ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

A Bayesian Replay Framework for Latent Fan Vote Estimation and Voting Rule Bias Analysis

Yu Pei[1,a,*], Yitong Ling[2,b], Xinyu Li[3,c]

[1]School of AI and Advanced Computing, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University, Suzhou, China
[2]School of Intelligent Manufacturing Ecosystem, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University, Suzhou, China
[3]School of Robotics, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University, Suzhou, China
[a]athenaraisa2004@outlook.com, [b]Eva.Ling27683255@outlook.com, [c]Kizer.5001@outlook.com
**Corresponding author*

Abstract: *Elimination competitions combine expert judgement with audience participation, yet public datasets usually report judge scores and elimination outcomes without revealing fan votes. This missing-observation structure makes it difficult to determine whether a contestant survived because of technical quality, latent popularity, or the rule used to aggregate both signals. This paper develops a Bayesian replay framework for latent fan-vote estimation and voting-rule bias analysis. A latent popularity utility is mapped to weekly fan-share probabilities through a softmax function, combined with standardized judge scores in an elimination likelihood, and smoothed over time to obtain posterior fan-support distributions. These posterior samples are then replayed under rank aggregation, percentage aggregation, and a bottom-two judges-save mechanism. The results indicate that the estimated latent fan signal is predictive and calibrated, with overall Hit@3 close to 0.69 and expected calibration error near 0.0047. Counterfactual replay further shows that rule choice changes placement and elimination timing, particularly for controversial contestants. The judges-save mechanism increases judge influence, weakens the connection between fan support and final placement, and generates heterogeneous winner/loser effects.*

Keywords: *Latent Fan Vote; Bayesian Inference; Posterior Replay; Voting Rule; Rule Bias; Elimination Competition*

1. Introduction

Elimination competitions are sequential social-decision systems in which expert scores and audience preferences jointly determine survival. Although such systems are presented as transparent entertainment procedures, the underlying aggregation problem is non-trivial. Classical social-choice theory demonstrates that no voting rule is free from structural compromises when individual preferences are transformed into a collective decision [1]-[3]. In an elimination competition, this problem becomes operational because a rule repeatedly changes the future set of contestants and therefore the entire trajectory of the competition.

The distinction between ordinal and cardinal aggregation is especially important. Rank-based rules reduce contestants to relative positions and can suppress score-scale effects, whereas percentage-based rules preserve magnitude information but may be sensitive to normalization across weeks. Research on rank aggregation and utilitarian social choice has shown that the same underlying preference profile can generate different collective outcomes when transformed by different rules [4], [5]. For competition data, the corresponding question is whether rank, percent and judges-save mechanisms transmit fan support differently when evaluated on the same latent inputs.

A second difficulty is empirical. Judge scores are usually observable, but fan votes are not. Public records typically include weekly judge totals, contestant identities, observed eliminations and final placements, while the decisive audience signal is hidden. Bayesian categorical-response modelling provides a principled way to infer unobserved utilities from observed discrete outcomes [6], [7]. Dynamic state-space models further motivate temporal smoothing because audience support is expected to evolve across adjacent weeks rather than reset independently [8], [9].

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

Model checking and predictive evaluation are also essential for this problem. A latent fan-vote model is useful only if it reproduces observed eliminations, quantifies uncertainty and produces calibrated probabilities. Posterior predictive assessment, Bayesian model-complexity measures and modern predictive-evaluation criteria provide the statistical basis for this validation step [10]-[12]. In addition, probabilistic forecasting research emphasizes that reliability, calibration and proper scoring rules should be reported when a model outputs event probabilities [13]-[15].

Existing competition-style analyses often attempt to estimate votes, explain contestant attributes, optimize new rules and issue policy recommendations in one broad framework. Such breadth can obscure the central inferential problem. This study adopts a narrower and more reproducible logic: latent fan votes are first inferred from observed judge scores and eliminations, and posterior fan-vote samples are then reused in a replay engine that compares voting rules under identical statistical inputs.

The paper makes three contributions. First, it formulates a Bayesian latent fan-vote estimation model that combines softmax fan shares, within-week judge-score standardization, temporal smoothing and an elimination likelihood. Second, it develops a posterior replay framework for rank, percent and bottom-two judges-save mechanisms. Third, it quantifies rule bias using placement change, elimination-week change, fan-orientation correlation and sensitivity among controversial contestants.

2. Data and problem formulation

The data are organized as a season-week panel. For season s and week t, $A_{s,t}$ denotes the active contestant set. The observable variables are contestant identifiers, weekly judge totals $J_{i,s,t}$ observed eliminated contestants $E_{s,t}$ and final placements. These variables define the empirical information available to a public analyst. The fan vote share $v_{i,s,t}$ is treated as latent because raw vote totals are not reported.

Because total fan-vote volume is unobserved, the model estimates normalized weekly shares rather than absolute counts. This choice is identifiable from elimination information and makes comparisons possible across seasons with different audience sizes. The inferential target is the posterior distribution of the latent share vector $v_{s,t}$ over active contestants. The replay target is the rule-specific transformation from the same posterior fan signal to counterfactual elimination trajectories. The variables and notation used throughout the Bayesian replay framework are summarized in Table 1.

Table 1. Variables and notation used in the Bayesian replay framework.

Symbol	Definition	Role
i, s, t	Contestant, season and week indices	Indexing
$A_{s,t}$	Active contestant set in week t	Risk set
$J_{i,s,t}$	Observed judge total	Observed input
$g^{J}_{i,s,t}$	Standardized judge signal	Model covariate
$eta_{i,s,t}$	Latent popularity utility	Unobserved state
$v_{i,s,t}$	Latent weekly fan-share	Posterior target
$E_{s,t}$	Observed eliminated contestant	Likelihood outcome
r	Counterfactual voting rule	Replay condition

3. Methodology

3.1 Bayesian Latent Fan Vote Estimation

The model first standardizes judge scores within each active set. This removes week-specific scale effects and places technical-performance signals on a comparable dimension across seasons and stages:

$$\mu^{J}_{s,t} = \frac{1}{|A_{s,t}|}\sum_{j\in A_{s,t}} J_{j,s,t},\quad g^{J}_{i,s,t} = \frac{J_{i,s,t}-\mu^{J}_{s,t}}{\sigma^{J}_{s,t}} \tag{1}$$

Fan support is represented by a latent popularity utility $eta_{i,s,t}$. Unknown vote shares are obtained through a softmax transformation, which ensures positivity and the simplex constraint: This formulation follows the latent-utility interpretation of Bayesian categorical-response models.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

$$v_{i,s,t} = \frac{\exp(\eta_{i,s,t})}{\sum_{j \in A_{s,t}} \exp(\eta_{j,s,t})}, \quad \sum_{i \in A_{s,t}} v_{i,s,t} = 1 \tag{2}$$

Temporal smoothing is introduced through a random-walk state equation. This assumption reflects the expectation that audience support changes gradually across adjacent weeks unless the elimination evidence suggests otherwise:

$$\eta_{i,s,t} = \eta_{i,s,t-1} + \varepsilon_{i,s,t}, \quad \varepsilon_{i,s,t} \sim \mathcal{N}\left(0, \sigma_{f,s}^2\right), \quad t \geq 2 \tag{3}$$

Judge performance and latent popularity are combined into a survival utility. A larger utility denotes stronger survival evidence, whereas elimination is modelled as a lower-tail categorical event:

$$u_{i,s,t} = w_t g_{i,s,t}^{J} + \left(1 - w_t\right) \eta_{i,s,t} \tag{4}$$

$$\Pr\left(E_{s,t} = i \mid u_{s,t}\right) = \frac{\exp\left(-\tau_t u_{i,s,t}\right)}{\sum_{j \in A_{s,t}} \exp\left(-\tau_t u_{j,s,t}\right)} \tag{5}$$

The posterior distribution combines the elimination likelihood, the softmax share mapping, the temporal prior and weakly informative parameter priors:

$$p\left(\eta, v, w, \tau, \sigma_f \mid E, J\right) \propto p\left(E \mid \eta, J, w, \tau\right) p\left(v \mid \eta\right) p\left(\eta \mid \sigma_f\right) p\left(w, \tau, \sigma_f\right) \tag{6}$$

Posterior summaries include weekly fan-share means, 95% credible intervals, posterior predictive elimination probabilities, Hit@k and reliability diagnostics. Posterior predictive checking and Bayesian model-evaluation principles are used to assess whether the latent fan signal is consistent with observed eliminations.

3.2 Posterior Replay Rule Comparison

The replay stage treats posterior samples as common inputs for all rules. For posterior draw m and rule r, the replay operator T_r maps active contestants, sampled fan shares and observed judge scores into an eliminated contestant:

$$E_{s,t}^{m,r} = T\left(A_{s,t}^{m,r}, v_{s,t}^{m}, J_{s,t}\right), \quad A_{s,t+1}^{m,r} = A_{s,t}^{m,r} \Big/ E_{s,t}^{m,r} \tag{7}$$

The rule-specific elimination sequence is converted into final placements through a deterministic placement map. Rule effects are then measured by differences in placement and elimination week:

$$\begin{aligned} \Delta\mathrm{Place}_{i,s}\left(r_1, r_2\right) &= \mathrm{Place}_{i,s}\left(r_1\right) - \mathrm{Place}_{i,s}\left(r_2\right) \\ \Delta\mathrm{Week}_{i,s}\left(r_1, r_2\right) &= \mathrm{Week}_{i,s}\left(r_1\right) - \mathrm{Week}_{i,s}\left(r_2\right) \end{aligned} \tag{8}$$

Fan orientation is measured by the correlation between mean posterior fan contribution and final placement. Since smaller placement values indicate better outcomes, a more negative value indicates stronger alignment with fan support:

$$\mathrm{FBI}_r = \mathrm{Corr}_i\left(\mathrm{mean}_t \mathbb{E}\left[v_{i,s,t} \mid E, J\right], \mathrm{Place}_{i,s}(r)\right) \tag{9}$$

Three rules are evaluated. The rank rule aggregates ordinal judge and fan ranks; the percent rule combines normalized judge and fan percentages; and the bottom-two judges-save mechanism first identifies a bottom-two set and then allows judges to determine which contestant is saved. The same posterior fan samples are used throughout, so differences are attributable to rule mechanics rather than simulation noise.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

4. Results and Discussion

4.1 Latent Fan-Vote Estimation

The Bayesian model recovered interpretable posterior fan-share estimates. In the selected results, Marie Osmond had posterior mean fan shares of 0.3648 and 0.3573 in the first two weeks of Season 5, whereas Donald Driver had posterior means of 0.2281 and 0.2323 in the first two weeks of Season 14. At the season level, the corresponding total fan-contribution estimates were 6.0667 and 3.9303. These results indicate that the model produces contestant-level popularity signals rather than only elimination probabilities.

Predictive diagnostics suggest that the inferred latent signal is informative. The average negative log-likelihood was 1.712. Overall Hit@1, Hit@2 and Hit@3 were approximately 0.400, 0.574 and 0.691, respectively. Thus, the observed eliminated contestant was included among the three highest-risk contestants in about 70% of weekly decisions. Reliability analysis also showed low calibration error, with ECE approximately 0.0047, consistent with accepted principles for evaluating probabilistic forecasts. The core posterior estimation and diagnostic results are summarized in Table 2. The reliability of posterior elimination-risk probabilities is illustrated in Figure 1, the Hit@k performance is shown in Figure 2, and the predicted-risk rank of the observed eliminated contestant is presented in Figure 3.

Table 2. Core posterior estimation and diagnostic results.

Category	Reported value	Interpretation
Weekly fan share	Marie Osmond: 0.3648 and 0.3573	High early latent fan support
Weekly fan share	Donald Driver: 0.2281 and 0.2323	Stable moderate latent support
Total contribution	6.0667; 3.9303	Aggregated contestant-level fan signal
Predictive fit	NLL = 1.712; Hit@3 = 0.691	Observed eliminations often ranked high-risk
Calibration	ECE = 0.0047	Predicted probabilities well calibrated

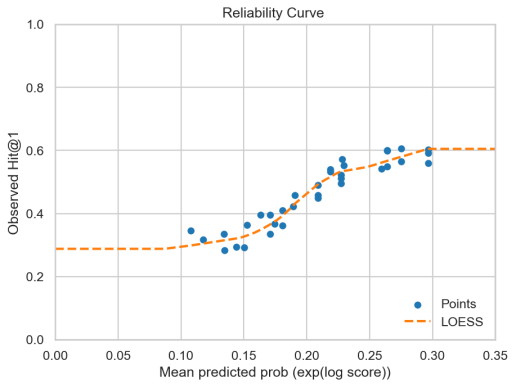

Figure 1. Reliability diagram for posterior elimination-risk probabilities.

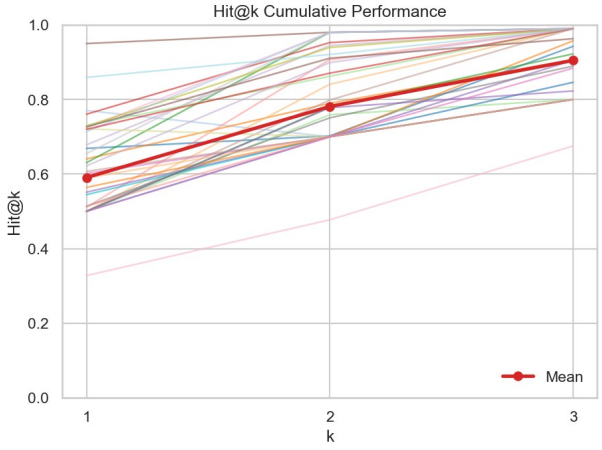

Figure 2. Hit@k performance of the latent fan-vote model.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

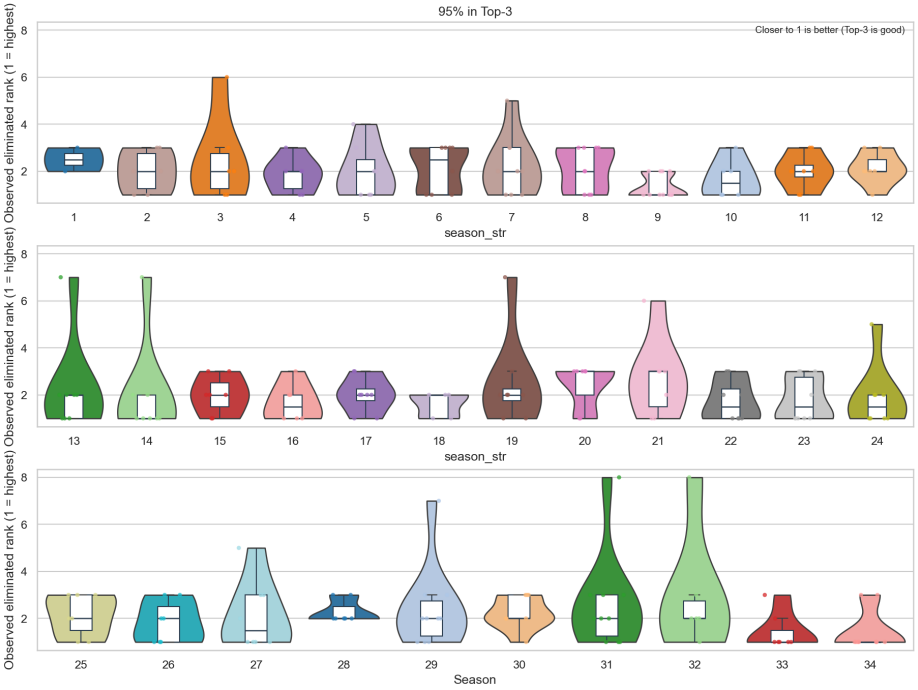

Figure 3. Predicted-risk rank of the observed eliminated contestant.

Uncertainty was stage dependent. Credible intervals widened in later weeks, when fewer contestants remained and the competition became more closely contested. The original analysis reported a strong positive association between week index and mean credible-interval width, with r approximately 0.952. This pattern is consistent with the interpretation that late-stage fan-vote inference is more difficult despite stronger cumulative information. The week-by-week credible-interval pattern is shown in Figure 4, while the main factors associated with fan-share uncertainty are summarized in Figure 5.

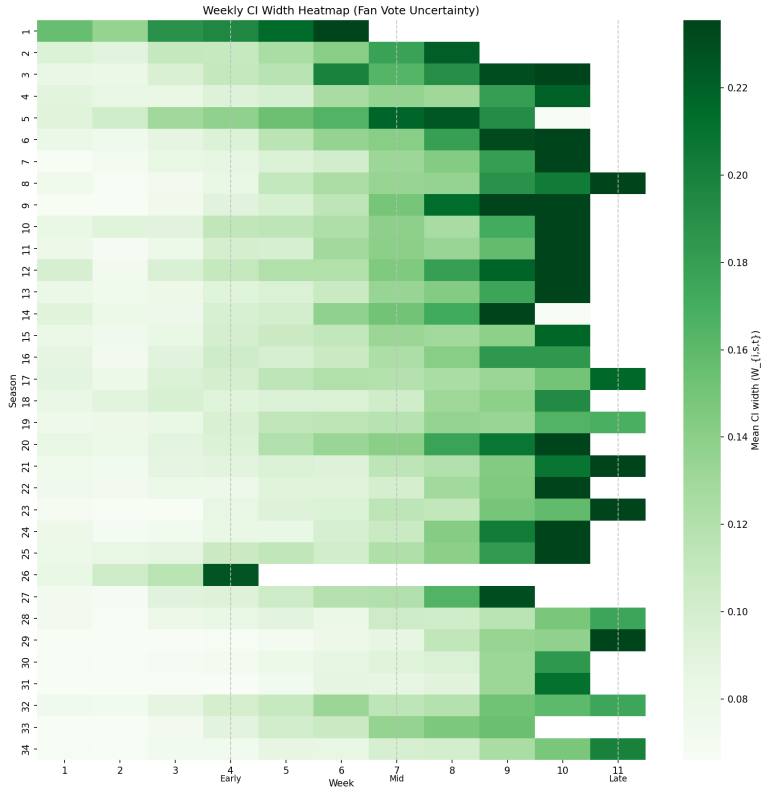

Figure 4. Weekly credible-interval width of latent fan-share estimates.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

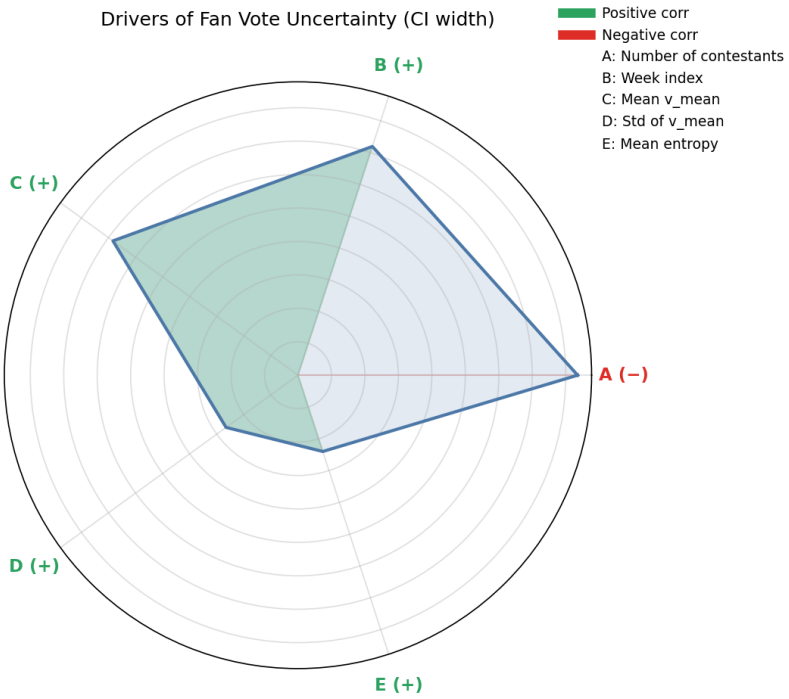

Figure 5. Factors associated with fan-share uncertainty.

4.2 Rank Rule versus Percent Rule

Posterior replay showed that rank and percent aggregation generated different competition paths. At the aggregate level, rank did not produce a one-sided placement advantage over percent: the mean placement difference was approximately -0.19 and the median was approximately 0.08. The effect on elimination timing was stronger. Rank tended to delay elimination, with mean DeltaWeek approximately 1.30, median approximately 1.05 and about 69% positive cases.

The fan-orientation comparison was not reducible to a single rule preference. Season-wise evidence suggested that percent often preserved fan support at least as strongly as rank. In the pooled evidence, however, the correlation between mean fan contribution and final placement was more negative for rank. This distinction indicates that rule bias is both season-specific and distributional; reporting only one aggregate measure would obscure important heterogeneity. Representative contestant-level differences between the two rules are reported in Table 3, and the season-wise fan-bias index difference is shown in Figure 6.

Table 3. Examples of counterfactual differences between rank and percent rules.

Season	Contestant	Delta placement	Delta elimination week
1	Evander Holyfield	-1.0000	0.0000
18	NeNe Leakes	+3.4200	-3.5000
34	Whitney Leavitt	-2.4286	3.1429

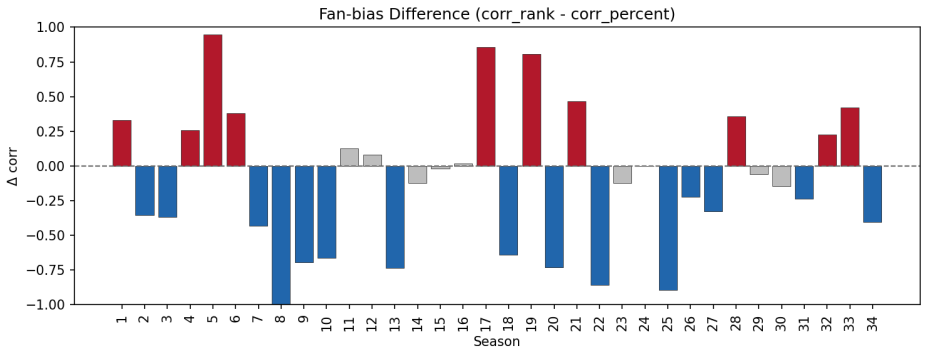

Figure 6. Season-wise fan-bias index difference between rank and percent rules.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

4.3 Controversial Contestants

Controversial contestants were markedly more sensitive to rule choice than the average contestant. Only 1 of 18 cases had an absolute placement change no larger than 0.5 between rank and percent. Elimination timing was also unstable, with only 4 of 18 cases showing an absolute elimination-week change no larger than 0.5. These findings imply that contestants located near the disagreement boundary between judges and fans are precisely those for whom rule design matters most.

Directionally, rank tended to improve placement and delay elimination for controversial contestants relative to percent. The mean placement difference was approximately -0.35, with a median of approximately -0.71; the mean elimination-week difference was approximately 1.26, with a median of approximately 1.07. These magnitudes are large enough to change the narrative of a season, not merely the marginal ordering of adjacent contestants. Representative large replay differences are listed in Table 4, while the placement and elimination-week shifts for controversial contestants are visualized in Figure 7.

Table 4. Large replay differences for controversial contestants under rank relative to percent.

Season	Contestant	Delta placement	Delta elimination week
18	Candace Cameron Bure	-3.10	3.02
9	Aaron Carter	-2.95	4.86
3	Mario Lopez	-2.62	2.58

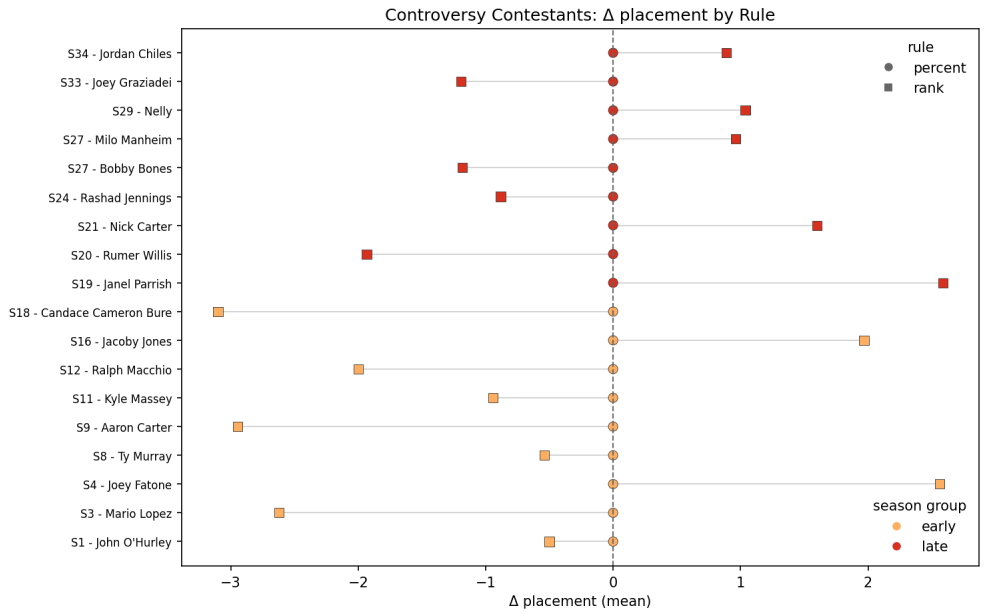

Figure 7. Placement and elimination-week shifts for controversial contestants.

4.4 Bottom-Two Judges-Save Mechanism

The bottom-two judges-save mechanism introduced an additional adjudication stage at the most consequential point of each week. Replay results showed a systematic aggregate shift toward slightly worse placements and earlier eliminations after the mechanism was introduced. Individual effects were heterogeneous: approximately 39.1% of contestants benefited, 44.7% were harmed and 16.2% were unchanged. The mechanism therefore generated a winner/loser redistribution rather than a neutral robustness correction. The macro-level outcome shift is illustrated in Figure 8, and the largest individual placement shifts are shown in Figure 9.

The strongest explanatory factor was judge-fan disagreement. The absolute correlations between the judge-fan gap and outcome changes were about 0.35-0.37. Judge score mean was also associated with placement change, with |corr| around 0.30, while controversy level had moderate correlations around 0.25. In contrast, fan-only variables, including mean fan share, fan-rank and uncertainty width, had weak correlations of about 0.00-0.04. This pattern indicates that the judges-save mechanism operates primarily by amplifying judge-side information at close elimination margins. The relative correlation strengths of these explanatory factors are compared in Figure 10.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

Overall fan-orientation indicators reinforced this interpretation. The correlations between mean fan contribution and final placement were approximately -0.605 for rank, -0.288 for percent and -0.231 for bottom-two judges-save. Because smaller placement is better, the weakest negative association belongs to judges-save. If preserving fan intent is a design objective, this mechanism should therefore not be adopted as the default rule. The overall fan-orientation indicators are summarized in Table 5, and the corresponding overall rule-bias comparison is presented in Figure 11.

Table 5. Overall fan-orientation indicators by rule.

Rule	Correlation with placement	Interpretation
Rank	approximately -0.605	Strongest pooled fan-orientation signal
Percent	approximately -0.288	Moderate fan-orientation signal
Bottom-two judges-save	approximately -0.231	Weakest fan-orientation signal

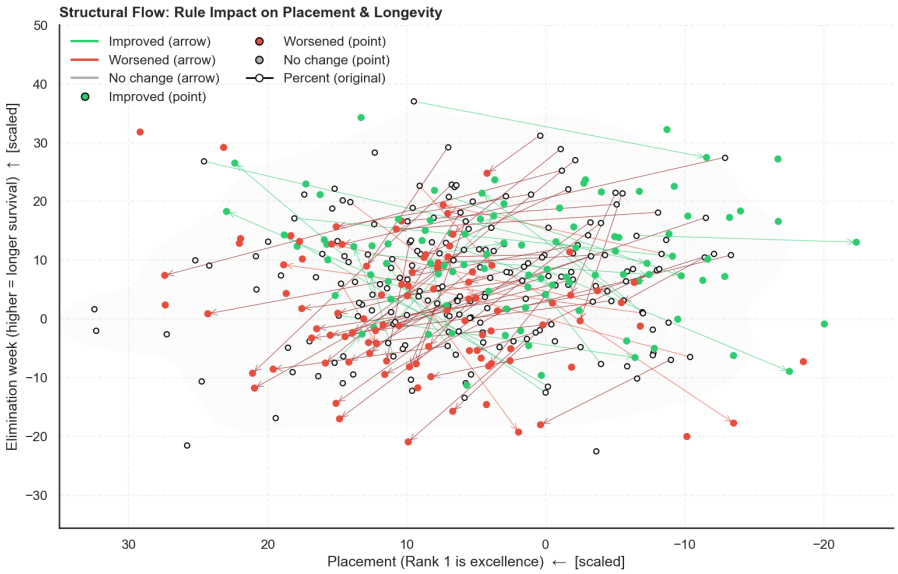

Figure 8. Macro-level outcome shift under the bottom-two judges-save mechanism.

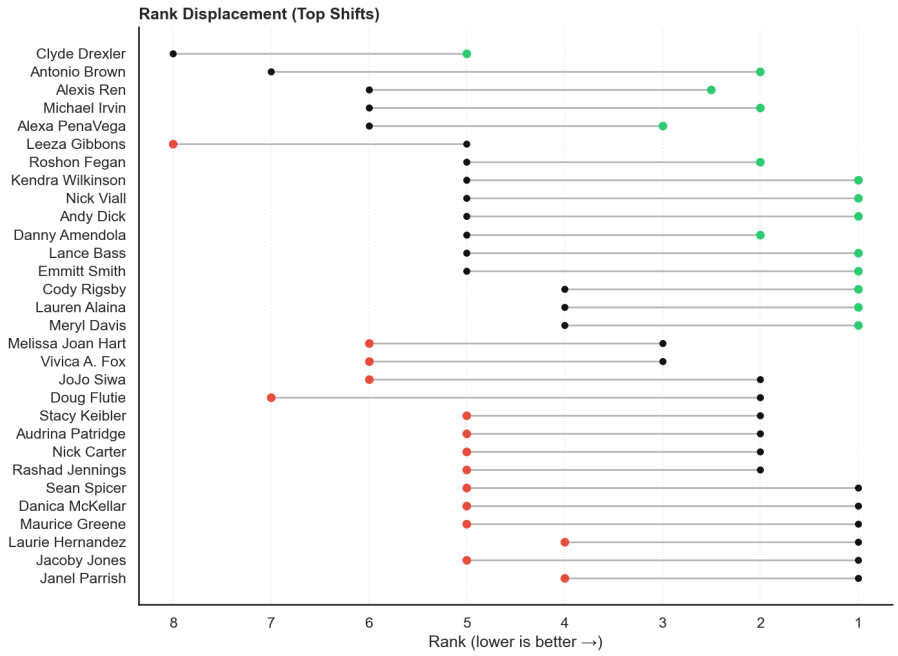

Figure 9. Largest individual placement shifts under judges-save replay.

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

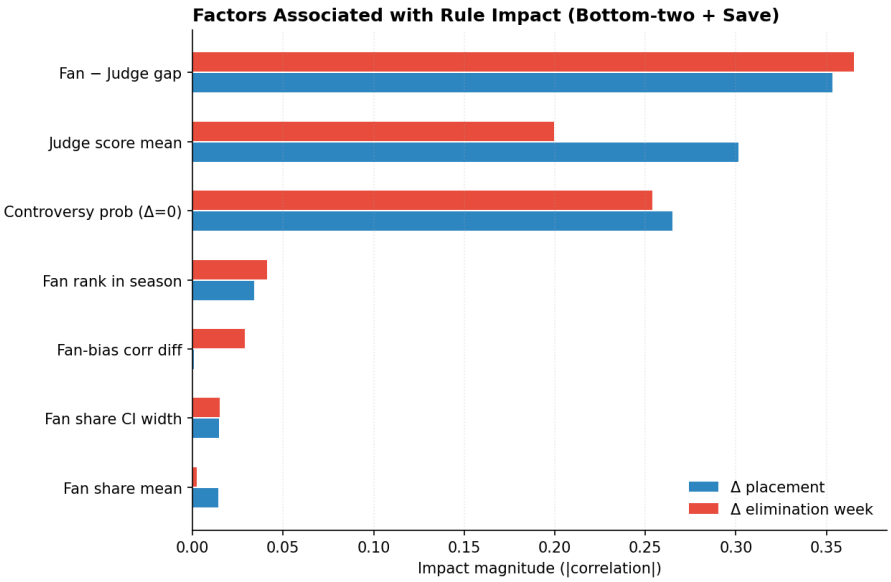

Figure 10. Correlation strengths explaining judges-save outcome changes.

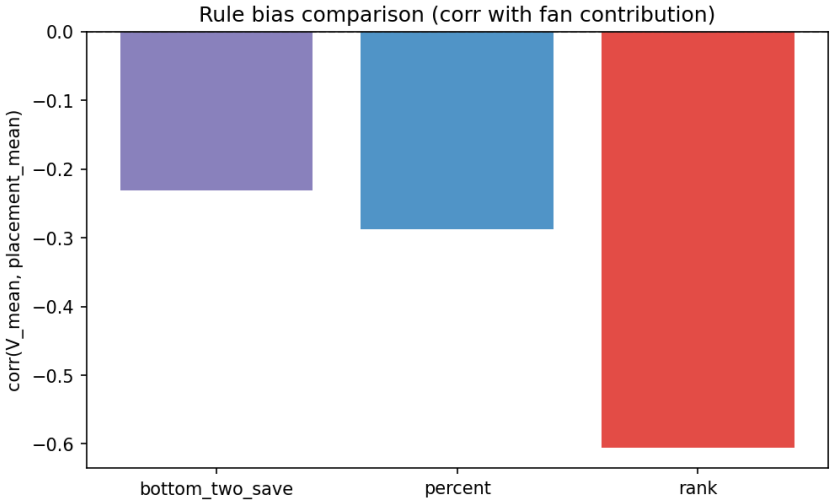

Figure 11. Overall rule-bias comparison across replayed mechanisms.

5. Conclusion

This study developed a Bayesian replay framework for analyzing latent fan support and voting-rule effects in elimination competitions. Instead of treating fan votes as directly observable quantities, the proposed approach inferred weekly latent fan-share distributions from observed judge scores and elimination outcomes. By integrating within-week score standardization, softmax fan-share mapping, temporal state smoothing and categorical elimination likelihoods, the model produced interpretable posterior fan-support signals together with uncertainty quantification and calibrated elimination probabilities.

The posterior predictive results demonstrated that the inferred latent fan signal contained substantial explanatory power for observed eliminations. The model achieved stable probabilistic performance across multiple diagnostic criteria, including negative log-likelihood, Hit@k accuracy and reliability calibration. In particular, the relatively low calibration error and consistent high-risk identification suggest that the Bayesian latent-utility formulation successfully captured the hidden audience-preference structure underlying elimination dynamics. The uncertainty analysis further revealed that fan-vote inference becomes increasingly difficult in later competition stages, when the contestant pool contracts and competitive margins narrow.

Using the inferred posterior fan shares as common statistical inputs, the replay framework enabled a

ISSN 2616-5775 Vol. 9, Issue 7: 100-109, DOI: 10.25236/AJCIS.2026.090713

controlled comparison of alternative aggregation rules. The results showed that voting-rule choice materially changes elimination trajectories, contestant placements and controversy sensitivity. Rank and percent aggregation produced noticeably different elimination timing patterns, while controversial contestants exhibited especially high sensitivity to rule transitions. These findings indicate that voting rules do not merely aggregate preferences passively; rather, they actively reshape the sequential evolution of the competition itself.

The analysis of the bottom-two judges-save mechanism further demonstrated that additional adjudication stages amplify judge influence at critical elimination boundaries. Although the mechanism may improve producer-side control and narrative stability, it weakens the transmission of audience preference signals and generates heterogeneous winner/loser redistribution effects across contestants. From the perspective of preserving fan-oriented outcomes, the replay evidence suggests that judges-save mechanisms should not be adopted as default competition rules unless stronger expert intervention is an explicit design objective.

More broadly, the study illustrates how Bayesian latent-variable inference and posterior replay analysis can be combined to evaluate collective-decision systems with partially unobservable preference structures. Although the present work focuses on entertainment competitions, the framework is extensible to broader sequential ranking and elimination environments involving hidden behavioral signals, repeated aggregation and rule-dependent outcomes. Future research may incorporate richer behavioral covariates, dynamic viewer-interaction data and causal preference-shift mechanisms to further improve the interpretability and external validity of latent fan-support inference.

References

[1] K. J. Arrow, "A difficulty in the concept of social welfare," Journal of Political Economy, vol. 58, no. 4, pp. 328-346, 1950.
[2] A. Gibbard, "Manipulation of voting schemes: A general result," Econometrica, vol. 41, no. 4, pp. 587-601, 1973.
[3] M. A. Satterthwaite, "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions," Journal of Economic Theory, vol. 10, no. 2, pp. 187-217, 1975.
[4] N. Ailon, M. Charikar, and A. Newman, "Aggregating inconsistent information: Ranking and clustering," Journal of the ACM, vol. 55, no. 5, Article 23, 2008.
[5] C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, and O. Sheffet, "Optimal social choice functions: A utilitarian view," Artificial Intelligence, vol. 227, pp. 190-213, 2015.
[6] J. H. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," Journal of the American Statistical Association, vol. 88, no. 422, pp. 669-679, 1993.
[7] S. Chib and E. Greenberg, "Analysis of multivariate probit models," Biometrika, vol. 85, no. 2, pp. 347-361, 1998.
[8] B. P. Carlin, N. G. Polson, and D. S. Stoffer, "A Monte Carlo approach to nonnormal and nonlinear state-space modeling," Journal of the American Statistical Association, vol. 87, no. 418, pp. 493-500, 1992.
[9] S. J. Koopman and J. Durbin, "A simple and efficient smoother for state space time series analysis," Biometrika, vol. 89, no. 3, pp. 603-616, 2002.
[10] A. Gelman, X. L. Meng, and H. Stern, "Posterior predictive assessment of model fitness via realized discrepancies," Statistica Sinica, vol. 6, no. 4, pp. 733-807, 1996.
[11] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," Journal of the Royal Statistical Society: Series B, vol. 64, no. 4, pp. 583-639, 2002.
[12] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," Statistics and Computing, vol. 27, no. 5, pp. 1413-1432, 2017.
[13] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," Journal of the American Statistical Association, vol. 102, no. 477, pp. 359-378, 2007.
[14] M. H. DeGroot and S. E. Fienberg, "The comparison and evaluation of forecasters," The Statistician, vol. 32, no. 1/2, pp. 12-22, 1983.
[15] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," Journal of the Royal Statistical Society: Series B, vol. 69, no. 2, pp. 243-268, 2007.