Psychometrika (2025), **90**, 904–931
doi:10.1017/psy.2025.10014

THEORY AND METHODS

Bayesian Rank-Clustering

Michael Pearce[1] and Elena A. Erosheva[2]

[1]Department of Mathematics and Statistics, Reed College, Portland, OR, USA; [2]Department of Statistics, School of Social Work, and the Center for Statistics and the Social Sciences, University of Washington, Seattle, WA, USA

Corresponding author: Michael Pearce; Email: michaelpearce@reed.edu

(Received 8 November 2024; revised 24 April 2025; accepted 28 April 2025; published online 16 June 2025)

Abstract

This article proposes a new statistical model to infer interpretable population-level preferences from ordinal comparison data. Such data is ubiquitous, e.g., ranked choice votes, top-10 movie lists, and pairwise sports outcomes. Traditional statistical inference on ordinal comparison data results in an overall ranking of objects, e.g., from best to worst, with each object having a unique rank. However, the ranks of some objects may not be statistically distinguishable. This could happen due to insufficient data or to the true underlying object qualities being equal. Because uncertainty communication in estimates of overall rankings is notoriously difficult, we take a different approach and allow groups of objects to have equal ranks or be *rank-clustered* in our model. Existing models related to rank-clustering are limited by their inability to handle a variety of ordinal data types, to quantify uncertainty, or by the need to pre-specify the number and size of potential rank-clusters. We solve these limitations through our proposed Bayesian *Rank-Clustered Bradley–Terry–Luce (BTL)* model. We accommodate rank-clustering via parameter fusion by imposing a novel spike-and-slab prior on object-specific worth parameters in the BTL family of distributions for ordinal comparisons. We demonstrate rank-clustering on simulated and real datasets in surveys, elections, and sports analytics.

Keywords: Bradley–Terry; fusion priors; item indifference; Plackett–Luce; rank aggregation; spike-and-slab

1. Introduction

In a traditional analysis of ordinal data, we assume I judges assess J objects by providing ordinal preferences, Π. The ordinal preferences of each judge, Π_i, may be provided in various forms, such as complete rankings, partial rankings, or pairwise comparisons among available objects or some subset thereof. Standard statistical model families for ranking data such as Mallows (Mallows, 1957) or Bradley-Terry-Luce (Bradley & Terry, 1952; Luce, 1959; Plackett, 1975) derive or estimate the rank of each object whereby each object receives a unique rank. An estimated *overall ranking* then orders all objects from best to worst. Analyses of this kind, often referred to as *rank aggregation* (Dwork et al., 2001), are used to rank candidates in ranked choice elections, (Gormley & Murphy, 2008; Mollica & Tardella, 2017), sports teams or players in a league using pairwise game outcomes (Barrientos et al., 2023; Tutz & Schauberger, 2015), or genes based on ordinal comparisons of genomics data (Eliseussen et al., 2023; Vitelli et al., 2018). In these scenarios we intentionally do not consider potential heterogeneity among judges. Our goal is to learn a single ranking which is the desired outcome, whether it is an ordering of candidates or an ordering of genes.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

However, requiring estimated ranks to be unique is not always useful or appropriate. For example, some objects may be equal or indistinguishable in their true quality or ability. Consider an election in which two candidates, both of the same political party, are running for an office. If voters express their preferences solely on the basis of party, the candidates are inherently equal in quality. In another situation, when the number of votes cast is small, estimated ranks assigned to each candidate could exhibit substantial uncertainty, suggesting the candidates are indistinguishable in quality based on the limited number of observed votes. In such situations, allowing for inference to estimate the candidates as having the same rank or be *rank-clustered* may improve interpretability, prediction, and decision-making when analyzing ordinal preferences.

In this article, we propose a Bayesian framework for ordinal data analysis that estimates an overall ranking of objects with rank-clusters, develop a computationally-efficient Gibbs sampler for estimation, and apply the model to real and simulated data. Specifically, we choose to model observed rank via the Bradley–Terry–Luce (BTL) family of distributions which permits analysis of ordinal preferences in many forms, such as complete rankings, partial rankings, pairwise comparisons, and groupwise comparisons. To induce rank-clusters, we place a novel spike-and-slab fusion prior on the object-specific parameters of BTL distributions. In contrast to existing work related to rank-clustering in the literature, our model requires neither the parameter order nor the number or size of rank-clusters to be known in advance. Instead, these quantities are treated as random variables and estimated simultaneously so that their corresponding uncertainty is naturally reflected in the resulting inferences.

The rest of the article is organized as follows. We first review literature related to rank-clustering in Section 2. Then, we propose the Partition-based Spike-and-Slab Fusion (PSSF) prior and apply it to a BTL model for ordinal data in Section 3. We develop a computationally-efficient Gibbs sampler based on reversible jump Markov chain Monte Carlo (RJMCMC) and demonstrate its accuracy on simulated data in Section 4. To demonstrate a wide variety of methodological benefits of our proposed framework, in Section 5, we apply the model to four real datasets: (i) complete rankings of sushi preferences provided by Japanese adults, (ii) partial rankings of 2021 Minneapolis mayoral candidates expressed by voters in a ranked choice election, (iii) complete and partial rankings of policy options from Eurobarometer 34.1, a survey which measures various European attitudes, and (iv) pairwise basketball game outcomes from the 2023–2024 season of the National Basketball Association (NBA). We conclude with a brief discussion in Section 6.

2. Background

Before reviewing the ordinal comparisons literature, it is helpful to introduce some basic terminology and notation. Rankings are a type of ordinal preference that denotes a relative ordering of objects from best to worst, potentially allowing ties. We use the operator '$\prec$' to denote a strict ordering of two objects; e.g., $A \prec B$ states that object A is strictly preferred to B. An object's *rank* is the place it receives in the ranking.[1] Rankings arise in different forms. Given a collection of objects, a ranking is called *complete* when all objects are ranked. In contrast, a ranking is called *partial* when only a subset of the most-preferred objects are ranked (e.g., a top-five ranking). In a partial ranking, we assume that unranked objects are less-preferred than those ranked, but also that the preference order among the unranked objects is unknown. Next, we call a ranking *incomplete* when a judge is asked only to rank a subset of the complete collection of objects. In incomplete rankings, no information can be gleaned regarding objects not considered. For example, if a voter is asked by an election pollster to rank candidates from a single political party, the ranking should provide no information regarding their preferences on candidates from other parties. We call incomplete rankings involving two objects (candidates in the above example) a *pairwise comparison*, and incomplete rankings involving more than two objects

[1] Although some authors have drawn a distinction between the terms "ranking" and "ordering," in this article we choose to use solely the former in accordance with its popular usage.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

a *groupwise comparison*. Rankings may be both partial and incomplete; e.g., a top-three ranking of mayoral candidates from a specific political party.

Next, we briefly review methods for estimating rank-clusters based on the BTL and Mallows families of ordinal data models in turn. For a more thorough review of these standard model families, see Marden (1996) and Alvo & Yu (2014).

2.1. Methods based on BTL distributions

Most work related to rank-clustering utilizes the BTL family, which comprises the Bradley–Terry and Plackett–Luce distributions and their extensions. The Bradley–Terry model, proposed by Zermelo (1929) and discovered independently by Bradley & Terry (1952), is parameterized by the vector $\omega \in \mathbb{R}^J_{>0}$, in which each ω_j corresponds to the *worth* of object j. Specifically, the Bradley-Terry model specifies the probability that object i will be ranked above object j in pairwise tournament as

$$P[i \prec j | \omega_i, \omega_j] = \frac{\omega_i}{\omega_i + \omega_j}. \tag{1}$$

The Plackett–Luce model (Plackett, 1975) extended the Bradley–Terry to allow for multiple comparisons, partial rankings, and incomplete rankings, and has been justified under Luce's Choice Axiom (Luce, 1959) and Thurstone's theory of comparative judgment (Thompson Jr. & Singh, 1967; Thurstone, 1927; Yellott Jr., 1977). In this model, a ranking $\pi = \{1 \prec 2 \prec \cdots \prec J\}$ of J objects is assigned probability

$$P[\Pi = \pi | \omega_1, \ldots, \omega_J] = \prod_{j=1}^{J} \frac{\omega_j}{\sum_{j'=j}^{J} \omega_{j'}}, \tag{2}$$

where often one sets $\sum_j \omega_j = 1$ for identifiability. Rankings drawn from the Plackett–Luce model may be interpreted as being created sequentially, where in the first stage an object is selected among all the options, in the second stage an object is selected among all the remaining, and so on. Extensions of distributions in the BTL family have been proposed to capture intricacies in ranked preferences such as order of presentation effects, ties, and covariates (Chapman & Staelin, 1982; Critchlow & Fligner, 1991; Gormley & Murphy, 2010; Rao & Kupper, 1967). Importantly, the BTL family can handle partial and incomplete rankings by exploiting its reliance on Luce's Choice Axiom.

Since BTL distributions have continuous parameters, rank-clusters may be estimated by employing *parameter fusion* or *shrinkage*. *Parameter fusion* is the process of simultaneously estimating parameter values and groups of parameters that should be set equal in value (i.e., "fusing" parameters together). Masarotto & Varin (2012) analyze pairwise comparison data from sports tournaments with parameter fusion techniques under the Bradley–Terry model. Masarotto & Varin (2012) estimate an overall ranking of teams with rank-clusters by applying the frequentist *fused lasso* (Tibshirani et al., 2005), in which the absolute difference between every pair of worth parameters is penalized after some data-driven normalization. In this approach, the fused parameters are made equal and thus create a rank-cluster among the corresponding objects. The approach of Masarotto & Varin (2012) was extended to additional datasets in sports (Tutz & Schauberger, 2015) and academic journal rankings (Vana et al., 2016; Varin et al., 2016). Jeon & Choi (2018) argued that shrinkage methods like those proposed by Masarotto & Varin (2012) and Tutz & Schauberger (2015) were developed specifically for pairwise comparisons, and thus have inappropriate penalty functions for application to richer kinds of ordinal data like partial or complete rankings. As a result, Jeon & Choi (2018) proposed a modified regularization penalty that may be applied to partial or complete rankings under the Plackett–Luce model. Relatedly, Hermes et al. (2024) consider sparse estimation of a Plackett–Luce model with object-level covariates under judge heterogeneity. In their setting, the number of heterogeneous preference groups and the group membership of each judge are assumed fixed and known. To improve efficiency of estimation across groups and predictive performance, they impose a lasso penalty on group-specific covariate coefficients and a simultaneous fused lasso penalty between each pair of group-specific covariate coefficients. We note that the setting studied by Hermes et al. (2024) is fundamentally different

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

to ours, in that they assume (known) preference heterogeneity among the judges and the presence of object-specific covariates.

Parameter fusion methods for rank-clustering exhibit four distinct disadvantages: First, maximum likelihood estimation of models in the BTL family, even in their simplest forms, often suffers from numerical instability and slow computational speed. As a result, numerous authors have proposed complex algorithms to improve estimation accuracy or speed (Hunter et al., 2004; Maystre & Grossglauser, 2015; Nguyen & Zhang, 2023; Turner et al., 2020). Second, uncertainty quantification is challenging and theoretically tenuous in lasso-based methods (Fan & Li, 2001; Tibshirani, 1996). Third, lasso penalty parameters may be difficult to select, requiring data-driven or *ad hoc* techniques (Masarotto & Varin, 2012; Tibshirani, 1996). Thus, interpretation of the resulting parameter estimates and associated uncertainty is reliant on the specific choice of penalty parameter. Fourth, prior knowledge on the amount and size of rank-clusters cannot be directly incorporated into the frequentist framework: Although the penalty parameter influences estimation of rank-clusters, the specific meaning of various possible choices is not directly interpretable.

Many of these disadvantages may be addressed using spike-and-slab priors, a Bayesian approach to variable selection (George & McCulloch, 1997; Ishwaran & Rao, 2005; Mitchell & Beauchamp, 1988). Spike-and-slab priors assign weight to both a point-mass at 0 ("spike") and a continuous density function ("slab"). Although the specific formulations of these priors vary, they estimate parameters which are precisely zero in a probabilistic framework that incorporates prior knowledge via interpretable hyperparameters, as opposed to opaque penalty parameters. However, we are aware of only one variant of this prior class for parameter fusion: Wu et al. (2021) apply spike-and-slab to differences in successive parameters in a linear regression. In their method, the order of parameters from least to greatest in coefficient value must be known in advance (as in the fused lasso). This is not practical in the canonical ordinal data setting because the parameter order is equivalent to the overall ranking, whose estimation is a primary goal. Thus, no Bayesian parameter fusions methods exist which may be directly applied to ordinal data analyses with rank-clustering. Alternatively, one may consider the class of continuous shrinkage priors, which include Bayesian variants of the lasso (Park & Casella, 2008) and fused lasso (Casella et al., 2010) among others (e.g., Bhattacharya et al., 2015; Carvalho et al., 2010; Griffin & Brown, 2005). However, continuous shrinkage priors do not place positive probability on coefficients (or their differences) being precisely zero. Thus, parameter fusion must be performed via thresholding the posterior distribution, which is often ad-hoc (Porwal & Rodriguez, 2024) and will not be considered in this work.

2.2. Methods based on Mallows distributions

Alternatively, one may consider rank-clustering under the Mallows family of ranking models (Mallows, 1957). The Mallows family is parameterized by the overall ranking, π_0, and a scale parameter $\theta \geq 0$ that dictates how likely rankings of a given distance to π_0 are to be drawn. Specifically, the probability of drawing a ranking π from a Mallows(π_0, θ) distribution is

$$P[\Pi = \pi | \pi_0, \theta] = \frac{e^{-\theta d(\pi, \pi_0)}}{\psi(\theta)}, \tag{3}$$

where $d(\cdot,\cdot)$ is a distance metric and $\psi(\theta)$ is a function which provides an appropriate normalizing constant. Foundational models in the family are defined by their distance metric, with common choices being the Kendall's τ (Kendall, 1938) and Spearman's ρ (Spearman, 1904).

To our knowledge, the Clustered Mallows Model (CMM) proposed by Piancastelli & Friel (2025) is the only rank-clustering method based on the Mallows model. Their work, proposed concurrently and independently to ours, models *item indifference* (i.e., rank-clusters) by permitting the overall ranking parameter π_0 to include groups of objects that are tied in rank. The model is estimated in a Bayesian framework from the observed ranking data. However, there are 3 major limitations to their work: First and most importantly, the model requires both the number of rank-clusters and the number of

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

objects per cluster to be pre-specified. Although the authors propose sensible and efficient tools for model selection, the requirement opens the possibility of model misspecification. For example, given seven objects there are 127 model specifications; given 10 objects there are 1,023 model specifications. In addition, pre-specifying the rank-clustering structure removes any uncertainty in the number of rank-clusters and their sizes from the inference task, which we believe to be of key interest in many applications. Second, Bayesian inference of a Clustered Mallows model is in the class of doubly-intractable problems since the proposed model's normalizing constant is not available in closed form. As a result, exact inference may be computationally slow, or approximation methods may need to be used that require an inexact pseudolikelihood approach. Third, the Mallows model is best suited for ordinal data in the form of complete or partial rankings, meaning the CMM cannot handle pairwise or groupwise comparisons. As will be shown in Section 3, our proposed model avoids all three issues by incorporating parameter fusion in the continuously-parameterized BTL model family.

3. The Rank-Clustered BTL model

In this section, we first develop a novel spike-and-slab prior for parameter fusion based on partitions. Then, we employ the prior in a model for rank-clustering based on the BTL family of ordinal data models.

3.1. PSSF prior

Suppose data are drawn exchangeably from a model, $\mathcal{M}$, parameterized by the vector ω. We suppose ω is of length J and let each $\omega_j \in \Omega$, $\Omega \subseteq \mathbb{R}$. Our goal is to estimate ω under the belief that some pairs or groups of parameters in ω may be clustered (i.e., *fused*). We say that two parameters $m, n \in \{1, \ldots, J\}$, $m \neq n$, are clustered precisely when $\omega_m = \omega_n$. Clustered parameters may take on any value in their domain, Ω.

Before specifying the prior, we provide some notation on partitions. A partition of an object set $\mathcal{J} = \{1, 2, \ldots, J\}$ is a collection $g = \{C(1), C(2), \ldots, C(K)\}$ of K disjoint nonempty subsets (henceforth referred to as "clusters") of $\mathcal{J}$ such that their union forms $\mathcal{J}$. Let $C^{-1}(j)$ represent the cluster that contains object $j \in \mathcal{J}$. We let $S(k) = \left|\{C(k)\}\right|$ be the size of the subset $C(k)$, and denote by K the number of clusters in g. To emphasize dependence on g, we often write K_g, $C_g(k)$, etc. Lastly, we let $\mathcal{G}$ represent the collection of all partitions g of $\mathcal{J}$, and let $\mathcal{G}_k = \{g \in \mathcal{G} : K_g = k\}$.

We are now ready to specify the PSSF prior. Under PSSF, ω is assumed to be generated via the following hierarchical model:

$$\begin{aligned} G &\sim f_G \\ v_k | G = g &\overset{iid}{\sim} f_v && k = 1, 2, \ldots, K_g \\ \omega_j &= v_{C_g^{-1}(j)} && j \in \mathcal{J}. \end{aligned} \tag{4}$$

In Equation (4), $f_G(\cdot)$ is a probability mass function on $\mathcal{G}$ and $f_v(\cdot)$ is a probability density function on Ω. In words, the prior generates a partition g, and then assigns a unique value v_k to each cluster $C(k) \in g$. Last, each parameter in ω is assigned the value of v corresponding to its cluster in g.

As an example, suppose $\mathcal{J} = \{1, 2, 3\}$ and we draw $g = \{C(1), C(2)\}$ such that $C(1) = \{2\}$ and $C(2) = \{1, 3\}$, and draw $v = [5, 10]$. Then, $\omega = [10, 5, 10]$ because,

$$\begin{aligned} \omega_1 &= v_{C_g^{-1}(1)} = v_2 = 10, \\ \omega_2 &= v_{C_g^{-1}(2)} = v_1 = 5, \text{ and} \\ \omega_3 &= v_{C_g^{-1}(3)} = v_2 = 10. \end{aligned}$$

3.1.1. Marginal prior probabilities

A useful feature of the PSSF prior is that, regardless of f_G, the marginal distribution of each ω_j follows f_v. This is because,

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

$$P[\omega_j] = \sum_{k=1}^{J} P[v_k | j \in C(k)] P[j \in C(k)] \tag{5}$$

$$= P[v_1] \sum_{k=1}^{J} P[j \in C(k)] \tag{6}$$

$$= f_v(\cdot). \tag{7}$$

Equation (5) holds as there cannot be more than J clusters and each object belongs to precisely one cluster, Equation (6) holds by the exchangeability of v_k, and Equation (7) holds since $P[v_1] = f_v(\cdot)$ by definition and the Law of Total Probability.

3.1.2. Relationship to spike-and-slab

We have not yet explained the proposed PSSF prior's relationship to the spike-and-slab. It is easiest to understand their connection by considering the joint prior distribution on two arbitrary component parameters, ω_m and ω_n, such that $m \neq n$. Due to the partitioning structure of parameters in the PSSF prior, there is prior probability associated with a parameter cluster. Thus, their joint prior distribution contains a "spike" component along the line $\omega_m = \omega_n$, with density of that line determined by f_v. Oppositely, given $\omega_m \neq \omega_n$ their joint prior distribution reflects independent draws from f_v.

Figure 1 gives examples of the PSSF prior under varying choices of f_G and f_v. In all panels, we let $\mathcal{J} = \{1,2\}$ and display the joint prior distribution of (ω_1, ω_2). In this setting, there are only two unique partitions, $g = \{1,1\}$ and $g = \{1,2\}$. Thus, we specify the prior f_G by stating the so-called "cluster probability," i.e., the probability that $g = \{1,1\}$. Columns correspond to cluster probabilities $0.1, 0.5$, and 0.9, respectively. Rows correspond to $f_v = \text{Normal}(0,1)$ and $\text{Gamma}(5,3)$, respectively. We notice that as the cluster probability increases, so does the density of points in the spike component. Regardless of f_G, marginal distributions of each parameter follow f_v. The marginal relationships seen in Figure 1 hold identically even as $\mathcal{J}$ grows.

Furthermore, we show the difference between parameters, $\omega_2 - \omega_1$, between different scenarios in Figure 2. The rows and columns are identical to that in Figure 1 and make clear the relationship between the PSSF prior and the traditional spike-and-slab, which has a spike component at 0 and a background slab density.

3.2. Rank-Clustered BTL model

We now introduce the Rank-Clustered BTL model for ordinal data. Let I be the number of judges who assess J objects. Let Π_i represent the ordinal preference provided by judge i, which may be a partial ranking, complete ranking, pairwise comparison, or groupwise comparison. Let R_i be the number of objects ranked by judge i, i.e., $R_i = |\Pi_i|$. When $R_i < J$, his/her ranking is partial. Let $\mathcal{S}_i$ denote the objects considered by judge i when forming his/her ranking, such that $\mathcal{S}_i \subseteq \mathcal{J}$. When $\mathcal{S}_i \subset \mathcal{J}$, his/her ranking is incomplete. R_i and $\mathcal{S}_i$ are assumed known.

Under the *Rank-Clustered BTL* model, we assume ordinal data is generated via the following Bayesian model:

$$\begin{aligned} \omega &\sim \text{PSSF}(f_G \propto \text{Poisson}(K_g|\lambda), f_v = \text{Gamma}(v_k|a_\gamma, b_\gamma)) \\ \Pi_i|\omega &\overset{iid}{\sim} \text{BTL}(\omega|\mathcal{S}_i, R_i) \qquad i = 1, \ldots, I. \end{aligned} \tag{8}$$

Rank-Clustered BTL applies the proposed PSSF prior under specific choices of f_G and fv to the BTL family of distributions for ordinal data. Note that the data-generating BTL distribution is identifiable up to scalar multiplication of ω. However, the proposed Bayesian model does not suffer from identifiability issues due to the non-uniform prior on ω (Johnson et al., 2022). We emphasize that unlike existing rank-clustering methods, the proposed model does not pre-specify the number of clusters, a specific

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

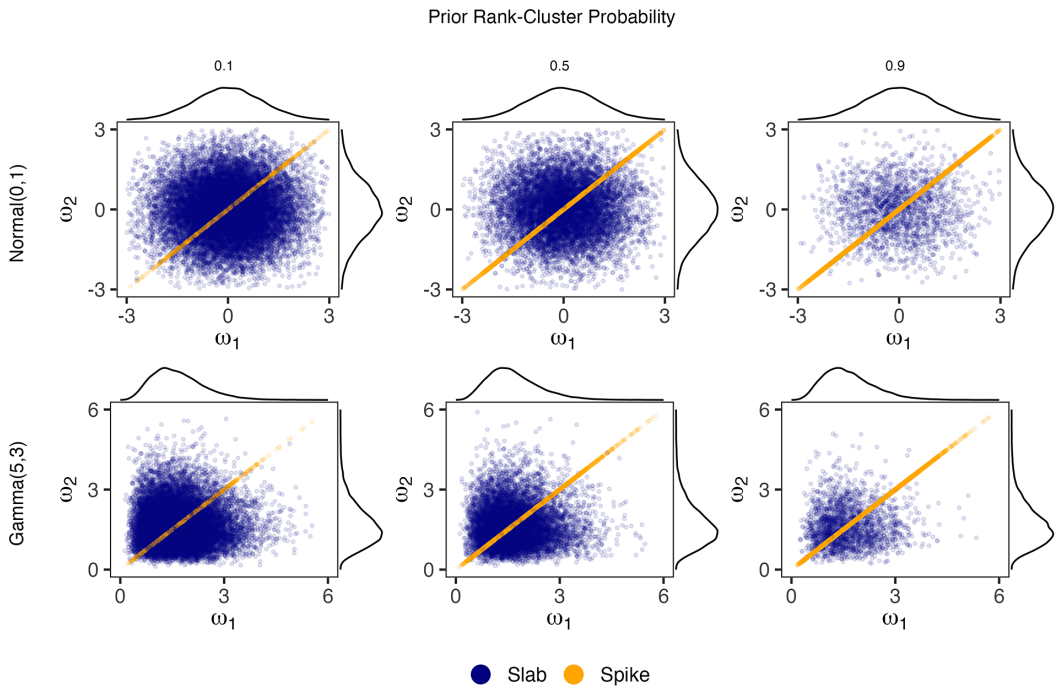

Figure 1. Joint distribution of (ω_1, ω_2) under the PSSF prior with varying combinations of f_G and f_ν.
Note: In all cases, $\mathcal{J} = \{1,2\}$, and plots show 20,000 sampled values with marginal density estimates along the axes. Rows correspond to the choice of f_ν and columns to f_G.

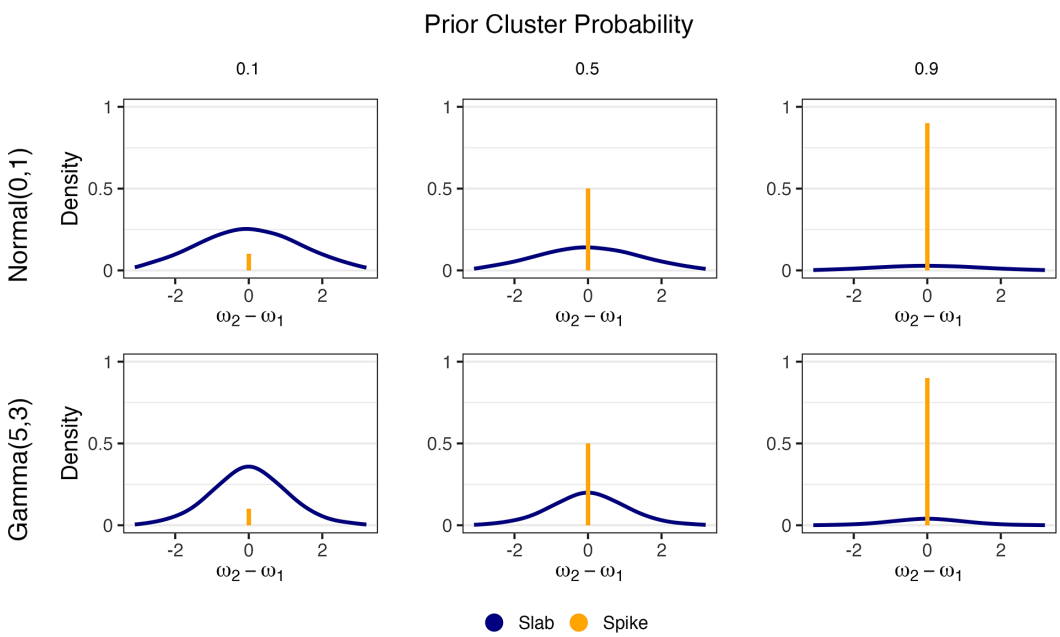

Figure 2. Distribution of $\omega_2 - \omega_1$ under the PSSF prior with varying combinations of f_G and f_ν.
Note: In all cases, $\mathcal{J} = \{1,2\}$. Rows correspond to the choice of f_ν and columns to f_G.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

rank-clustering structure, or the order of objects. These are treated as random variables and estimated simultaneously.

3.2.1. Prior selection

We now discuss the selection of priors and hyperparameters. We set f_G according to

$$f_G(g) \propto \text{Poisson}(K_g|\lambda). \tag{9}$$

In words, the prior probability of drawing a specific partition g depends only on how many unique clusters, K_g, it contains. This prior is intentionally vague to permit a variety of rank-clustering patterns. Note that every partition with the same K_g has equal prior probability. As a consequence, cluster sizes do not explicitly impact the prior probability of each g.[2] Still, there is an implicit connection between cluster size and K_g. For example, if $K_g = J$, every cluster must be a singleton. In this setup, one could set $\lambda \approx 1$ to encourage rank-clustering, or $\lambda \approx J$ to discourage rank-clustering. Next, we set f_v according to

$$f_v(v_k) = \text{Gamma}(v_k|a_\gamma, b_\gamma). \tag{10}$$

This Gamma prior has been used in Bayesian estimation of BTL models as it allows for closed-form Gibbs sampling via data augmentation (Caron & Doucet, 2012; Mollica & Tardella, 2017). The hyperparameters a_γ and b_γ control the prior distribution on the worth parameters. Since ω is invariant to multiplicative transformations, a_γ and b_γ are generally non-influential. Nonetheless, because the ratios between worth parameters could become very large when one object is strongly preferred over another, (a_γ, b_γ) should be chosen to give some density to values near 0 to allow for such extreme ratios.

3.2.2. Goodness-of-fit

To assess the adequacy of an estimated Bayesian model to observed data, we use a *posterior predictive p-value* (Gelman et al., 2013, p. 146),

$$p = P\Big(T(\Pi^{rep}; g, v) \geq T(\Pi^{obs}; g, v) \;\Big|\; \Pi^{obs}\Big),$$

where Π^{rep} is a draw from the posterior predictive distribution, Π^{obs} is the observed data, $T(\Pi; g, v)$ is a *discrepancy measure* chosen to test a specific quality of the assumed model, and the probability is taken over the posterior distribution of parameters g, v and the posterior predictive distribution of Π. Based on Yao & Böckenholt (1999) and Mollica & Tardella (2017), we employ a discrepancy measure that considers the number of times item j beats item j', denoted $\tau_{jj'}$, for $j, j' = 1, \ldots, J$. Specifically,

$$T(\Pi; g, v) = \sum_{j<j'} \frac{(\tau_{jj'} - \tau^*_{jj'})^2}{\tau^*_{jj'}},$$

where $\tau^*_{jj'}$ is the theoretical frequency expected under an assumed model with parameters g, v. Under a well-fitting model, the posterior predictive p-value would be near 0.5, with small values indicating inadequate model fit.

4. Bayesian estimation

In this section, we develop a Gibbs sampler for Bayesian estimation of Rank-Clustered BTL models and provide simulations to demonstrate its performance under varying numbers of observations and rank-clusters.

[2] It is possible to specify f_G such that cluster sizes explicitly impact the prior probability of each g. For example, one could set $f_G \sim \text{Poisson}(K_g^{(1)}|\lambda)$, where $K_g^{(1)}$ is the size of the first-place rank-cluster in g. In this case, model estimation would be unchanged beyond a substitution of the new prior likelihood for f_G in Equation (14).

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Algorithm 1 Gibbs sampler for Rank-Clustered BTL models

1. Initialize $g^{(0)}, v^{(0)}$ at random, ensuring that $|v^{(0)}| = K_{g^{(0)}}$.
2. For $t = 1, 2, \ldots, T_1$,
   (a) Sample $g^{(t)}$ via its full conditional using RJMCMC in order to traverse the space of partitions of varying numbers of clusters.
   (b) Sample $v^{(t)}$ via its full conditional T_2 times, which is possible via closed-form Gibbs sampling with data augmentation.

4.1. Gibbs sampler

Equation (4) defines ω by the pair (v, g). Thus, to estimate ω, we sample from the joint posterior distribution of (v, g). We do so using a RJMCMC Gibbs sampler that alternates between updating g and v via their full conditionals after data augmentation. The sampler is summarized in Algorithm 1.

Based on our experience fitting Rank-Clustered BTL models to real and simulated data, we recommend initializing $g^{(0)} = \{1, 2, \ldots, J\}$ (and thus $K_{g^{(0)}} = J$) as it allows rank-clusters to be formed during the estimation process (as opposed to being imposed by the analyst during initialization). For Step 2, T_1 should be sufficiently large to allow for convergence of the MCMC chain, although specific choices are context-dependent. Step 2(a) performs RJMCMC on clusters of objects. Since RJMCMC can be slow to converge in high dimensions, it is important to run multiple chains and assess for mixing and convergence (Gelman et al., 2013). Step 2(b) relies on a closed-form Gibbs sampler. We find $T_2 \leq 5$ is usually sufficient for posterior sampling.

4.1.1. Details of Step 2(a)

We now detail Step 2(a), which proposes a new partition g' based on the current partition g. Since (v, g) are intricately tied, v must simultaneously be updated to an appropriate v'. The sampling of discrete partitions is challenging to perform efficiently. In a seminal paper on RJMCMC, Green (1995) provided a method for sampling partitions. We adapt that work for the Rank-Clustered BTL model.

Following Green (1995), we only propose g' which are slight modifications of g: Precisely, we allow only for "births" splitting one cluster into two, or "deaths" merging two clusters into one. Since all partitions have positive probability, this process is irreducible, as required. There is no need to propose g' that shuffle the partitions but maintain the number of clusters, as these partitions may be obtained by successive birth and death moves.

Births are attempted with probability $b_g = 0.5$.[3] In this case, we select a cluster k at random among those with at least two objects. The cluster is split "binomially", meaning that each object is placed independently into one of the "child" subgroups, k_1 or k_2, with equal probability, conditional on each subgroup ultimately containing at least one object. Deaths are attempted with probability $d_g = 1 - b_g = 0.5$. In a death, two adjacent clusters are merged at random. Adjacency means that $\nexists\, k : v_k \in (v_{k_1}, v_{k_2})$.

Births and deaths require updating v by increasing or decreasing its dimension by 1, respectively. In a birth, we split a cluster's worth v_k into (v'_{k_1}, v'_{k_2}) using,

$$v'_{k_1} = u v_k, \quad v'_{k_2} = u^{-1} v_k, \tag{11}$$

where $u \sim \text{Unif}(0.5, 1.5)$. The corresponding death solves these equations simultaneously:

$$v_k = \sqrt{v'_{k_1} v'_{k_2}}. \tag{12}$$

For reversibility, we automatically reject proposed births where v'_{k_1}, v'_{k_2} are not adjacent.

[3] One could specify an alternative $b_g \in (0, 1)$ or make b_g a function of K_g (as in Green, 1995). For simplicity, we fix $b_g = 0.5$.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Per Green (1995), the Metropolis–Hastings probabilities for a birth and death, respectively, are $\min(1,A)$ and $\min(1,A^{-1})$, where

$$A = \frac{P(v',g'|\Pi)}{P(v,g|\Pi)} \times \frac{q(v,g|v',g')}{q(v',g'|v,g)P(u)} \times \left|\frac{\partial(v'_{k_1},v'_{k_2})}{\partial(u,v_k)}\right|, \tag{13}$$

where $q(v',g'|v,g)$ is the transition probability of sampling (v',g') given current parameter set (v,g). We now calculate each term in A. First,

$$\begin{aligned}\frac{P(v',g'|\Pi)}{P(v,g|\Pi)} &= \frac{P(\Pi|v',g')P[v'|g']P[g']}{\sum_{g''}\int_{v''}P(\Pi|v'',g'')P[v''|g'']dv''P[g'']}\frac{\sum_{g''}\int_{v''}P(\Pi|v'',g'')P[v''|g'']dv''P[g'']}{P(\Pi|v,g)P[v|g]P[g]}\\ &= \frac{P(\Pi|v',g')P[v'|g']P[g']}{P(\Pi|v,g)P[v|g]P[g]}\\ &= \frac{P(\Pi|v',g')}{P(\Pi|v,g)} \times \frac{\text{Gamma}(v'_{k_1}|a_\gamma,b_\gamma)\text{Gamma}(v'_{k_2}|a_\gamma,b_\gamma)}{\text{Gamma}(v_k|a_\gamma,b_\gamma)} \times \frac{P[g']}{P[g]},\end{aligned} \tag{14}$$

where $P(\Pi|v,g)$ and $P[g]$ are defined by Equation (8). Second,

$$\begin{aligned}\frac{q(v,g|v',g')}{q(v',g'|v,g)P(u)} &= \frac{d_{g'} \times \frac{1}{K_{g'}-1}}{\left(b_g \times \frac{1}{\#\{l:S_l(g)\geq 2\}} \times \frac{2}{2^{S_g(k)}-2}\right)\left(\frac{1}{1.5-0.5}\right)}\\ &= \frac{d_{g'}\#\{l:S_g(l)\geq 2\}(2^{S_g(k)-1}-1)}{b_g(K_{g'}-1)}.\end{aligned} \tag{15}$$

The numerator in Equation (15) is the death probability, $d_{g'}$, times the probability of selecting a pair of adjacent partitions given $K_{g'}$ total partitions after a split (there are $K_{g'}-1$ such pairs). The denominator is the birth probability, b_g, times the probability of selecting a specific cluster k among those with at least two members. This term also includes the probability of dividing the $S_g(k)$ objects in cluster k into two non-empty subsets. There are $(2^{S_g(k)}-2)/2$ such subsets, since there are $2^{S_g(k)}$ total possible partitions, two empty partitions, and two ways to obtain each two-way split. Third and last,

$$\begin{aligned}\left|\frac{\partial(v'_{k_1},v'_{k_2})}{\partial(u,v_k)}\right| &= \left|\begin{bmatrix}\frac{\partial}{\partial u}v'_{k_1} & \frac{\partial}{\partial v_k}v'_{k_1}\\ \frac{\partial}{\partial u}v'_{k_2} & \frac{\partial}{\partial v_k}v'_{k_2}\end{bmatrix}\right| = \left|\begin{bmatrix}\frac{\partial}{\partial u}uv_k & \frac{\partial}{\partial v_k}uv_k\\ \frac{\partial}{\partial u}v_k/u & \frac{\partial}{\partial v_k}v_k/u\end{bmatrix}\right| = \left|\begin{bmatrix}v_k & u\\ -v_k/u^2 & 1/u\end{bmatrix}\right|\\ &= \frac{2v_k}{u}.\end{aligned} \tag{16}$$

4.1.2. Details of Step 2(b)

To update v conditional on a partition g and our data, Π, we turn to a clever data augmentation trick for Bayesian estimation of Plackett–Luce models as seen in Caron & Doucet (2012) and Mollica & Tardella (2017). Here, we adapt their trick to account for the more general BTL family of distributions and rank-clustering. Let $Y = \{Y_{ir}\}$ be a collection of independent random variables, $i = 1,\ldots,I$ and $r = 1,\ldots,R_i$, sampled according to

$$Y_{ir} \sim \text{Exponential}\Big(\sum_{j\in\mathcal{S}_i} v_{g^{-1}(j)} - \sum_{s=0}^{r-1} v_{g^{-1}(\pi_i(s))}\Big). \tag{17}$$

The exponential rates are precisely the denominator terms from BTL densities that are burdensome to calculate. The full conditional posterior probability $P[v|Y,\Pi,g]$ is then,

$$\begin{aligned}P[v|Y,\Pi,g] &\propto P[Y|\Pi,g,v]P[\Pi|g,v]P[g|v]P[v]\\ &\propto P[Y|\Pi,g,v]P[\Pi|g,v]P[v]\end{aligned}$$

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

$$
\begin{aligned}
&= \prod_{i=1}^{I}\prod_{r=1}^{R_i}\Big(\sum_{j\in\mathcal{S}_i} v_{g^{-1}(j)} - \sum_{s=0}^{r-1} v_{g^{-1}(\pi_i(s))}\Big) e^{-y_{ir}\left(\sum_{j\in\mathcal{S}_i} v_{g^{-1}(j)} - \sum_{s=0}^{r-1} v_{g^{-1}(\pi_i(s))}\right)} \times \\
&\quad \prod_{i=1}^{I}\prod_{r=1}^{R_i} \frac{v_{g^{-1}(\pi_i(r))}}{\sum_{j\in\mathcal{S}_i} v_{g^{-1}(j)} - \sum_{s=0}^{r-1} v_{g^{-1}(\pi_i(s))}} \times \prod_{k=1}^{K} v_k^{a_\gamma - 1} e^{-b_\gamma v_k} \\
&= \prod_{i=1}^{I}\prod_{r=1}^{R_i} v_{g^{-1}(\pi_i(r))} e^{-y_{ir}\left(\sum_{j\in\mathcal{S}_i} v_{g^{-1}(j)} - \sum_{s=0}^{r-1} v_{g^{-1}(\pi_i(s))}\right)} \times \prod_{k=1}^{K} v_k^{a_\gamma - 1} e^{-b_\gamma v_k}. \qquad (18)
\end{aligned}
$$

Given these cancellations, we notice a closed-form expression for the posterior:

$$
\begin{aligned}
P[v|Y,\Pi,g] &\propto \prod_{i=1}^{I}\prod_{k=1}^{K} v_k^{c_{ki}} e^{-v_k \sum_{r=1}^{R_i} y_{ir}\delta_{irk}} \times \prod_{k=1}^{K} v_k^{a_\gamma - 1} e^{-b_\gamma v_k} \\
&= \prod_{k=1}^{K} v_k^{a_\gamma + \sum_{i=1}^{I} c_{ki} - 1} e^{-v_k\left(b_\gamma + \sum_{i=1}^{I}\sum_{r=1}^{R_i} y_{ir}\delta_{irk}\right)} \\
&\propto \prod_{k=1}^{K} \mathrm{Gamma}\Big(v_k \,\Big|\, a_\gamma + \sum_{i=1}^{I} c_{ki}, b_\gamma + \sum_{i=1}^{I}\sum_{r=1}^{R_i} y_{ir}\delta_{irk}\Big), \qquad (19)
\end{aligned}
$$

where

$$
c_{ki} = \left|\left\{j : j \in \pi_i, g^{-1}(j) = k\right\}\right| \qquad (20)
$$

$$
\delta_{irk} = \left|\left\{j : j \in \mathcal{S}_i, j \notin \{\pi_i(1),\ldots,\pi_i(r-1)\}, g^{-1}(j) = k\right\}\right|. \qquad (21)
$$

Thus, we can sample v from a closed-form Gamma distribution after augmentation of the conditioning data Π and random variable g with Y.

Now that we have developed an efficient estimation algorithm for Rank-Clustered BTL models, we turn to a numerical simulation to demonstrate estimation accuracy under different rank-clustering regimes.

4.2. Numerical simulation

We now demonstrate accurate estimation of worth parameters and rank-clusters via a Rank-Clustered BTL model in a numerical simulation. We assume there are $J = 8$ objects which form K=1, 2, 4, or 8 rank-clusters. When $K = J = 8$, every object is independent; there are only singleton rank-clusters. In the true worth parameter vector, ω_0, rank-clustered objects have identical values and successive rank-clusters are separated in value by a factor of 4 (see Table 1 for specific values). Fourfold increases induce strong but not absolute separation between objects: For demonstration, in a pairwise tournament between an object with $\omega_1 = 1$ and $\omega_2 = 4$, the probability of selecting object 2 is,

$$
P[2 \prec 1 | \omega_1 = 1, \omega_2 = 4] = \frac{\omega_2}{\omega_1 + \omega_2} = \frac{4}{4+1} = 0.8.
$$

We also vary the Poisson hyperparameter on the number of rank-clusters, $\lambda \in \{0.1, 2, 4, 8\}$, which encourages rank-clustering to different extents and allows us to measure robustness of results when λ is somewhat misspecified. To assess consistency in the number of observations, we vary the number of judges $I \in \{50, 200, 800\}$. Finally, to assess the influence of partial and incomplete rankings, we vary the tuple $(R,S) \in \{(2,2),(4,4),(2,8),(4,8),(8,8)\}$, where R is the number of ranked objects and S is the number of objects considered by each judge. When $R < 8$ the ranking is partial, when $S < 8$ the ranking is incomplete. The set of considered objects, $\mathcal{S}_i$ for each judge i, is selected independently and uniformly at random.

For each combination of K, λ, (R,S), and I, we generate 20 independent datasets and fit a Rank-Clustered BTL distribution to each, under hyperparameters $a_\gamma = 5$ and $b_\gamma = 3$. We set

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Table 1. Simulation settings for ω_0 under varying numbers of true rank-clusters, K

Setting:	ω_0
$K = 1$	$\{4^0, 4^0, 4^0, 4^0, 4^0, 4^0, 4^0, 4^0\}$
$K = 2$	$\{4^0, 4^0, 4^0, 4^0, 4^1, 4^1, 4^1, 4^1\}$
$K = 4$	$\{4^0, 4^0, 4^1, 4^1, 4^2, 4^2, 4^3, 4^3\}$
$K = 8$	$\{4^0, 4^1, 4^2, 4^3, 4^4, 4^5, 4^6, 4^7\}$

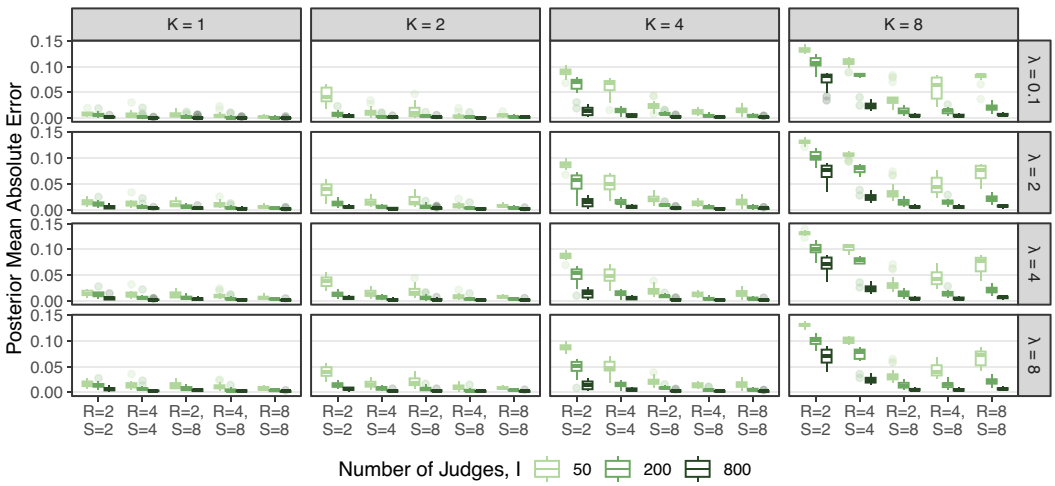

Figure 3. Boxplots of posterior mean absolute error for ω_0 across combinations of the number of judges I, true number of rank-clusters K, hyperparameter λ, number of ranked objects R, and number of assessed objects S.
Note: Errors are calculated after normalization of posterior samples such that $\sum_j \omega_{0j} = 1$.

$T_1 = 5{,}000$ and $T_2 = 2$ to obtain 10,000 posterior samples in each MCMC chain and remove the first half as burn-in. We note that no MCMC chain of length 10,000 took longer than 20 minutes to run (~0.12 seconds/iteration); many ran in under 2 minutes. For identifiability, posterior estimates of ω_0 are normalized *post-hoc* such that $\sum_j \omega_{0j} = 1$.

We first examine the accuracy of estimation for ω_0 across simulation settings. Figure 3 displays boxplots of mean absolute error (MAE) for ω_0 by number of judges I, true number of rank-clusters K, and the choice of hyperparameter λ. In general, estimation is quite accurate. We see that for any specific combination of K and λ, MAE decreases as I increases. Estimation error is higher when K is large and I is small, most likely the result of error estimating a complex rank-clustering structure.

Figure 4 displays the mean posterior probability of rank-clustering across object pairs which are truly rank-clustered (navy) or independent (gold) in ω_0.

Results are further separated by the number of judges, I, true number of clusters, K, and hyperparameter λ. For rank-clustered pairs, accuracy of recovery is generally high and increases with the number of judges, I. Accuracy is best when hyperparameter $\lambda \approx K$, which occurs when prior belief regarding the number of rank-clusters is approximately correct. If there is limited prior knowledge on the number of rank-clusters, we suggest specifying a vague hyperparameter setting such as $\lambda = \frac{I}{2}$ and assessing sensitivity of results to various choices of λ. The posterior probability of rank-clustering independent object pairs is near 0 in all simulations, indicating excellent recovery accuracy of objects with distinct worth parameters.

The numerical simulations in this section indicate that the proposed Rank-Clustered BTL model is able to accurately estimate the relative worth of objects in a collection, including in the presence of

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

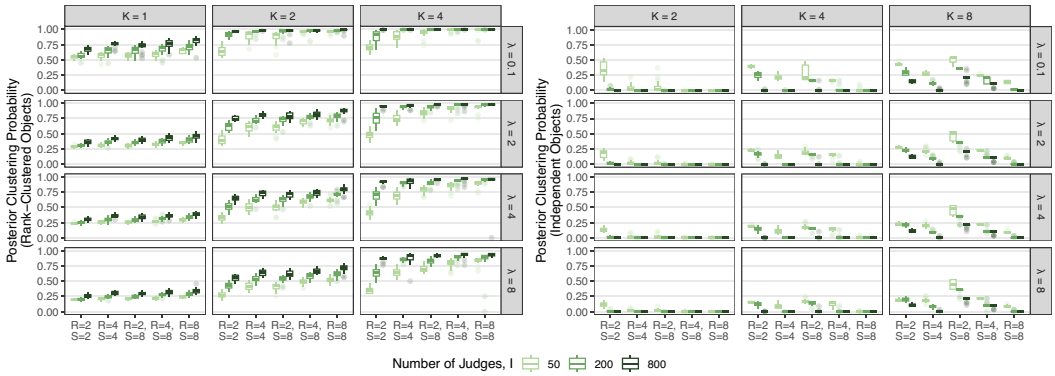

Figure 4. Boxplots of the mean posterior probability of rank-clustering object pairs which are truly rank-clustered (left) or independent (right) across combinations of I, K, λ, R, and S.

rank-clustering or partial/incomplete observed rankings. Estimation error decreases to 0 as the number of observations increases. Overall, the model correctly identifies rank-clustered and independent object pairs.

5. Applications

In this section, we apply the Rank-Clustered BTL model to four real datasets involving ordinal comparisons. These four applications were chosen to highlight the applicability of our method to various ordinal data types and domain areas and illustrate methodological values of our approach which are summarized in Table 2. The data sets are comprised of sushi preferences of Japanese adults (Kamishima, 2003), ranked-choice votes in a Minneapolis mayoral election (Minneapolis Elections and Voter Services, 2021), policy preferences of respondents from Great Britain in a Eurobarometer survey (Reif & Melich, 1993), and pairwise game outcomes among teams in the US NBA (National Basketball Association, 2024).

5.1. Sushi preferences in Tohoku

We first study complete preference rankings of 10 sushi types from a benchmarking dataset by Kamishima (2003). To allow our results to be comparable with an analysis of the sushi data by Piancastelli & Friel (2025), we analyze the preferences of survey respondents who lived in Japan's Tohoku region until at least 15 years of age. There were 280 such respondents. We fit a Rank-Clustered BTL distribution to the data with $a_\gamma = 5$, $b_\gamma = 3$, and $\lambda = 2$ to encourage rank-clustering but permit a wide variety of outcomes. We ran a total of 32,000 MCMC iterations, which took approximately 12 minutes (0.02 seconds/iteration). Figure 5 displays posterior rank-clustering probabilities (left) and parameter posteriors (right). In the left panel, the color of the (i,j) square of the clustering matrix represents the posterior probability that sushi types i and j are equal in rank at the population level. Additional results, including goodness-of-fit and convergence diagnostics, are provided in the Appendices.

Sushi types are ordered according to posterior median worth. Based on the left panel in Figure 5, fatty tuna appears to be strictly most preferred in this population, followed by tuna and shrimp rank-clustered in second place. Salmon roe and sea eel exhibit high posterior probability of rank-clustering, as do sea urchin, tuna roll, and squid; these two groups may themselves be rank-clustered. Egg and cucumber roll are rank-clustered in last place. Our results demonstrate the proposed model's ability to rank-cluster objects with uncertainty under complete rankings in survey data.

We compare our results to those found by Piancastelli & Friel (2025) in a CMM. They estimate the following ranking: fatty tuna $\prec$ tuna $\prec$ shrimp $\prec$ {salmon roe, sea urchin}$\prec${sea eel, tuna roll, squid}$\prec$

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Table 2. Summary of applications by subsection

	Setting	Data type	Methodological value
5.1	Sushi preferences in Tohoku	Complete rankings of 10 sushi types	Rank-clusters sushi types by preferences. Comparing inferred overall ranking with that of the Clustered Mallows Model.
5.2	Minneapolis mayoral election votes	Top-3 partial rankings of 17 candidates	Interpretable overall ranking captures the winner's mandate in ranked-choice elections. Comparing inferred overall ranking with those from a standard BTL model and two election procedures.
5.3	Eurobarometer survey policy preferences	Partial rankings of 7 policy options	Inferred overall ranking permits identification of similarly preferred options to aid policymakers.
5.4	Basketball game outcomes	Pairwise comparisons (game winners) among 30 teams	Inferred overall order captures similarly-performing teams. Setting with limited information and low signal.

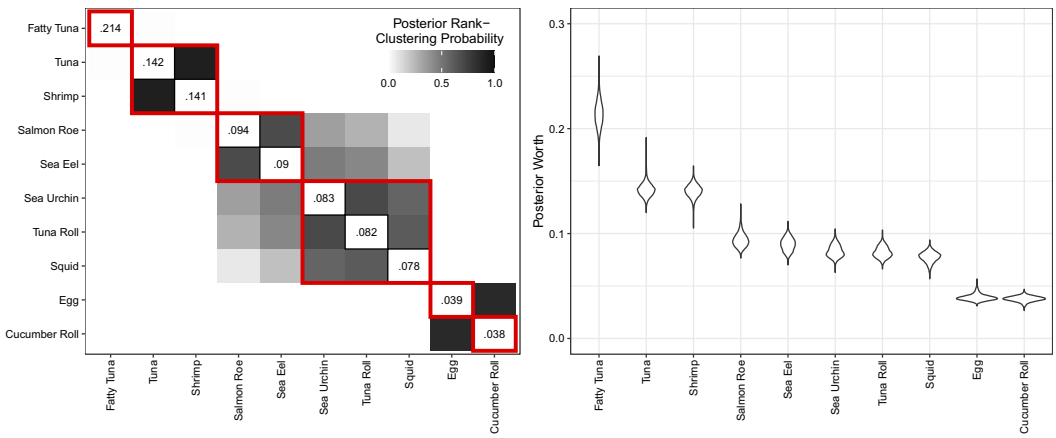

Figure 5. Primary results from Rank-Clustered BTL analysis of Tohoku sushi data.
Note: *Left*: Posterior rank-clustering probabilities. Main diagonal displays posterior median estimate of worth parameter after normalization. Red squares indicate maximum *a posteriori* rank-clusters. *Right*: Posterior distributions of sushi-specific worth parameters.

{egg, cucumber roll}. Our results are, unsurprisingly, similar, but differ in illuminating ways. Tuna and shrimp are rank-clustered in our model. The rank-clusters {salmon roe, sea eel} and {sea urchin, tuna roll, squid} swap the rank of sea eel and sea urchin. These two rank-clusters exhibit some posterior probability of rank-clustering themselves. These differences showcase how the model pre-specification required by CMM limits the flexibility of results and may not fully show what the data has to offer or fully account for uncertainty in the estimated ranks and rank-clusters. The Rank-Clustered BTL model requires no pre-specification and permits complex posterior summaries of rank-clustering, including uncertainty in the number of rank-clusters and their respective sizes.

5.2. 2021 Minneapolis mayoral election

Our second example analyzes real rank-choice votes from the 2021 mayoral election in Minneapolis, Minnesota (Minneapolis Elections and Voter Services, 2021). This election included 17 candidates (excluding write-ins and one who received no votes) and asked voters to rank their top-three choices, in order. A total of 145,337 votes were cast in this election. To mimic exit polling data, we randomly sample 1000 valid votes for analysis, which we treat as a random sample of preferences from the

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

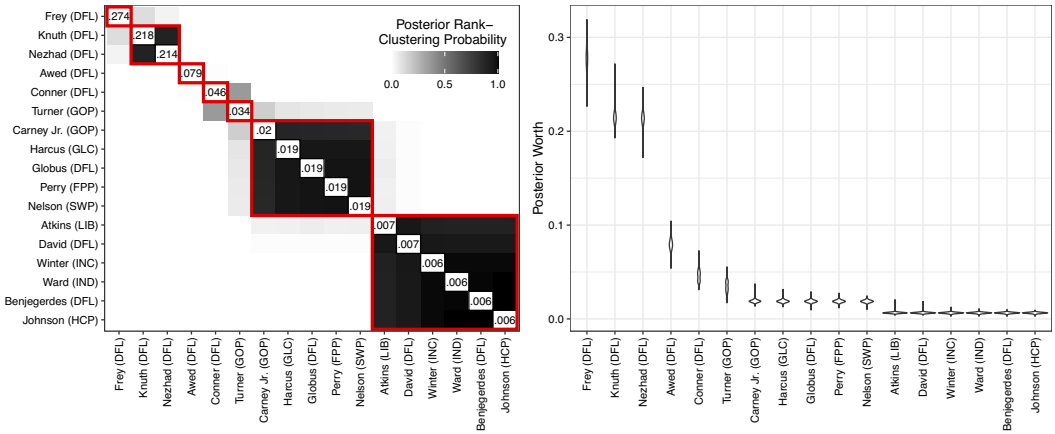

Figure 6. Primary results from Rank-Clustered BTL analysis of mayoral votes.
Note: Party abbreviations are in parentheses after candidate surnames. *Left*: Posterior rank-clustering probabilities. Main diagonal displays posterior median estimate of worth parameter after normalization. Red squares indicate maximum *a posteriori* rank-clusters. *Right*: Posterior distributions of candidate-specific worth parameters.

population of Minneapolis voters. We want to estimate the overall preferences of Minneapolis voters regarding mayoral candidates and learn which candidates, if any, are rank-clustered at the population level. Clustering candidates may be of interest to political scientists or local political organizations for the purpose of understanding voter preferences (Dimock et al., 2014; Gunther & Diamond, 2003). For example, if the winner of the election is deemed to be rank-clustered with other candidate(s), their mandate may be considered weak. Conversely, if the winner is a singleton first-place rank-cluster—clearly ranked above all other candidates—their mandate may be considered strong. We fit a Rank-Clustered BTL to the data with $a_\gamma = 5$, $b_\gamma = 3$, and $\lambda = 2$ to encourage few rank-clusters. We ran a total of 80,000 MCMC iterations, which took approximately 72.5 minutes (approximately 0.05 seconds/iteration). Figure 6 displays posterior rank-clustering probabilities (left) and parameter posteriors (right). In the left panel, the color of the (i,j) square of the clustering matrix represents the posterior probability that candidates i and j are equal in rank at the population level. Additional results, including goodness-of-fit and convergence diagnostics, are provided in the Appendices.

In Figure 6, candidates are ordered by their posterior median estimate of worth. Cluster 1 consists of Jacob Frey, the winner and incumbent. We note that Frey is not rank-clustered with other candidates with high posterior probability, suggesting a relatively strong mandate. Cluster 2 consists of Kate Knuth and Sheila Nezhad, both female, non-incumbent DFL candidates. Last, Cluster 7 consists of 6 candidates with minimal support.

Figure 7 compares point estimates of rank for each candidate across four methods. The first and second rows display assigned ranks from ranked choice and "first-past-the-post" (FPP) election procedures, respectively. We calculate FPP ranks by ordering candidates by the number of first place votes he/she received (ignoring all second and third place votes).[4] The third and fourth rows display maximum *a posteriori* ranks from a standard Bayesian BTL and our Rank-Clustered BTL, respectively. Frey wins the election in all methods. The BTL and Rank-Clustered BTL models roughly reflect the deterministic algorithms, although we notice some swaps in candidate ranks which may be attributed to differences between first place and second or third place votes. For example, Conner received fewer first place votes than Turner, but far more second and third place votes (see Appendices for vote totals). As a result, deterministic algorithms rank Turner above Conner, while the BTL model takes

[4]If the actual election had utilized FPP tabulation, results may have been different based on the differing voter strategies encouraged by ranked choice and FPP elections.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

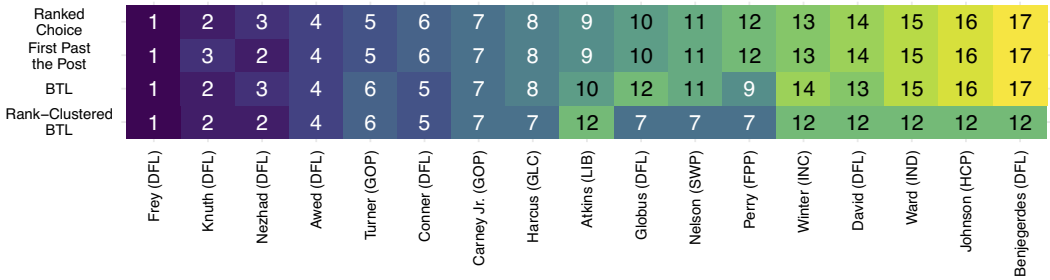

Figure 7. Comparison of estimated rank for each candidate across four aggregation methods: Ranked Choice, First-Past-the-Post (FPP), BTL, and Rank-Clustered BTL (RC BTL).
Note: Candidates are ordered by their rank in the actual ranked choice election.

into account the additional preference information and ranks Conner above Turner. In summary, the overall ordering estimated by the Rank-Clustered BTL differs from a standard BTL model and two deterministic election procedures. Furthermore, our model confirms that Frey is strictly preferred over the remaining candidates by voters.

5.3. Eurobarometer 34.1 survey data

We analyze data from the Eurobarometer 34.1 survey (Reif & Melich, 1993), which included the following question:

> Question 28: There are various actions that could be taken to eliminate the drugs problem. In your opinion, what is the first priority? And the next most urgent? (Ask respondent to rank all 7, with 1 as the most urgent.)
>
> 1. Information campaigns about the dangers of drugs.
> 2. Hunting down drug pushers and distributors.
> 3. Legal penalty for drug taking.
> 4. Looking after and treating drug addicts and rehabilitating them.
> 5. Funding research into drug substitutes, and into the treatment of drug addiction.
> 6. Fighting the social causes of drug addiction.
> 7. Reinforcing the control or distribution and usage of addictive medicines.

We subset the data to respondents from Great Britain to avoid heterogeneity and non-proportional sampling among respondents from different European countries. There were 1005 valid responses among this group (out of 1,031 total surveyed), of which 970 were complete rankings and the rest ranked between one and five items (a top-six ranking is inherently equivalent to a complete ranking since all survey options were presented). We seek to identify a population-level ordering of the priorities that accounts for potential equality or indistinguishability among the options based on the survey data. These data were previously studied by Wang et al. (2017) with a mixed-membership model to learn about heterogeneity of opinions among survey respondents. Our analysis, although a simplification of the diverse population's heterogeneous preferences, provides a simpler interpretation to policy-makers interested in understanding rank-ordering of policy preferences.

We fit a Rank-Clustered BTL model to the data with $a_\gamma = 5$, $b_\gamma = 3$, and $\lambda = 2$ to encourage rank-clustering. We ran a total of 16,000 MCMC iterations, which took approximately 12.5 minutes (0.047 seconds/iteration). Figure 8 displays posterior rank-clustering probabilities (left) and parameter posteriors (right). In the left panel, the color of the (i,j) square of the clustering matrix represents the posterior probability that policies i and j are equal in rank at the population level. Additional results, including goodness-of-fit and convergence diagnostics, are provided in the Appendices.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

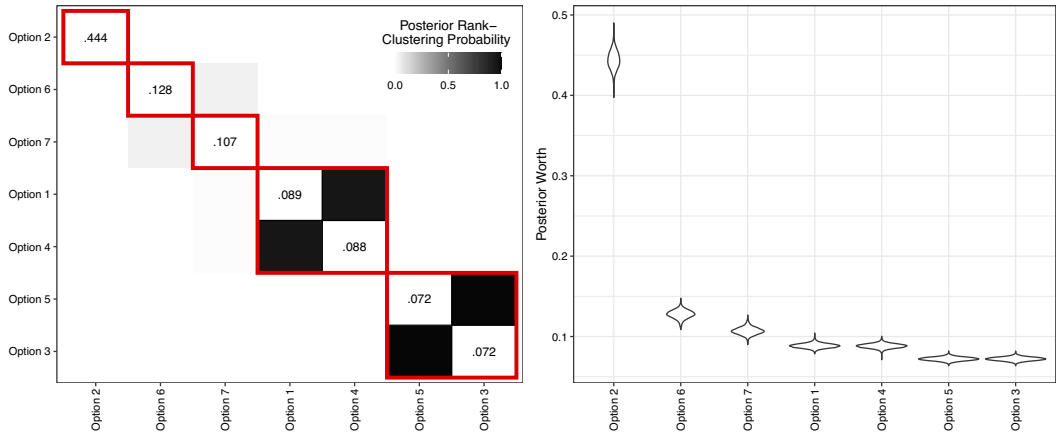

Figure 8. Primary results from Rank-Clustered BTL analysis of Eurobarometer 34.1 data.
Note: *Left*:Posterior rank-clustering probabilities. Main diagonal displays posterior median estimate of worth parameter after normalization. Red squares indicate maximum *a posteriori* rank-clusters. *Right*: Posterior distributions of policy-specific worth parameters.

Policy option 2 (*hunting drug pushers*) is strictly preferred to the rest among the population of survey respondents from Great Britain, whereas options 5 (*funding research*) and 3 (*legal penalty*) are rank-clustered last. The results indicates to policymakers that respondents in Great Britain strongly prioritize Option 2 in comparison to the rest, while pairs of Options 1 and 4 and Options 3 and 5, are, respectively, indistinguishable within each pair, with 1 and 4 being strongly preferred to 3 and 5. By rank-clustering similarly-preferred options, interpretation of constituent preferences is simplified for policymakers.

5.4. 2023–2024 NBA game outcomes

Last, we analyze outcomes of 1,230 games from the 2023–2024 season of the National Basketball Association (NBA) of the United States of America (National Basketball Association, 2024). In this season, 30 teams each played 82 games, including between two and five games against every other team. We seek to estimate an overall ranking of teams that allows for potential equality in ranking.

We fit a Rank-Clustered BTL model to the data with $a_\gamma = 5$, $b_\gamma = 3$, and $\lambda = 1$ to encourage rank-clustering given the limited ordinal comparison data provided by pairwise matchups. We ran a total of 320,000 MCMC iterations, which took approximately 22.6 hours (approximately 0.25 seconds/iteration). Figure 9 displays posterior rank-clustering probabilities (left) and parameter posteriors (right). In the left panel, the color of the (i,j) square of the clustering matrix represents the posterior probability that teams i and j are equal in rank at the population level. Additional results, including goodness-of-fit and convergence diagnostics, are provided in the Appendices.

In this setting, the Rank-Clustered BTL model estimates an ordering of professional basketball teams with uncertain rank-clustering patterns. Uncertain rank-clustering may result from two aspects of this application. First, pairwise comparisons provide little information in relation to partial or complete rankings, by construction. Second, game outcomes provide low signal measurements of team ability (Baumer et al., 2023). That is because many factors influence game outcomes, such as skill, home advantage, injuries, roster changes, and luck (Cai et al., 2019). Consistent with the low signal and limited information setting, an 80% posterior credible interval indicates that there are between 6 and 9 rank-clusters. Every team has less than 0.037 posterior probability of belonging to a singleton rank-cluster.

As seen in the left panel of Figure 9, four teams (Boston Celtics, Oklahoma City Thunder, Denver Nuggets, and Minnesota Timberwolves) appear to be rank-clustered for first place. Based on regular season data alone, our model suggests that these 4 teams were of roughly indistinguishable ability.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

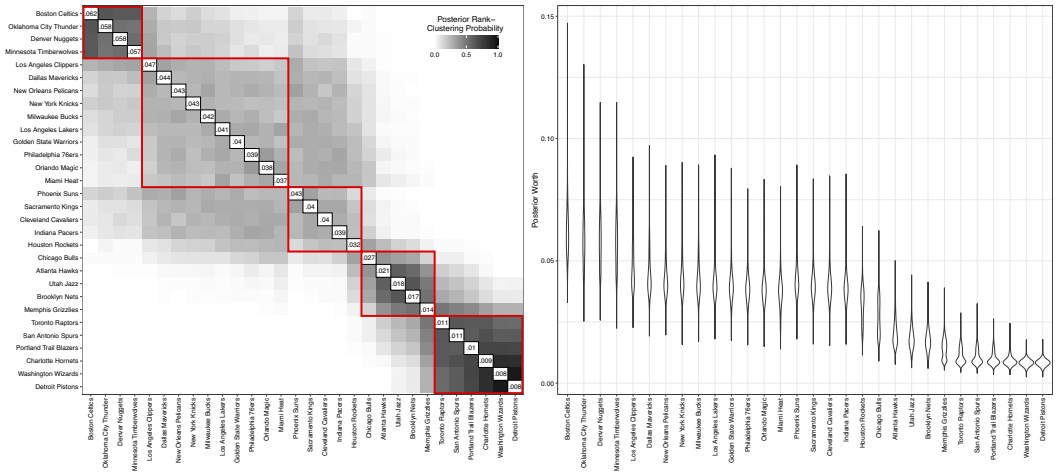

Figure 9. Primary results from Rank-Clustered BTL analysis of 2023–2024 NBA data.
Note: *Left*: Posterior rank-clustering probabilities. Main diagonal displays posterior median estimate of worth parameter after normalization. *Right*: Posterior distributions of team-specific worth parameters.

Conversely, we observe that 6 teams (Toronto Raptors, San Antonio Spurs, Portland Trail Blazers, Charlotte Hornets, Washington Wizards, and Detroit Pistons) all have a high posterior probability of rank-clustering in last place. Instead of reporting the uncertain ranking of these teams with some granularity, we recommend to infer that these teams were the worst teams of the league in this season. These rank-clusters, despite not accounting for the complexities of the sport, provide useful and interpretable summaries of the teams' abilities across the regular season. A similar analysis could be used in the future to predict postseason performance.

6. Discussion

In this article, we proposed the Rank-Clustered BTL model for estimating an overall ranking of objects with rank-clusters. The model employs the BTL family of distributions for ordinal comparisons. We proposed PSSF prior to estimates model parameters in a Bayesian framework. The model requires neither pre-specification of the number or size of the rank-clusters (improving upon Piancastelli & Friel, 2025), nor specification of lasso-based penalty parameters (improving upon Hermes et al., 2024; Jeon & Choi, 2018; Masarotto & Varin, 2012). In a simulation study, we demonstrated the model's ability to accurately and consistently estimate the relative worth of objects in a collection while simultaneously estimating rank-clusters. We used Rank-Clustered BTL on four real datasets under different types of ordinal comparison data.

In contrast to the only other spike-and-slab based prior for parameter fusion Wu et al. (2021), PSSF prior we developed does not require a known parameter order. Visual inspection of the prior distribution makes obvious its connection to spike-and-slab: "spike" components correspond to parameter clusters and "slab" components correspond to independent parameters. Estimation of parameters under this model requires reversible jump MCMC. To overcome potentially slow or computationally-burdensome estimation in this setting, we proposed a computationally efficient Gibbs sampler. The sampler alternates between updating the partition of objects, based on the seminal work of Green (1995), and updating object-level worth parameters following a data augmentation trick for standard Plackett–Luce models by Caron & Doucet (2012) that was later adapted for Plackett–Luce mixtures by Mollica & Tardella (2017).

The proposed PSSF prior requires selecting hyperpriors for partitions, f_G, and the continuous values for each unique parameter, f_ν. In this work, we specified $f_G \propto \text{Poisson}(K_g|\lambda)$ to be intentionally vague

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

over the large space of partitions and $f_v \propto \text{Gamma}(a_\gamma, b_\gamma)$ based on conjugacy. However, alternative hyperpriors are available. A Negative Binomial or Beta Negative Binomial distribution for f_G may be more appropriate when stronger prior knowledge of K_g is available. If PSSF were to be applied to linear regression for parameter fusion, a Normal or $t-$distribution may be substituted for f_v.

A useful benefit of estimating parameter values and clusters in a single Bayesian framework is the avoidance of issues associated with *selective inference* (Taylor & Tibshirani, 2015) or, more colloquially, *double dipping* (Kriegeskorte et al., 2009). Selective inference occurs when the same data is used twice in the process of model selection and/or estimation, e.g., to estimate some latent structure underlying the data and subsequently to estimate parameters conditional on that estimated structure. In our context, selective inference would occur if ordinal preference data was used first to identify rank-clusters and then used again to estimate worth parameter values conditional on those clusters. We note that selective inference occurs in the estimation of the related CMM by Piancastelli & Friel (2025), which requires selecting the number and size of rank-clusters among objects before fitting the model. Selective inference often leads to invalid inference in part because uncertainty regarding the estimated clustering structure is not taken into account. However, Rank-Clustered BTL models do not perform selective inference because parameter values and rank-clusters are estimated simultaneously. As such, we believe our parameter estimates to be more credible than those from the aforementioned methods in the literature because they rely on a fully Bayesian approach that incorporates uncertainty across the posterior distributions of both the rank-clustering structure and the specific parameter values (Gelman et al., 2013, p. 24).

Results from Rank-Clustered BTL models are useful in a variety of inferential contexts. As noted in other fusion literatures on rankings, estimated overall rankings may be easier to understand and interpret when rank-clusters of objects are identified, as rank-clusters lead to fewer rank levels of objects to distinguish (Masarotto & Varin, 2012). In contexts where model results are used for prediction, such as in sports, estimating rank-clusters may improve predictive accuracy (Tutz & Schauberger, 2015). Similarly, estimating rank-clusters is important in the context of decision-making: In peer review, for example, rank-clusters can be beneficial for communicating uncertainty in the assessment of preferences and for better transparency in funding decisions. We might imagine a scenario where a government agency is only able to fund two grants, however, two grant proposals are rank-clustered in second place. In this case, rank-clustering can be used to communicate uncertainty in the relative quality of the top proposals. A potential danger is that under this uncertainty, decision makers may be tempted to resort to unfair tie-breaking methods, e.g., selecting the proposal with the most famous author. Instead, tie-breaking should occur based on a fairer or more principled method, such as a partial lottery (Fang & Casadevall, 2016; Heyard et al., 2022; Roumbanis, 2019).

We list a few possible directions for future research. First, in this work we have not considered the level of interconnectedness among the assessed objects (e.g., if separate groups of judges assess completely distinct sets of objects). This is particularly relevant in the case of pairwise comparison data, in which some pairs of objects may never experience a head-to-head match-up. Second, the PSSF prior could be imposed as a prior for more complex BTL models or to other models entirely. In the former, the PSSF prior could be applied to preference learning via BTL distributions that incorporate covariates (e.g., Baldassarre et al., 2023; Chapman & Staelin, 1982; Gormley & Murphy, 2010; Hermes et al., 2024) or ties in the observed ordinal comparison data (e.g., Rao & Kupper (1967)). In that case, the prior may be modified to permit covariate parameter estimation in addition to rank-clustering. In the latter case, the PSSF prior may be applied to regression for variable fusion, and its performance may be compared to other existing Bayesian variable fusion methods (e.g., Casella et al., 2010; Shimamura et al., 2019; Song & Cheng, 2020). Third, we notice that the PSSF prior bears some resemblance to a Dirichlet process prior (Escobar & West, 1995). Specifically, we may consider f_v in PSSF as a base distribution in a Dirichlet process. However, the Dirichlet process' concentration parameter is related to but distinct from f_G in PSSF. Thus, the connection between Bayesian nonparametrics and Bayesian parameter fusion requires further study. Fourth, the proposed model could be studied in the framework of a latent class mixture model in order to introduce clustering among both objects (i.e. rank-clusters) and judges (i.e., preference

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

heterogeneity) simultaneously. Doing so would result in a novel form of biclustering. However, the identifiability of such a model is not clear and would require theoretical investigation.

The proposed Rank-Clustered BTL model accurately estimates rank-clusters, permitting complex summaries beyond the traditional overall ranking and allowing for improved interpretability of the results. The Bayesian Rank-Clustered BTL model relies on a novel, spike-and-slab type prior for parameter fusion, and is estimated in a computationally-efficient manner. The applications in survey data, voting, and sports to aid informed inference and decision-making illustrate methodological versatility and broad applicability of our proposed rank-clustering approach.

Supplementary material. The supplementary materials include R code and data required to reproduce our simulations and data analysis.

Data availability statement. An R implementation of the Rank-Clustered BTL is publicly available at https://github.com/pearce790/rankclust. Furthermore information on the package can be found at https://pearce790.github.io/rankclust/index.html.

Acknowledgements. The authors thank the editor, associate editor, and three anonymous referees for their insightful comments that greatly improved this work. The authors also thank Drs. T. Brendan Murphy and I. Claire Gormley for inspiring conversations on rank data modeling during the early stages of this work.

Funding statement. The authors were supported by the National Science Foundation under Grant No. 2019901.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Alvo, M. and Yu, P. L. (2014). *Statistical methods for ranking data*. (Vol. 1341). Springer.

Baldassarre, A., Dusseldorp, E., D'Ambrosio, A., Rooij, M. D., & Conversano, C. (2023). The Bradley–Terry regression trunk approach for modeling preference data with small trees. *Psychometrika*, *88*(4), 1443–1465.

Barrientos, A. F., Sen, D., Page, G. L., & Dunson, D. B. (2023). Bayesian inferences on uncertain ranks and orderings: Application to ranking players and lineups. *Bayesian Analysis*, *18*(3), 777–806.

Baumer, B. S., Matthews, G. J., & Nguyen, Q. (2023). Big ideas in sports analytics and statistical tools for their investigation. *Wiley Interdisciplinary Reviews: Computational Statistics*, *15*(6), e1612.

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, *110*(512), 1479–1490.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3–4), 324–345.

Cai, W., Yu, D., Wu, Z., Du, X., & Zhou, T. (2019). A hybrid ensemble learning framework for basketball outcomes prediction. *Physica A: Statistical Mechanics and its Applications*, *528*, 121461.

Caron, F., & Doucet, A. (2012). Efficient Bayesian inference for generalized Bradley–Terry models. *Journal of Computational and Graphical Statistics*, *21*(1), 174–196.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480.

Casella, G., Ghosh, M., Gill, J., & Kyung, M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, *5*(2), 369–411.

Chapman, R. G., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, *19*(3), 288–301.

Critchlow, D. E., & Fligner, M. A. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, *56*(3), 517–533.

Dimock, M., Doherty, C., Kiley, J., & Krishnamurthy, V. (2014). Beyond red vs. blue: The political typology. Pew Research Center.

Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th international conference on world wide web* (pp. 613–622).

Eliseussen, E., Frigessi, A., & Vitelli, V. (2023). Rank-based Bayesian clustering via covariate-informed Mallows mixtures. arXiv preprint, arXiv:2312.12966.

Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, *90*(430), 577–588.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fang, F. C., & Casadevall, A. (2016). Research funding: The case for a modified lottery. *mBio*, *7*(2), e00422.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.
George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, *7*(2), 339–373.
Gormley, I. C., & Murphy, T. B. (2008). A mixture of experts model for rank data with applications in election studies. *The Annals of Applied Statistics*, *2*(4), 1452–1477.
Gormley, I. C., & Murphy, T. B. (2010). Clustering ranked preference data using sociodemographic covariates. In H. Stephane and D. Andrew (Eds.), *Choice modelling: The state-of-the-art and the state-of-practice*. Emerald Group Publishing Limited.
Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732.
Griffin, J., & Brown, P. (2005). *Alternative prior distributions for variable selection with very many more variables than observations*. Technical report, University of Warwick.
Gunther, R., & Diamond, L. (2003). Species of political parties: A new typology. *Party Politics*, *9*(2), 167–199.
Hermes, S., van Heerwaarden, J., & Behrouzi, P. (2024). Joint learning from heterogeneous rank data. arXiv preprint, arXiv:2407.10846.
Heyard, R., Ott, M., Salanti, G., & Egger, M. (2022). Rethinking the funding line at the Swiss national science foundation: Bayesian ranking and lottery. *Statistics and Public Policy*, *9*(1), 110–121.
Hunter, D. R., et al. (2004). MM algorithms for generalized Bradley–Terry models. *The Annals of Statistics*, *32*(1), 384–406.
Ishwaran, H., & Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, *33*(2), 730–773.
Jeon, J.-J., & Choi, H. (2018). The sparse Luce model. *Applied Intelligence*, *48*, 1953–1964.
Johnson, S. R., Henderson, D. A., & Boys, R. J. (2022). On Bayesian inference for the extended Plackett-Luce model. *Bayesian Analysis*, *17*(2), 465–490.
Kamishima, T. (2003). Nantonac collaborative filtering: Recommendation based on order responses. In *Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 583–588).
Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1–2), 81–93.
Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.
Luce, R. D. (1959). *Individual choice behavior*. John Wiley and Sons, Inc.
Mallows, C. L. (1957). Non-null ranking models. I. *Biometrika*, *44*(1–2), 114–130.
Marden, J. I. (1996). *Analyzing and modeling rank data*. CRC Press.
Masarotto, G., & Varin, C. (2012). The ranking lasso and its application to sport tournaments. *The Annals of Applied Statistics*, *6*(4), 1949–1970.
Maystre, L., & Grossglauser, M. (2015). Fast and accurate inference of Plackett–Luce models. In *Advances in neural information processing systems* (Vol. 28).
Minneapolis Elections, & Voter Services. (2021). 2021 mayor results. https://vote.minneapolismn.gov/results-data/election-results/2021/mayor/.
Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032.
Mollica, C., & Tardella, L. (2017). Bayesian Plackett–Luce mixture models for partially ranked data. *Psychometrika*, *82*(2), 442–458.
National Basketball Association (2024). NBA 2023-24 regular season standings.
Nguyen, D., & Zhang, A. Y. (2023). Efficient and accurate learning of mixtures of Plackett–Luce models. In *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. *37* (pp. 9294–9301). arXiv:2302.05343.
Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.
Piancastelli, L. S., & Friel, N. (2025). The clustered Mallows model. *Statistics and Computing*, *35*(21). arXiv preprint, arXiv:2403.12880.
Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *24*(2), 193–202.
Porwal, A., & Rodriguez, A. (2024). Laplace Power-Expected-Posterior Priors for Logistic Regression. *Bayesian Analysis*, *19*(4), 1163–1186.
Rao, P., & Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley–Terry model. *Journal of the American Statistical Association*, *62*(317), 194–204.
Reif, K., & Melich, A. (1993). *Euro-barometer 34.0: Perceptions of the european community, and employment patterns and child rearing, October-November, 1990*. Inter-university Consortium for Political and Social Research.
Roumbanis, L. (2019). Peer review or lottery? A critical analysis of two different forms of decision-making mechanisms for allocation of research grants. *Science, Technology, & Human Values*, *44*(6), 994–1019.
Shimamura, K., Ueki, M., Kawano, S., & Konishi, S. (2019). Bayesian generalized fused lasso modeling via NEG distribution. *Communications in Statistics–Theory and Methods*, *48*(16), 4132–4153.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Song, Q., & Cheng, G. (2020). Bayesian fusion estimation via t shrinkage. *Sankhya A*, *82*(2), 353–385.
Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*. *15*(1), 72–101.
Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, *112*(25), 7629–7634.
Thompson Jr., W., & Singh, J. (1967). The use of limit theorems in paired comparison model building. *Psychometrika*, *32*(3), 255–264.
Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *58*(1), 267–288.
Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(1), 91–108.
Turner, H. L., van Etten, J., Firth, D., & Kosmidis, I. (2020). Modelling rankings in R: The PlackettLuce package. *Computational Statistics*, *35*(3), 1027–1057.
Tutz, G., & Schauberger, G. (2015). Extended ordered paired comparison models with application to football data from German Bundesliga. *Advances in Statistical Analysis*, *99*, 209–227.
Vana, L., Hochreiter, R., & Hornik, K. (2016). Computing a journal meta-ranking using paired comparisons and adaptive lasso estimators. *Scientometrics*, *106*, 229–251.
Varin, C., Cattelan, M., & Firth, D. (2016). Statistical modelling of citation exchange between statistics journals. *Journal of the Royal Statistical Society, Series A(Statistics in Society)*, *179*(1), 1.
Vitelli, V., Sørensen, Ø., Crispino, M., Frigessi Di Rattalma, A., & Arjas, E. (2018). Probabilistic preference learning with the Mallows rank model. *Journal of Machine Learning Research*, *18*(158), 1–49.
Wang, Y. S., Matsueda, R. L., & Erosheva, E. A. (2017). A variational EM method for mixed membership models with multivariate rank data: An analysis of public policy preferences. *The Annals of Applied Statistics*, *11*(3), 1452–1480.
Wu, S., Shimamura, K., Yoshikawa, K., Murayama, K., & Kawano, S. (2021). Variable fusion for Bayesian linear regression via spike-and-slab priors. In *Intelligent decision technologies: Proceedings of the 13th KES-IDT 2021 conference* (pp. 491–501). Springer.
Yao, G., & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, *52*(1), 79–092.
Yellott Jr., J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*(2), 109–144.
Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *29*(1), 436–460.

Appendix A. Additional application results

Appendix A.1. Additional results from Section 5.1

Figure A.1 displays stacked bar charts of ranks received by each sushi type by the survey respondents from Tohoku. Figure A.2 shows a comparison among results obtained using the proposed Rank-Clustered BTL, a standard BTL, and the Clustered Mallows model.

Table A.1 contains posterior predictive p-values based on the discrepancy measure defined in Section 3.2.2. A p-value is calculated for both a standard BTL distribution and a Rank-Clustered BTL distribution fit to the observed data. All statistics are well above a standard 0.05 threshold, indicating acceptable fit to the observed data. We recall that posterior predictive p-values are used as tools to assess potential model misfits, and not to compare or choose among the models (Gelman et al., 2013, p. 150).

Figures A.3 and A.4 contain trace plots for K and ω after burn-in for each chain. We find the trace plots to demonstrate satisfactory mixing and convergence.

Appendix A.2. Additional results from Section 5.2

Figure A.5 displays stacked bar charts of the sampled votes by rank level for each candidate.

Candidates are ordered by their final placement according to the official ranked choice voting algorithm. The incumbent, Jacob Frey, receives the largest share of first place votes, although Kate Knuth and Sheila Nezhad also receive substantial support. The remaining candidates receive comparatively few votes. Most candidates are associated with the Democratic–Farmer–Labor (DFL) party, which is affiliated with the national Democratic Party. Laverne Turner and Bob "Again" Carney Jr. are the only Republicans (GOP) in the race. The remaining candidates represent Grassroots–Legalize Cannabis (GLC), Libertarian (LIB), Socialist Workers Party (SWP), For the People Party (FPP), Independence (INC), Independent (IND), and Humanitarian–Community Party (HCP).

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Table A.1. Posterior predictive *p*-values based on a standard BTL and Rank-Clustered BTL (RC-BTL) to assess goodness-of-fit in the Sushi data analysis

Model	BTL	RC-BTL
p-value	0.30	0.24

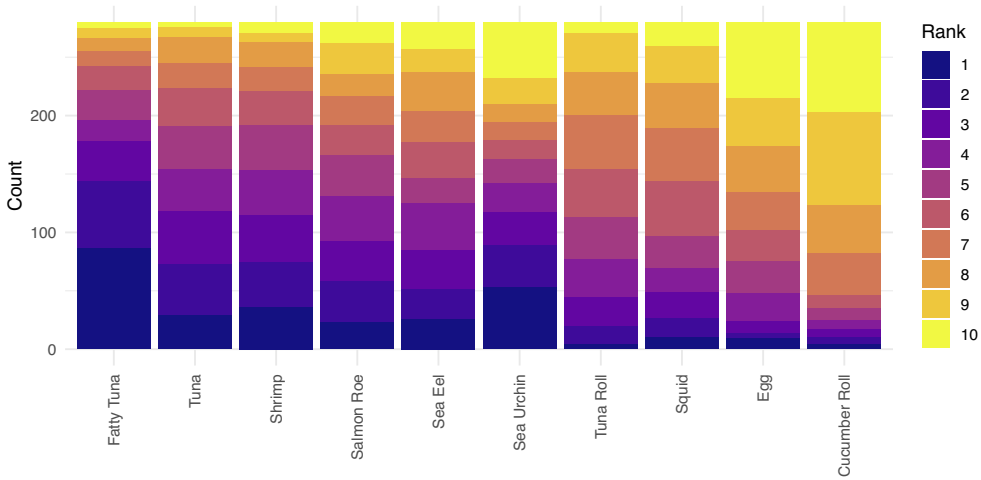

Figure A.1. Stacked bar charts of ranks received by each sushi type.

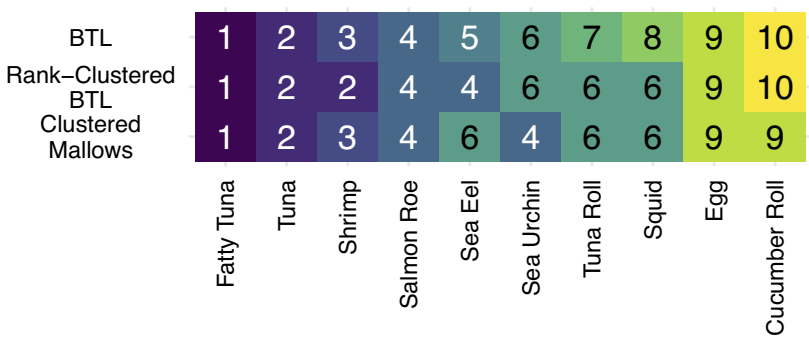

Figure A.2. Comparison of results among comparator methods for the Sushi data analysis.

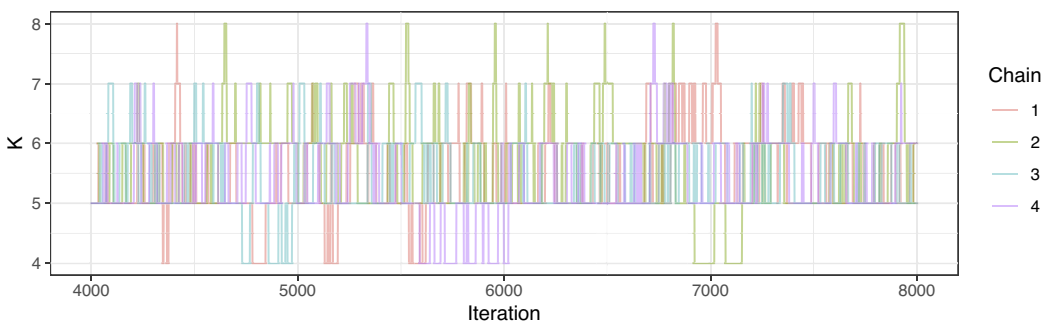

Figure A.3. Trace plot of K in the Sushi data analysis.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

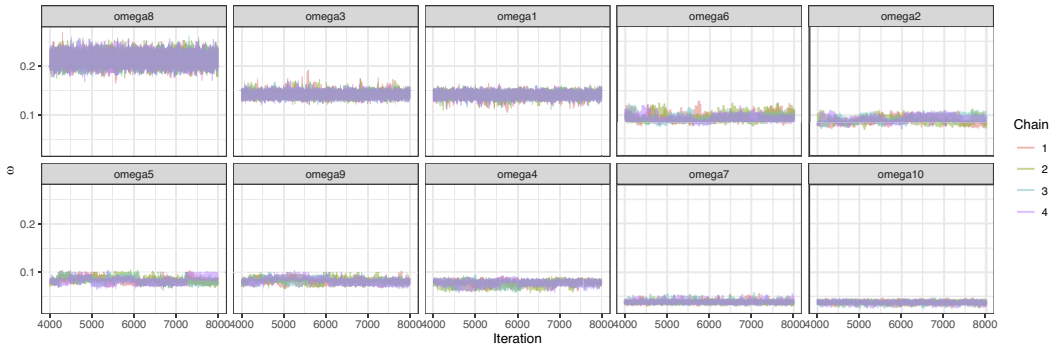

Figure A.4. Trace plots of ω in the Sushi data analysis.

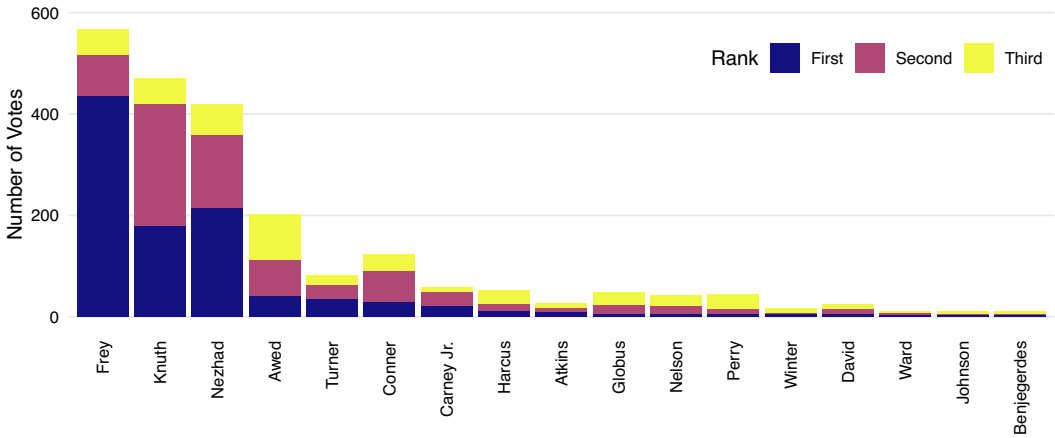

Figure A.5. Number of votes by rank level and candidate. Candidates are ordered by their position in the official ranked choice election. *Note*: Acronyms on the tops of bars represent each candidate's political party.

Table A.2. Posterior predictive p-values based on a standard BTL and Rank-Clustered BTL (RC-BTL) to assess goodness-of-fit in the Minneapolis mayoral election data analysis

Model	BTL	RC-BTL
p-value	0.41	0.30

Table A.2 contains posterior predictive p-values based on the discrepancy measure defined in Section 3.2.2. A p-value is calculated for both a standard BTL distribution and a Rank-Clustered BTL distribution fit to the observed data. All statistics are well above a standard 0.05 threshold, indicating acceptable fit to the observed data. We recall that posterior predictive p-values are used as tools to assess potential model misfits, and not to compare or choose among the models (Gelman et al., 2013, p. 150).

Figures A.6 and A.7 contain trace plots for K and ω after burn-in for each chain. We find the trace plots to demonstrate satisfactory mixing and convergence.

Appendix A.3. Additional Results from Section 5.3

Figure A.8 displays stacked bar charts of ranks received by each policy option by the survey respondents from Great Britain. Figure A.9 shows a comparison between results obtained using the proposed Rank-Clustered BTL and a standard BTL model.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

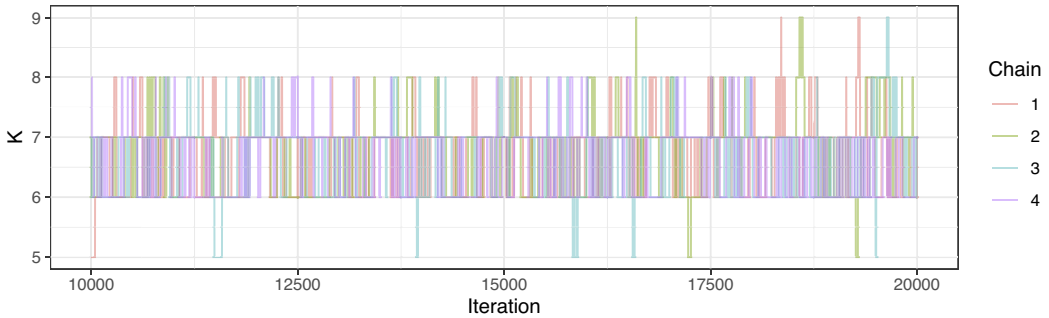

Figure A.6. Trace plots of K in the Minneapolis mayoral election data analysis.

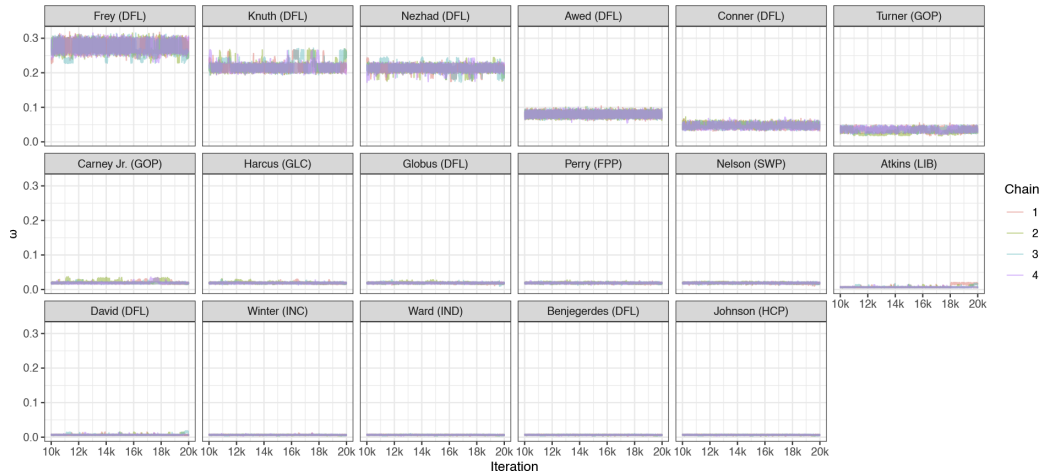

Figure A.7. Trace plots of ω in the Minneapolis mayoral election data analysis.

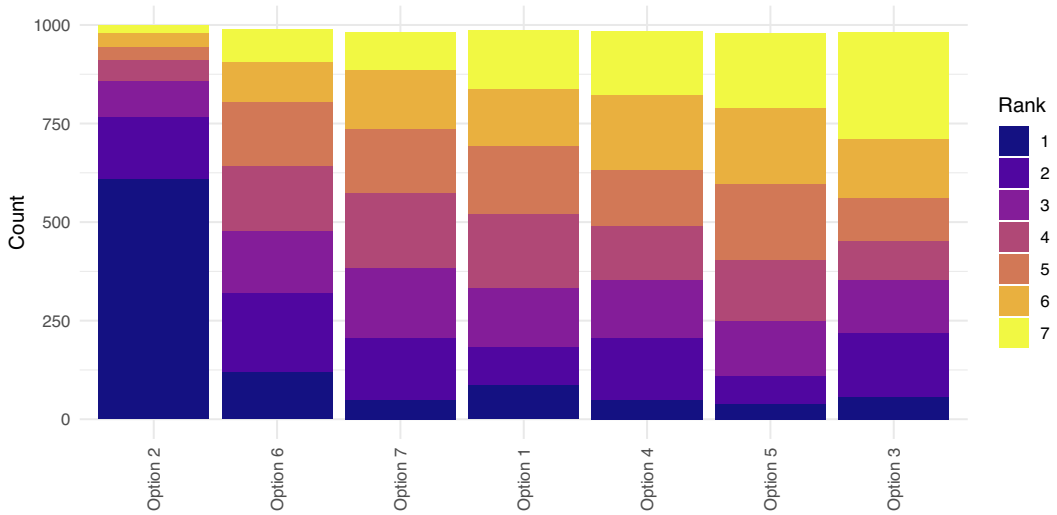

Figure A.8. Stacked bar charts of ranks received by each policy option.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Table A.3. Posterior predictive *p*-values based on a standard BTL and Rank-Clustered BTL (RC-BTL) to assess goodness-of-fit in the Eurobarometer survey data analysis

Model	BTL	RC-BTL
p-value	0.32	0.35

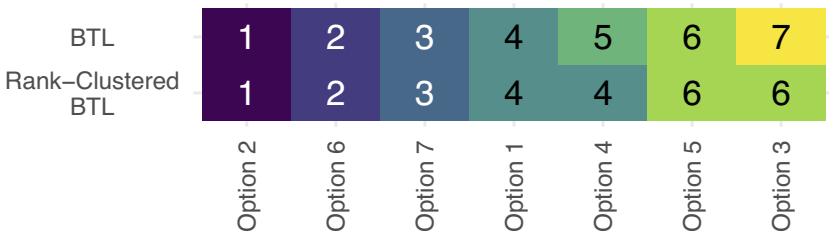

Figure A.9. Comparison of results among comparator methods for the Eurobarometer survey data analysis.

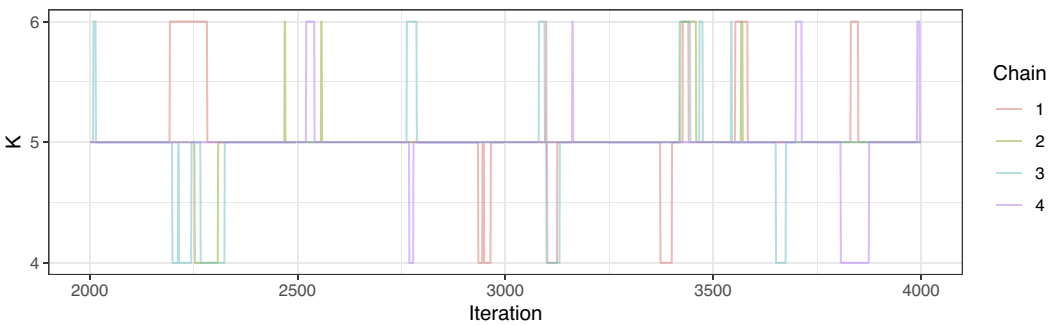

Figure A.10. Trace plot of K in the Eurobarometer survey data analysis.

Table A.3 contains posterior predictive *p*-values based on the discrepancy measure defined in Section 3.2.2. A *p*-value is calculated for both a standard BTL distribution and a Rank-Clustered BTL distribution fit to the observed data. All statistics are well above a standard 0.05 threshold, indicating acceptable fit to the observed data. We recall that posterior predictive *p*-values are used as tools to assess potential model misfits, and not to compare or choose among the models (Gelman et al., 2013, p. 150).

Figures A.10 and A.11 contain trace plots for K and ω after burn-in for each chain. We find the trace plots to demonstrate satisfactory mixing and convergence.

Appendix A.4. Additional results from Section 5.4

Figure A.12 displays stacked bar charts of the season record of each NBA team across the 2023–2024 season. Figure A.13 shows a comparison between results obtained using the proposed Rank-Clustered BTL and a standard BTL model.

Table A.4 contains posterior predictive *p*-values based on the discrepancy measure defined in Section 3.2.2. A *p*-value is calculated for both a standard BTL distribution and a Rank-Clustered BTL distribution fit to the observed data. All statistics are well above a standard 0.05 threshold, indicating acceptable fit to the observed data. We recall that posterior predictive *p*-values are used as tools to assess potential model misfits, and not to compare or choose among the models (Gelman et al., 2013, p. 150).

Figures A.14 and A.15 contain trace plots for K and ω after burn-in for each chain. We find the trace plots to demonstrate satisfactory mixing and convergence.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

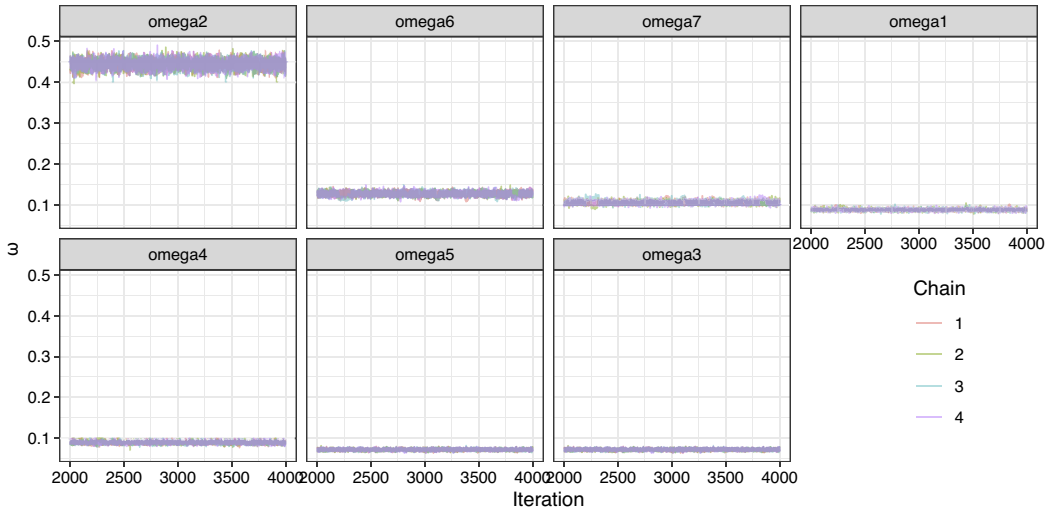

Figure A.11. Trace plot of ω in the Eurobarometer survey data analysis.

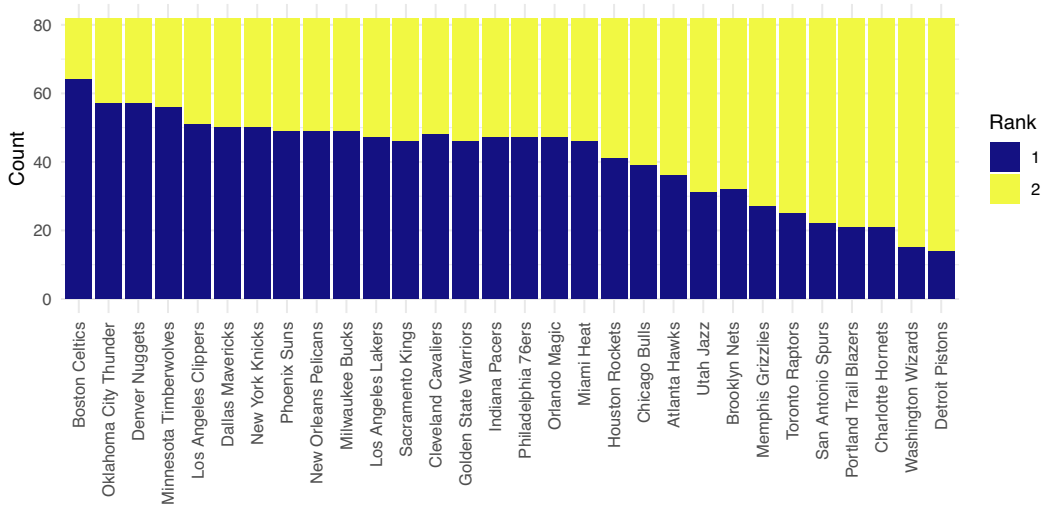

Figure A.12. Stacked bar charts of ranks received by each NBA team across the 2023–2024 season.
Note: Winning = rank 1; losing = rank 2.

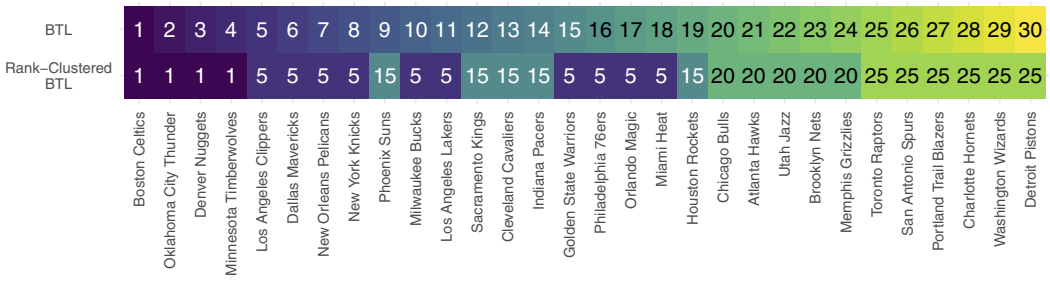

Figure A.13. Comparison of results among comparator methods for the 2023–2024 NBA season analysis.

https://doi.org/10.1017/psy.2025.10014 Published online by Cambridge University Press

Table A.4. Posterior predictive *p*-values based on a standard BTL and Rank-Clustered BTL (RC-BTL) to assess goodness-of-fit in the 2023–2024 NBA season analysis

Model	BTL	RC-BTL
p-value	0.61	0.60

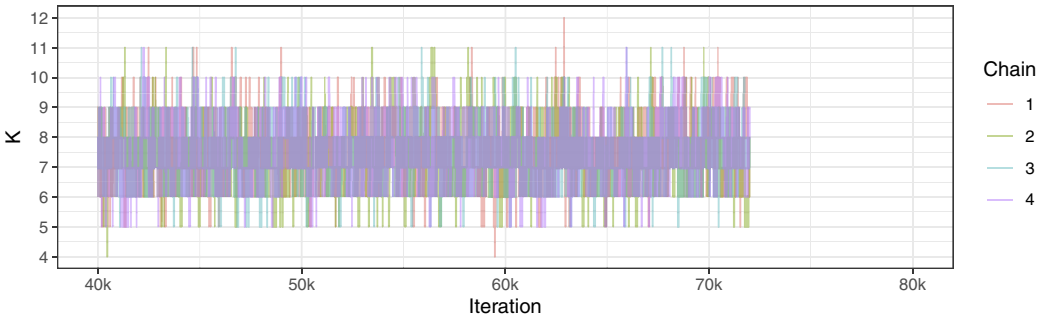

Figure A.14. Trace plot of K in the 2023–2024 NBA season analysis.

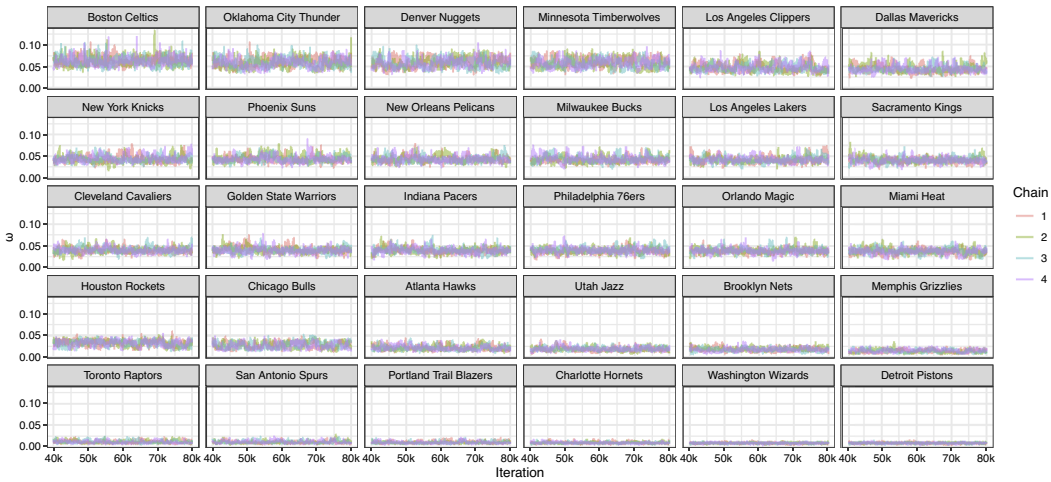

Figure A.15. Trace plot of ω in the 2023–2024 NBA season analysis.

Cite this article: Pearce, M. and Erosheva, E. A. (2025). Bayesian Rank-Clustering. *Psychometrika*, 904–931. https://doi.org/10.1017/psy.2025.10014