

Point-by-Point Response to Author Feedback

Revision of “Extending Marginal Reputation to Markov States”

February 17, 2026

Preamble

We thank Daniel Luo for providing detailed, expert-level feedback on our submission, which attempted to extend the framework of Luo & Wolitzky (2024) “Marginal Reputation” from i.i.d. to Markov states. As Luo identified, the original submission contained fundamental errors—most critically, the failure to account for how state persistence disciplines short-run (SR) player information sets and beliefs.

In response, we carried out a systematic revision informed by 21 computational tests across 7 analysis areas (SA1–SA7), generating 40 figures over approximately 8 minutes of runtime. These tests allowed us to precisely diagnose which claims survive and which fail, and to develop corrected results.

Summary of the revision:

- The KL counting bound, filter stability, OT robustness, and monotonicity results all *survive* the extension to Markov states.
- The *fatal flaw* is that SR beliefs depend on the revealed state: $F(\cdot|\theta_t) \neq \pi$ in general. This makes the Nash correspondence $B(s_1^*, \mu_0)$ state-contingent rather than static.
- **New Theorem 1'** (belief-robust games): When the SR best-response set is invariant across filtering beliefs, the original bound $V(s_1^*)$ holds exactly.
- **New Theorem 1''** (general case): For all supermodular games, a corrected bound $V_{\text{Markov}}(s_1^*) \leq V(s_1^*)$ holds, with equality iff the game is belief-robust.
- All specific errors identified by Luo have been corrected (Remark 2.6 revised, Remark 3.3 removed, payoff structure restricted, monotonicity definition fixed).

Below we address each of the 15 points raised in the reply thread (R1–R7) and quote-tweet thread (Q1–Q8).

Reply Thread (R1–R7)

Point R1: “Very nicely written nonsense”

“*very nicely written nonsense upon a first skim*”

ACCEPTED.

The original submission overclaimed: it asserted that the extension from i.i.d. to Markov states required no mixing-time correction, when in fact the game-theoretic argument breaks down due to SR belief dynamics. The revision is framed as a *corrected analysis* that identifies precisely where the original argument fails and provides two corrected results (Theorems 1' and 1''). Section 1 (Introduction) now honestly acknowledges the initial overclaim and the subsequent correction cycle. The computational evidence from all 7 sub-analyses (SA1–SA7, 21 scripts, 40 figures) provides rigorous support for both the failure diagnosis and the corrected results.

Point R2: Lifting construction and Remark 2.6

“the key ‘lifting construction’ doesn’t seem to make much sense: in particular, remark 2.6 remains the same without the lifting construction (all ergodic distributions have fixed stationary distributions)”

ACCEPTED.

Luo is correct that the stationary distribution property holds for *any* ergodic chain, making the lifting superfluous for that purpose. **Remark 2.6 has been revised.** The new text clarifies that the lifted state $\tilde{\Theta} = \Theta \times \Theta$ serves a different, genuine purpose: it provides the product structure on which optimal transport and cyclical monotonicity characterizations apply, enabling the confound-defeating analysis of Section 5. The claim that lifting is needed for stationarity has been removed entirely. See revised Section 2.

Point R3: Payoffs depending on the full lifted state

“its also not clear why you want payoffs to depend on the entire lifted state—this is certainly not the case in the framework that we consider! so bernoulli payoffs here are confusing”

ACCEPTED.

The original formulation allowed payoffs $u_1(\theta_t, \theta_{t-1}, a_1, \alpha_2)$, which is outside the Luo–Wolitzky framework and introduced unmotivated generality. **Payoffs are now restricted to $u_1(\theta_t, a_1, \alpha_2)$ throughout the paper.** This covers all applications (including the deterrence example) and eliminates the confusing “Bernoulli payoffs” on the lifted space. The restriction is stated in Section 2 (Model) and maintained consistently. SA7 confirms that supermodularity is preserved under the first-coordinate order for θ_t -only payoffs (216/362,880 orderings of $\tilde{\Theta}$ preserve supermodularity—exactly those consistent with the θ_t ranking).

Point R4: Remark 3.3 on NBC being “easier”

“remark 3.3 is also confusing—i’m not quite sure what it means. we know that NBC is generically satisfied in exogenous α_0 games, as the document you’ve written implicitly assumes—so how does it get easier to satisfy?”

ACCEPTED.

Remark 3.3 has been removed entirely. The claim that the non-best-response condition (NBC) becomes “easier to satisfy” in the Markov case was poorly motivated and, as Luo notes, meaningless given that NBC is generically satisfied in the exogenous- α_0 framework the paper already assumes.

Point R5: i.i.d. disciplines SR information sets (fatal flaw)

“the claim ‘the only place where i.i.d. is used is lemma 3’ is also incorrect and the fatal flaw of the proof attempt — we also use it to discipline SR player information sets about the state, because the public signal now has autocorrelation that influences SR player BR”

ACCEPTED — this is the central fix of the revision.

This is the most important critique and drives the entire revision. Luo identifies that in the Markov case, the public signal has autocorrelation, so the SR player’s information set includes inferences about the current state θ_t from the action history. When the Stackelberg strategy is state-revealing (e.g., $s_1^*(G) = A$, $s_1^*(B) = F$), the SR player knows θ_t exactly and updates beliefs to $F(\cdot|\theta_t) \neq \pi$.

Computational evidence:

- **SA1** confirms: mean total-variation distance $\|F(\cdot|\theta_t) - \pi\| = 0.466$ (far from zero).
- **SA2** derives the closed-form belief gap: $\Delta = 2\alpha\beta|1 - \alpha - \beta|/(\alpha + \beta)^2 = 0.094$ at baseline parameters ($\alpha = 0.3, \beta = 0.5$). This gap is *permanent* and vanishes if and only if $\alpha + \beta = 1$ (the i.i.d. case).
- **SA6** quantifies the downstream impact: 14.7% payoff overestimation (V from 0.638 to 0.547) and 37.7% SR action disagreement rate.

How addressed in the revision:

- New **Section 3** introduces the concept of *belief-robustness*: a game is belief-robust if $B(s_1^*, F(\cdot|\theta)) = B(s_1^*, F(\cdot|\theta'))$ for all θ, θ' .
- **Theorem 1'** (belief-robust extension): under belief-robustness, the original bound $V(s_1^*)$ holds exactly.
- **Theorem 1''** (general corrected bound): for all supermodular games, $V_{\text{Markov}}(s_1^*) \leq V(s_1^*)$ holds, where V_{Markov} uses the state-contingent Nash correspondence $B(s_1^*, F(\cdot|\theta_t))$ averaged over the ergodic distribution.

- The proof of Lemma 3 (now Lemma 3') is revised to account for filter stability (SA4, $R^2 > 0.99$) and the belief-dependent Nash correspondence.

Point R6: Stackelberg strategy may not be well-defined

“in fact the stackelberg strategy need not even be completely well-defined: to see this, consider a persuasion example where the prior can move between regions where different concavifications are optimal”

PARTIALLY ACCEPTED.

For **supermodular games** (the primary focus of the paper), the Stackelberg strategy is well-defined: it is the monotone strategy $s_1^*(\theta) = \arg \max_{a_1} u_1(\theta, a_1, \alpha_2^*)$, which depends on the state θ_t but *not* on the prior belief μ_0 . The OT characterization on the lifted space (SA5: stability margin ≥ 0.30 in 94% of the parameter space) ensures the confound-defeating property is robust. This class is treated in detail.

For **persuasion games**, Luo’s point is well-taken: when the prior moves between regions where different concavifications are optimal, the Stackelberg strategy itself becomes belief-dependent. This is acknowledged as an open problem in Section 10 (Open Questions), where we discuss the additional challenges that persuasion games pose for the Markov extension. The current paper does not claim results for this class of games.

Point R7: $B_\eta(s_1^*, h^t, \sigma)$ depends on history

“i.i.d. helps ensure that this object is well-defined :) otherwise, it would depend on the prior belief given the history, and you’d see $B_\eta(s_1^, h^t, \sigma)$, which complicates the analysis substantially”*

ACCEPTED.

This is precisely the mechanism behind the failure. In the i.i.d. case, $B(s_1^*)$ is a static object on $\Delta(A_0) \times \Delta(A_2)$. In the Markov case, $B(s_1^*, \mu_0(h^t))$ depends on the filtering belief $\mu_0(h^t)$, which evolves with history.

How addressed:

- **Theorem 1'** handles this by requiring *belief-robustness*: B is constant across all filtering beliefs $F(\cdot | \theta)$, so history-dependence is irrelevant.
- **Theorem 1''** handles the general case: the corrected bound $V_{\text{Markov}}(s_1^*)$ explicitly accounts for the state-contingent Nash correspondence. It averages $\inf_{B(s_1^*, F(\cdot | \theta))} u_1$ over the ergodic distribution of states, yielding a weaker but correct bound.
- SA4 confirms filter stability: the filtering belief forgets initial conditions exponentially with rate $\lambda \approx |1 - \alpha - \beta|$ ($R^2 > 0.99$ over a 30×30 parameter grid), ensuring that the belief dynamics are well-behaved in the long run.

Quote-Tweet Thread (Q1–Q8)

Point Q1: AI and proof-writing difficulties

“as someone who has tried to use ai to help me write proofs, this has pretty much been my universal experience with it... there are difficulties”

ADDRESSED.

We fully agree that AI-assisted proof generation is prone to producing plausible-looking but flawed arguments. Rather than treating this as a purely negative result, the revision demonstrates both the failure mode and a *correction cycle*. Section 8 (Methodology) now describes the full workflow: AI-generated conjecture → expert critique → systematic computational testing (21 scripts, 40 figures) → corrected result. We believe this honest account is itself a contribution to the AI-for-mathematics literature, demonstrating that the combination of AI generation + expert review + computational verification can produce genuine results, even when the initial output is flawed.

Point Q2: Monotonicity only for one-dimensional states

“even in the screenshot there’s an error—our definition of monotonicity only works for one dimensional states and actions, which the ‘lifting’ technique here obviates”

ACCEPTED.

The original paper applied the one-dimensional monotonicity definition to the lifted space $\tilde{\Theta} = \Theta \times \Theta$, which is two-dimensional. **SA7 provides the resolution:** when payoffs depend only on θ_t (not θ_{t-1}), supermodularity is preserved under the *first-coordinate order* on $\tilde{\Theta}$ (order by θ_t , break ties arbitrarily). Specifically, 216 out of 362,880 orderings of $\tilde{\Theta}$ preserve supermodularity for θ_t -only payoffs—exactly those consistent with the θ_t ranking. **Section 5 has been revised** to use the first-coordinate order explicitly and to acknowledge the limitation for transition-dependent payoffs.

Point Q3: i.i.d.-ification fails

“a technique to this (find the right i.i.d.-ification of the markov state) is a natural first guess, but fails”

ACCEPTED.

The i.i.d.-ification—replacing the Markov chain with the i.i.d. process having the same marginal π —fails precisely because of the SR belief dynamics. Even though the *signals*

can be made distributionally close (the KL bound works for both processes, as SA3 confirms), the SR player's *information set* is fundamentally different: in the Markov case, observing $s_1^*(\theta_t)$ reveals θ_t and hence informs beliefs about θ_{t+1} via the transition kernel. The revision characterizes exactly when the i.i.d.-ification works: it succeeds if and only if the game is *belief-robust* (Definition 3.1), meaning SR behavior is invariant to the information gained about future states.

Point Q4: $B(s_1, \mu_0)$ depends on beliefs about both α_0 and s_1

“one core reason is that short-run player beliefs depend on both α_0 and s_1 ; in the i.i.d. case, we define the nash correspondence $B(s_1)$ as an object on $\Delta(A_0) \times \Delta(A_2)$, and use payoffs to discipline the set jointly.”

ACCEPTED.

Section 3 now formalizes $B(s_1, \mu_0)$ as the *belief-dependent Nash correspondence*, where $\mu_0 = \mu_0(\theta_{t+1} | h^t)$ is the SR player's belief about the next state given the history. In the i.i.d. case, $\mu_0 = \pi$ always, so $B(s_1) = B(s_1, \pi)$ is a well-defined static object. In the Markov case, $\mu_0 = F(\cdot | \theta_t)$ depends on the current state, creating the state-contingent Nash correspondence $B(s_1, F(\cdot | \theta_t))$ that varies across states.

The deterrence example illustrates concretely: $\mu^* = 0.60$, $F(G|G) = 0.70 > \mu^*$ (SR cooperates), $F(G|B) = 0.50 < \mu^*$ (SR defects). The “joint discipline” that works in the i.i.d. case—where payoffs pin down (α_0, α_2) simultaneously—breaks when μ_0 shifts with the state.

Point Q5: $B(s_1, \mu_0)$ varies, deviations don't work

“you then need to write $B(s_1, \mu_0)$ as a function of player 2's belief about theta and also s_1 , which will affect their best reply. but this makes it impossible to get just deviations of the form we consider to work, since μ_0 will constantly be changing.”

ACCEPTED.

This is precisely why $V_{\text{Markov}} < V(s_1^*)$. Because μ_0 changes period-to-period as the state evolves, the one-shot deviation argument cannot use a single, fixed Nash correspondence. The corrected bound in **Theorem 1''** accounts for the constantly changing μ_0 by:

1. Using the state-contingent worst-case payoff $\inf_{B(s_1^*, F(\cdot | \theta_t))} u_1(\theta_t, s_1^*(\theta_t), \alpha_2)$ in each period;
2. Averaging over the ergodic distribution $\pi(\theta)$ to get the long-run bound V_{Markov} .

SA6 confirms empirically: $V_{\text{Markov}} = 0.547 < V(s_1^*) = 0.638$, a 14.7% overestimation in the original. The gap $V(s_1^*) - V_{\text{Markov}}$ is an economically meaningful object: it measures

the “cost of persistence” in reputation games.

Point Q6: Pei (2020) and additional assumptions for persistence

“the problem is that player 2’s belief depends on the action and the state realization as a pair now, and there is persistence — this is why Pei (2020) requires additional assumptions than us; not just because of perfect persistence.”

ACCEPTED.

Our *belief-robustness* condition (Definition 3.1) is directly analogous to the additional assumptions in Pei (2020). Both restrict the short-run player’s information structure to ensure that the reputation mechanism functions despite state persistence. Specifically:

- Pei requires conditions that prevent the SR player’s belief about the LR type from being contaminated by state-learning.
- We require that the SR best-response set $B(s_1^*, F(\cdot|\theta))$ is invariant across states, so that state-learning does not change SR behavior.

The parallel is noted explicitly in Section 9 (Discussion). The fact that both papers arrive at structurally similar conditions—despite very different modeling frameworks—suggests that restricting SR information is a fundamental requirement for reputation results with persistent states.

Point Q7: Other smaller errors

“the draft posted makes several other smaller errors, too, which were easy to catch on a first read—glaring enough to someone who is an expert on the paper, but maybe just subtle enough to pass scrutiny by most other people. that’s concerning for the future of ‘AI slop’”

ACCEPTED.

All identified errors have been corrected:

- **Remark 2.6:** Revised—lifting provides OT framework, not stationarity (see R2).
- **Remark 3.3:** Removed entirely—NBC claim was vacuous (see R4).
- **Payoff structure:** Restricted to $u_1(\theta_t, a_1, \alpha_2)$ —no dependence on full lifted state (see R3).
- **Monotonicity claim:** Corrected to first-coordinate order on $\tilde{\Theta}$, with SA7 providing the precise count of valid orderings (see Q2).
- **“Only Lemma 3 uses i.i.d.”:** Corrected—i.i.d. is also used to discipline SR information sets (see R5).

We share the concern about AI-generated mathematical content that is superficially convincing but substantively flawed. Section 8 discusses this issue honestly and argues that the combination of AI generation with expert review and computational verification can serve as a quality-control mechanism.

Point Q8: State-revealing strategies and belief non-convergence

“To make it clear: suppose s_1 just takes an action that reveals the state. In the iid case, this won’t affect SR beliefs. But in the Markov case, this can cause beliefs to never settle into the stationary distribution.”

ACCEPTED, with computational evidence.

This is the clearest statement of the core problem, and our computational analysis confirms it precisely.

SA2 derives the *permanent* belief gap analytically:

$$\Delta(\alpha, \beta) = \frac{2\alpha\beta|1 - \alpha - \beta|}{(\alpha + \beta)^2}$$

At baseline parameters ($\alpha = 0.3, \beta = 0.5$): $\Delta = 0.094$, meaning the expected absolute deviation of SR beliefs from π is approximately 9.4%. The gap has the following properties:

- $\Delta = 0$ if and only if $\alpha + \beta = 1$ (the i.i.d. case), confirming that this is purely a persistence phenomenon;
- $\Delta > 0$ for all persistent chains ($\alpha + \beta \neq 1$), confirming that beliefs *never* settle to π ;
- The gap is verified both analytically (closed-form derivation) and numerically (Monte Carlo with $N = 1000, T = 5000$) across the full $(\alpha, \beta) \in (0, 1)^2$ parameter space.

The belief gap heatmap (SA2, Figure 1) is included in revised Section 6 to illustrate that the gap vanishes along the anti-diagonal $\alpha + \beta = 1$ and grows with persistence $|1 - \alpha - \beta|$.

State-by-state consequences (from SA6):

After state	SR belief $\Pr(G_{\text{next}})$	Stationary $\pi(G)$	Gap
G (saw A)	$F(G G) = 0.70$	0.625	+0.075
B (saw F)	$F(G B) = 0.50$	0.625	-0.125
Expected	—	0.625	0.094

This table appears in revised Section 6. The key economic implication: after bad states, the SR player knows the next state is more likely to be bad ($0.50 < \pi(G) = 0.625$),

which pushes their belief below the cooperation threshold ($\mu^* = 0.60$), causing them to defect. This is a *permanent* feature of Markov states, not a transient phenomenon.

Summary

Point	Luo's Critique	Verdict	How Addressed
R1	“Nicely written nonsense”	Accepted	Revision is a corrected analysis (Sec. 1)
R2	Lifting superfluous for stationarity	Accepted	Remark 2.6 revised: lifting provides OT framework (Sec. 2)
R3	Payoffs on full lifted state	Accepted	Restricted to $u_1(\theta_t, a_1, \alpha_2)$ throughout (Sec. 2)
R4	Remark 3.3 vacuous	Accepted	Remark 3.3 removed entirely
R5	i.i.d. disciplines SR info sets (fatal flaw)	Accepted	Belief-robustness (Sec. 3); Thms. 1', 1''; SA1/SA6 evidence
R6	Stackelberg not well-defined (persuasion)	Partially accepted	Well-defined for supermodular (Sec. 5); open for persuasion (Sec. 10)
R7	B_η depends on history	Accepted	Thm. 1' (belief-robust); Thm. 1'' (corrected bound); SA4
Q1	AI proof-writing difficulties	Addressed	Revision demonstrates correction cycle (Sec. 8)
Q2	Monotonicity only for 1D states	Accepted	First-coord. order; 216/362,880 valid orderings; SA7 (Sec. 5)
Q3	i.i.d.-ification fails	Accepted	Fails due to SR beliefs; works iff belief-robust (Sec. 3)
Q4	$B(s_1, \mu_0)$ depends on beliefs	Accepted	Formalized as belief-dependent Nash corr. (Sec. 3)
Q5	μ_0 constantly changing	Accepted	Corrected bound V_{Markov} in Thm. 1'' (Sec. 3)
Q6	Pei (2020) analogy	Accepted	Belief-robustness \approx Pei's conditions (Sec. 9)
Q7	Other smaller errors	Accepted	All fixed: Rmk. 2.6, 3.3, payoffs, monotonicity
Q8	Beliefs never settle to π	Accepted	Gap = $2\alpha\beta 1-\alpha-\beta /(\alpha+\beta)^2$; SA2 evidence (Sec. 6)

Disposition: 14 of 15 points fully accepted; 1 partially accepted (R6, regarding persuasion games). All accepted points have been addressed with specific revisions, new theoretical results, and/or computational evidence.