 0.090) than by region, whereas the cultural contrast varies more by region (-0.058 to 0.028) and is smaller across decades (-0.042 to 0.001). Most intervals include zero. The earliest decades contain almost exclusively text-supported cases, so their contrasts rely strongly on partial pooling and should not be read as direct within-decade comparisons. The results rule out neither contextual differences nor uncertainty, but they make the possible cancellation visible rather than relying on one pooled point estimate.

Robustness of election-year estimates

I also re-estimated the model under sensitivity specifications focused on the released election-year data product. These checks ask whether the main released estimates change substantially when a major input source, an important preprocessing rule, or a final item-screening decision is changed. First, I estimated a source-ablation specification that removes V-Party. V-Party is selected for single-source ablation because it is the largest expert dataset in the model and because, unlike real-time contemporaneous surveys (CHES, POPPA, GPS), it re-

Comparison class	Category n	Economic difference (95% confidence interval)	Cultural difference (95% confidence interval)
Text only direct or nearby	1,658	0.015 (−0.023, 0.053)	0.018 (−0.022, 0.058)
Expert only direct or nearby	237	−0.015 (−0.083, 0.052)	0.040 (0.005, 0.074)
Temporal propagation	84	−0.051 (−0.165, 0.064)	−0.045 (−0.201, 0.111)

Table 11: Evidence-composition balance in the election-year panel. Entries are adjusted differences in posterior means relative to election years with both text and expert evidence direct or nearby, with country-clustered 95 percent confidence intervals. Models include country and decade fixed effects and log vote share; model $n = 4,918$.

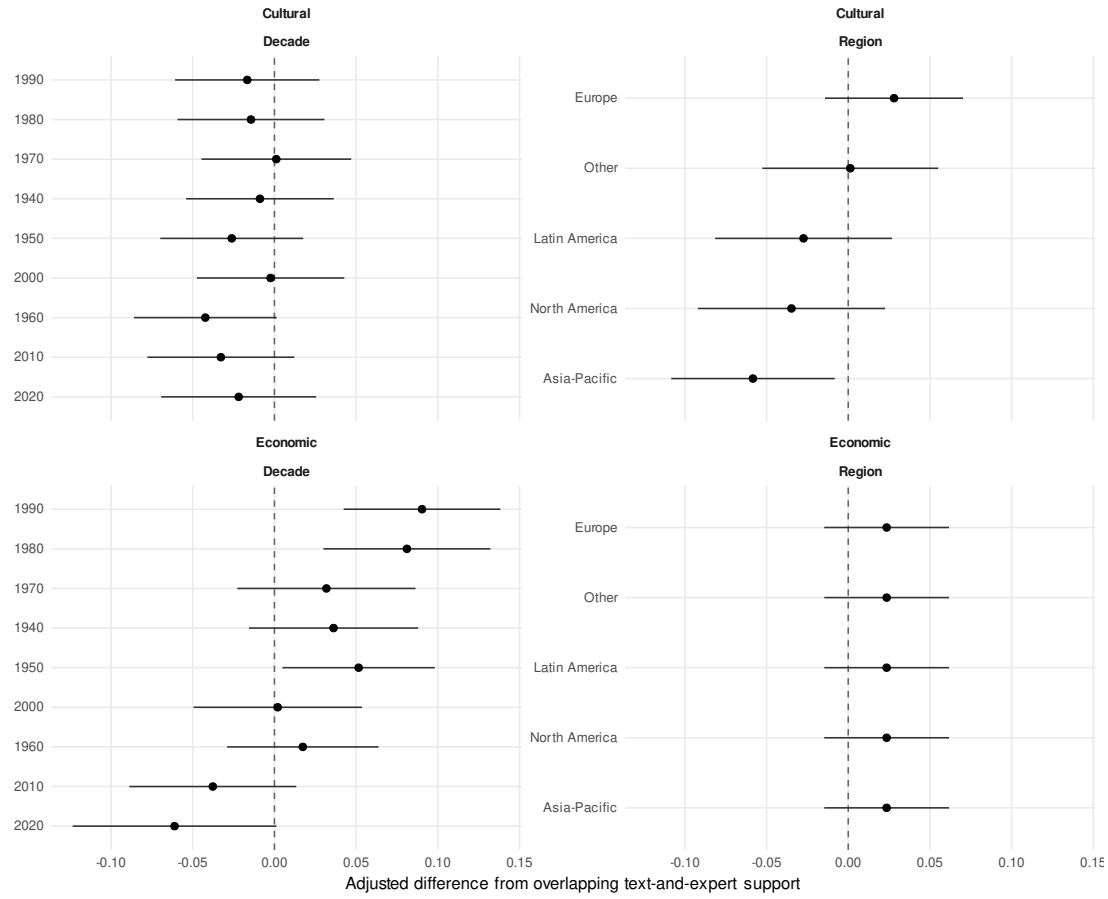


Figure 7: Partially pooled adjusted text-only contrasts by decade and region. Points compare text-only election years with rows having nearby text and expert evidence; bars are approximate 95 percent conditional intervals. Models retain country structure, log vote share and pooled indicators for the sparse expert-only and temporal-propagation classes. Early-decade comparisons have little overlap and are correspondingly pooling-dependent.

lies on retrospective expert recall across multi-decade coding windows. Second, I changed the party history splitting threshold from the production value of seven years to five years and ten years. Third, I compared the final release to the preceding two-dimensional production run that still included one PolDem economic item later excluded because its item response was flat. All comparisons use the released election-year rows as the target set and match estimates by PartyFacts identifier, country, and year, so that changes in segment numbering do not mechanically prevent matching.

The gap threshold sensitivities produce nearly identical election-year estimates. Correlations with the production estimates exceed 0.999 on both dimensions, median absolute differences are at most 0.006, and posterior interval widths are essentially unchanged. This shows that the production estimates are not an artifact of the seven-year segmentation threshold. The final item-screening check is even smaller: correlations with the preceding run are 0.999985 on both dimensions, mean absolute differences are below 0.003, and no matched election-year estimate differs by more than 0.01.

The recovered row-level V-Party ablation permits a more specific assessment. Rematching it to the final release yields 4,629 party-election rows. Correlations are 0.948 economically and 0.891 culturally; mean absolute differences are 0.054 and 0.079, and 95th-percentile absolute differences are 0.167 and 0.231. Mean credible-interval widths change from 0.229 to 0.239 economically and from 0.207 to 0.339 culturally. Sensitivity is not evenly distributed. Latin America has the largest regional economic mean absolute change (0.117), while Asia-Pacific has the largest cultural change (0.173). At country level, Australia is most sensitive culturally among countries with at least ten matched rows; Mexico and Colombia are among the most sensitive economically. The 1990s show the largest decade-average economic change (0.064), while the 1970s show the largest cultural change (0.099). By source-support class, the small temporal-propagation group ($n = 62$) has the largest mean absolute changes (0.102 economic and 0.109 cultural), compared with 0.049 and 0.071 for text-only rows and 0.056 and 0.083 for rows with both source types nearby. Agrarian and conservative families have the largest mean cultural changes (0.105 and 0.098). Figure 8 and the complete row-, country-, decade-, region-, family- and support-class tables allow users to identify estimates that depend materially on this source.

The largest displayed cases were checked against the party identifiers and labels in the final panel. They include the Mexican Green Ecologist Party in 2006, whose economic estimate changes from 0.783 to 0.303 without V-Party, and the Australian Democrats in 2001, whose cultural estimate changes from 0.339 to 0.853. These are sensitivity diagnostics, not evidence that either specification is substantively correct. The aggregate common signal persists, but individual scores and cultural uncertainty can depend heavily on V-Party in the contexts identified above.

The segmentation and item-screening results show strong stability. The V-Party ablation, however, reveals material case-level and regional dependence despite preserving the aggregate ordering, especially for cultural estimates. Users studying the highlighted contexts should inspect the case-level sensitivity table and source-support metadata. Table 12 reports the sensitivity results.

Specification	Dimension	Matched n	r	Mean $\lvert\Delta\rvert$	Median $\lvert\Delta\rvert$	P95 $\lvert\Delta\rvert$
Source ablation: no V-Party	Economic	4,629	0.948	0.054	0.037	0.167
Source ablation: no V-Party	Cultural	4,629	0.891	0.079	0.060	0.231
Final item screening	Economic	4,924	1.000	0.003	0.003	0.005
Final item screening	Cultural	4,924	1.000	0.003	0.003	0.005
Gap threshold 5 years	Economic	4,244	0.999	0.003	0.002	0.007
Gap threshold 5 years	Cultural	4,244	0.999	0.002	0.001	0.006
Gap threshold 10 years	Economic	4,265	0.999	0.006	0.006	0.011
Gap threshold 10 years	Cultural	4,265	1.000	0.003	0.002	0.006

Table 12: Sensitivity of election-year estimates to source exclusion, segmentation choices, and final item screening. The source-ablation rows remove V-Party, the largest expert source in the model. The final item-screening rows compare the release to the preceding production run before removal of one weak PolDem economic item. P95 denotes the 95th percentile.

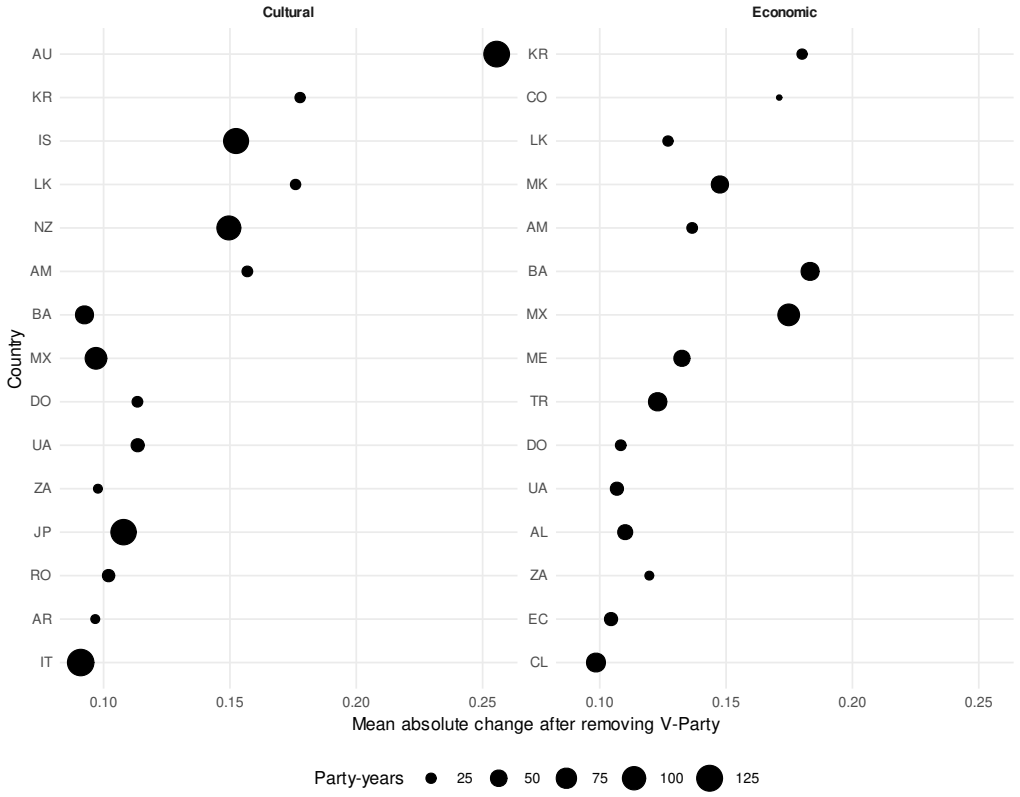


Figure 8: Countries with the largest mean absolute election-year changes after removing V-Party, among countries with at least ten matched rows. Point size represents the number of matched party-election rows. The complete machine-readable table includes every country, decade, region, party family and source-support class.

Usage Notes

The panel is designed for comparative election-year analyses across 708 parties in 65 countries from 1944 to 2025. It supports analyses of party-system change, long-run ideological movement, cross-national party competition, and the relationship between economic left-right and cultural cosmopolitan–traditionalist positioning using a common model-estimated latent scale. It is especially suited to designs that require party-level ideological measures to be linked to election results, regional conditions, voter preferences, institutions, or other contextual data across many party systems. The estimates are model-based latent positions interpreted as posterior summaries on that common scale, not as raw manifesto scores or raw expert survey scores. Estimates from all source-composition classes are placed on the same model-estimated latent scale. Posterior intervals summarize how tightly the full evidence base constrains each estimate, and the source-composition fields document the text and expert observations around each election year.

The annual latent trajectory is a secondary model output for research designs requiring annual covariates or inter-election estimates. Unlike simple linear interpolation between election-year endpoints, it combines irregularly timed text and expert evidence through the same dynamic measurement model used for the primary panel. The annual values remain posterior estimates, and their intervals widen when direct evidence is temporally distant, preserving information about the strength of the temporal link.

The cultural dimension is best interpreted as a broad cosmopolitan–traditionalist axis in the GAL–TAN tradition. It includes national identity, multiculturalism, immigration, moral conservatism, civil liberties, internationalism, defense, security, and law and order. Early decades rely more heavily on manifesto data than expert survey data, especially on the cultural dimension. Users should therefore inspect the source-support class, nearest-observation distances and credible-interval width rather than assume that every party-year has the same evidentiary basis. The common output scale makes model-conditional comparisons possible, while the metadata preserve important differences in support.

The party-blocked validation complements the random holdout by withholding the entire expert history of selected parties while retaining their text histories. It is consequently the more relevant test for the eligible text-supported parties it contains, but it is not evidence about parties or periods with no comparable expert criterion, nor does it turn the released estimates into directly observed historical positions. Users studying text-only or temporally propagated estimates should treat the source-support metadata and interval width as part of the design, report sensitivity to evidence support where possible, and avoid generalizing the validation result beyond the blocked sample.

The validation results also imply specific cautions. Nominal 95 percent posterior predictive intervals for expert ratings cover 89.8 percent of economic and 84.8 percent of cultural observations, with misses concentrated in coarse V-Party items; those predictive intervals should not be treated as exactly calibrated across items. This finding is distinct from the released latent-position credible intervals. Removing V-Party preserves strong aggregate association but produces large changes for some parties and countries and materially widens cultural intervals, so applications centered on the highlighted contexts should consult the case-level ablation table. Partially pooled source-support contrasts are generally modest but not uniform across region and decade, and the earliest text-dominant decades have little

within-period source-class overlap.

Analyses that require unmodeled raw source scores or source-specific measurement traditions are better served by the original source projects. The model assumes that parties have latent economic left-right and cultural cosmopolitan–traditionalist positions that generate observable expert ratings and text-coded policy signals with error. Cross-country and over-time comparisons are therefore model-based comparisons on a common latent scale. Source-specific parameters, country deviations, temporal smoothing, and uncertainty estimates are designed to absorb many differences in source behavior, party-system context, and temporal coverage.

Data Availability

The dataset described in this Data Descriptor is deposited at Harvard Dataverse [25] under a CC0 waiver and DOI 10.7910/DVN/CDNTYJ. The deposit contains the primary election-year party-position panel, the secondary annual model output, the metadata dictionaries, the versioned country-coverage file, the validation summaries and the complete code tree. Inference is released as posterior summaries; the full posterior bundle is available from the Codeberg release. The versioned release files are additionally available from the public `v1` release. The public project repository contains the model-ready inputs, metadata, documentation, validation artifacts and source-setup workflow used to reproduce the released files. The up-to-date machine-readable country list for this release is the versioned country-coverage file (`metadata/country_coverage_v1.csv`) in the deposit and repository; its release field ties it to `v1`, and the full list is also reproduced in Data Records.

The model uses third-party input datasets from existing text and expert-survey projects. Where redistribution is restricted by the original providers, the repository documentation records the source, version, variables used, access route, and processing steps required to reconstruct the model inputs from the original data.

Code Availability

The code required to reproduce data construction, model estimation, posterior processing, validation analyses, diagnostics report generation, and release-file construction is available from the public project repository, and a frozen code snapshot is deposited alongside the data at Harvard Dataverse [25]. The repository contains the Stan model, R and shell scripts for source setup and model-ready input reconstruction, Julia scripts for Stan data preparation, model fitting, post-estimation, enrichment, and validation, and documentation for reproducing the final results from the committed model-ready inputs. The source setup workflow downloads script-accessible inputs, checks locally supplied restricted inputs, rebuilds model-ready files under a local workspace, and compares them with the committed model-ready inputs without overwriting the committed files.

Some preprocessing scripts require third-party input datasets that cannot be redistributed under their original licences. For these sources, the repository provides version information, access instructions, variable mappings, and processing scripts.

Acknowledgements

Artificial intelligence (AI)-assisted tools, including OpenAI ChatGPT and Anthropic Claude Code, were used during manuscript preparation and project development to support language editing, manuscript organization, code drafting and debugging, and data-management workflows. All data harmonization decisions, model specifications, dimensional mappings, variable selections, validation analyses, interpretations, and final wording were reviewed, tested where applicable, and approved by the author. No AI tool is listed as an author, no model training or fine-tuning was performed, and the author takes full responsibility for the manuscript, code, data release, and conclusions.

Author Contributions

Armin Seimel designed the dataset, assembled and harmonized the input data, developed the Bayesian measurement model, ran the estimation and validation pipeline, and wrote the manuscript.

Competing Interests

The author declares no competing interests.

Funding

The author received no specific funding for this work.

Ethics

No new human participant data were collected for this study. The dataset reuses aggregate, secondary party-position data and text-coded political data from existing sources.

References

[1] Pola Lehmann, Tobias Burst, Theres Matthieß, Sven Regel, Andrea Volkens, Bernhard Weßels, and Lisa Zehnter. The manifesto data collection, 2025. URL `https://manifesto-project.wzb.eu/`. Version 2025a.

[2] Seth Jolly, Ryan Bakker, Liesbet Hooghe, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. Chapel Hill expert survey trend file, 1999–2019. *Electoral Studies*, 75:102420, 2022. doi: 10.1016/j.electstud.2021.102420.

[3] Staffan I. Lindberg, Anna Lührmann, Lisa Gastaldi, Sandra Grahn, Sebastián Gutierrez Laufer, Laura Maxwell, Valeriya Mechkova, Richard Morgan, Natalia Natsika,

Shreeya Pillai, Johannes von Römer, and Steven Wilson. V-Party dataset v2, 2022. URL `https://www.v-dem.net/`.

[4] Kenneth Benoit and Michael Laver. Estimating party policy positions: Comparing expert surveys and hand-coded content analysis. *Electoral Studies*, 26(1):90–107, 2007. doi: 10.1016/j.electstud.2006.04.008.

[5] Gary Marks, Liesbet Hooghe, Marco R. Steenbergen, and Ryan Bakker. Crossvalidating data on party positioning on european integration. *Electoral Studies*, 26(1):23–38, 2007. doi: 10.1016/j.electstud.2006.03.007.

[6] Ryan Bakker and Sara Hobolt. Measuring party positions. In Geoffrey Evans and Nan Dirk de Graaf, editors, *Political Choice Matters: Explaining the Strength of Class and Religious Cleavages in Cross-National Perspective*, pages 27–45. Oxford University Press, Oxford, 2013. doi: 10.1093/acprof:oso/9780199663996.003.0002.

[7] Alejandro Ecker, Marcelo Jenny, and Wolfgang C. Müller. How and why party position estimates from manifestos, expert, and party elite surveys diverge: A comparative analysis of the 'left–right' and the 'European integration' dimensions. *Party Politics*, 28(3): 501–517, 2022. doi: 10.1177/1354068821990298.

[8] Anna-Sophie Kurella and Milena Rapp. Combining voter preferences with party position estimates from different sources for studying voting behavior and representation. *Electoral Studies*, 87:102734, 2024. doi: 10.1016/j.electstud.2023.102734.

[9] Theresa Gessler and Swen Hutter. Measuring party positions and issue salience with mass media and manifesto data: Comparing PolDem with manifesto project data. *Party Politics*, 2026. doi: 10.1177/13540688261415753. OnlineFirst.

[10] Patrick Parschan and Charlott Jakob. Computational measurement of political positions: A review of text-based ideal point estimation algorithms. *Quality & Quantity*, 60 (2):5663–5690, 2026. doi: 10.1007/s11135-025-02500-4.

[11] Liesbet Hooghe, Gary Marks, and Carole J. Wilson. Does left/right structure party positions on European integration? *Comparative Political Studies*, 35(8):965–989, 2002. doi: 10.1177/001041402236310.

[12] Gary Marks, Liesbet Hooghe, Moira Nelson, and Erica Edwards. Party competition and european integration in the east and west: Different structure, same causality. *Comparative Political Studies*, 39(2):155–175, 2006. doi: 10.1177/0010414005281932.

[13] Herbert Kitschelt. *The Transformation of European Social Democracy*. Cambridge Studies in Comparative Politics. Cambridge University Press, Cambridge, 1994. doi: 10.1017/CBO9780511622014.

[14] Holger Döring and Sven Regel. Party facts: A database of political parties worldwide. *Party Politics*, 25(2):97–109, 2019. doi: 10.1177/1354068818820671.

[15] Holger Döring and Philip Manow. Parliaments and governments database (ParlGov): Information on parties, elections and cabinets in established democracies, 2024. URL `https://parlgov.org/`. University of Bremen.

[16] Hanspeter Kriesi, Roman Makarov, Jasmine Brügger, Swen Hutter, and Laurent Bernhard. PolDem – national election campaign dataset, version 1, 2020. URL `https://poldem.eui.eu/`.

[17] Jan Rovny, Ryan Bakker, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Marco Steenbergen, and Milada Anna Vachudova. The 2024 Chapel Hill Expert Survey. *Electoral Studies*, 97:102981, 2025. doi: 10.1016/j.electstud.2025.102981.

[18] Ruth Dassonneville and Jelle Koedam. The Chapel Hill Expert Survey – Canada. *Canadian Journal of Political Science*, 2024. doi: 10.1017/S0008423924000337.

[19] Cecilia Martínez-Gallardo, Jan Rovny, Ryan Bakker, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, and Marco Steenbergen. Revisiting party system structuration in Latin America and Europe. *Party Politics*, 2022. doi: 10.1177/13540688221090604.

[20] Roi Zur and Ryan Bakker. The Israeli parties' positions in comparative perspective. *Party Politics*, 2023. doi: 10.1177/13540688231218917.

[21] Maurits J. Meijers and Andrej Zaslove. Measuring populism with expert surveys: POPPA, 2020.

[22] Pippa Norris. Global party survey, 2019, 2020.

[23] Michael-John Morgan. *The Modelling of Governmental Coalition Formation: A Policy-Based Approach with Results from European Simulation Data*. PhD thesis, University of Michigan, 1976.

[24] Michael Laver and John Garry. Estimating policy positions from political texts. *American Journal of Political Science*, 44(3):619–634, 2000. doi: 10.2307/2669268.

[25] Armin Seimel. Replication data for: A two-dimensional party-position panel from text and expert data (v1), 2026. URL `https://doi.org/10.7910/DVN/CDNTYJ`. Versioned release files additionally available at the Codeberg v1 release.

[26] Brigid Fowler. Concentrated orange: Fidesz and the remaking of the Hungarian centre-right, 1994–2002. *Journal of Communist Studies and Transition Politics*, 20(3):80–114, 2004. doi: 10.1080/1352327042000260814.

[27] Zsolt Enyedi. Populist polarization and party system institutionalization: The role of party politics in de-democratization. *Problems of Post-Communism*, 63(4):210–220, 2016. doi: 10.1080/10758216.2015.1113883.

[28] Donald T. Campbell and Donald W. Fiske. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2):81–105, 1959. doi: 10.1037/h0046016.