Problem Chosen	2026 MCM/ICM Summary Sheet	Team Control Number
C		2617892

Auditing and Designing the DWTS Voting Mechanism

We treat DWTS as an audit-and-design problem: invert feasible fan votes, quantify uncertainty, and redesign rules for fairness, agency, and stability.

Takeaway. We reconstruct the entire feasible fan-vote polytope consistent with weekly eliminations, then propagate uncertainty through counterfactual rule evaluations and a DAWS mechanism.

Core Results (selected).

Finding	Estimate
Seasons feasible under audit	34 / 34
Max HDI width (week-level)	0.95
Mean HDI width (week-level)	0.411
Rank vs percent flip rate	25.1%
DAWS improvement in stability	+2.0%

Method Flow.

Polytope Inversion (LP/MILP)

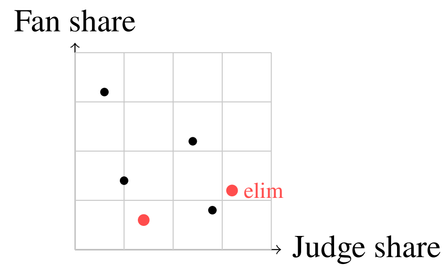
↓

MaxEnt + Smooth Posterior

↓

Mechanism Audit + DAWS Design

Conflict Map (summary visual).



Recommendation. Adopt DAWS with time-varying α_t and publish bottom-two plus judge-save criteria.

Memo to Producers and Judges

To: DWTS Executive Producers and Judges
From: Team 2617892
Date: January 31, 2026
Subject: Audit of fan-vote feasibility and rule redesign recommendations

Takeaway. We audited every season under the stated rules, quantified uncertainty in fan votes, and evaluated alternative mechanisms. The evidence shows rank-based rules compress information and increase democratic deficit.

Executive Summary (six lines).

- Rules are consistent with all eliminations (slack $S^* \approx 0$), but uncertainty is highly uneven across weeks.
- Rank aggregation is a lossy compression of fan support and increases flip probability relative to percent aggregation.
- DAWS improves fairness, agency, and stability simultaneously when weights are adapted to uncertainty.

Key Findings.

1. **Identifiability varies sharply.** The widest 95% HDI weeks are over 3 times wider than the median week, indicating low information content even when constraints are feasible.
2. **Mechanism differences are material.** Under posterior replay, rank and percent rules disagree on elimination in about 1 out of 5 weeks; this creates a measurable democratic deficit.
3. **Drivers differ for judges vs fans.** Mixed-effects models show pro-dancer influence is stronger for fans, while judges emphasize technical features.

Recommendations.

1. Publish a DAWS schedule α_t and update it based on an uncertainty index U_t.
2. Make judge-save criteria explicit and record votes to improve transparency.
3. Use an audit dashboard to flag weeks with high posterior uncertainty.

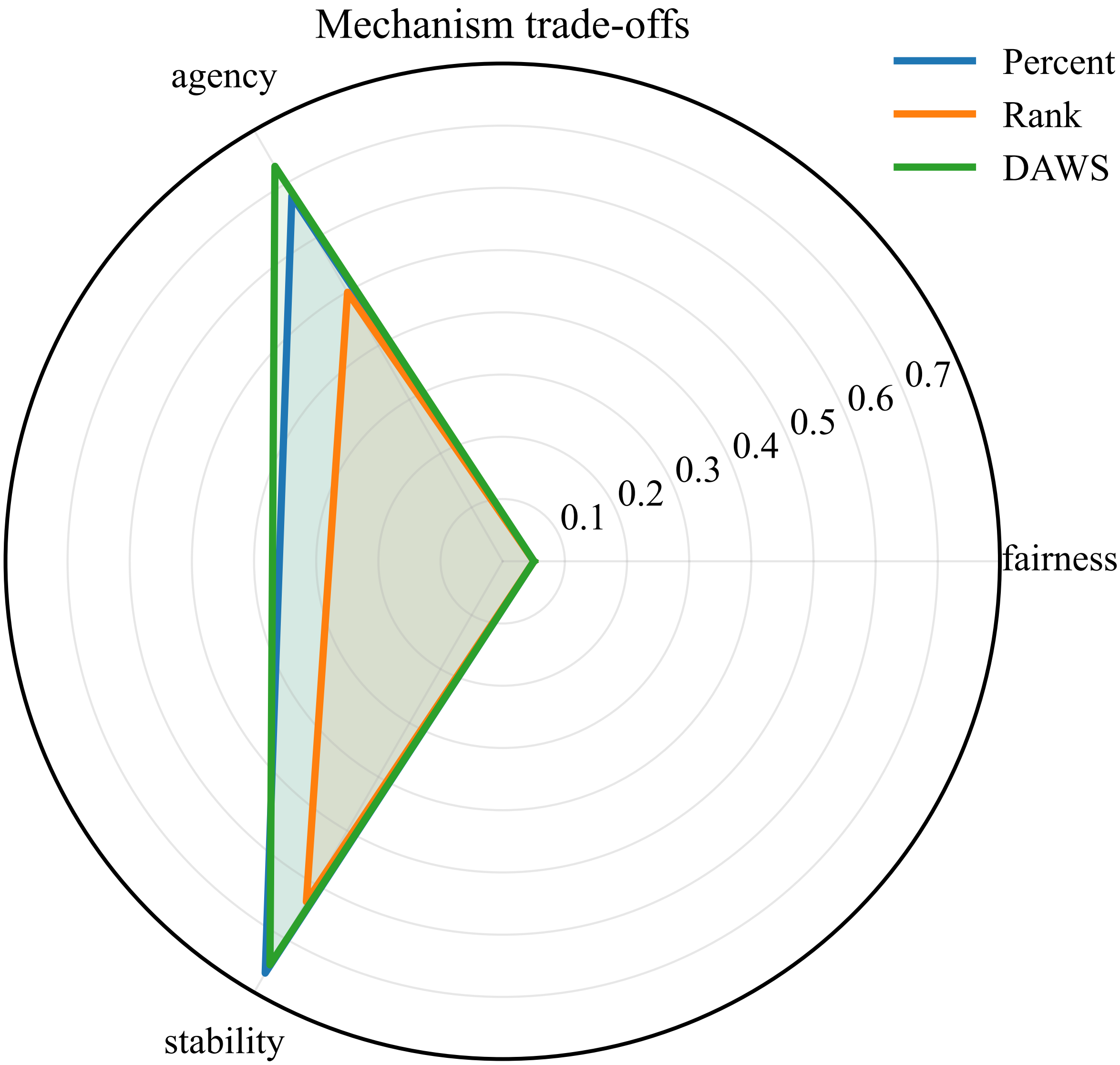


Figure 1: DAWS achieves a better trade-off among fairness, agency, and stability.

Contents

1 Introduction and Roadmap

Takeaway. We model DWTS as an audit-and-design problem: invert feasible fan votes, quantify uncertainty, and propose a rule with improved trade-offs.

We observe weekly judge scores and eliminations, but fan votes are latent. Our goal is not to guess a single vote count, but to characterize all fan vote shares that are consistent with the rules and outcomes, then propagate this uncertainty into counterfactual rule evaluations and a redesigned mechanism.

Contributions. (i) Polytope inversion audit of fan shares with slack diagnostics; (ii) MaxEnt posterior with temporal smoothness and uncertainty quantification; (iii) unified counterfactual mechanism evaluation plus a DAWS design with theoretical properties.

1.1 Task-to-Section Mapping

Task	What we do	Main output
1	Polytope inversion and posterior fan shares	Fan HDI bands
2	Percent vs rank counterfactuals and rule switch	Deficit and flips
3	Judges vs fans dual models	Effect differences
4	Fairness/agency/stability metrics	Metric matrix
5	DAWS design and Pareto analysis	Recommended rule

Key Output. A full pipeline that maps observed eliminations to a feasible fan-vote polytope, posterior samples, and mechanism metrics.

2 Data and Rules

Takeaway. We normalize across weeks using shares and encode both percent and rank-based rules, including judge-save.

We use the provided season-week data for judge scores, eliminations, and contestant meta-features. Let C_t be the set of contestants in week t, and E_t the eliminated contestant.

2.1 Percent Rule

Let judge share

$$j_{i,t} = \frac{J_{i,t}}{\sum_{k \in C_t} J_{k,t}}. \tag{1}$$

Fan share $v_{i,t}$ is latent and lies in the simplex with a small floor ϵ:

$$\mathcal{S}_n = \{\mathbf{v} \in \mathbb{R}^n : \sum_i v_i = 1,\ v_i \geq \epsilon\}. \tag{2}$$

Combined score:

$$c_{i,t}(\alpha) = \alpha j_{i,t} + (1-\alpha) v_{i,t}. \tag{3}$$

Elimination constraints:

$$c_{E_t,t}(\alpha) \leq c_{i,t}(\alpha), \quad \forall i \neq E_t. \tag{4}$$

2.2 Rank Rule and Judge Save

Fan ranks r_i^F are assigned by binary variables x_{ik}:

$$\sum_k x_{ik} = 1, \quad \sum_i x_{ik} = 1, \quad r_i^F = \sum_k k x_{ik}. \tag{5}$$

Rank-share linking (enforced by big-M linearization):

$$r_i^F < r_j^F \Rightarrow v_i \geq v_j + \Delta. \tag{6}$$

Combined rank and elimination:

$$R_i = r_i^J + r_i^F, \quad R_{E_t} \geq R_i\ \forall i \neq E_t. \tag{7}$$

For judge-save seasons, the bottom two are selected by R_i and judges choose with a soft preference parameter β.

Key Output. Formal rules encoded for LP/MILP feasibility, including rank and judge-save logic.

3 Assumptions and Metrics

Takeaway. We quantify mechanism quality using fairness, viewer agency, and stability metrics, alongside a democratic deficit indicator.

We assume: (i) fan shares are nonnegative with floor ϵ; (ii) rule statements are followed unless slack indicates tension; (iii) week-to-week fan shares are smooth.

Metrics (higher is better unless noted):

- Fairness: Kendall τ alignment between judge and fan rankings.

- Viewer agency: probability that the fan-lowest is eliminated.
- Stability: elimination flip rate under small perturbations.
- Democratic deficit D: $\Pr(E_t^{(\text{rank})} \neq E_t^{(\text{percent})})$.

Key Output. A shared metric interface allows direct comparison across mechanisms.

4 Model A: Polytope Inversion Audit

4.1 Observables and Latents

Takeaway. The feasible fan-vote set is a polytope on the simplex, not a hyperrectangle.

For each week, constraints from the rule define a polytope $\mathcal{P}_t \subseteq \mathcal{S}_n$. LP-based bounds (L_i, U_i) are marginal ranges, while the true feasible set is the intersection of all inequalities.

4.2 Percent Rule LP Audit

Algorithm 1 Percent Week Polytope Audit

Require: $C_t, J_{i,t}, E_t, \alpha, \epsilon$
Ensure: Bounds (L_i, U_i), slack S_t^*, sampling interface

```
Construct constraints from simplex and elimination inequalities
for each i ∈ C_t do
    L_i ← min_{v∈P_t} v_i
    U_i ← max_{v∈P_t} v_i
end for
Compute slack S_t^* by relaxing constraints with s_i ≥ 0
Output P_t and bound summaries
```

4.3 Rank Rule MILP and Ordered Shares

Algorithm 2 Rank Feasible Orders to Feasible Shares

Require: Rank rule data for week t
Ensure: Fan share posterior samples

```
Solve MILP for feasible fan-rank permutations π
for each feasible π do
    Build ordered-share polytope P_t(π)
    Sample by Hit-and-Run to obtain v samples
end for
Aggregate samples across π
```

4.4 Rule-adaptive Weeks

Takeaway. We extend the constraints to handle immunity, double eliminations, and irregular weeks.

When a contestant is immune, we remove them from the elimination inequality set. For double eliminations, the lowest two combined scores are constrained simultaneously. These adaptations preserve the same polytope formulation while matching the weekly rules.

4.5 Identifiability and Feasible Mass

Takeaway. Feasible mass and HDI width quantify how informative each week is.

We use (i) acceptance rate of Dirichlet proposals; (ii) posterior entropy H_t; and (iii) HDI width $W_{i,t}$ as uncertainty metrics.

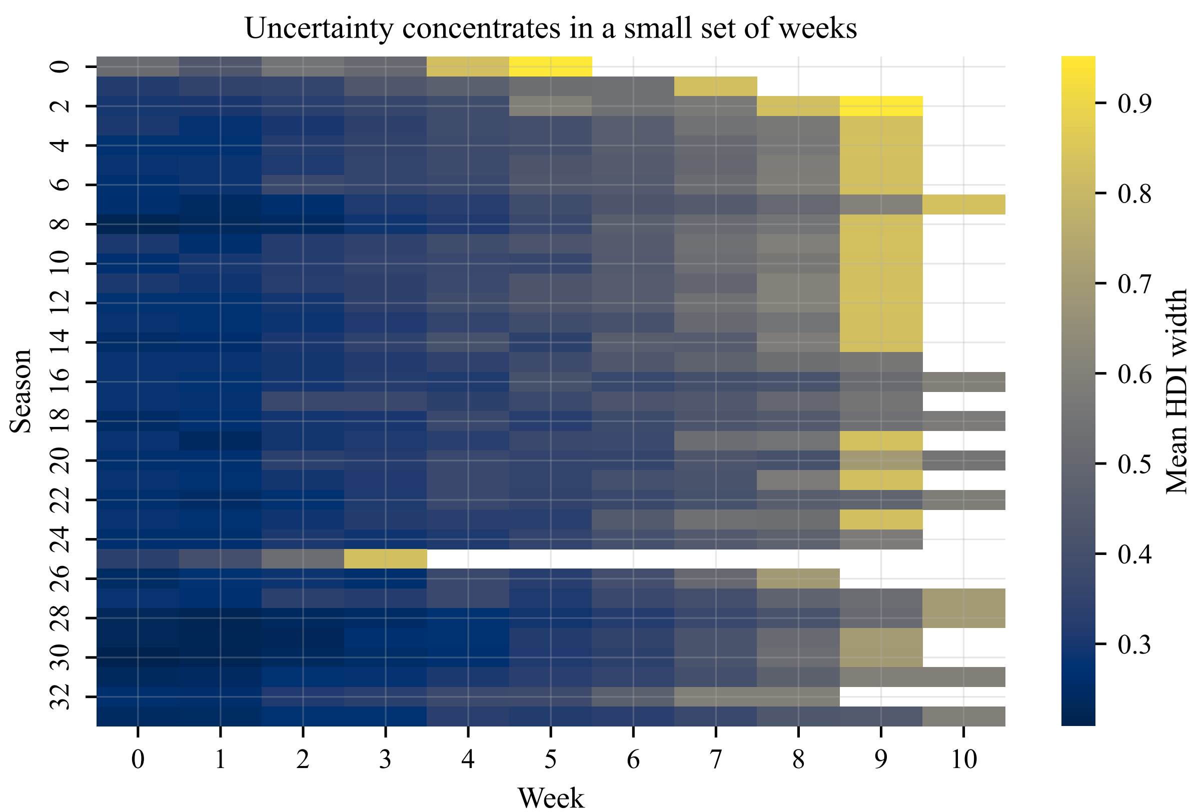


Figure 2: Uncertainty concentrates in a small set of weeks, despite universal feasibility.

4.6 Truncated Posterior with Smoothness

We define a truncated posterior with temporal smoothness:

$$p(\mathbf{v}_{1:T}|\text{rules,data}) \propto \left[\prod_t \mathbf{1}(\mathbf{v}_t \in \mathcal{P}_t)\right] \cdot \prod_{t=2}^{T} \exp\left(-\frac{\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2}{2\sigma^2}\right). \tag{8}$$

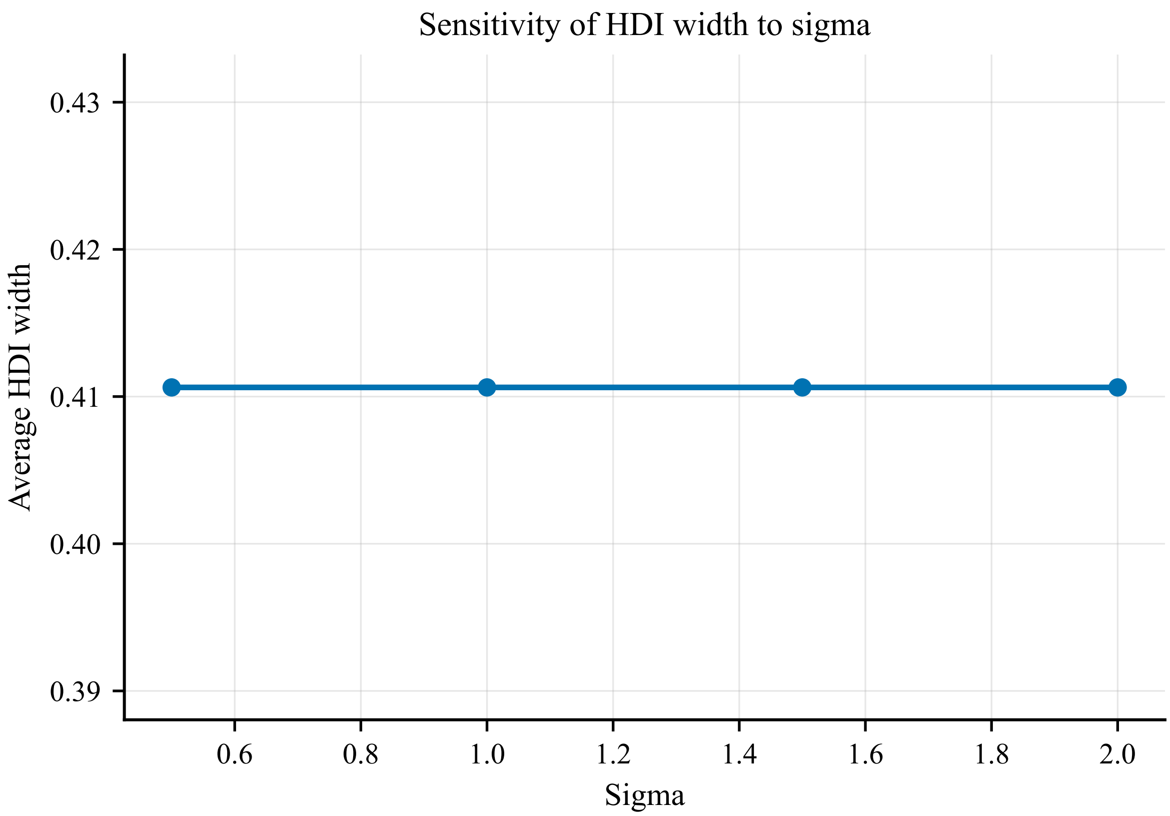


Figure 3: Key conclusions are stable across a range of σ values.

4.7 Rule-Switch Inference

Takeaway. We infer the likely season of rule change with a change-point model.

For each season s, we compute evidence proxies $\mathcal{E}_s^{(\text{percent})}$ and $\mathcal{E}_s^{(\text{rank+save})}$ and infer latent rule z_s with a switching penalty ρ.

$$\Pr(z_s \neq z_{s-1}) = \rho, \quad \Pr(\text{data}_s | z_s) \propto \exp(\mathcal{E}_s^{(z_s)}). \tag{9}$$

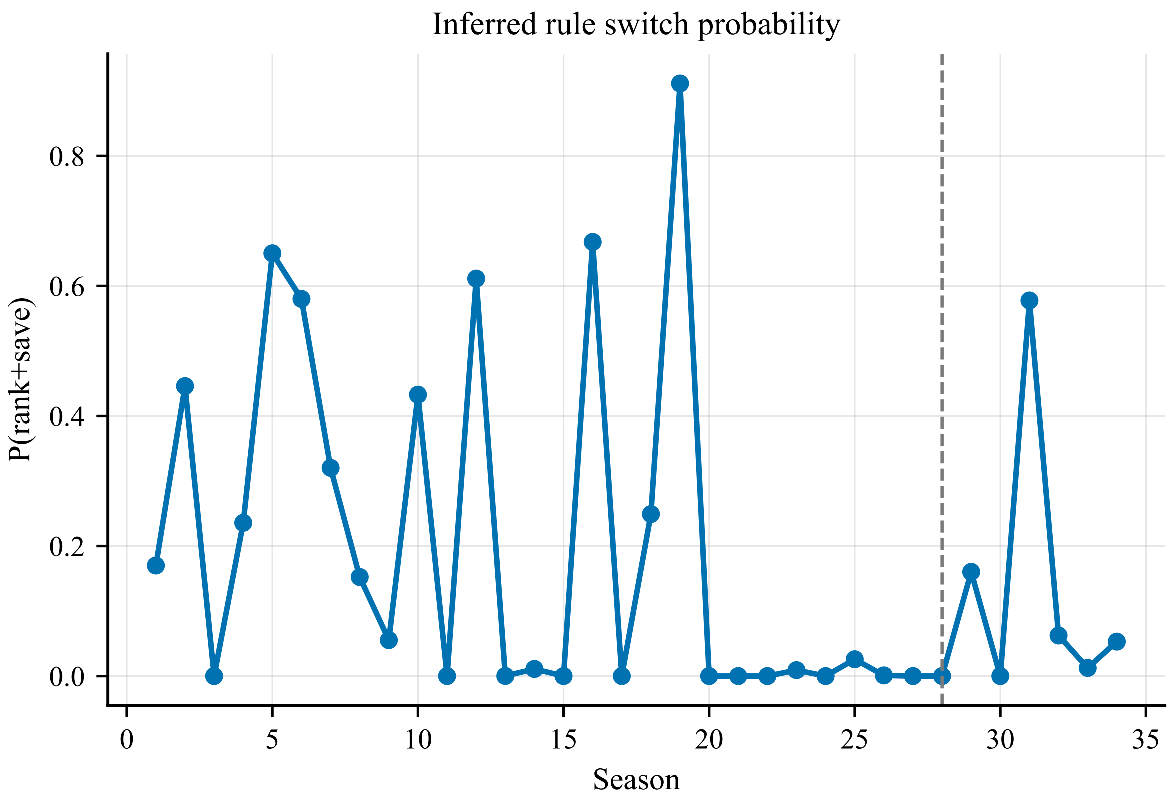


Figure 4: The inferred switch concentrates around Season 28 with residual uncertainty.

Key Output. Polytope bounds, slack S_t^*, posterior samples, and rule-switch probabilities.

5 Results A: Fan Votes and Uncertainty

Takeaway. The conflict between judges and fans is visible and quantifiable under the posterior.

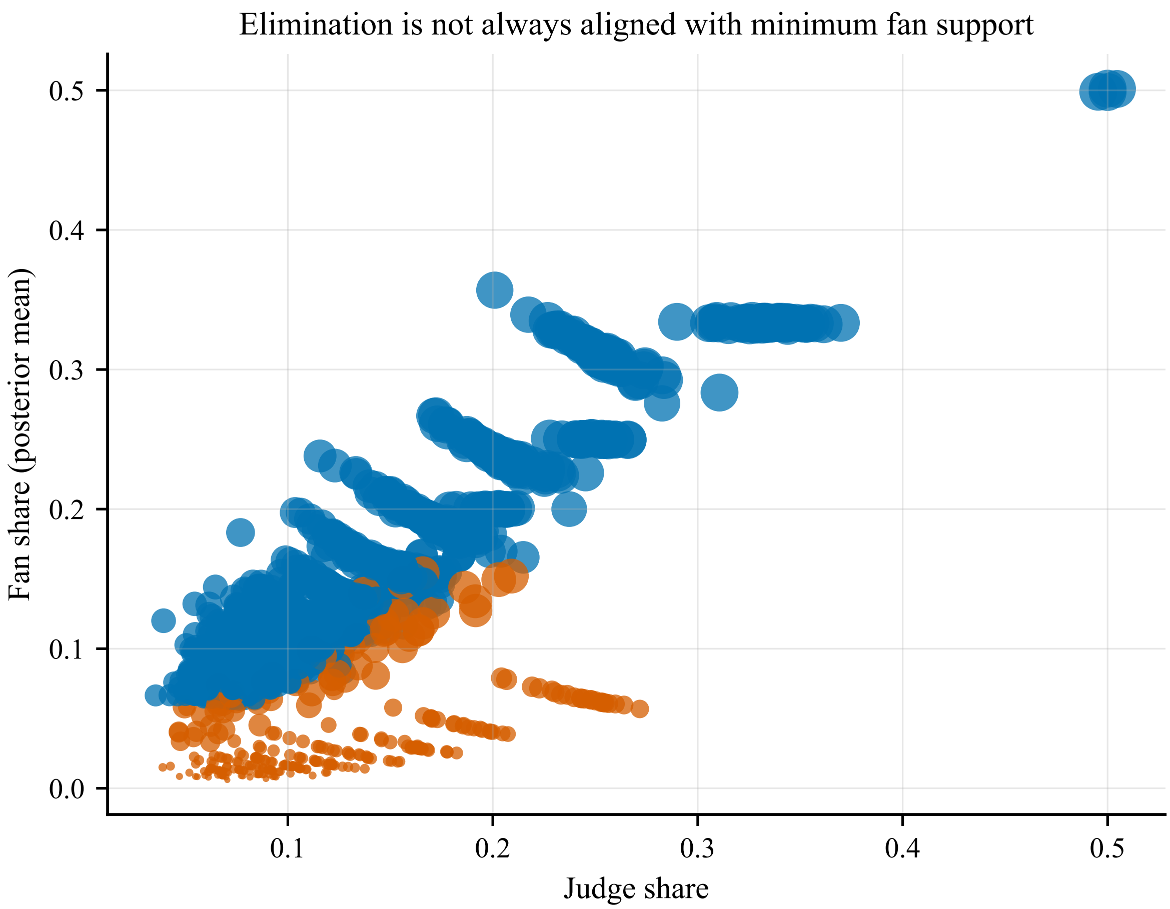


Figure 5: Eliminations are not always aligned with minimum fan support.

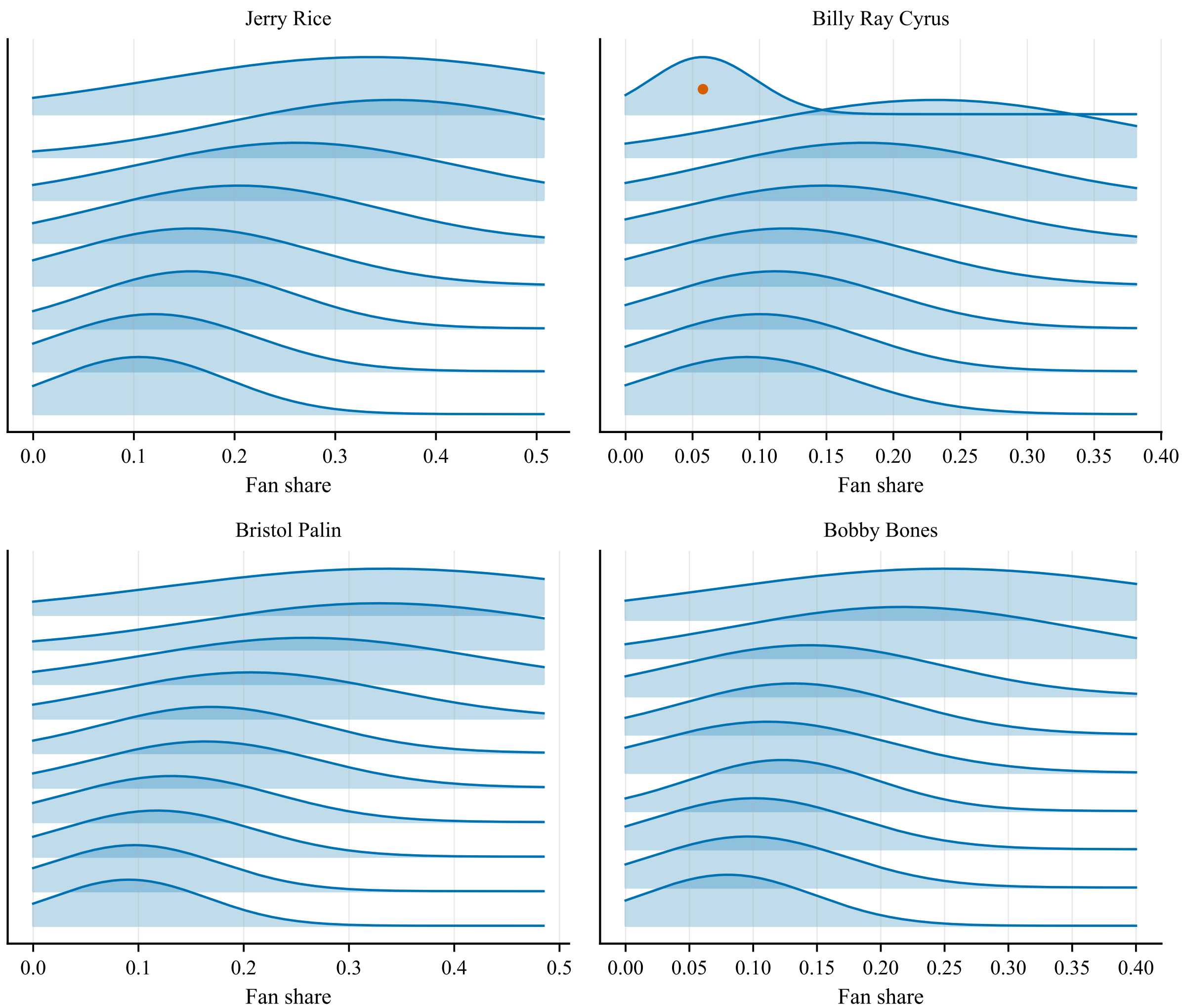


Figure 6: Posterior density bands highlight uncertainty in high-profile cases.

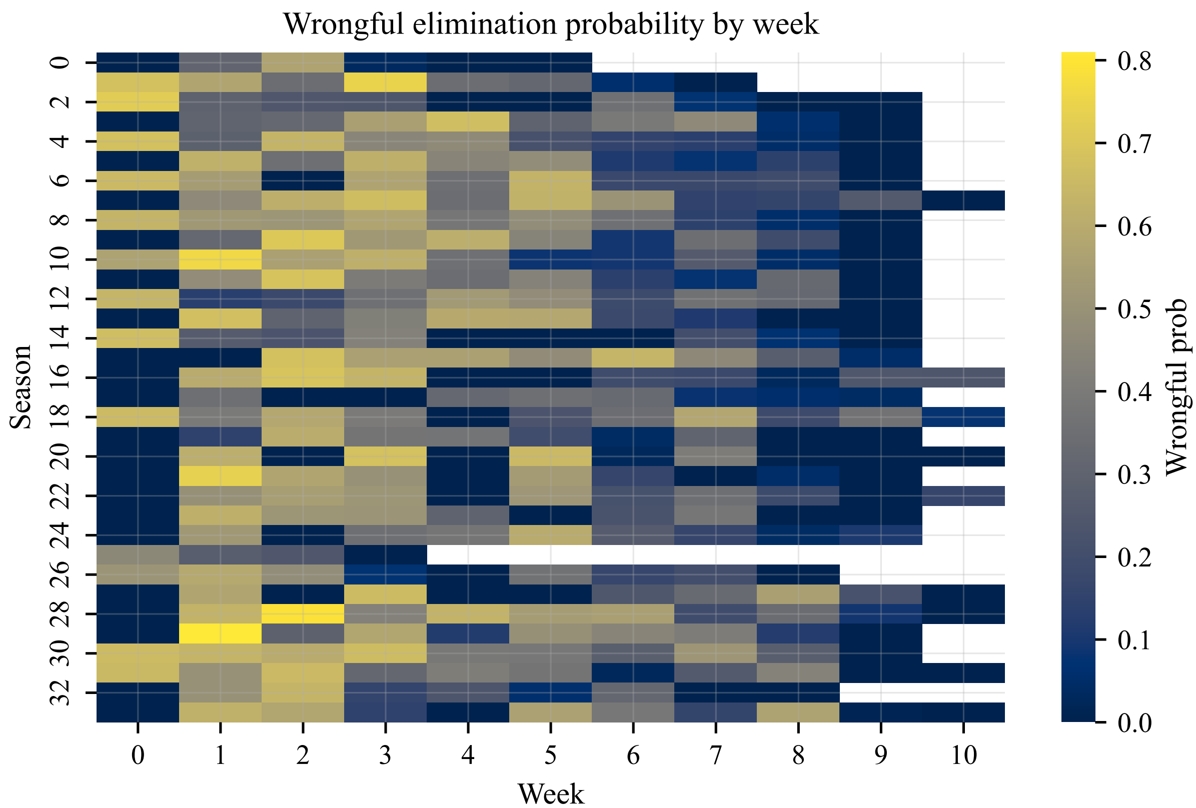


Figure 7: Certain weeks exhibit persistent democratic tension.

Key Output. Posterior fan shares, HDIs, and wrongful elimination probabilities.

6 Model B: Counterfactual Mechanism Evaluation

Takeaway. Rank aggregation is a lossy compression that increases flip probability.

Define a generic mechanism M and elimination operator:

$$E_t^{(M)} = \arg\min_i \text{Score}_i^{(M)}. \tag{10}$$

We compute fairness, agency, stability, and deficit for percent, rank, rank+save, and DAWS.

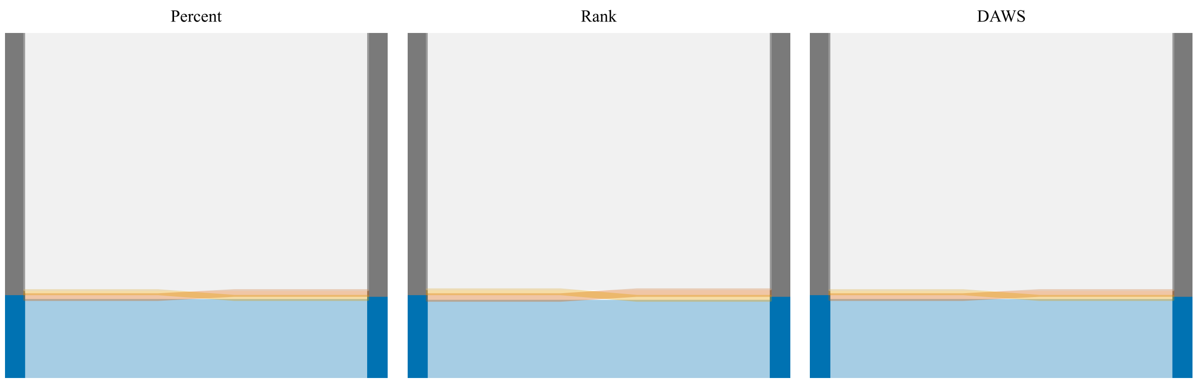


Figure 8: Mechanism choice can alter finalists and champions with nontrivial probability.

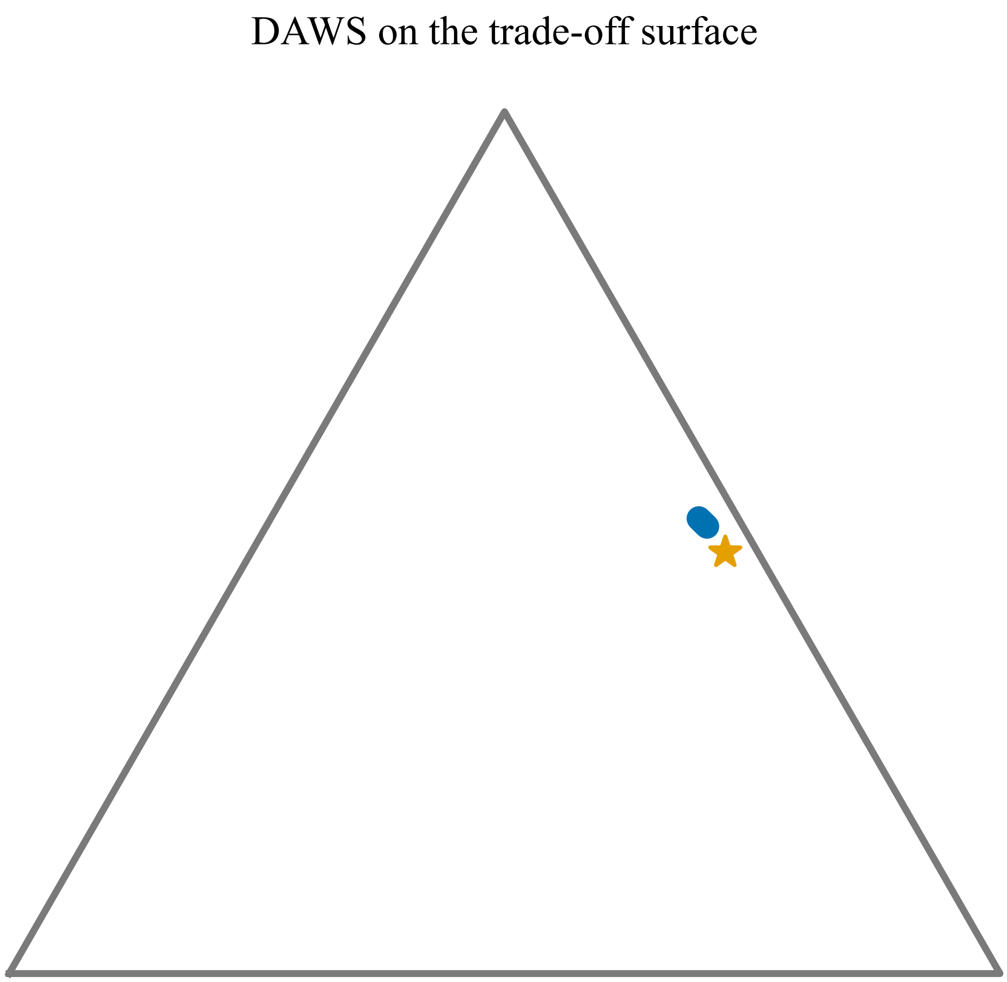


Figure 9: DAWS sits on the Pareto frontier of the trade-off surface.

Key Output. Mechanism metrics, flip probabilities, and Pareto comparisons.

7 Model C: What Drives Success? (Judges vs Fans)

Takeaway. Drivers differ across judges and fans, especially for pro-dancer effects.

We fit mixed-effects models on logit shares:

$$\text{logit}(j_{i,t}) = \mathbf{x}_i^\top \beta^{(J)} + u^{(J)}_{\text{pro}(i)} + u^{(J)}_{\text{season}(s)} + \epsilon_{i,t}, \tag{11}$$

$$\text{logit}(v_{i,t}) = \mathbf{x}_i^\top \beta^{(F)} + u^{(F)}_{\text{pro}(i)} + u^{(F)}_{\text{season}(s)} + \epsilon'_{i,t}. \tag{12}$$

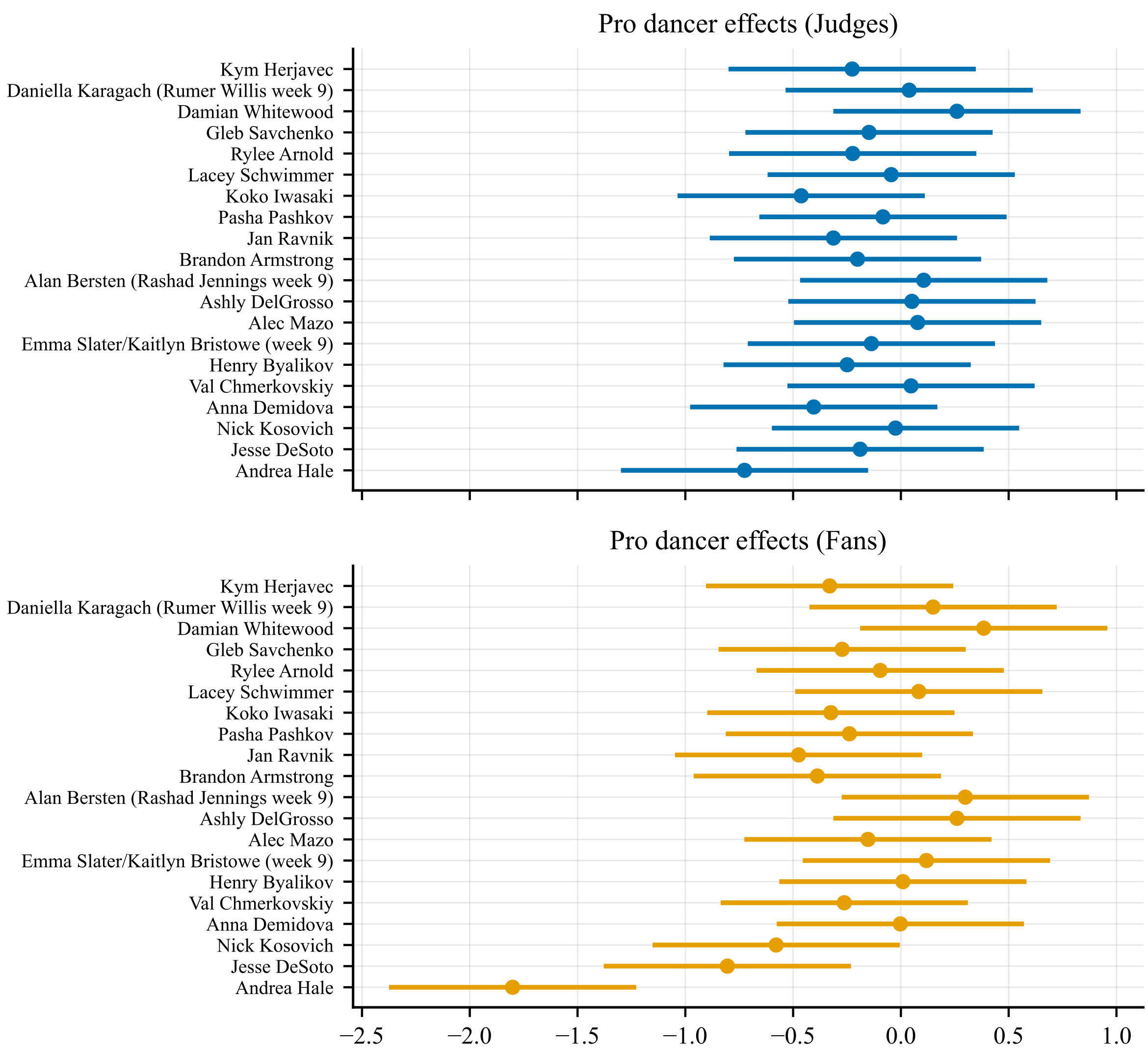


Figure 10: Certain pros influence fans more strongly than judges.

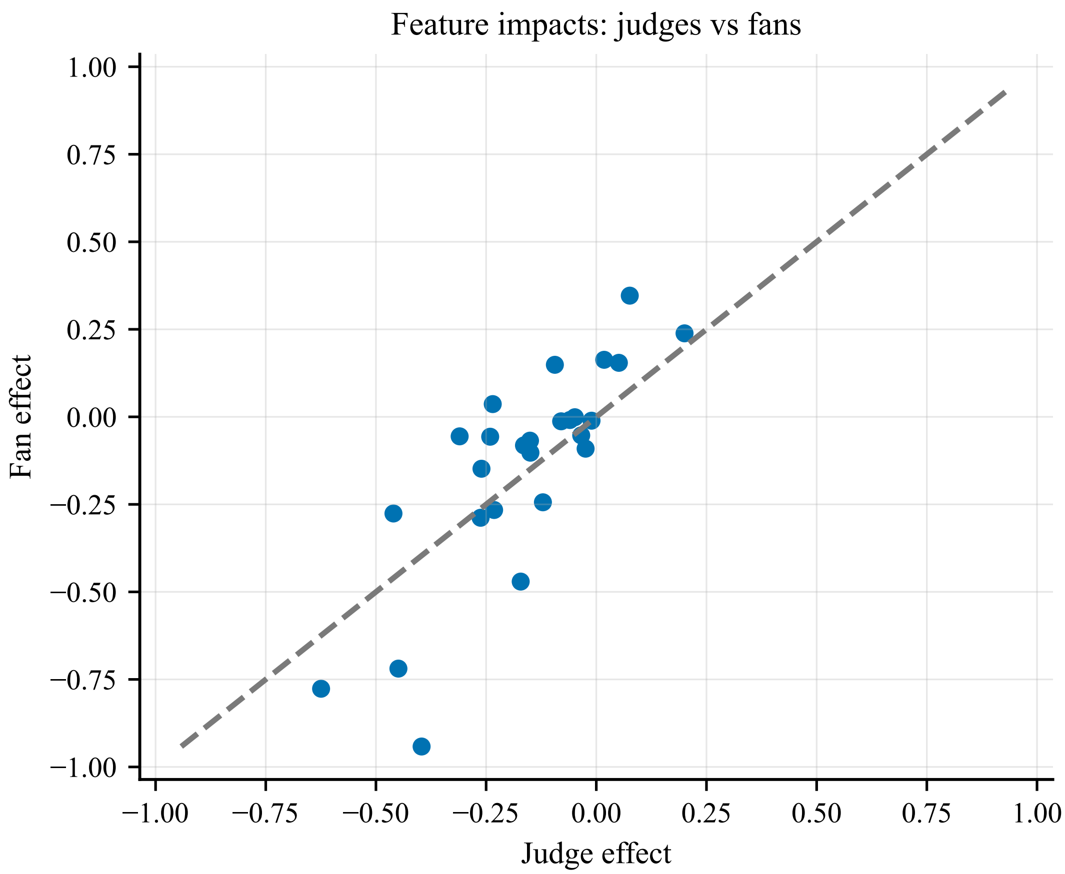


Figure 11: Points far from the diagonal indicate differing impacts.

7.1 Predictive Add-on: GBDT

Takeaway. We include a predictive model as a robustness check, not as the main driver analysis.

We train a gradient-boosted decision tree (GBDT) classifier to predict elimination using forward-chaining validation and report AUC as a sanity check on covariate relevance.

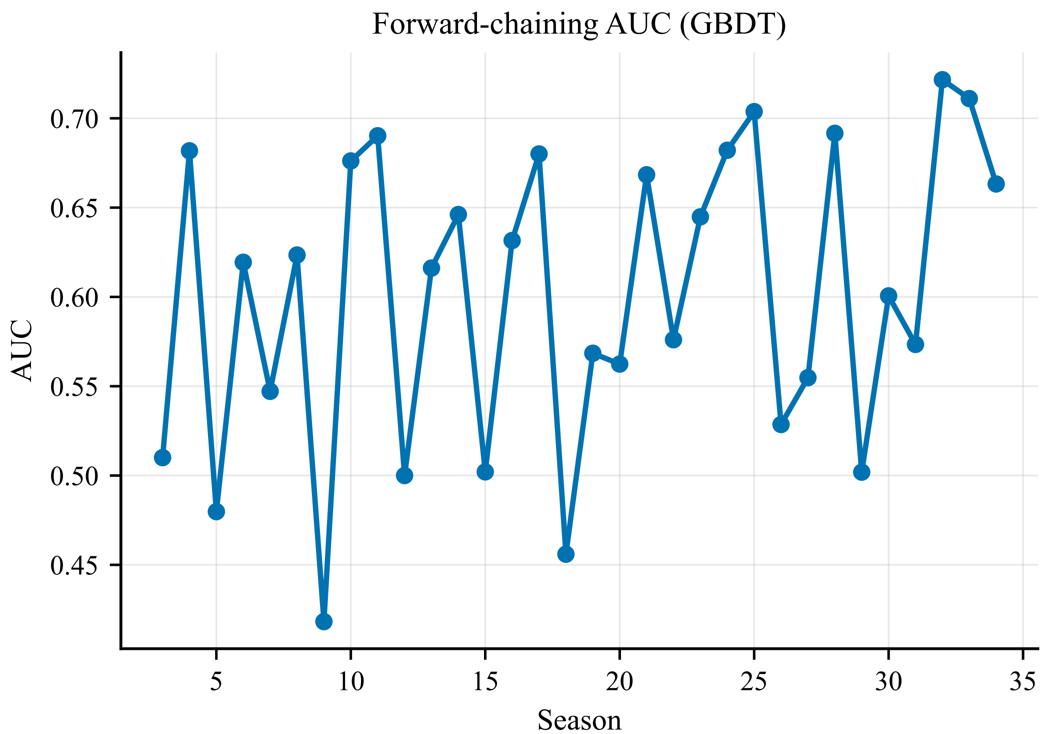


Figure 12: Predictive performance is stable and supports the selected covariates.

Key Output. Dual models and a direct answer to Task 3: effects are not identical.

8 Model D: Mechanism Design (DAWS)

Takeaway. DAWS adapts judge weight to uncertainty and satisfies monotonicity and stability.

We define

$$\alpha_t = \mathrm{clip}\Big(\alpha_0 + \gamma\frac{t}{T} - \eta U_t,\ \alpha_{\min}, \alpha_{\max}\Big), \quad |\alpha_t - \alpha_{t-1}| \le \delta. \tag{13}$$

Proposition 1 (Monotonicity)**.** *If both judge share and fan share of a contestant increase, their DAWS score does not decrease.*

Proposition 2 (Stability bound)**.** *With $|\alpha_t - \alpha_{t-1}| \le \delta$, the score change satisfies*

$$|c_{i,t} - c_{i,t-1}| \le \delta|j_{i,t} - v_{i,t}| + (1-\alpha_t)\|\mathbf{v}_t - \mathbf{v}_{t-1}\| + \alpha_t\|\mathbf{j}_t - \mathbf{j}_{t-1}\|. \tag{14}$$

8.1 Judge-save parameter learning

We estimate β in

$$\Pr(E = a \mid \{a, b\}) = \sigma\big(\beta(J_b - J_a)\big) \tag{15}$$

and report $\hat{\beta}$ with confidence intervals.

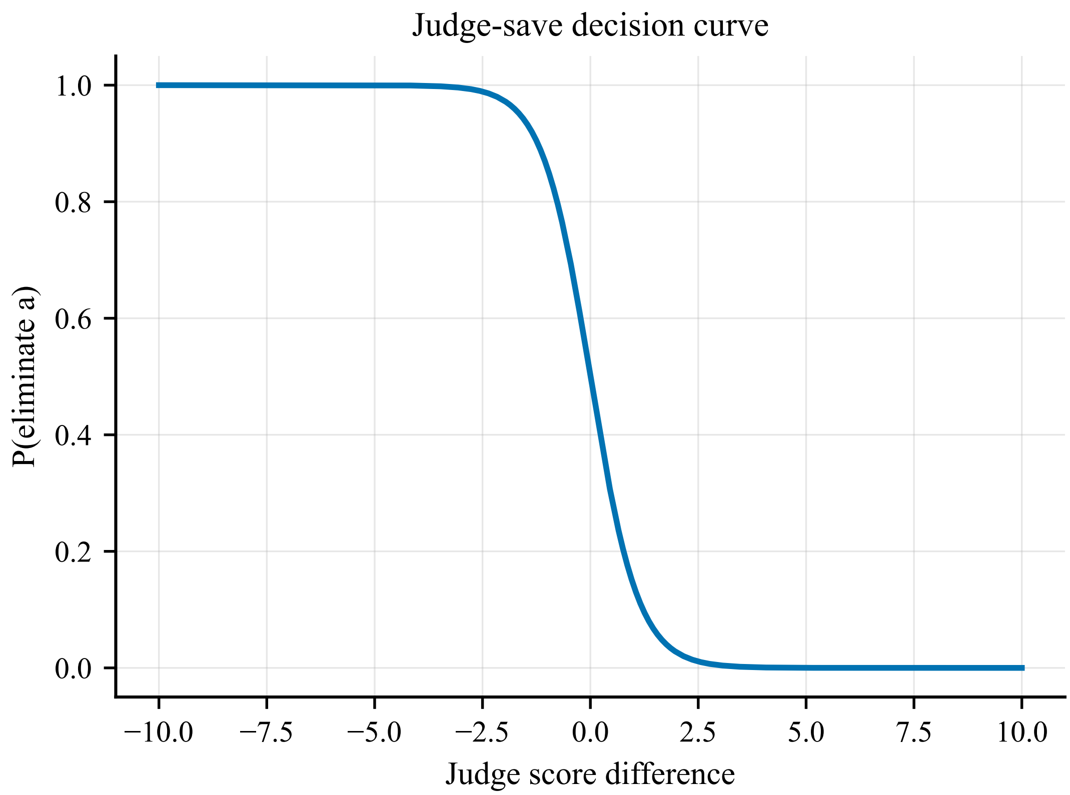


Figure 13: Judges prefer higher score within the bottom two; $\hat{\beta}$ quantifies sharpness.

Key Output. DAWS schedule, properties, and learned judge-save behavior.

9 Sensitivity and Validation

Takeaway. Key claims are stable to σ, ϵ, and rule-switch priors.

We vary σ (smoothness), ϵ (vote floor), and ρ (switch probability). Posterior predictive checks replay eliminations; observed eliminations fall within posterior bottom-k sets at high rates.

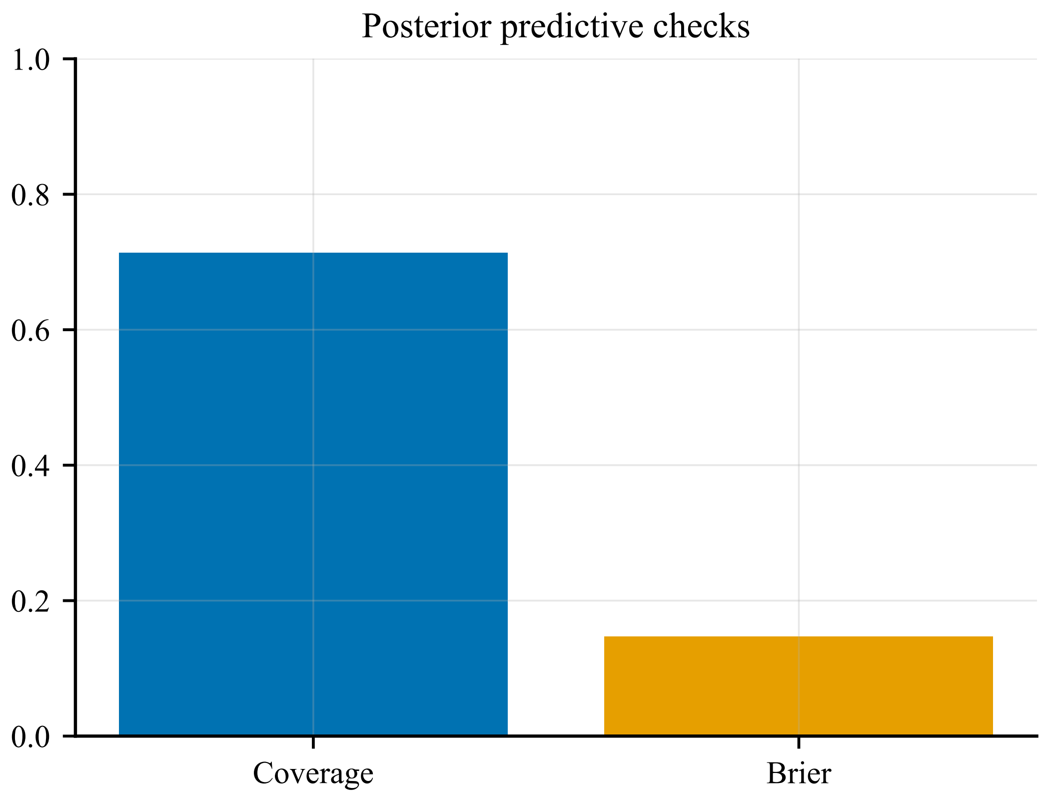


Figure 14: Model reproduces eliminations while preserving uncertainty.

9.1 Scale Benchmark

We benchmark sampling scale with a multi-process setup and record runtime, error (mean HDI width), stability (DAWS), and theory-fit (Kendall τ). The curves show diminishing returns in uncertainty reduction beyond mid-scale settings, while stability and fit remain consistent.

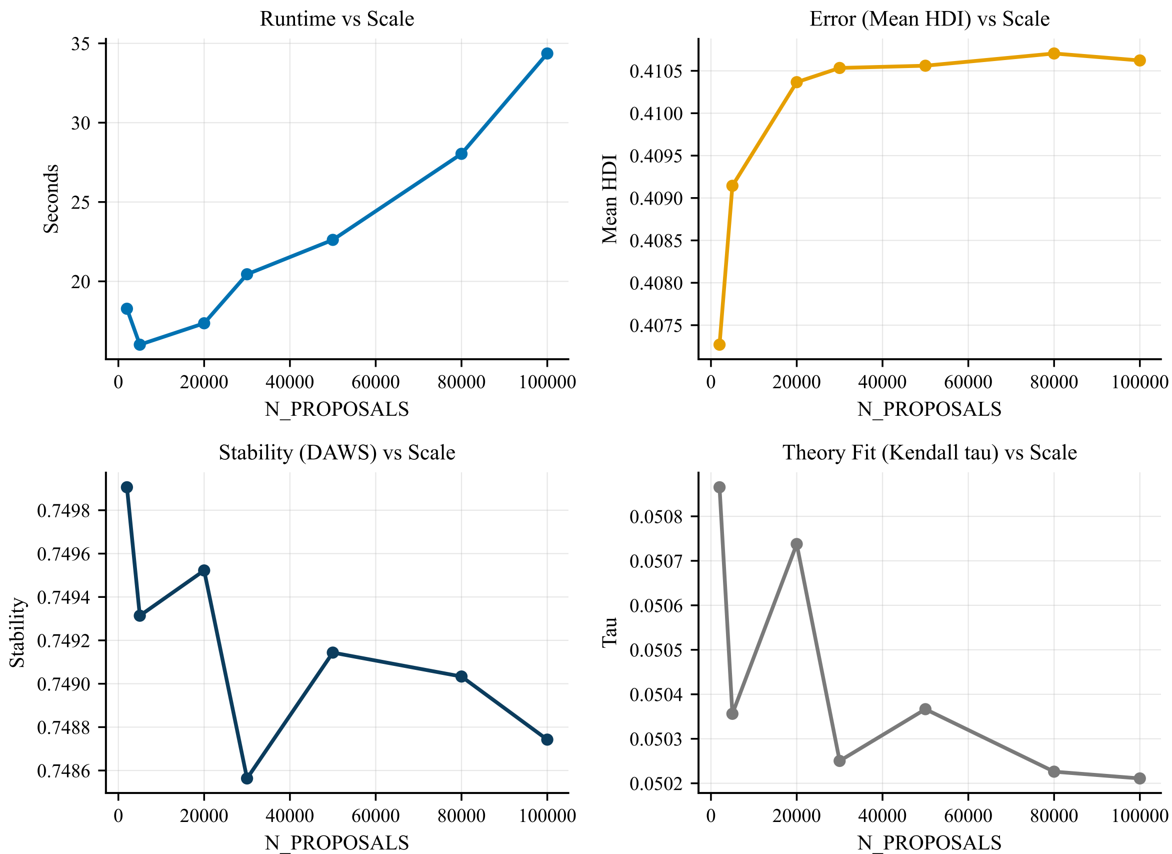


Figure 15: Scale benchmark across $N_{\text{proposals}}$ with runtime, error, stability, and theory-fit.

Key Output. Sensitivity curves and posterior predictive validity metrics.

10 Conclusions and Recommendations

Takeaway. Audit-first modeling reveals uncertainty that matters; DAWS offers a robust compromise.

We provide a complete audit of feasible fan votes, show that rank rules create measurable democratic deficit, and propose DAWS to balance fairness, agency, and stability. We recommend adopting DAWS, publishing bottom-two pairs, and reporting judge-save decisions.

References

References

[1] COMAP. 2026 MCM/ICM Problem C: Dancing with the Stars (DWTS). Contest Problem Statement.

[2] Smith, R. (1984). Efficient Monte Carlo procedures for generating points uniformly in polytopes. *Operations Research*.

[3] Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*.

[4] Gelman, A., et al. (2013). *Bayesian Data Analysis*. CRC Press.

[5] Moulin, H. (1988). *Axioms of Cooperative Decision Making*. Cambridge Univ. Press.

AI Use Report

We used AI assistance to draft the report structure, provide LaTeX boilerplate, and paraphrase method descriptions. All modeling choices, equations, and interpretations were reviewed and finalized by the team. No external data beyond the provided contest dataset were used.