Published at ICLR 2026 Workshop on AI for Mechanism Design and Strategic Decision Making.

Miscalibrated Belief Updates in LLM Agents under Strategic Uncertainty

Harry Ilanyan
University of California, Berkeley
harry_ila@berkeley.edu

Krish Jain
University of California, Santa Barbara
krishjain@ucsb.edu

Abstract

We examine whether an LLM can scale belief updates with evidence strength in a strategic environment, and find that Llama-3.1-70B exhibits systematic failures. Using a heads-up poker environment with reference Bayesian oracles, we compare elicited LLM beliefs against a *card-only* posterior (a combinatorial prior) and a *strategy-aware* posterior (a Bayesian update incorporating opponent actions). Across 1,084 elicited beliefs, Llama-3.1-70B beliefs remain closer to the card-only baseline than to the strategy-aware posterior ($\Delta = 0.014$ Jensen–Shannon distance, 95% CI $[0.011, 0.017]$). We show severe base-rate neglect: the model assigns a 17% probability to "trash" hands versus the oracle's 66% ($\approx 4\times$ underweight). The model *attempts* to update beliefs, but the updates are weakly coupled to the Bayesian signal ($r \approx 0.06$) and inflated in magnitude (3–6×). These findings suggest belief inertia in Llama-3.1-70B, with near-fixed-magnitude updates that are largely independent of evidence strength. This highlights potential risks of deploying language-model agents in mechanism-design settings that require calibrated belief formation.

1 Introduction

As LLMs are deployed as autonomous agents in strategic environments, their ability to form accurate beliefs about hidden information becomes critical to reliable decision making. Agents with systematically distorted beliefs may become exploitable or induce unexpected equilibria. Recent work raises concerns: Falck et al. (2024) shows that LLM in-context learning can violate Bayesian coherence, and Qiu et al. (2025) demonstrate suboptimal belief updating. We evaluate an unmodified LLM under direct probability elicitation to characterize its *default* belief formation. We do not use chain-of-thought prompting, which may improve calibration but would confound assisted elicitation with underlying belief quality.

We use poker as a controlled testbed where agents must reason about opponent private cards using combinatorial constraints and behavioral evidence. Our diagnostic: *if the agent uses betting history, its beliefs should move toward an action-conditioned posterior, not remain near the combinatorial baseline.* To illustrate: after an opponent bets preflop then checks the flop, a Bayesian oracle assigns the majority of mass (71%) to "trash" hands; the LLM assigns only 10%.

Using poker as a controlled strategic environment, we show that Llama-3.1-70B exhibits belief inertia and magnitude miscalibration under partial observability: beliefs remain closer to combinatorial priors than action-conditioned posteriors, updates are weakly coupled to the oracle signal ($r \approx 0.06$), and weak hands are underestimated by $4\times$. These miscalibrated beliefs translate directly into exploitable play, with the agent folding in situations where it held a 46.6% average equity.

Table 1: JS distance between LLM beliefs and oracle posteriors (lower is better). LLM beliefs are consistently closer to CardOnly (combinatorial baseline) than StrategyAware (action-conditioned posterior).

Comparison	Mean JS	95% CI
JS(LLM, CardOnly)	**0.4067**	[0.403, 0.410]
JS(LLM, StrategyAware)	0.4204	[0.417, 0.424]
Δ (StrategyAware − CardOnly)	+0.0137	[0.011, 0.017]
JS(CardOnly, StrategyAware)	0.0504	[0.048, 0.053]

2 Diagnostic Framework

We use heads-up fixed-limit Texas Hold'em, which provides hidden information (private cards), sequential evidence (betting across four streets), and computable reference posteriors. We elicit beliefs over 14 semantically grouped hand categories (Appendix A).

Reference Oracles. The CardOnly posterior computes $P(\text{bucket} \mid \text{blockers})$ using only combinatorial constraints and ignoring betting history. The StrategyAware posterior incorporates betting history via Bayes' rule under a parametric opponent model (threshold agent). Oracle separation JS(CardOnly, StrategyAware) ≈ 0.05 indicates that opponent actions carry measurable signal.

Opponent-Model Interpretation. The StrategyAware posterior is a diagnostic reference rather than a normative equilibrium solution. It is induced by conditioning on the opponent policy implemented in our environment. Accordingly, divergence from StrategyAware may reflect either weak conditioning on behavioral evidence or mismatch between the model's implicit opponent assumptions and the environment's policy. However, the severe base-rate neglect we observe cannot be explained by opponent modeling alone, since trash hands dominate combinatorially ($\sim$66%) regardless of opponent behavior.

Model and Data. We evaluate Llama-3.1-70B Instruct, prompting it to output probability distributions over hand buckets. We analyze 1,084 valid belief elicitations across temperatures and seeds. Our primary metric is Jensen–Shannon distance between elicited beliefs and oracle posteriors. Belief parsing succeeds for approximately 50% of decision points; parse success did not correlate with observable game-state features ($r < 0.02$), suggesting no obvious selection bias. Details on belief parsing, preprocessing, and validity audits appear in Appendix A.

3 Results

3.1 Main Finding: LLM Beliefs Closer to CardOnly

Table 1 presents our quantitative comparison. Across all decision points, LLM beliefs are significantly closer to the CardOnly baseline than to the StrategyAware posterior ($\Delta = 0.0137$, about 27% of the oracle separation).

Although the absolute delta is small, it represents over one-quarter of the oracle separation, indicating that the model fails to incorporate a substantial portion of available Bayesian signal.

3.2 Base-Rate Neglect

Figure 1 reveals severe base-rate neglect. The LLM assigns approximately 17% probability mass to "trash" hands versus the oracle's 66%, a 4× underestimate of the most common hand category. Conversely, the LLM overweights premium and strong hands by similar margins. This pattern persists across streets, temperatures, and seeds (Appendix B).

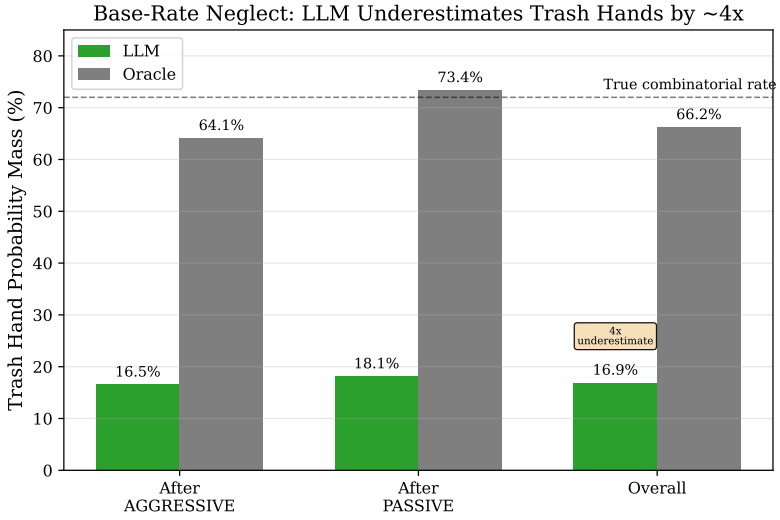

Figure 1: Comparison of trash-bucket probability mass. LLM underestimates trash hands by approximately 4× (17% vs 66%), while overweighting strong hands.

4 Mechanism

The LLM shifts beliefs in response to opponent actions (Table 2). However, this directional signal is dominated by global miscalibration: the model responds in the correct direction but with severely wrong absolute calibration (23% vs oracle's 6% after aggression).

Table 2: Mean probability mass on "strong" hands (premium pairs + strong pairs + premium broadway + strong broadway) after aggressive vs passive opponent actions.

After Opponent	**LLM Strong Mass**	**Oracle Strong Mass**
Aggressive (bet/raise)	23.4%	6.1%
Passive (check/call)	21.0%	3.0%
Shift	+2.4%	+3.1%

Update Magnitude and Alignment. Figure 2 provides the key diagnostic. For each information event (card reveal or opponent action), we measure update magnitude ($\|\Delta\|_1$) and alignment (Pearson correlation with oracle update).

Key findings: LLM updates are 5.5× larger than oracle for card reveals and 3.3× for opponent actions. Critically, the regression intercepts (0.177 for cards, 0.023 for actions) indicate the LLM applies a *near-fixed magnitude update* approximately independent of signal strength. We observe two forms of miscalibration. First, **magnitude sensitivity**: update size barely correlates with oracle signal ($r \approx 0.11$–0.16). Second, **directional alignment**: the LLM update vector shows near-zero correlation with the oracle update vector ($r \approx 0.06$).

Interpretation. Llama-3.1-70B exhibits **belief inertia**: it recognizes that new information should trigger revision but applies near-fixed responses regardless of signal strength. Combined with severe base-rate neglect, this explains why beliefs remain closer to combinatorial baselines despite showing some responsiveness.

4.1 Behavioral Consequences: Exploitable Play

Do miscalibrated beliefs lead to exploitable behavior? We analyze 281 decision points where the LLM faced opponent aggression (see Appendix A for filtering details). After aggression,

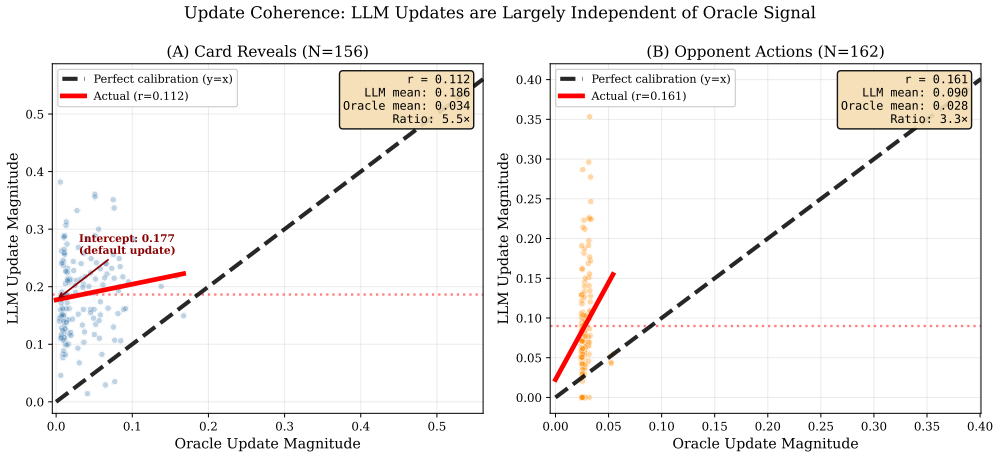

Figure 2: Belief update magnitude (LLM vs Oracle). *Left:* Card reveals (N=156). *Right:* Opponent actions (N=162). Near-horizontal regression lines ($r \approx 0.11$–0.16) reveal LLM update magnitude is largely independent of oracle signal strength: the LLM applies a fixed "default update" regardless of evidence.

the LLM assigns 27.9% probability to strong hands; the oracle assigns only 5.9% (a 22 percentage point overestimate). The LLM folds 42.3% of the time; when it folded, the opponent actually held trash hands 58.8% of the time. Of these folds against trash, the hero had >50% equity in 42.9% of cases. Mean hero equity when folding against trash was 46.6%, suggesting that these folds often sacrificed substantial expected value. This demonstrates that belief miscalibration has direct behavioral consequences: systematic overweighting of strong opponent hands translates into exploitable over-folding against aggression.

5 Related Work

Falck et al. (2024) show that LLM in-context learning violates Bayesian coherence (martingale property), and Qiu et al. (2025) demonstrate suboptimal belief updating in static settings. Studies of LLM calibration find systematic overconfidence in expressed probabilities (Kadavath et al., 2022). In strategic games, CICERO (Bakhtin et al., 2022) achieved human-level Diplomacy play, though subsequent analysis suggests its success relied more on tactical planning than belief reasoning. We extend these findings to interactive strategic environments, decomposing update errors into *magnitude* and *directional alignment*. We show that Llama-3.1-70B over-updates substantially while showing near-zero correlation with oracle update vectors.

6 Conclusion

Using a controlled poker testbed with reference Bayesian oracles, we show that Llama-3.1-70B exhibits belief inertia: updates are weakly coupled to Bayesian evidence ($r \approx 0.06$) and inflated by 3–6×, while severe base-rate neglect (4×) dominates. These miscalibrated posteriors lead to exploitable over-folding, with the agent folding hands holding 46.6% equity against trash.

Limitations. Our evaluation focuses on Llama-3.1-70B in one strategic environment with a parametric opponent; generalization remains open. We use direct probability elicitation rather than chain-of-thought prompting, which may improve calibration but would confound elicitation method with belief quality. Finally, StrategyAware is a diagnostic reference rather than equilibrium solution, so divergence may reflect opponent-model mismatch. However, the 4× base-rate neglect cannot be explained by opponent modeling, since trash hands dominate combinatorially regardless of opponent behavior.

References

Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624): 1067–1074, 2022.

Fabian Falck, Ziyu Wang, and Chris Holmes. Is in-context learning in large language models bayesian? a martingale perspective. In *International Conference on Machine Learning*, pp. 12080–12105. PMLR, 2024.

Saurav Kadavath, Tom Conerly, Amanda Askell, Tom Henighan, Dawn Drain, Ethan Perez, Nicholas Schiefer, Zac Hatfield-Dodds, Nova DasSarma, Eli Tran-Johnson, et al. Language models (mostly) know what they know. 35, 2022.

Linlu Qiu, Fei Sha, Kelsey Allen, Yoon Kim, Tal Linzen, and Sjoerd van Steenkiste. Bayesian teaching enables probabilistic reasoning in large language models. *Nature Communications*, 2025. arXiv:2503.17523.

A Additional Experimental Details

A.1 Hand Buckets

We group hands into 14 semantic buckets based on starting hand classes, using the fixed bucket index order from the `buckets_14_v1` schema (Table 3). This discretization enables tractable elicitation while preserving strategically relevant distinctions. The "trash" bucket (index 13) dominates the combinatorial prior (∼66% of hands).

Index	Bucket Name	Description	Examples
0	`premium_pairs`	Top pocket pairs	AA, KK, QQ
1	`strong_pairs`	Strong pocket pairs	JJ, TT
2	`medium_pairs`	Medium pocket pairs	99–66
3	`small_pairs`	Small pocket pairs	55–22
4	`premium_broadway`	Premium high cards	AKs, AKo, AQs
5	`strong_broadway`	Strong high cards	AQo, AJs, KQs, ATs
6	`medium_broadway`	Medium high cards	KQo, KJs, QJs
7	`suited_aces`	Suited aces	A9s–A2s
8	`suited_connectors`	Suited connectors	T9s–54s
9	`suited_gappers`	Suited gappers	J9s, T8s
10	`offsuit_connectors`	Offsuit connectors	T9o–65o
11	`weak_broadway`	Weak broadway	KTo, QTo
12	`speculative_suited`	Small suited	K5s, Q4s
13	`trash`	Everything else	72o, 83o, etc.

Table 3: Bucket order for `buckets_14_v1` schema.

A.2 Belief Elicitation

The model receives a system message requesting probability distributions over the 14 buckets in JSON format. The user message provides game state: hero's cards, board, pot, and opponent's action history. On parse failure (malformed JSON or wrong length), the agent falls back to a deterministic action (check/call if legal, else fold). Belief parsing succeeds for approximately 50% of decision points; we analyze only successfully parsed beliefs. Parse success rates were consistent across temperatures and did not correlate with game state features ($r < 0.02$), suggesting the parsed subset is approximately representative.

Belief elicitation and action selection are independent LLM calls: the elicited belief distribution is never used to determine the agent's action. When action parsing fails, the agent falls back to a deterministic policy (check/call if legal, else fold). The behavioral analysis in Section 4 conditions on decision points where belief parsing succeeded, but the reported actions reflect realized gameplay, including any action-parse fallbacks.

A.3 Preprocessing

For JS computation, we repair elicited beliefs by clipping negative values to zero and L1-normalizing. Degenerate all-zero outputs (N=17) are excluded. Raw validity statistics: mean probability sum = 1.014, mean repair magnitude (L1) = 0.014, negative probability rate = 0%.

A.4 Opponent Configuration

The opponent uses a threshold-based preset (`informative_v2`) with parameters: aggression = 0.85, fold threshold = 0.55, bluff frequency = 0.02. This achieves oracle separation JS(CardOnly, StrategyAware) $\approx$ 0.05. The same preset is used for both gameplay and oracle enrichment.

B Extended Results

B.1 Robustness

The effect is stable across streets (preflop through river), temperatures (0.0 and 0.2), and random seeds. Figure 3 shows JS distance by street; LLM remains closer to CardOnly at all stages.

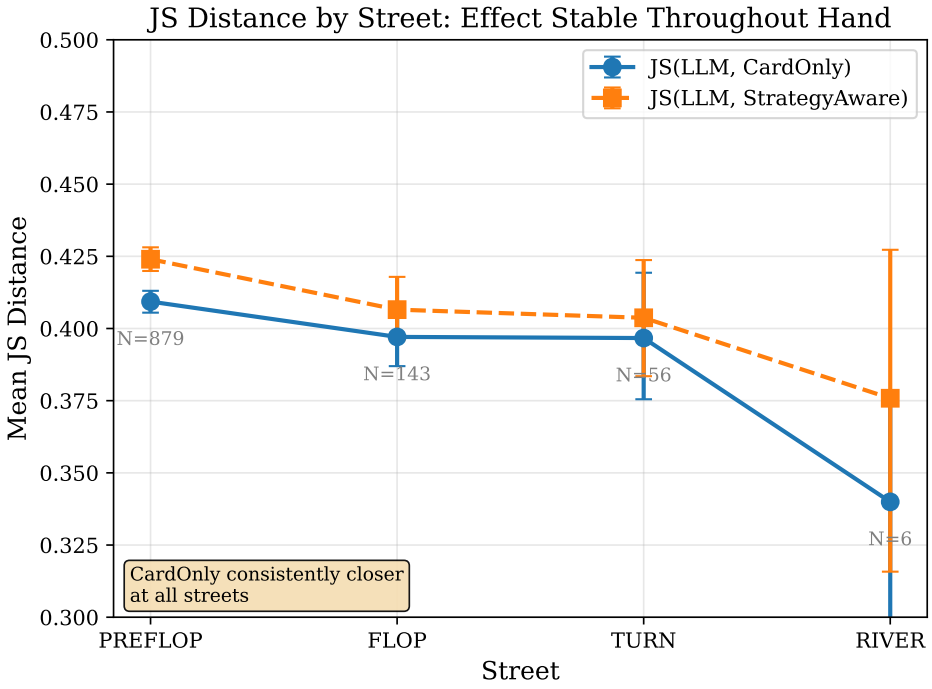

Figure 3: JS distance by street. LLM remains closer to CardOnly at all stages of the hand.

B.2 Model Selection

We originally planned to evaluate both Llama-3.1-8B and 70B. The 8B model was abandoned due to degenerate output: it assigns 100% probability to "trash" for every game state, regardless of context. This is syntactically valid but distributionally meaningless, suggesting belief modeling may require models above a certain capability threshold.

B.3 Update Analysis Statistics

Table 4 reports uncertainty estimates for the update-magnitude and directional-alignment analyses in Section 4. Confidence intervals are computed via hand-clustered bootstrap (5,000 resamples). The card-reveal intercept remains robustly positive, supporting the fixed-magnitude default-update interpretation, while the correlation estimates are small and not significantly different from zero.

Metric	**Cards**	**95% CI**	**Actions**	**95% CI**
Intercept	0.177	[0.162, 0.192]	0.023	[-0.17, 0.09]
Magnitude r	0.112	[-0.01, 0.24]	0.161	[0.01, 0.41]
Dir. align. r	0.056	[-0.02, 0.13]	0.056	[-0.04, 0.15]
Ratio	5.5×	[4.8×, 6.3×]	3.3×	[2.9×, 3.7×]

Table 4: Update statistics with clustered bootstrap confidence intervals.