

Estimating Dynamic Dimensionality in Roll Call Data: An Application to UNGA Voting

WU, TUNG-WEN

1 UNGA Voting Patterns and Their Significance in International Relations Research

Understanding the dimensional structure of ideological preferences is a core challenge in both political science and item response theory (IRT). The voting behavior of states in the United Nations General Assembly (UNGA) has long served as a framework for estimating these ideological dimensions, providing insights into global political alignments and shifts over time. UNGA voting research has predominantly employed ideal point estimation models rooted in IRT, which assume that voting behavior reflects latent ideological positions within a low-dimensional space.

While previous models offer valuable insights into voting behavior, they often assume fixed dimensionality, overlooking the dynamic and evolving nature of ideological preferences. Additionally, these models typically focus on estimating ideal points for individual states, but they may fail to identify important clusters of countries that share similar ideological profiles. Such clusters, which reflect substantively significant groupings of countries with aligned voting patterns, can reveal deeper insights into the dynamics of global politics.

To address these limitations, this paper introduces a flexible nonparametric Bayesian model for analyzing the temporal and multidimensional evolution of global voting patterns in the UNGA. By allowing both the dimensionality of ideological preferences to evolve over time and identifying clusters of countries with similar ideological profiles, this approach offers a more nuanced understanding of shifting alliances and global political dynamics. This research provides an adaptable and accurate framework for capturing these complex patterns, which is especially useful during periods of rapid geopolitical change. By incorporating clustering into the model, it contributes significantly to understanding not only individual country positions but also the broader alignment trends that define international relations.

Moreover, this research provides valuable insights for policymakers and scholars of international relations. By analyzing the evolving dimensions of ideological preferences and the clusters, decision-makers can better anticipate shifts in global alliances and craft more informed

diplomatic strategies. The methodology introduced in this paper also has broad applicability across various fields, including political science, economics, and computer science. For instance, the model could be adapted to analyze voting patterns in other legislative bodies, enhance discrete choice models for economic decision-making by automatically detecting unobserved changes in utility functions, or address complex challenges related to model complexity in machine learning and artificial intelligence. This includes applications in areas like recommendation systems and language models, which often assume a fixed and static number of dimensions but could benefit from the flexibility offered by the proposed approach.

2 Methodology and Broader Implications

Early studies, predating the widespread use of IRT, such as Russett (1966), utilized static factor analysis to analyze countries' voting patterns, with the aim of moving beyond the conventional Cold War dichotomy. Subsequent work, such as Voeten (2000), introduced dynamic models like *NOMINATE* to examine changes in ideological dimensions over time, estimating separate models for the periods 1946–1988 and 1991–1996. More recent research, such as Bailey, Anton Strezhnev, and Voeten (2017), has further refined this approach by estimating time-varying ideal points, providing a more nuanced understanding of the evolution of ideological preferences.

Despite these advancements, most existing approaches in both political science and IRT impose rigid parametric assumptions about dimensionality. Many studies assume a fixed number of ideological dimensions—often one (Bailey, Anton Strezhnev, and Voeten (2017))—or rely on cross-validation techniques Voeten (2000) to determine the appropriate number of dimensions. These approaches require multiple model estimations and raise concerns about overfitting. This method may lack the flexibility necessary to capture the complex and evolving nature of ideological structures, particularly when the salience of political dimensions shifts over time.

Additionally, while much of the literature has focused on domestic legislative voting behavior, researchers have also explored methods for identifying voting clusters among legislators. These clusters can reveal previously unnoticed or significant patterns of ideological alignment. Spirling and Quinn (2010) applied the Dirichlet process to cluster the ideological vectors of Members of the UK Parliament and visualize voting patterns within political parties. In a similar vein, Navarro et al. (2006) used the Dirichlet process in psychology to model the distribution of individual parameters, aiming to categorize behavior patterns across individuals.

However, in the formulations of Spirling and Quinn (2010) and Navarro et al. (2006), the parameters of the subjects being analyzed are directly sampled from the Dirichlet process, implying that all individuals share a small set of common parameters. This approach constrains the ability to identify meaningful similarity patterns, as everyone within a group shares the same parameters. When comparing politically significant cases—such as identifying “representatives similar to Representative A” or “countries similar to Country A”—this formulation limits the ability to draw meaningful comparisons between individuals or countries within the

same group. While such grouping may suffice for some research objectives, visualizing the differences among representatives or countries offers a more nuanced analysis.

Moreover, as Ghosal and Van der Vaart (2017) point out, the discrete nature of the Dirichlet process makes it unsuitable for estimating probability densities. Estimating the probability density of ideologies or preferences is crucial when visualizing the spread of ideological vectors in the posterior distribution. Furthermore, when predicting the voting behavior of newly elected representatives, using probability densities enables more flexible and accurate predictions than relying on a finite set of fixed points.

In the context of UNGA voting research, IRT models often treat latent dimensions as underlying “voting groups” that explain voting behavior. However, it is essential to distinguish latent dimensions from the concept of voting groups. For instance, although Russett (1966) employed factor analysis, he directly equated the estimated factors with “groupings.” A latent dimension refers to a continuous, unobserved ideological axis that captures variations in countries’ preferences on political issues. In contrast, a voting group refers to a discrete set of countries that vote similarly on specific resolutions, often due to shared political or ideological interests. While both latent dimensions and voting groups can be used to interpret voting patterns, the former represents a continuous, unidimensional construct, whereas the latter reflects clusters of countries whose voting behavior may align along multiple dimensions, potentially without corresponding to a single, fixed axis. This distinction is critical for understanding the limitations of models that conflate these concepts, as it oversimplifies the complex, multidimensional nature of global political alignments.

To address these limitations, this paper proposes a nonparametric Bayesian model that allows for both time-invariant ideological vectors and time-varying dimensional salience. Unlike traditional IRT models, which assume a known number of dimensions or rely on ad hoc selection methods, our approach enables the data to determine the appropriate number of dimensions in a probabilistic framework. By fixing ideological positions while allowing the salience of different political dimensions to evolve, our model offers a novel approach to understanding global voting patterns, particularly during periods of rapid geopolitical transformation such as the Cold War and its aftermath.

3 A Flexible Nonparametric Bayesian Model for Analyzing UNGA Voting

To address the limitations outlined in Section 2, we introduce a nonparametric Bayesian model that provides both flexibility in the dimensionality of ideological preferences and the ability to identify clusters of countries with similar voting patterns. This model is designed to capture the temporal and multidimensional evolution of voting behavior in the UNGA, allowing for both the salience of political dimensions and the alignment of countries to shift over time.

The proposed approach builds on the strengths of IRT but expands on its static assumptions by incorporating dynamic element that adapt to the data. In particular, we allow the dimensionality, or salience, of ideological preferences to evolve dynamically, while also identifying latent clusters of countries with similar voting patterns. This flexibility enables the model to capture complex geopolitical shifts, such as the changing alignment of countries in response to global events or policy changes.

Unlike previous studies that use IRT models with dynamic structures, such as Martin and Quinn (2002) and Bailey, Anton Strezhnev, and Voeten (2017), which allow ideal points to evolve over time through an autoregressive process—with prior distributions based on the previous year’s values—our model assumes that the ideological vector remains fixed over time. Although this assumption may initially seem restrictive, ideological change can still be captured through variations in the salience vectors. Furthermore, since countries exhibit different values across multiple dimensions in the ideological vectors, their relative positioning—reflected in the revealed ideological vectors—can shift over time. In the case of significant ideological shifts, as we will demonstrate in the empirical analysis, the proposed model is capable of introducing a new dimension to accommodate such changes. Thus, while the ideological vector may appear static, this approach functions as a strong regularization mechanism. It ensures that only substantial historical events that provoke significant ideological shifts trigger the introduction of new dimensions, preventing undue flexibility in the model.

Before going to the model description, it is important to stress the differences between “salience” and “importance” of issues. As defined in Bartle and Laycock (2012),

In the following subsections, we describe the key components of the model, starting with the underlying assumptions and the data structure, followed by the probabilistic formulation and the inference procedure.

3.1 Utility Functions

The utility function for countries employed in this study is based on the ordered regression models widely utilized in the existing literature, such as in Bailey, Anton Strezhnev, and Voeten (2017). However, this model extends those formulations by incorporating multiple dimensions and salience vectors.

Each country is assumed to have an infinite-dimensional ideological vector $\theta_S = \{\theta_{S,d}\}_{d=1}^{\infty}$, where the index S denotes the country and d represents the dimension. Similarly, each resolution is represented by an infinite-dimensional resolution vector $\beta_R = \{\beta_{R,d}\}_{d=1}^{\infty}$, and each year is associated with an infinite-dimensional salience vector $\rho_Y = \{\rho_{Y,d}\}_{d=1}^{\infty}$.

The resolution vectors are assigned independent standard normal prior distributions, while the prior distributions for the ideological vectors and salience vectors will be detailed in the following section.

To define the ordered logistic regression model, we first need to establish the spatial preference, or the utility, that country S attaches to resolution R . Drawing from the literature on ideal point estimation and machine learning (see Gopalan et al. (2014)), the spatial preference is represented as:

$$\mu_{S,R} = \sum_{d=1}^{\infty} \theta_{S,d} \beta_{R,d} \rho_{year_R,d}$$

For each resolution, there exists a two-dimensional threshold vector $\gamma_R = (\gamma_{R,i})_{i \in \{1,2\}}$, and country S 's voting decision on resolution R is determined by the following rule:

$$\text{Result}_{S,R} = \text{Nay} \quad \text{if} \quad \mu_{S,R} < \gamma_{R,1}$$

$$\text{Result}_{S,R} = \text{Abstain} \quad \text{if} \quad \gamma_{R,1} < \mu_{S,R} < \gamma_{R,2}$$

$$\text{Result}_{S,R} = \text{Yea} \quad \text{if} \quad \mu_{S,R} > \gamma_{R,2}$$

Thus, the model can be viewed as an ordered logistic regression with an infinite number of latent covariates. Moreover, since μ 's are constructed by summing an infinite number of elements, it is not immediately clear whether this summation will be finite or even summable. We will demonstrate the summability and finiteness of μ after introducing the prior distribution structures for the model parameters.

3.2 Ideological Vectors

The concept of an ideological regime refers to the prior distribution of ideological vectors for each country. This distribution is denoted as G .

While assuming a known distribution, such as the normal distribution, is a standard approach, it may not sufficiently capture the complexity of national ideologies. These ideologies are often multimodal due to latent ideological clusters, which are typically interpreted as indicators of ideological alignment—a phenomenon of central interest in international relations and political science. Consequently, in this study, we estimate G nonparametrically using a Dirichlet process mixture.

Following prior research such as Shiraito, Lo, and Olivella (2023), we construct the Dirichlet process using the stick-breaking process in the way proposed by Sethuraman (1994).

Specifically, we first sample the global hyperparameter α from a gamma distribution:

$$\alpha \sim \text{Gamma}(0.001, 0.001)$$

For each regime p within an infinite number of possible regimes, we sample the necessary variables as follows:

$$\pi_p \sim \text{Beta}(1, \alpha)$$

$$p_p = \pi_p \prod_{l=1}^{p-1} (1 - \pi_l)$$

The central ideological vector of regime $\bar{\theta}$ and its dispersion $\bar{\sigma}$ are sampled as follows:

$$\bar{\theta}_p \sim \text{Normal}(0, 1)$$

$$\bar{\sigma}_p \sim \text{Gamma}(0.001, 0.001)$$

When sampling the ideological vector for a given country S , we first draw an index from a categorical distribution parameterized by the stick-breaking probabilities:

$$\eta_S \sim \text{Categorical}(p)$$

Based on the sampled η_S , we then sample the ideological vector:

$$\theta_S \sim \text{Normal}(\bar{\theta}_{\eta_S}, \bar{\sigma}_{\eta_S})$$

As evident from this formulation, each θ_S is drawn from the normal distribution associated with the η_S th regime, implying that ideological positions will be concentrated around a finite number of cluster centers, contingent on the number of regimes effectively utilized by the model. However, unlike parametric models that impose a normal prior, our approach permits the distribution of ideological vectors to deviate from predefined parametric forms, such as the normal distribution, when justified by the data. Moreover, in contrast to the models proposed by Spirling and Quinn (2010) and Navarro et al. (2006), the sampled ideological vectors in our framework do not perfectly coincide with those of other countries in the same regime, allowing for greater flexibility in capturing ideological variation.

This entire sampling process defines the distribution G , whose posterior distribution will be examined in the empirical analysis to illustrate the ideological alignments of countries since the inception of the UNGA, as inferred from their voting behavior within the assembly.

3.3 Saliency Vectors

Efforts to estimate the number of dimensions have been extensively examined in prior research, both within political methodology and across other quantitative disciplines. One of the seminal works, Poole and Rosenthal (1991), highlighted the critical role of dimensionality estimation, offering visualizations that demonstrate how model classification performance varies with the number of dimensions. More recently, Kim, Londregan, and Ratkovic (2018) introduced a method utilizing a Bayesian Lasso prior, which shrinks specific dimensions to zero. Additionally, in the context of biological research on factor models for gene expression, formulations have been developed that drive the variance of normal distributions toward zero as the number of dimensions increases (Bhattacharya and Dunson (2011)).

In this study, we adopt an approach that utilizes a saliency vector, or weight vector derived from the stick-breaking process, as proposed by Gopalan et al. (2014) in the context of recommender systems. This method facilitates the suppression of all but the first few dimensions, not through binary classification of dimensions as useful or not, but by assigning a continuous weight between zero and one to each dimension, which can be interpreted as its saliency.

To account for temporal variation, we model the weights produced by the stick-breaking process in an autoregressive manner, using the framework of the hierarchical Dirichlet process (Yee Whye Teh and Blei (2006)).

Yee Whye Teh and Blei (2006) pointed out that if a separate Dirichlet process is applied to each group (e.g., each year), the dimensions across groups would inherently differ. This discrepancy arises because the Dirichlet process samples discrete values from a continuous probability distribution, making the probability of overlap between values drawn from different Dirichlet processes effectively zero.

To address this issue, Yee Whye Teh and Blei (2006) introduced the hierarchical Dirichlet process, which accommodates dependencies between groups while preserving flexibility. We adopt this regularization framework to allow the year-specific saliency vectors to evolve over time, while simultaneously mitigating excessive deviations.

First, we sample the importance-weighted vector for the initial year using a stick-breaking process. Specifically, we first sample:

$$\tau \sim \text{Gamma}(0.001, 0.001)$$

Then, the importance-weighted vector for year 1, denoted as $\rho_{1,d}$, is obtained through the stick-breaking process:

$$\delta_{1,d} \sim \text{Beta}(1, \tau)$$

$$\rho_{1,d} = \delta_{1,d} \prod_{l=1}^{d-1} (1 - \delta_{1,l})$$

This procedure is iterated for $d = 1$ to $d = \infty$. Next, we sample the parameter that governs the degree of temporal variation:

$$\zeta \sim \text{Gamma}(0.001, 0.001)$$

For year Y , the importance-weighted vector is sampled from a hierarchical stick-breaking process, using the vector from year $Y-1$ as the prior:

$$\delta_{Y,d} \sim \text{Beta} \left(\zeta \rho_{Y-1,d}, \zeta \left(1 - \sum_{l=1}^d \rho_{Y-1,l} \right) \right)$$

$$\rho_{Y,d} = \delta_{Y,d} \prod_{l=1}^{d-1} (1 - \delta_{Y,l})$$

This process is repeated for all $d = 1$ to $d = \infty$.

By incorporating hierarchical structure into the stick-breaking process, our approach allows for dynamic shifts in dimensional importance over time while maintaining consistency across periods. This framework provides a more flexible and theoretically grounded approach to modeling evolving ideological structures.

3.4 Finiteness and Summability of μ

In the last part, let us return to the issue of whether the $\mu_{S,R}$, which is constructed as the infinite summation of the sequence $\{\theta_{S,d} \beta_{R,d} \rho_{year_R,d}\}_{d=1}^{\infty}$.

As already explained, the sequence $\mu_{S,R}$ as follows:

$$\mu_{S,R} = \sum_{d=1}^{\infty} \theta_{S,d} \beta_{R,d} \rho_{year_R,d}$$

where $\theta_{S,d}$, $\beta_{R,d}$, and $\rho_{year_R,d}$ represent the elements of the ideological vector, the elements of the resolution vector, and the weights in the salience vector constructed by stick-breaking process, respectively.

The goal of this proof is to demonstrate that the infinite sum defining μ is well-defined and summable. The argument proceeds in three main steps:

1. **Resolution Vectors:** We begin by addressing the resolution vectors $\beta_{R,d}$, which are independently drawn from standard normal distributions. Since the standard normal distribution is known to have finite variance, this guarantees that the resolution vectors exhibit controlled behavior in expectation. This step does not require further elaboration, as the properties of the normal distribution are well-established.
2. **Ideological Vectors:** The primary focus is on the ideological vectors $\theta_{S,d}$, which are derived from an infinite mixture of normal distributions. To ensure the well-definedness of the infinite sum, we must demonstrate that these vectors have finite variance. Specifically, we aim to show that the contributions of $\theta_{S,d}$ decay sufficiently quickly as $d \rightarrow \infty$, thus ensuring that the sum converges.
3. **Salience Vectors:** Finally, we address the salience vectors $\rho_{year_R,d}$, which follow a stick-breaking process in the first year and a hierarchical stick-breaking process in subsequent years. We demonstrate that the elements of this sequence—both for the initial year, constructed using a stick-breaking process, and for subsequent years, modeled via a hierarchical stick-breaking process—decay at a sufficiently fast rate to guarantee the convergence of the infinite sum. This step is essential in ensuring that the salience vectors do not introduce instability, thereby maintaining the summability of μ .

By addressing these three components—resolution vectors, ideological vectors, and salience vectors—we establish the sufficient conditions for the well-definedness and summability of the infinite sum μ .

Since $\theta_{S,d}$ and $\beta_{R,d}$ originate from an infinite mixture of normals and a normal distribution, respectively, and have finite variance, their contributions remain controlled in expectation. Therefore, to ensure the well-definedness of the infinite sum, it is sufficient to demonstrate that $\rho_{year_R,d}$ converges to zero at a sufficiently fast rate as $d \rightarrow \infty$, ensuring absolute summability.

First, since both the construction of ideological vectors and salience vectors are governed by stick-breaking processes, we demonstrate that the elements of these vectors, derived from the stick-breaking process, decay exponentially fast.

3.4.1 Proposition 1: Exponential Decay of Stick-Breaking Weights

As introduced in the prior distribution part, typical stick-breaking process generates a sequence ρ_d as follows:

$$\rho_1 = v_1, \quad \rho_d = v_d \prod_{l=1}^{d-1} (1 - v_l), \quad d \geq 2$$

where $v_d \sim \text{Beta}(1, \alpha)$ independently.

Taking expectations, we examine:

$$\mathbb{E}[\rho_d] = \mathbb{E} \left[v_d \prod_{l=1}^{d-1} (1 - v_l) \right].$$

Since the v_d are independently sampled, the expectation of the product can be decomposed into the product of individual expectations, allowing us to analyze each term separately.

$$\mathbb{E} \left[v_d \prod_{l=1}^{d-1} (1 - v_l) \right] = \mathbb{E} [v_d] \prod_{l=1}^{d-1} \mathbb{E} [1 - v_l]$$

Since $v_d \sim \text{Beta}(1, \alpha)$, its expectation is:

$$\mathbb{E}[v_d] = \frac{1}{1 + \alpha}.$$

For $d \geq 2$,

$$\mathbb{E}[1 - v_l] = \frac{\alpha}{1 + \alpha}.$$

So,

$$\prod_{l=1}^{d-1} \mathbb{E}[1 - v_l] = \left(\frac{\alpha}{1 + \alpha} \right)^{d-1}.$$

Thus,

$$\mathbb{E}[\rho_d] = \frac{1}{1 + \alpha} \left(\frac{\alpha}{1 + \alpha} \right)^{d-1}.$$

This demonstrates that the sequence exhibits exponential decay at a rate of $\frac{\alpha}{1+\alpha}$, which lies strictly between zero and one for any positive and finite α .

□

3.4.2 Proposition 2: Finiteness of the Variance of Ideological Vectors

As already shown, $\theta_{S,d}$ are the ideological vector components, drawn from a Dirichlet process mixture of normal distributions, which can be written in the following way:

$$\theta_{S,d} \sim \sum_{k=1}^{\infty} p_k \text{Normal}(\bar{\theta}_k, \sigma_k^2)$$

As previously established, the components of the ideological vector, $\theta_{S,d}$, are drawn from a Dirichlet process mixture of normal distributions, which can be expressed as follows:

To establish the finiteness of $\text{Var}(\theta_{S,d})$, we observe that, as shown above, the distribution can be represented as an infinite summation of normally distributed random variables. Since each mixture component has finite variance, denoted as $\bar{\sigma}_k^2$, and the expectation of the total variance under the Dirichlet process prior is given by:

$$\mathbb{E}[\text{Var}(\theta_{S,d})] = \sum_{k=1}^{\infty} p_k^2 \bar{\sigma}_k^2.$$

Since the expectation of p_k^2 decays quadratically due to the stick-breaking construction as shown in as shown in Proposition 1, and $\bar{\sigma}_k^2$ is finite due to its prior distribution of $\text{Gamma}(0.001, 0.001)$, this sum converges and is finite.

□

3.4.3 Proposition 3: Exponential Decay of Hierarchical Stick-Breaking Weights

As shown in Section 3.3, for the second year, the salience vectors are constructed as

$$\delta_{2,d} \sim \text{Beta} \left(\zeta \rho_{1,d}, \zeta \left(1 - \sum_{l=1}^d \rho_{1,l} \right) \right)$$

$$\rho_{2,d} = \delta_{2,d} \prod_{l=1}^{d-1} (1 - \delta_{2,l}).$$

As in Proposition 1, the expectation of $\rho_{2,d}$ can be expressed as

$$\mathbb{E}[\rho_{2,d}] = \mathbb{E} \left[\delta_{2,d} \prod_{l=1}^{d-1} (1 - \delta_{2,l}) \right].$$

Since the variables $\delta_{2,d}$ are sampled independently, we can evaluate the expectation term by term:

$$\mathbb{E} \left[\delta_{2,d} \prod_{l=1}^{d-1} (1 - \delta_{2,l}) \right] = \mathbb{E} [\delta_{2,d}] \prod_{l=1}^{d-1} \mathbb{E} [1 - \delta_{2,l}].$$

For a Beta-distributed variable $\delta_{2,d} \sim \text{Beta} \left(\zeta \rho_{1,d}, \zeta \left(1 - \sum_{l=1}^d \rho_{1,l} \right) \right)$, its expectation is given by:

$$\mathbb{E}[\delta_{2,d}] = \frac{\zeta \rho_{1,d}}{\zeta \rho_{1,d} + \zeta \left(1 - \sum_{l=1}^d \rho_{1,l} \right)}.$$

Simplifying, we obtain

$$\mathbb{E}[\delta_{2,d}] = \frac{\rho_{1,d}}{1 - \sum_{l=1}^{d-1} \rho_{1,l}}.$$

Similarly, for the expectation of $1 - \delta_{2,l}$, we have:

$$\mathbb{E}[1 - \delta_{2,l}] = 1 - \mathbb{E}[\delta_{2,l}].$$

Since $\rho_{1,d}$ is constructed from a stick-breaking process, from Proposition 1, its expectation is:

$$\mathbb{E}[\rho_{1,d}] = \frac{1}{1 + \alpha} \left(\frac{\alpha}{1 + \alpha} \right)^{d-1}.$$

Substituting this into our previous expression, we obtain:

$$\mathbb{E}[\delta_{2,d}] = \frac{\frac{1}{1+\alpha} \left(\frac{\alpha}{1+\alpha} \right)^{d-1}}{1 - \sum_{l=1}^{d-1} \frac{1}{1+\alpha} \left(\frac{\alpha}{1+\alpha} \right)^{l-1}}.$$

Using the geometric series formula,

$$\sum_{l=1}^{d-1} \frac{1}{1 + \alpha} \left(\frac{\alpha}{1 + \alpha} \right)^{l-1} = 1 - \left(\frac{\alpha}{1 + \alpha} \right)^{d-1},$$

we obtain

$$\mathbb{E}[\delta_{2,d}] = \frac{1}{1+\alpha},$$

which is of the same form as in the non-hierarchical stick-breaking weights.

Thus, as in Proposition 1, for $d \geq 2$,

$$\mathbb{E}[1 - \delta_{2,d}] = \frac{\alpha}{1+\alpha}.$$

So,

$$\prod_{l=1}^{d-1} \mathbb{E}[1 - \delta_{2,d}] = \left(\frac{\alpha}{1+\alpha} \right)^{d-1}.$$

Thus,

$$\mathbb{E}[\rho_{2,d}] = \frac{1}{1+\alpha} \left(\frac{\alpha}{1+\alpha} \right)^{d-1}.$$

This demonstrates that the sequence exhibits exponential decay at a rate of $\frac{\alpha}{1+\alpha}$, which lies strictly between zero and one for any positive and finite α .

The same result trivially applies to the subsequent years.

□

3.4.4 Conclusion: Well-Definedness and Summability of μ

Having established the necessary conditions for the finiteness and summability of the infinite sum defining $\mu_{S,R}$, we can now summarize the key results:

1. **Exponential Decay of Stick-Breaking Weights:** We demonstrated that the expectation of the stick-breaking weights $\rho_{year_R,d}$ exhibits exponential decay at a rate of $\frac{\alpha}{1+\alpha}$, ensuring that their contributions to the sum diminish sufficiently fast as $d \rightarrow \infty$.
2. **Finiteness of the Variance of Ideological Vectors:** The ideological vector components $\theta_{S,d}$ were shown to have finite variance, derived from a Dirichlet process mixture of normal distributions, ensuring that their contributions remain controlled in expectation.
3. **Hierarchical Stick-Breaking Process Stability:** The hierarchical stick-breaking construction of salience vectors for subsequent years retains the exponential decay property, ensuring that $\rho_{year_R,d}$ continues to diminish at a sufficient rate.

With these properties established, we conclude that the infinite sum defining $\mu_{S,R}$ is absolutely summable and well-defined. This guarantees the theoretical soundness of our modeling approach, allowing for stable inference and meaningful interpretations of the latent ideological and salience structures over time.

4 Empirical Findings from UNGA Voting Data

In the following analysis, we draw upon data from Voeten, Strezhnev, and Bailey (2009). While a full Markov Chain Monte Carlo (MCMC) approach would be the most theoretically robust method for estimating the model, the large scale of the dataset—comprising over 700,000 country-resolution pair observations—and the complex hierarchical structure of the model necessitate an alternative approach. As a result, we implement variational inference, following Kucukelbir et al. (2017), using the Stan programming language (Stan Development Team (2024)). This method transforms the computationally intensive sampling process into a more tractable optimization problem.

Variational inference has gained traction among political scientists working with large datasets, as it enables the extraction of meaningful quantities while preserving the Bayesian framework (Grimmer (2011)). However, despite its advantages—such as scalability—variational inference has notable limitations, particularly its tendency to underestimate posterior variance (Blei, Kucukelbir, and McAuliffe (2017)).

Theoretically, the number of ideological regimes and salience dimensions in our model could be infinite. However, given computational constraints, handling infinite-dimensional arrays is infeasible. Instead, we approximate infinity with a sufficiently large finite value. For the empirical analysis, we set this value to 20 for both ideological regimes and salience dimensions. Even with 20 dimensions, many ideological regimes and salience dimensions remain effectively unused, justifying this choice.

To evaluate the model’s predictive accuracy, we randomly held out 1,000 ‘yea’, 1,000 ‘nay’, and 1,000 ‘abstain’ cases.

The variational inference algorithm, leveraging distributed evaluation of the log-likelihood function, converges in approximately 38 minutes on an M1 MacBook Air with 16 GB of memory and 8 cores.

4.1 Prediction Accuracy

To assess whether the model effectively captures meaningful patterns in UNGA voting data, we evaluate its predictive performance on a held-out set of 3,000 observations. Following the posterior predictive framework proposed by Meng (1994), we measure predictive accuracy using the F1 score—a widely adopted metric in classification tasks within the machine learning literature—as the test statistic.

Rather than computing a posterior predictive p-value against a fixed benchmark (e.g., 0.5), we visualize the entire posterior distribution of the F1 score. This approach provides a more comprehensive assessment of the model’s predictive capability and its alignment with observed voting behavior.

More formally, we compute the posterior distribution of the F1 score for each class, as well as the macro F1 score—the arithmetic mean of the per-class F1 scores—conditioned on the observed data. Mathematically, this is expressed as:

$$p(\text{F1 score} \mid x) = \int p(\text{F1 score} \mid \theta) p(\theta \mid x) d\theta,$$

where x represents the set of training data, and θ denotes the parameters of the model. This formulation accounts for parameter uncertainty by integrating over the posterior distribution, ensuring a robust evaluation of the model’s predictive performance.

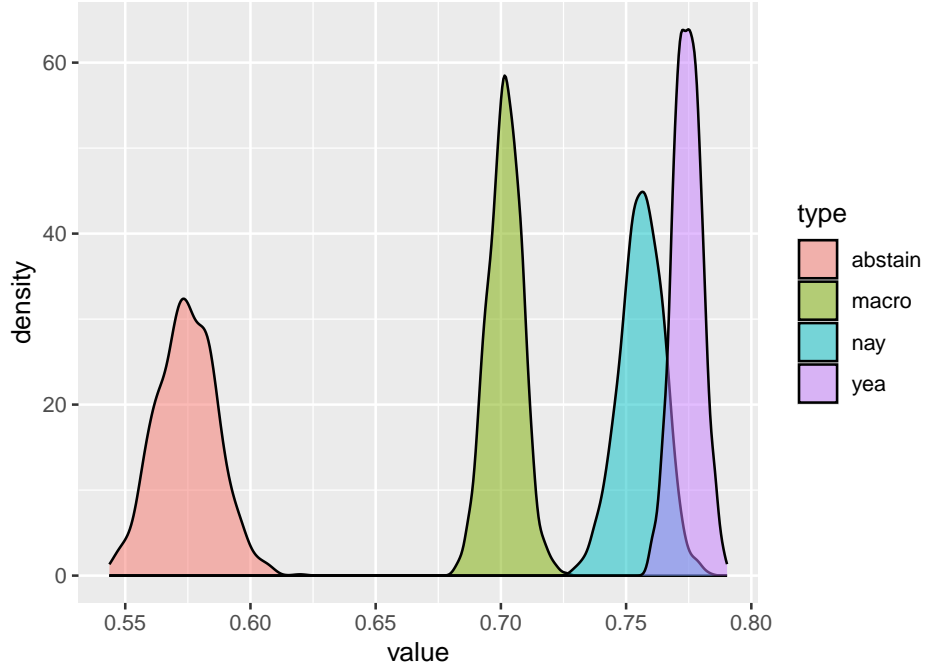


Figure 1: Estimated Saliency Vectors

Figure 1 presents the posterior distributions of the F1 scores for the three voting categories—‘yea’, ‘nay’, and ‘abstain’—as well as the macro F1 score, which provides an overall measure of predictive accuracy. The results indicate that the F1 scores for the ‘yea’ and ‘nay’ classes are significantly higher than 0.7, demonstrating strong predictive performance. In contrast, the F1 score for the ‘abstain’ class is lower, ranging from approximately 0.55 to 0.6.

The relatively lower predictive accuracy for the ‘abstain’ class is likely not a deficiency of the model itself but rather a reflection of the inherent nature of abstentions, which convey less clear information compared to affirmative or negative votes. Nevertheless, the F1 score for the ‘abstain’ class remains significantly above 0.5, indicating that the model performs substantially better than random chance.

The macro F1 score distribution is centered around 0.7, suggesting that the model achieves a high overall predictive accuracy across all voting categories. This result implies that the model effectively captures underlying patterns in UNGA voting behavior while maintaining balanced performance across different classes.

Overall, these findings suggest that the model successfully identifies meaningful patterns in UNGA voting behavior, particularly for decisive votes (‘yea’ and ‘nay’), while also demonstrating reasonable predictive capability in more ambiguous cases (‘abstain’).

4.2 Overview of the Estimated Saliency Vectors

To gain a comprehensive understanding of the patterns identified by the model in the UNGA voting data, we begin by visualizing the estimated saliency vectors. The model effectively utilizes six saliency dimensions, while the remaining dimensions exhibit negligible values across all years, accompanied by near-zero credible intervals. Consequently, our analysis is concentrated on these six significant dimensions.

To capture the temporal shifts in saliency, we visualize the estimated saliency vectors, ρ_{year} , for each year. The saliency vectors are plotted vertically, with connections drawn between the same dimension across different years, facilitating an examination of the evolution of saliency over time. Specifically, we focus on six dimensions identified by the model, each representing a key global issue: human rights, north-south relations, nuclear weapons, the Cold War, the Yom Kippur War, and membership in the United Nations. The interpretation and naming of these dimensions will be addressed in the subsequent discussion.

Accompanying the line plot is a shaded ribbon, which represents the 5th and 95th percentiles of the saliency vector estimates, visually indicating the uncertainty surrounding the mean values. A vertical dashed red line marks the year 1989, signaling the end of the Cold War, a pivotal geopolitical event.

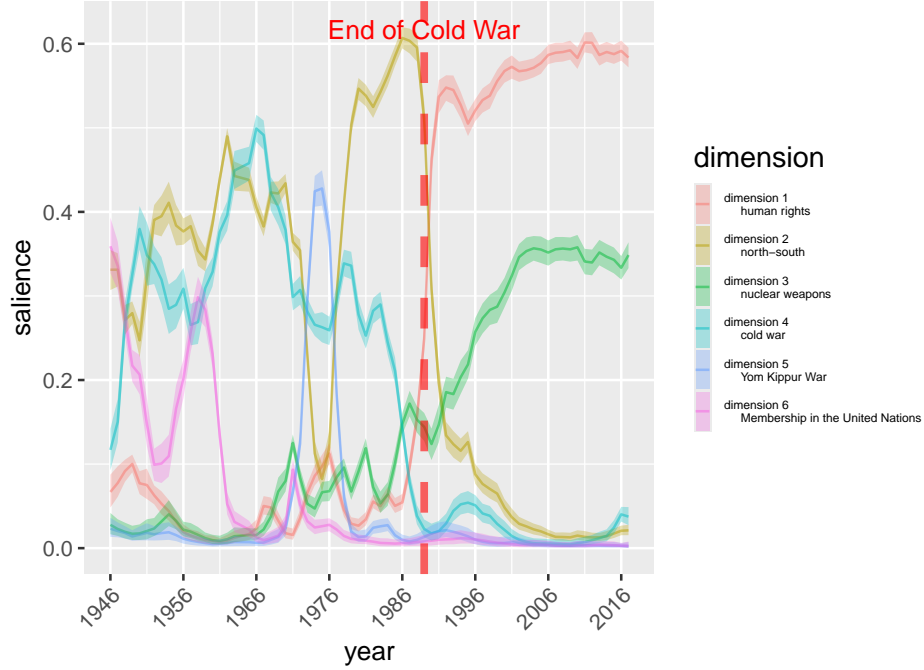


Figure 2: Estimated Saliency Vectors

As illustrated in Figure 2, prior to the end of the Cold War, the movement of the saliency vectors was pronounced, with many active dimensions reflecting significant shifts in global issues. However, following the end of the Cold War, the movement of the saliency dimensions became more stable. By the year 2000, only two dimensions remained actively influencing the model, and the relative values of these two dimensions remained relatively constant thereafter.

To quantify the variability of saliency vectors across years more effectively, we compute the entropy of the saliency vectors for each year Y using the following formula:

$$entropy_Y = - \sum_{d=1}^{\infty} \rho_{Y,d} \log(\rho_{Y,d})$$

Entropy is a key concept for quantifying uncertainty and disorder in a probability distribution. For instance, consider an unfair coin that always lands heads when tossed—its outcome is highly predictable, resulting in low entropy. In contrast, a fair coin that has an equal probability of landing heads or tails introduces more uncertainty, leading to a higher entropy. Although the saliency vector in this context is not explicitly used as a probability distribution but rather as a weighting vector, which assigns higher weights to relevant dimensions while

shrinking unnecessary ones to zero, its construction allows it to effectively represent a distribution. Consequently, we adopt the concept of entropy to visualize the uncertainty associated with the salience vectors.

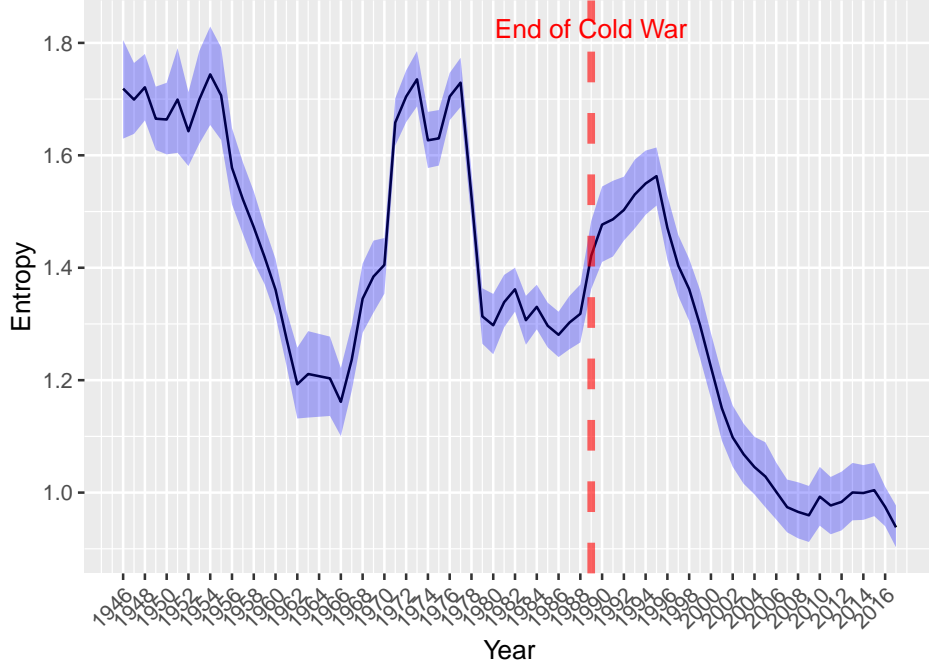


Figure 3: Entropy of Salience Vectors

Figure 3 highlights significant shifts following the conclusion of the Cold War. After the Cold War, entropy experienced a slight increase, peaking in 1995. However, it subsequently began a steady decline, ultimately reaching levels lower than the minimum entropy observed during the Cold War, specifically between 1962 and 1966. This ongoing downward trend indicates that, in terms of issue diversity measured by the entropy of salience vectors, the issue configuration within the UNGA resolutions is lower than at any point in its history.

This conclusion aligns closely with the findings of Voeten (2000), who suggests that, following the Cold War, voting patterns in the UNGA became increasingly driven by a single dimension. In contrast, our model reveals that post-Cold War voting is influenced by one dominant dimension, with a salience value of approximately 0.6, and a second, less significant dimension with a salience value around 0.35. In other words, our model indicates that approximately ‘1.5’ dimensions govern post-Cold War voting behavior. Despite the differences in the precise number of dimensions, both studies reach the same overarching conclusion: UNGA voting during the Cold War was more complex compared to the post-Cold War period.

Table 1: Distinctive Words for Each Dimension

dimension_1	dimension_2	dimension_3	dimension_4	dimension_5	dimension_6
rights_sudan	israel_south	res_alrev	development_organization	res_alrev	nepal_amendment
rights_cuba	res_alrev	nationals	appropriations_year	exercise_palestinian	council_admission
democratic_republic	rights_sudan	disarmament_general	moratorium	condemn_israels	palestine_question
moratorium	vote_operative	reducing_nuclear	budget_appropriations	moratorium	al_general
moratorium_use	possible_offer	nuclear_danger	appropriations_biennium	continuing_increasing	adopt_trusteeship
use_death	offer_necessary	legality	biennium_amount	increasing_collaboration	approved_th
penalty_resolution	resources_meet	legality_threat	appropriations	korea_continue	reject
rights_democratic	meet_needs	danger_resolution	nepal_amendment	israel_racist	consider_proposals
armaments_resolution	near_ea	arrangements_strengthening	expel	crime_humanity	section_ad
rights_iraq	us_motion	reducing	un_expel	rights_population	weapons_testing

4.3 Interpretation of the Estimated Saliency Dimensions and Their Substantive Implications

In this section, we interpret the substantive meaning of the estimated saliency vector dimensions introduced in the previous subsection. Since the model relies solely on resolution IDs, country IDs, year IDs, and voting outcomes as inputs, it derives these dimensions purely from the underlying correlational structure in the data. Consequently, external information is required to meaningfully interpret the estimation results.

A promising direction for future research would be to incorporate textual data into the proposed model, allowing for a more direct association between estimated dimensions and substantive political themes. The estimation of ideal points, or ideological vectors as referred to in our model, has been explored in the machine learning literature. For example, Vafa, Naidu, and Blei (2020) propose a model that estimates ideal points using speech data, highlighting the potential for integrating text-based information into such analyses.

To interpret the estimated dimensions, we fit an additional model using the distributed multinomial regression approach proposed by Taddy (2015). This model incorporates resolution metadata along with the estimated resolution and saliency vectors. Specifically, we first weight the resolution vectors by their corresponding saliency vectors for the year of the vote. The resulting weighted resolution vectors serve as covariates in a distributed multinomial regression model, predicting the frequency of words in the resolution metadata provided by Voeten, Strezhnev, and Bailey (2009). This approach allows us to identify words that are strongly associated with each saliency dimension.

Bibliography

- Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. “Estimating Dynamic State Preferences from United Nations Voting Data.” *Journal of Conflict Resolution* 61 (2): 430–56.
- Bartle, John, and Samantha Laycock. 2012. “Telling More Than They Can Know? Does the Most Important Issue Really Reveal What Is Most Important to Voters?” *Electoral Studies* 31 (4): 679–88.

- Bhattacharya, Anirban, and David B Dunson. 2011. “Sparse Bayesian Infinite Factor Models.” *Biometrika* 98 (2): 291–306.
- Blei, David M, Alp Kucukelbir, and Jon D McAuliffe. 2017. “Variational Inference: A Review for Statisticians.” *Journal of the American Statistical Association* 112 (518): 859–77.
- Ghosal, Subhashis, and Aad W Van der Vaart. 2017. *Fundamentals of Nonparametric Bayesian Inference*. Vol. 44. Cambridge University Press.
- Gopalan, Prem, Francisco J Ruiz, Rajesh Ranganath, and David Blei. 2014. “Bayesian Non-parametric Poisson Factorization for Recommendation Systems.” In *Artificial Intelligence and Statistics*, 275–83. PMLR.
- Grimmer, Justin. 2011. “An Introduction to Bayesian Inference via Variational Approximations.” *Political Analysis* 19 (1): 32–47.
- Kim, In Song, John Londregan, and Marc Ratkovic. 2018. “Estimating Spatial Preferences from Votes and Text.” *Political Analysis* 26 (2): 210–29.
- Kucukelbir, Alp, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M Blei. 2017. “Automatic Differentiation Variational Inference.” *Journal of Machine Learning Research* 18 (14): 1–45.
- Martin, Andrew D, and Kevin M Quinn. 2002. “Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999.” *Political Analysis* 10 (2): 134–53.
- Meng, Xiao-Li. 1994. “Posterior Predictive p -Values.” *The Annals of Statistics* 22 (3): 1142–60.
- Navarro, Daniel J, Thomas L Griffiths, Mark Steyvers, and Michael D Lee. 2006. “Modeling Individual Differences Using Dirichlet Processes.” *Journal of Mathematical Psychology* 50 (2): 101–22.
- Poole, Keith T, and Howard Rosenthal. 1991. “Patterns of Congressional Voting.” *American Journal of Political Science*, 228–78.
- Russett, Bruce M. 1966. “Discovering Voting Groups in the United Nations.” *American Political Science Review* 60 (2): 327–39.
- Sethuraman, Jayaram. 1994. “A Constructive Definition of Dirichlet Priors.” *Statistica Sinica*, 639–50.
- Shiraito, Yuki, James Lo, and Santiago Olivella. 2023. “A Nonparametric Bayesian Model for Detecting Differential Item Functioning: An Application to Political Representation in the US.” *Political Analysis* 31 (3): 430–47.
- Spirling, Arthur, and Kevin Quinn. 2010. “Identifying Intraparty Voting Blocs in the UK House of Commons.” *Journal of the American Statistical Association* 105 (490): 447–57.
- Stan Development Team. 2024. *Stan Reference Manual, Version 2.34.1*. <https://mc-stan.org>.
- Taddy, Matt. 2015. “Distributed multinomial regression.” *The Annals of Applied Statistics* 9 (3): 1394–414. <https://doi.org/10.1214/15-AOAS831>.
- Vafa, Keyon, Suresh Naidu, and David Blei. 2020. “Text-Based Ideal Points.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 5345–57. Online: Association for Computational Linguistics.
- Voeten, Erik. 2000. “Clashes in the Assembly.” *International Organization* 54 (2): 185–215.

- Voeten, Erik, Anton Strezhnev, and Michael Bailey. 2009. “United Nations General Assembly Voting Data.” Harvard Dataverse. <https://doi.org/10.7910/DVN/LEJUQZ>.
- Yee Whye Teh, Matthew J Beal, Michael I Jordan, and David M Blei. 2006. “Hierarchical Dirichlet Processes.” *Journal of the American Statistical Association* 101 (476): 1566–81.