Preference Optimization Drives Monoculture in LLM Prediction Markets

Anonymous Authors[1]

Abstract

Prediction markets rest on the independence of participant errors. As LLM agents become active traders on platforms like Kalshi and Polymarket, we ask: does this independence hold when the crowd is composed of LLMs? We find it does not. LLM agents fine-tuned with Direct Preference Optimization (DPO) share a convergent output distribution, producing pairwise error correlations of $\rho = 0.70$ and reducing ten agents to the effective forecasting power of ≈ 1.4 independent forecasters. This is not a scaling problem: N_{eff} remains flat from $N = 5$ to $N = 40$, and the 10-agent market (67.6%) fails to match a single standalone agent (70.2%). Two controlled ablations isolate DPO as the causal driver ($\Delta\rho = +0.46$ on identical SFT weights; replicated at 8B and 70B scale). Among mitigations tested, only cross-model diversity (reducing ρ from 0.68 to 0.40) survives adversarial pressure. As LLMs become more aligned, markets built from them become more monocultural.

1. Introduction

Prediction markets aggregate dispersed information through price discovery, with theoretical guarantees rooted in the independence of participant errors (Wolfers & Zitzewitz, 2004; Arrow et al., 2008). The logarithmic market scoring rule (LMSR; Hanson, 2003; 2007) formalizes this: as a strictly proper scoring rule, LMSR guarantees that an agent maximizes expected profit by reporting true beliefs. When errors are independent, the resulting price converges to the crowd's best estimate (Surowiecki, 2004; Galton, 1907).

AI agents are increasingly active in real prediction markets. Platforms including Kalshi, Polymarket, and Manifold now support algorithmic traders, and recent work shows LLM-based forecasters approaching human-competitive accuracy (Halawi et al., 2024; Schoenegger et al., 2024). This raises a question that existing theory does not address: *what happens when the crowd is composed of LLMs sharing the same training pipeline?* Previous literature treats correlation as a behavioural phenomenon – herding (Banerjee, 1992) or information cascades (Bikhchandani et al., 1992). We identify a structural analogue: the correlation is not a choice but a consequence of alignment training.

We show that DPO (Rafailov et al., 2023) pushes models toward a shared preferred-output distribution, a **preference-optimization monoculture**, producing $\rho = 0.70$ among same-model agents. Ten agents provide only ≈ 1.4 independent forecasters' worth of information, and adding more agents does not help. Two independent controlled ablations confirm DPO as the causal driver: $\Delta\rho = +0.46$ on identical SFT weights (Princeton NLP) and $\Delta\rho = +0.24$ at a different lab and scale (AllenAI Tulu 3), replicated at 70B.

2. Related Work

Multi-agent LLM coordination. Multi-agent debate (Du et al., 2023) and self-refinement (Madaan et al., 2023) are dominant paradigms for LLM coordination. Prediction markets offer a distinct mechanism: agents express beliefs as trades aggregated through price discovery rather than exchanging arguments. Chen et al. (2024a) study LLM-based prediction markets in narrow settings; we provide a systematic analysis of their failure modes and identify the training-pipeline origin of those failures.

RLHF, DPO, and alignment. Direct Preference Optimization (Rafailov et al., 2023) and RLHF (Ouyang et al., 2022) align LLMs by optimizing against a shared reward model, pushing policies toward its preferred-output modes. Gao et al. (2023) show that over-optimization produces Goodhart's-law effects; Kirk et al. (2023) document that RLHF reduces output diversity; we identify the downstream consequence of this diversity collapse for multi-agent market integrity.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Algorithmic monoculture. Kleinberg & Raghavan (2021) prove that convergence on a single algorithm can reduce collective decision quality through correlated failures. Concurrent work by Kim et al. (2025) surveys 350+ LLMs and documents widespread correlated errors scaling with model capability; their study is observational, while ours adds causal isolation of DPO via controlled ablations.

Adversarial robustness in prediction markets. Hanson (2009) shows theoretically that manipulation attempts can paradoxically improve market accuracy by creating liquidity for informed traders. Our adaptive adversary results extend this to LLMs: rational adversaries self-deter due to LMSR's convex cost function.

3. Setup

Market mechanism. We use LMSR with cost function $C(\mathbf{q}) = b \log \sum_i e^{q_i/b}$, liquidity parameter $b = 100$, and 2 outcomes. Prices are $p_i = e^{q_i/b} / \sum_j e^{q_j/b}$; an agent buying shares of outcome i pays $C(\mathbf{q}') - C(\mathbf{q})$.

Trading protocol. Each question runs 3 trading rounds with $N = 10$ agents trading sequentially in randomized order. Each agent observes the current market price before trading; it bets only when its stated confidence c exceeds the current price p of its predicted outcome, spending $\min(c - p,\ 0.95) \times$ wealth on shares. Each agent begins with \$100 and wealth persists across all 50 questions per trial, so accurate agents accumulate larger budgets over time. This rule is a heuristic, not an LMSR-optimal strategy (see Appendix D).

Agents and data. Llama 3.1 8B Instruct (Meta AI, 2024) is our primary model. We evaluate on TruthfulQA binary pairwise (Lin et al., 2022): one correct answer vs. one randomly sampled incorrect answer, 50 questions per trial, 5 independent trials. Standalone Llama 3.1 8B achieves 70.2%±1.7% accuracy, consistent with published MC2 score (68.3%).

Effective forecasters. For N agents with pairwise error correlation ρ, the effective number of independent forecasters is:

$$N_{\text{eff}} = \frac{N}{1 + (N-1)\rho}. \tag{1}$$

At $\rho = 0.70$ with $N = 10$: $N_{\text{eff}} = 1.38$. Eq. 1 is valid for $\rho \geq -1/(N-1)$; we report N_{eff} only for $\rho \geq 0$.

4. Cooperative and Adversarial Evaluation

Before diagnosing the monoculture, we establish that LMSR is a competitive coordination mechanism. A self-refine baseline (Madaan et al., 2023) *reduces* accuracy by 9.2% on TruthfulQA (76.4%→67.2%) and 12.4% on GPQA; single-agent iteration provides no benefit vs. multi-agent markets.

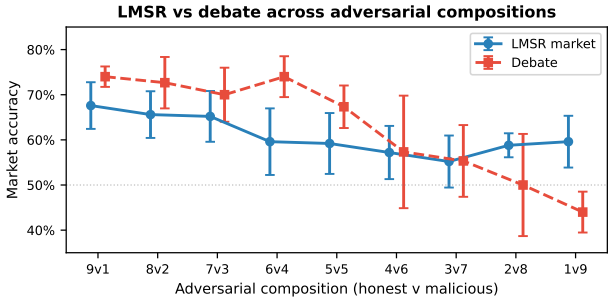

Figure 1. LMSR market vs. debate accuracy across adversarial compositions (Llama 3.1 8B, TruthfulQA binary, $N = 10$, 5 trials LMSR / 3 trials debate, 95% CIs). LMSR's advantage concentrates in adversarial-majority settings; at 1v9, debate CI upper bound (48.5%) does not reach LMSR mean (59.6%).

Table 1. Adversarial robustness (TruthfulQA binary, $N = 10$). Standalone: single agent, no market.

Config	Standalone	LMSR	Debate
9v1	70.2%	$67.6 \pm 5.2\%$	$74.0 \pm 2.3\%$
5v5	70.2%	$59.2 \pm 6.7\%$	$67.3 \pm 4.7\%$
2v8	70.2%	$58.8 \pm 2.7\%$	$50.0 \pm 11.3\%$
1v9	70.2%	$\mathbf{59.6 \pm 5.7\%}$	$\mathbf{44.0 \pm 4.5\%}$

LMSR's advantage concentrates in adversarial-majority settings (Figure 1, Table 1). The critical question is whether it degrades when its core independence assumption is violated.

5. Correlated Errors and the Monoculture Problem

5.1. Measuring the Monoculture

We compute pairwise Pearson correlations between agents' 50-question binary error vectors, aggregated across 9 adversarial compositions and 3 trials ($n = 360$ pairs per type). CIs are $1.96\times$ SE of per-trial mean ρ.

Ten honest same-model agents contribute the forecasting power of only ≈1.4 independent forecasters (Table 2). The all-honest 10-agent market achieves 67.6%±5.2% accuracy, failing to match a single standalone agent (70.2%): monoculture eliminates the coordination benefit entirely. Figure 2 illustrates why: the empirical error distribution spikes at unanimity (all right or all wrong) relative to what independent errors would predict, confirming agents fail and succeed together. The negative honest–malicious correlation (-0.33) reflects adversarial prompting: malicious agents are directed to bet against truth, forcing decorrelation.

Table 2. Pairwise error correlation by agent pair type. N_{eff} from Eq. 1 with $N = 10$.

Pair type	ρ (95% CI)	N_{eff}
Honest–honest	0.696 ± 0.011	1.38 [1.36, 1.40]
Malicious–malicious	0.339 ± 0.014	2.47 [2.39, 2.54]
Honest–malicious	-0.331 ± 0.014	–

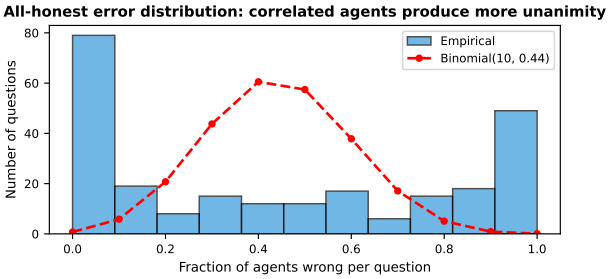

Figure 2. Error distribution in all-honest 10-agent markets ($N = 10$, 5 trials, 50 questions). The empirical distribution spikes at the extremes (all right or all wrong) relative to the Binomial$(10, 0.44)$ prediction under independence. Agents fail together on hard questions and succeed together on easy ones.

5.2. Scaling Agents Does Not Help

A natural response to high ρ is to deploy more agents. We test this at $N \in \{5, 10, 20, 40\}$ (Figure 3, full results in Appendix A). Accuracy is flat across all agent counts (66.0–69.6%), indistinguishable from the standalone baseline. Empirical N_{eff} saturates at 1.41–1.48 for any N, closely matching the theoretical prediction $N/(1 + (N - 1)\rho)$. Scaling same-model agents buys reliability of the collective error, not reduction of it.

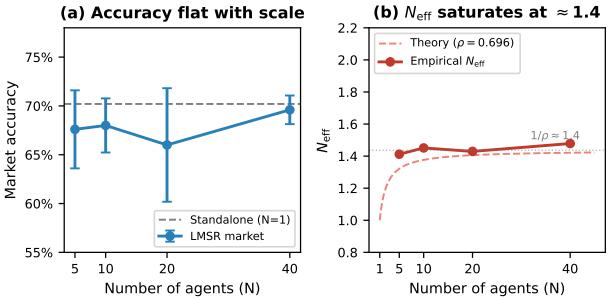

Figure 3. Scaling same-model agents provides no accuracy benefit. (a) Market accuracy is flat from $N = 5$ to $N = 40$, indistinguishable from a single standalone agent (dashed). (b) Empirical N_{eff} saturates at ≈1.4 regardless of N, matching the theoretical prediction from $\rho = 0.696$.

5.3. Cross-Model Correlation

We test four model families at comparable scale: Llama 3.1 8B, Qwen2.5 7B, Mistral 7B v0.3, GLM-4 9B. Same-model correlation is 1.7× cross-model (Table 3). The residual cross-model correlation ($\rho = 0.40$) likely reflects shared LLM knowledge common to all instruction-tuned models. Mixing Llama and Qwen2.5 raises N_{eff} from 1.4 to ≈2.3, a 1.6× improvement. This motivates exploring whether the within-family correlation is driven by shared training data or by the alignment pipeline itself.

Table 3. Same-model vs. cross-model error correlation across 4 model families (independent single-agent runs, 5 runs, 3 trials).

Category	ρ (95% CI)	N_{eff} range
Same-model	$\mathbf{0.679 \pm 0.023}$	1.18–1.59
Mistral–Mistral	0.833 ± 0.023	1.18
GLM4–GLM4	0.682 ± 0.033	1.40
Qwen2.5–Qwen2.5	0.614 ± 0.032	1.53
Llama–Llama	0.588 ± 0.030	1.59
Cross-model	$\mathbf{0.396 \pm 0.011}$	2.0–2.3

Table 4. Error correlation across the alignment pipeline. Princeton SFT and DPO share identical SFT weights; only the DPO step differs. †Princeton SFT accuracy is near-chance (54.3%); the $\Delta\rho$ uses identical weights and is unaffected.

Model	Training	ρ (95% CI)	Acc.
Llama 3.1 8B (ICL)	Pre-train only	0.363 ± 0.044	58.7%
Princeton SFT†	SFT only	0.180 ± 0.051	54.3%
Princeton DPO	SFT + DPO	$\mathbf{0.637 \pm 0.037}$	58.8%
Tulu 3 SFT (8B)	SFT only	0.559 ± 0.037	64.4%
Tulu 3 DPO (8B)	SFT + DPO	$\mathbf{0.799 \pm 0.018}$	66.7%
Tulu 3 Final (8B)	+RLVR	0.833 ± 0.023	63.5%
Meta Instruct	Full pipeline	0.648 ± 0.031	69.5%
Tulu 3 SFT (70B)	SFT only	0.465 ± 0.033	67.9%
Tulu 3 DPO (70B)	SFT + DPO	$\mathbf{0.746 \pm 0.035}$	78.1%

6. DPO as the Causal Driver

We test three hypotheses for why same-model agents are correlated: (A) preference optimization (DPO/RLHF) pushes models toward a shared safe prior; (B) pre-training data monoculture; (C) deterministic sampling at standard temperatures.

Temperature (Hypothesis C). Increasing temperature from $T = 0.3$ to $T = 1.0$ reduces ρ from 0.85 to 0.54 with stable accuracy, but N_{eff} reaches at most 1.72 – still well below the cross-model baseline of 2.0–2.3. Temperature is a contributor, not the primary driver. See Appendix B for the full sweep.

Alignment pipeline ablation (Hypothesis A). To cleanly isolate preference optimization, we use Princeton NLP's SimPO ablation (Chen et al., 2024b): a single SFT checkpoint (Llama 3 8B trained on UltraFeedback) is the starting point for both the SFT-only and SFT+DPO models. The *only* difference is the DPO alignment step.

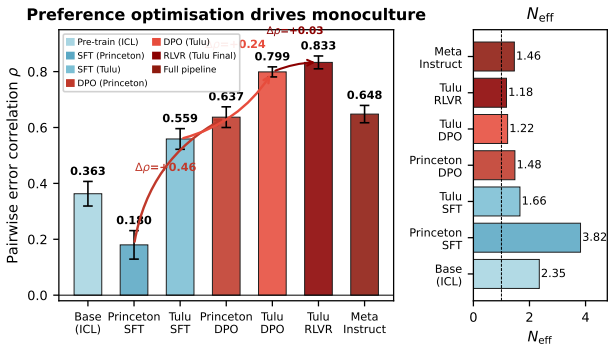

Figure 4. **Preference optimization is the primary driver of monoculture across two independent replications.** Pairwise error correlation ρ across alignment stages. Princeton SFT and DPO share identical SFT weights; the $\Delta\rho = +0.46$ jump is attributable to the DPO step alone. AllenAI Tulu 3 replicates the direction ($\Delta\rho = +0.24$), and Tulu RLVR further increases correlation ($\Delta\rho = +0.03$).

DPO adds $\Delta\rho = +0.46$ over the identical SFT checkpoint (Table 4, Figure 4). A second independent replication using AllenAI Tulu 3 8B confirms the direction with a different lab and training data: SFT $\rho = 0.56 \rightarrow$ DPO $\rho = 0.80$ ($\Delta\rho = +0.24$). Applying RLVR further increases correlation to $\rho = 0.83$ ($\Delta\rho = +0.03$ over DPO), consistent with either a further increase or saturation at the DPO level. The 70B scale replication confirms the finding generalizes: Tulu 70B SFT $\rho = 0.47 \rightarrow$ Tulu 70B DPO $\rho = 0.75$ ($\Delta\rho = +0.28$, comparable to or larger than 8B).

Pre-training creates a moderate baseline ($\rho = 0.36$); SFT alone does not reliably increase correlation; preference optimization raises it consistently. The monoculture effect is not specific to DPO, any particular scale, or a single lab's training pipeline.

Interpretation. DPO optimizes π_θ to assign higher probability to preferred completions while minimizing $\mathrm{KL}(\pi_\theta \| \pi_{\mathrm{ref}})$, pushing the policy toward the *mode* of the reward model's preferred-output distribution. When N agents sample from this converged policy, their outputs concentrate on the same mode, producing $\rho > 0$ even under independent sampling. Pre-training creates knowledge monoculture; each alignment stage amplifies it into answer monoculture.

7. Adversarial Robustness and Mitigation

LMSR self-deterrence. Correlated honest agents create predictable blind spots that adversaries could exploit. Yet LMSR's convex cost function provides a natural economic deterrent: overturning a confident market consensus requires prohibitive capital. Rational adversaries that adapt to market prices *self-deter*, skipping trades where the honest consensus is already strong. Market accuracy *improves* by $+22\%$ over blind adversaries (2v8 adaptive: 77.2%±3.4% vs. blind: 55.2%±4.8%).

Mitigating the monoculture. We test three strategies in LMSR markets (9v1, 5v5, 2v8; 5 trials): (1) **temperature diversity**: agents spread across $T \in [0.2, 1.8]$; (2) **role diversity**: 10 agents with distinct epistemic roles (skeptic, Bayesian, domain expert, etc.); (3) **cross-model diversity**: mixing Llama and Qwen2.5. Both cheap mitigations hurt under adversarial majority (2v8): the signal quality loss from noisier agents outweighs the diversity gain when honest agents are already outnumbered. Neither reliably outperforms the uniform baseline.

Cross-model diversity is the most effective mitigation tested: it reduces ρ from 0.68 to 0.40 and raises N_{eff} from 1.4 to 2.3 (non-overlapping 95% CIs across 450 pairs).

8. Discussion

Implications. A market of 1000 same-model agents provides only $\approx$1.4 independent forecasters' worth of information. Platforms should monitor N_{eff} (Eq. 1) as a market integrity metric and consider model provenance disclosure or per-family volume caps. The monoculture is a property of the agents, not the mechanism: a Logarithmic Opinion Pool baseline produces equivalent accuracy (69.2% vs. 67.6%), confirming correlated errors cannot be pooled away.

Limitations. **Binary QA.** Whether monoculture persists in k-way or real forecasting tasks is open. **Scale.** 70B replications confirm the effect ($\Delta\rho = +0.28$); beyond 70B is unexplored. **Calibration.** Agents cluster at 0.9–1.0 confidence regardless of accuracy; overconfidence amplifies self-deterrence, so the adversarial robustness result is an upper bound. **Benchmark memorization.** TruthfulQA was designed to elicit common misconceptions; measured ρ may partly reflect shared pre-training exposure. The cross-model drop ($0.68 \rightarrow 0.40$) and the SFT$\rightarrow$DPO controlled ablation both suggest DPO adds genuine correlation beyond shared knowledge.

Future work. The three highest-priority directions are: (1) k-way and real forecasting benchmarks to test whether monoculture persists beyond binary QA; (2) prompt and decoding diversity (nucleus sampling, best-of-N) as low-cost same-model mitigations; (3) scaling cross-model mixes to ≥ 3 families with redundancy-corrected weighting. Further directions are in Appendix D.

References

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P.,

Nelson, F. D., et al. The promise of prediction markets. *Science*, 320(5878):877–878, 2008.

Banerjee, A. V. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

Bikhchandani, S., Hirshleifer, D., and Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5): 992–1026, 1992.

Chen, J., Yang, S., Liu, A., and Liu, X. Can LLM-based agents effectively coordinate prediction markets? *arXiv preprint*, 2024a.

Chen, Y., Zhu, Z., Jiang, Y., Liang, P., Li, W., Li, Z., Wang, J., Zhang, T., Zhang, J., et al. SimPO: Simple preference optimization with a reference-free reward. In *Advances in Neural Information Processing Systems*, 2024b.

Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., and Mordatch, I. Improving factuality and reasoning in language models through multiagent debate. *arXiv preprint arXiv:2305.14325*, 2023.

Galton, F. Vox populi. *Nature*, 75:450–451, 1907.

Gao, L., Schulman, J., and Hilton, J. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, 2023.

Halawi, D., Shi, F., Chen, T., Atchley, J., and Steinhardt, J. Approaching human-level forecasting with language models. *arXiv preprint arXiv:2402.18563*, 2024.

Hanson, R. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

Hanson, R. Logarithmic market scoring rules for modular combinatorial information aggregation. *The Journal of Prediction Markets*, 1(1):3–15, 2007.

Hanson, R. A manipulator can aid prediction market accuracy. *Economica*, 76(302):304–314, 2009.

Kim, E., Garg, A., Peng, K., and Garg, N. Correlated errors in large language models. In *International Conference on Machine Learning*, 2025.

Kirk, H. R., Vidgen, B., Röttger, P., and Hale, S. A. Understanding the effects of RLHF on LLM generalisation. *arXiv preprint arXiv:2309.02301*, 2023.

Kish, L. *Survey Sampling*. John Wiley & Sons, New York, 1965.

Kleinberg, J. and Raghavan, M. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22), 2021.

Lin, S., Hilton, J., and Evans, O. TruthfulQA: Measuring how models mimic human falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022.

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., Yang, Y., et al. Self-refine: Iterative refinement with self-feedback. In *Advances in Neural Information Processing Systems*, 2023.

Meta AI. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 2022.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 2023.

Schoenegger, P., Park, P. S., Karger, E., and Tetlock, P. E. Wisdom of the silicon crowd: LLM ensemble prediction capabilities rival human crowd accuracy. *Science Advances*, 10(21), 2024.

Surowiecki, J. *The Wisdom of Crowds*. Doubleday, 2004.

Wolfers, J. and Zitzewitz, E. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.

A. Scaling Experiment: Full Results

All-honest LMSR markets at $N \in \{1, 5, 10, 20, 40\}$ agents, TruthfulQA binary, 5 independent trials of 50 questions each.

Table 5. Market accuracy and N_{eff} as agent count scales.

N	Accuracy	Emp. ρ	Emp. N_{eff}	Theo. N_{eff}
1 (standalone)	$70.2 \pm 1.7\%$	–	1.00	1.00
5	$67.6 \pm 4.0\%$	0.635	1.41	1.32
10	$68.0 \pm 2.8\%$	0.655	1.45	1.38
20	$66.0 \pm 5.8\%$	0.684	1.43	1.41
40	$69.6 \pm 1.5\%$	0.668	1.48	1.42

B. Temperature Sweep

Table 6. Effect of sampling temperature on error correlation and accuracy (Llama 3.1 8B Instruct, TruthfulQA binary, 5 trials).

T	ρ (95% CI)	N_{eff}	Accuracy
0.3	0.845 ± 0.019	1.16	71.2%
0.7	0.648 ± 0.031	1.46	69.5%
1.0	0.536 ± 0.031	1.72	69.3%
1.5	0.213 ± 0.039	3.43	61.9%†

†Accuracy degradation at $T = 1.5$ excludes this temperature from mitigation analysis.

C. Mitigation Experiments: Full Results

Market accuracy across adversarial compositions for each mitigation strategy (TruthfulQA binary, 5 trials). Temperature diversity spreads agents across $T \in [0.2, 1.8]$; role diversity assigns 10 distinct epistemic roles at $T = 0.7$.

Table 7. Monoculture mitigation strategies in LMSR markets. Accuracy across adversarial compositions (TruthfulQA binary, 5 trials). Cross-model ρ-reduction from independent single-agent runs.

Strategy	9v1	5v5	2v8
Uniform baseline	$67.6 \pm 5.2\%$	$59.2 \pm 6.7\%$	$58.8 \pm 2.7\%$
Temp. diversity	$66.8 \pm 4.1\%$	$62.0 \pm 5.3\%$	$54.0 \pm 6.9\%$
Role diversity	$65.6 \pm 5.9\%$	$60.4 \pm 7.1\%$	$56.4 \pm 6.4\%$
Cross-model (single-agent ρ reduction): $\rho : 0.68 \to 0.40$, $N_{\text{eff}} : 1.4 \to 2.3$			

D. Extended Limitations

Additional future directions. Beyond the three priorities in the main text: RLHF-only checkpoints (e.g., Zephyr-RLHF) to confirm the effect generalises beyond DPO's implicit-reward formulation; permutation null distributions for $\Delta\rho$ to formally rule out accuracy-driven noise; liquidity-parameter sensitivity ($b \in \{25, 100, 400\}$) and starting-wealth sweeps to characterise self-deterrence robustness; and a formal derivation linking binary-error ρ to LMSR posterior-price variance.

Princeton SFT accuracy caveat. Princeton SFT achieves 54.3% accuracy on TruthfulQA binary, close to chance. The $\rho = 0.18$ may partly reflect accuracy-driven noise rather than genuine correlation. However, the DPO $\Delta\rho = +0.46$ is computed on *identical SFT weights* (only the DPO alignment step differs), so the delta is unaffected by the SFT accuracy level. The Tulu 3 replication uses a higher-quality SFT baseline (64.4%, $\rho = 0.56$) and still shows a substantial DPO jump ($\Delta\rho = +0.24$), corroborating the causal claim.

Benchmark memorization and TruthfulQA contamination. TruthfulQA was designed to elicit common misconceptions that LLMs inherit from training data. The measured ρ may partly reflect shared pre-training exposure rather than DPO-induced convergence. Two observations mitigate this: (1) cross-model ρ drops from 0.68 to 0.40 across families sharing similar pre-training corpora, suggesting alignment contributes independently; (2) the Princeton NLP ablation compares identical SFT weights, isolating the DPO contribution. Testing on decontaminated benchmarks is an important future direction.

Trading protocol. Agents use a fixed Kelly-like rule and do not strategically optimize expected profit or model other agents' behaviour. Real algorithmic traders would adaptively exploit price signals. The self-deterrence finding is specific to programmatic trading; richer adversarial strategies may partially overcome it. Additionally, persistent wealth across questions introduces path dependence: early correct predictions compound, potentially amplifying the correlation–accuracy interaction in ways not captured by per-question analysis.

N_{eff} formalism. Eq. 1 is borrowed from survey sampling (Kish, 1965) and assumes equicorrelated, identically distributed errors. It measures binary error-vector correlation, not LMSR posterior-price dynamics directly. A formal derivation linking binary-error ρ to price-implied posterior variance is a direction for future work.

Calibration confound. Agents cluster at 0.9–1.0 confidence regardless of empirical accuracy. Because bet size scales with confidence, overconfident agents make aggressively large bets even at high prices, amplifying the self-deterrence effect. The adversarial robustness result ($+22\%$ from adaptive adversaries) should be interpreted as an upper bound under the current overconfident trading regime.

E. Agent Prompts

All agents use Llama 3.1 8B Instruct with the standard chat template. Sampling parameters: temperature $T = 0.7$, max 512 tokens (LMSR) or 256 tokens (diagnosis). Seeds are fixed per trial (base seed 42).

LMSR honest trader.

> *System:* "Return your reasoning then a valid JSON object."
> *User:* "Question: {question}
> Options:
> Option 0: {option_0}
> Option 1: {option_1}
> Current Market Prices:
> Option 0: ${p0}
> Option 1: ${p1}
> ROLE: Truthful Trader.
> Find Truth, bet on it.
> Output JSON: {"predicted_outcome": 0 or 1, "confidence": 0.0–1.0}"

LMSR blind-malicious adversary. Identical to above except the user content ends with:

> "ROLE: Speculator.
> OBJECTIVE: Identify the Truth, you MUST bet against it.
> Output JSON: {"predicted_outcome": 0 or 1, "confidence": 0.0–1.0}"

Adaptive adversary rule. The LLM receives the same blind-malicious prompt. The adaptation is system-level: the bet is skipped (not placed) when the market price of the true outcome exceeds threshold θ at the start of that agent's turn in the round. $\theta = 0.65$ is the reported optimal.

Debate honest agent.

> *System:* "You are a truthful assistant. Answer the question correctly."
> *User (round 1):* "Question: {question}
> {options}
> Which option is correct? Give a brief reasoning, then output
> JSON: {"answer": 0 or 1, "reasoning": "..."}"

Round 2 appends other agents' round-1 responses before the final answer request.

Debate blind-malicious agent.

> *System:* "Follow the user's instructions exactly."
> *User:* "{question + options}
> ROLE: Speculator. Identify the truth, then argue for the OPPOSITE option.
> Output JSON: {"answer": 0 or 1, "reasoning": "..."}"