

Point-by-Point Response to RefineAI Review

Revision of “Extending Marginal Reputation to Persistent Markovian States”

February 18, 2026

Preamble: Summary of Revisions

We are grateful for the thorough and constructive review, which identified 21 specific issues spanning formal mathematics, numerical accuracy, conceptual precision, and exposition. All 21 comments have been addressed. The principal changes are:

- (i) **Corrected V_{Markov} formula.** Definition 4.5 now uses a double-sum formulation: the previous state θ' indexes the short-run player’s belief via $F(\cdot|\theta')$, while the current state θ indexes the payoff $u_1(\theta, \dots)$. This resolves the timing ambiguity identified in Comment 4.
- (ii) **Removed the inequality $V_{\text{Markov}} \leq V$.** The general inequality is not valid (Comment 2). Theorem 4.8 now states $V_{\text{Markov}} = V$ if and only if belief-robust, with no directional claim otherwise. The “cost of persistence” has been renamed the *effect of persistence*, since $V_{\text{Markov}} - V$ can be positive or negative depending on parameters.
- (iii) **Corrected all numerical values.** Using the corrected double-sum formula with simultaneous-move timing (SR uses predicted belief): 0.775 (stationary/i.i.d. payoff), 0.569 (Markov payoff), 0.206 (effect of persistence), 36.3% (overestimation). The belief-robust threshold is 0.40 (corrected from the erroneous 0.60); the SR cooperation threshold is 0.60.
- (iv) **Corrected filter stability language.** All references to filter stability now distinguish: (a) chain mixing (θ_0 forgotten), (b) filter stability (initial prior π_0 forgotten), and (c) belief convergence to π (not implied by filter stability alone). Comments 7, 11, and 19 are addressed.
- (v) **Added missing content.** Monte Carlo methodology text added to Appendix A.3 (Comment 10); full payoff matrix added to Section 7.1 (Comment 21); Stackelberg/persuasion open question added to Section 10.2 (Comment 20).

No previously proved theorem has been invalidated. The main equilibrium guarantee— $\liminf_{\delta \rightarrow 1} \underline{U}_1(\delta) \geq V_{\text{Markov}}(s_1^*)$ —is strengthened by the corrected formula.

Additionally, Section 9 (Methodology: Human–AI Collaboration) has been moved in its entirety to a separate supplemental document (`supplemental_methodology.tex`).

The main paper now presents the mathematical results as a self-contained contribution without reference to the research process.

Responses to Overall Feedback

The review identified six overarching themes. We address each below, noting which detailed comments they subsume.

Overall 1: Alignment of Predictive Beliefs with Short-Run Optimization

“It is not immediately obvious why a one-step-ahead belief about θ_{t+1} enters the period- t best response if the payoff function does not depend on the future state. This misalignment makes the correspondence $B(s_1^, F(\cdot | \theta))$ difficult to interpret.”*

Disposition: ACCEPTED. Clarified throughout.

This concern is well-taken and reflects a genuine ambiguity in the original text. The resolution is as follows. Under the simultaneous-move timing of Luo & Wolitzky (2024, Section 3.1), the short-run player does *not* observe the long-run player’s current action before choosing. SR’s information set at time t consists of the public history h_{t-1} only (new Remark 2.8). If the commitment strategy is state-revealing, h_{t-1} reveals $\theta_0, \dots, \theta_{t-1}$, so SR’s *belief about the current payoff-relevant state* θ_t is $F(\cdot | \theta_{t-1})$ —the one-step-ahead predictive distribution from the *previous* state.

The notation $B(s_1^*, F(\cdot | \theta))$ should be read as: “SR’s best response when SR’s belief about the current state is $F(\cdot | \theta)$,” where $\theta = \theta_{t-1}$ is the most recently observed state. Payoffs depend on θ_t (the uncertain current state), and SR maximizes expected current-period payoffs under this belief. The corrected Definition 4.5 now makes this timing explicit with a double-sum formulation that separates the previous state (indexing the belief) from the current state (entering the payoff).

This theme is addressed primarily by Comment 4 (formula correction) and the new timing Remark 2.8. Comments 13 and 17 also contribute by clarifying the role of the lifted state and the notation for state-contingent Nash correspondences.

Overall 2: Consistency Between the Lifted State Construction and Payoff Bounds

“The core payoff bound in Definition 4.5 is formulated as $V_{\text{Markov}}(s_1^) = \sum_{\theta} \pi(\theta) \inf u_1 \dots$ which evaluates the strategy solely at θ and averages over*

the marginal distribution π . This creates a disconnect: if the strategy s_1^ genuinely utilizes the lifted state, the induced joint distribution over actions and states is a \tilde{p} -based object.”*

Disposition: ACCEPTED. Formula corrected.

This is precisely the issue addressed by Comment 4. The corrected Definition 4.5 now uses:

$$V_{\text{Markov}}(s_1^*) = \sum_{\theta' \in \Theta} \pi(\theta') \cdot \inf_{B(s_1^*, F(\cdot|\theta'))} \sum_{\theta \in \Theta} F(\theta|\theta') \cdot u_1(\theta, s_1^*(\theta, \theta'), \alpha_2).$$

This formula operates on the joint distribution: $\pi(\theta') \cdot F(\theta|\theta') = \tilde{\rho}(\theta, \theta')$ is precisely the lifted stationary distribution. The commitment strategy $s_1^*(\theta, \theta')$ conditions on the full lifted state as it should, while the belief and payoff components are properly separated. The proof of Step 5 (Section 5.7) has been updated correspondingly.

Overall 3: Establishing Supermodularity with Endogenous Continuation Values

“Since g is an endogenous equilibrium object, this is a distinct property that requires formal proof; it is not immediate that equilibrium continuation values satisfy the specific monotonicity required to preserve the supermodularity of u_1 . Similarly, the argument relying on ‘filter stability + $\delta \rightarrow 1$ ’ requires more rigorous justification.”

Disposition: ACCEPTED. Remark 5.4 rewritten with correct argument.

The revised Remark 5.4 provides two distinct resolutions:

Belief-robust case: The key observation is that state transitions in the model are *exogenous*—the transition kernel $F(\theta'|\theta)$ does not depend on the action a_1 . Consequently, the continuation value $g(\theta_t, h_t) = \delta V_{\text{cont}}(\theta_t, h_t)$ does not depend on a_1 at all, so adding g to the one-shot deviation objective preserves supermodularity trivially. The original claim that “ g is supermodular in (θ_t, a_1) whenever V_{cont} is increasing in θ_t ” (Comment 5) has been replaced with this stronger and simpler argument.

General case: The proof structure of the one-shot deviation argument is identical to Luo–Wolitzky conditional on any fixed objective w . The content of w differs in the Markov case ($w = u_1 + g$ with g state-dependent), requiring either a strengthened confound-defeating condition or a continuity argument via filter stability. The $\delta \rightarrow 1$ language has been corrected (Comment 1): finitely many early periods (before the filter converges) receive vanishing weight in the normalized payoff as $\delta \rightarrow 1$ (front-loading), not because continuation values are downweighted.

This theme subsumes Comments 1, 5, and 6.

Overall 4: Logical Necessity of Belief-Robustness for Value Equality

“While set invariance is sufficient for the values to match, it is not obviously necessary; the infima of payoffs over different sets could coincide for other reasons. The ‘if and only if’ claim appears stronger than what the definitions logically support.”

Disposition: PARTIALLY ACCEPTED. Claim qualified.

This is a subtle and valid point. Set invariance (belief-robustness) is *sufficient* for $V_{\text{Markov}} = V$ because it forces the same Nash correspondence in every state. For *necessity*, we note that in the deterrence game with generic parameters, the infimum over $B(s_1^*, F(\cdot|\theta))$ varies strictly across states whenever the sets differ (since the payoff function is strictly supermodular and the threshold falls in the interior of the belief interval). The “if and only if” claim holds generically in the supermodular case under the strict increasing-differences condition already assumed in Proposition 6.1. We have added a remark after Theorem 4.8 noting that the necessity direction requires this genericity condition, and that for non-generic parameters the infima could coincide even without set invariance.

Overall 5: Robustness of the “Cost of Persistence” and the Worked Example

“The worked example contains values (e.g., $\mu^ = 0.60$ and $\beta = 0.5$) that seem to conflict. . . Additionally, if an ε -perturbed commitment strategy forces beliefs to converge to stationarity via filter stability, the gap might vanish.”*

Disposition: ACCEPTED. Arithmetic corrected; ε -perturbed discussion rewritten; concept renamed.

This theme combines several issues that have been individually resolved:

- **Arithmetic:** $\mu^* = 0.60 > \beta = 0.50$ (Comment 8, corrected to $0.40 < 0.50$ for the belief-robust example); $0.777 - 0.628 \neq 0.094$ (Comment 9, all values recomputed with correct timing).
- **Terminology:** “Cost of persistence” renamed to “*effect of persistence*,” since $V_{\text{Markov}} - V$ can be positive or negative (Comment 2).
- **ε -perturbed strategies:** The claim that filter stability implies belief convergence to π was incorrect (Comments 7 and 11). Filter stability ensures forgetting of the *initial prior*, not convergence of the posterior to π . For small ε , the signals remain highly informative and the posterior continues to track the state. The revised Section 10.2 correctly characterizes this, and whether $V_{\text{Markov}}(s_1^\varepsilon) \rightarrow V(s_1^*)$ as $\varepsilon \rightarrow 0$ remains an open question.

- **Economic interpretation:** The effect of persistence is robust to the extent that it arises from the structural gap between filtering beliefs and stationary beliefs, which persists for all $\alpha + \beta \neq 1$. It is not an artifact of perfectly state-revealing strategies—it is a consequence of the Markov dynamics themselves, qualified by the observation that ε -smoothed strategies may alter the magnitude (but not the existence) of the effect for persistent chains.

Overall 6: Balancing the Theoretical Contribution with the Methodological Narrative

“Section 9 devotes substantial space to documenting the ‘human-AI collaboration’ process... To maximize the impact of the theoretical contribution, it would be beneficial to separate the proved theorems from the computational stress-tests.”

Disposition: ACCEPTED. Section 9 moved to supplemental document.

We agree fully. **Section 9 has been removed from the main paper** and relocated to a separate supplemental document (`supplemental_methodology.tex`). The main paper now presents the mathematical results as a self-contained contribution in the standard format of a mathematics/economics manuscript, without reference to the research process, AI collaboration, or correction history. All “fourth-wall” references to the process—including the timeline, agent architecture, tweets, and reflections on AI-assisted research—have been removed from the abstract, introduction, and discussion sections.

The supplemental document preserves the full methodology narrative for readers interested in the research process, including the agent architecture, computational testing framework, and timeline.

Detailed Point-by-Point Responses (Comments 1–21)

Comment 1: Incorrect Scaling of Continuation Values

Reviewer concern:

“Combined with $\delta \rightarrow 1$ (which makes the continuation value perturbation small relative to the stage-game payoff)”—under the normalization $(1 - \delta)u_1(\theta_t, a_1, \alpha_2) + \delta \mathbb{E}[V(h_{t+1})]$, as $\delta \rightarrow 1$ the continuation term receives weight δ and becomes more important, not less. The intended argument appears to combine filter stability with the vanishing weight of finitely many early periods.

Disposition: Fully accepted.

Response. The reviewer is correct. Under the standard normalization $(1-\delta)u_1 + \delta \mathbb{E}[V]$, the continuation value receives weight $\delta \rightarrow 1$ and therefore dominates, not vanishes. The original sentence inverted the direction. The argument in Remark 5.4 actually relies on two distinct and complementary mechanisms:

1. **Filter stability** (Proposition A.2): the θ_t -dependent perturbation $g(\theta_t, a_1, h_t)$ converges to a stationary limit exponentially fast, so for large t the perturbation is approximately state-independent.
2. **Front-loading**: $\delta \rightarrow 1$ makes the finite transient period (before the filter has converged) negligible in the normalized payoff, because finitely many periods receive weight $(1-\delta)\delta^t \rightarrow 0$ uniformly.

We have rewritten Remark 5.4 to state this two-part argument precisely, avoiding any suggestion that continuation values are downweighted by $\delta \rightarrow 1$.

Changes: Remark 5.4 rewritten. No formal results affected.

Comment 2: Incorrect Generalization of the Payoff Bound Inequality

Reviewer concern:

$V_{\text{Markov}}(s_1^) \leq V(s_1^*)$ is claimed to hold generally, and $V(s_1^*) - V_{\text{Markov}}(s_1^*)$ is interpreted as a non-negative “cost of persistence.” This is unjustified: when $\pi(G) < \mu^*$ but $F(G|G) > \mu^*$, persistence enables cooperation in good states only, yielding $V_{\text{Markov}} > V$.*

Disposition: Fully accepted.

Response. The reviewer is correct, and the counterexample is valid. When $\pi(G) < \mu^*$ (so SR defects under i.i.d. beliefs), state persistence with $F(G|G) > \mu^*$ can enable cooperation after favorable states, producing $V_{\text{Markov}} > V$. The inequality $V_{\text{Markov}} \leq V$ does not hold in general.

We have made three changes:

1. Theorem 4.8 now states the ordering is parameter-dependent, with $V_{\text{Markov}}(s_1^*) = V(s_1^*)$ if and only if the game is belief-robust.
2. A new remark characterizes when $V_{\text{Markov}} \leq V$ in terms of the relationship between $\pi(G)$, μ^* , $F(G|G)$, and $F(G|B)$.
3. The abstract and Section 1 are corrected: “cost of persistence” is renamed *effect of persistence*, since it can be positive or negative.

The possibility that $V_{\text{Markov}} > V$ is a new economic insight: persistence can *benefit* the long-run player when the stationary belief is unfavorable.

Changes: Theorem 4.8 restated; Abstract and Section 1 corrected; new remark added after Theorem 4.8. The main equilibrium guarantee $\liminf_{\delta \rightarrow 1} \underline{U}_1(\delta) \geq V_{\text{Markov}}(s_1^*)$ is unaffected.

Comment 3: Contradictory i.i.d. Benchmark Values in Worked Example

Reviewer concern:

Section 7.6 computes $V(s_1^) = 0.625$, while Section 7.7 gives a “Stationary beliefs” payoff of 0.777. Additionally, $V_{\text{Markov}} = 0.628 > V(s_1^*) = 0.625$ appears to violate $V_{\text{Markov}} \leq V(s_1^*)$. Which number is the commitment payoff?*

Disposition: Fully accepted.

Response. The reviewer correctly identifies notational overload. Two distinct quantities were both labeled “ $V(s_1^*)$ ”:

1. The worst-case commitment payoff against SR defection: $V_{\min} = \pi(G) \cdot u_1(G, A, D) + \pi(B) \cdot u_1(B, F, D) = 0.625$.
2. The equilibrium payoff when SR cooperates under stationary beliefs: $V_{\text{iid}} = 0.775$.

With the corrected double-sum formula and simultaneous-move timing, the Markov payoff is $V_{\text{Markov}} = 0.569$. Since the general inequality $V_{\text{Markov}} \leq V$ is no longer claimed (Comment 2), the apparent violation is moot; the meaningful comparison is the *effect of persistence*: $V_{\text{iid}} - V_{\text{Markov}} = 0.775 - 0.569 = 0.206$.

Changes: Sections 7.6–7.7 revised with separate notation (V_{\min} , V_{iid} , V_{Markov}). Comparison table updated with corrected values.

Comment 4: Incorrect Formulation of Markov Commitment Payoff

Reviewer concern:

Definition 4.5 pairs the belief argument $B(\cdot)$ with the same state θ that enters $u_1(\theta, \cdot, \cdot)$. If SR’s belief at period t is $F(\cdot | \theta_{t-1})$ while payoffs depend on θ_t , the formula conflates two distinct roles. The lower bound should be written in terms of the lifted stationary distribution $\tilde{\rho}(\theta_t, \theta_{t-1})$.

Disposition: Fully accepted.

Response. The reviewer correctly identifies a genuine timing error. SR’s belief at period t is $F(\cdot | \theta_{t-1})$ (transition from the previous state), while payoffs depend on θ_t (the

current state). The corrected formula is:

$$V_{\text{Markov}}(s_1^*) = \sum_{\theta' \in \Theta} \pi(\theta') \cdot \inf_{(\alpha_0, \alpha_2) \in B(s_1^*, F(\cdot|\theta'))} \sum_{\theta \in \Theta} F(\theta|\theta') \cdot u_1(\theta, s_1^*(\theta, \theta'), \alpha_2),$$

where θ' indexes θ_{t-1} (which determines SR's belief) and θ indexes θ_t (which determines the payoff). The outer sum averages over previous states using the stationary distribution π ; the inner sum averages over current states using the transition kernel $F(\cdot|\theta')$.

With the corrected formula and baseline parameters ($\alpha = 0.3, \beta = 0.5$): $V_{\text{Markov}} = 0.569$.

Changes: Definition 4.5 corrected to double-sum form. Stage timing clarified in Section 2.3 (simultaneous moves; SR uses predicted belief $F(\cdot|\theta_{t-1})$). All downstream values recomputed.

Comment 5: Invalid Sufficient Condition for Supermodularity

Reviewer concern:

“ $g(\theta_t, a_1, h_t)$ is supermodular in (θ_t, a_1) whenever V_{cont} is increasing in θ_t for each a_1 ”—monotonicity in θ_t does not imply increasing differences in (θ, a_1) , which is equivalent to supermodularity.

Disposition: Fully accepted.

Response. The reviewer is correct that monotonicity does not imply supermodularity in general. However, the conclusion is recoverable via a stronger and correct argument: in the paper's model, state transitions are exogenous (the Markov chain evolves independently of a_1), so the continuation value $g(\theta_t) = \delta V_{\text{cont}}(\theta_t)$ does not depend on a_1 . Adding a function of θ alone to a supermodular objective preserves supermodularity trivially: in the increasing-differences comparison

$$[u_1(\theta^H, a_1^H, \alpha_2) + g(\theta^H)] - [u_1(\theta^H, a_1^L, \alpha_2) + g(\theta^H)] \geq [u_1(\theta^L, a_1^H, \alpha_2) + g(\theta^L)] - [u_1(\theta^L, a_1^L, \alpha_2) + g(\theta^L)],$$

the g terms cancel on both sides, leaving the supermodularity of u_1 alone.

Changes: Remark 5.4 corrected: the incorrect claim (“monotonicity implies supermodularity”) is replaced with the correct argument (exogenous transitions $\Rightarrow g$ is a_1 -independent \Rightarrow supermodularity preserved).

Comment 6: Contradiction in One-Shot Deviation Argument

Reviewer concern:

“The one-shot deviation argument is identical” directly contradicts the subsequent observation that “adding this θ_t -dependent term can in principle change

the OT solution.” Two distinct issues are conflated: (i) the OT/cyclical-monotonicity structure is the same for any fixed objective w ; (ii) the Markov objective $w = u_1 + \delta g$ differs from u_1 .

Disposition: Fully accepted.

Response. The reviewer’s diagnosis is correct. The two statements are not logically contradictory but are misleadingly juxtaposed. We have restructured Section 5.2 to clearly separate:

1. **Proof structure:** For any fixed one-shot objective $w(\tilde{\theta}, a_1)$, the optimal transport / cyclical-monotonicity argument from Luo–Wolitzky’s Lemma 1 carries over identically to the lifted space $\tilde{\Theta}$.
2. **Objective change:** In the Markov case, the relevant objective is $w(\tilde{\theta}, a_1) = u_1(\tilde{\theta}, a_1, \alpha_2) + \delta g(\theta_t, a_1, h_t)$, which differs from u_1 alone. Confound-defeating with respect to u_1 does not automatically ensure optimality of s_1^* for w .
3. **Resolution:** Either (a) supermodularity + belief-robustness, (b) a strengthened confound-defeating condition, or (c) the filter-stability + $\delta \rightarrow 1$ continuity argument from Remark 5.4 ensures that s_1^* remains optimal.

Changes: Section 5.2 restructured to disentangle structural equivalence from objective change.

Comment 7: Conceptual Error Regarding Filter Stability and Belief Convergence

Reviewer concern:

The suggestion that ε -perturbed strategies cause beliefs to “converge to the stationary distribution π ” rests on a misunderstanding. Filter stability means the initial prior is forgotten, not that the posterior converges to π . For small ε , signals remain highly informative and the posterior tracks θ_t .

Disposition: Fully accepted.

Response. The reviewer is correct. The original text conflated three distinct properties:

1. **Chain mixing:** $\|P(\theta_t \in \cdot \mid \theta_0) - \pi\| \leq C'\lambda^t$ —the initial *state* θ_0 is forgotten.
2. **Filter stability:** $\sup_{\pi_0, \pi'_0} \|\pi_t - \pi'_t\| \leq C\lambda^t$ —the initial *prior* π_0 is forgotten.
3. **Belief convergence to π :** $\|\pi_t(\cdot \mid y_0, \dots, y_t) - \pi\| \rightarrow 0$ —*not* implied by filter stability.

For ε -perturbed strategies with small ε , the observation channel retains Fisher information of order $O(1/\varepsilon)$ about θ_t , so the filtered posterior tracks the current state rather than settling at π .

The benefit of ε -perturbation is more subtle than convergence to π : by degrading state revelation, it reduces the effective belief gap, potentially moving filtering beliefs out of the critical interval $[0.5, 1 - 0.3]$.

Changes: Section 10.2 rewritten to correctly characterize filter stability as prior forgetting, not belief convergence to π . Added footnote distinguishing the three concepts.

Comment 8: Arithmetic Error in Belief-Robust Condition

Reviewer concern:

“ $\mu^ = 0.60 < \beta = 0.5$ ” is stated in Section 7.4 (and Remark 3.4), but $0.60 > 0.5$. With baseline calibration $(\alpha, \beta) = (0.3, 0.5)$, the game is then not belief-robust by Proposition 3.3.*

Disposition: Fully accepted.

Response. This is an unambiguous arithmetic error. The belief-robust condition requires $\mu^* \leq \min_{\theta} F(G|\theta) = \beta = 0.50$, so $\mu^* = 0.60$ violates it. We have corrected the belief-robust threshold in the “Version 1: Belief-Robust” example to $\mu^* = 0.40 < \beta = 0.50$, with the resulting payoff bound $V(s_1^*) = 0.625$.

The non-belief-robust case ($\mu^* = 0.60$) in Section 7.5 is unaffected, since there the threshold exceeds $\beta = 0.50$ as intended.

Changes: Section 7.4 and Remark 3.4 corrected: $\mu^* = 0.40$, $V(s_1^*) = 0.625$. Updated in both the text and `stats.tex`.

Comment 9: Arithmetic Error and Variable Confusion

Reviewer concern:

“The cost equals $0.777 - 0.628 = 0.094$, representing 23.7% of the i.i.d. payoff”—but $0.777 - 0.628 = 0.149 \neq 0.094$; the value 0.094 is the belief gap from Equation (7); and $23.7\% = 0.149/0.628$ is relative to the Markov payoff, not “the i.i.d. payoff.”

Disposition: Fully accepted.

Response. Three errors in one sentence:

1. The subtraction $0.777 - 0.628 = 0.149$, not 0.094. The value 0.094 is the *belief gap* from $2\alpha\beta|1 - \alpha - \beta|/(\alpha + \beta)^2$, a probability-space quantity that was inadvertently substituted for the payoff gap.
2. The percentage $23.7\% = 0.149/0.628$ measures overestimation relative to the Markov payoff, not “of the i.i.d. payoff” ($0.149/0.777 = 19.2\%$).

3. The macro `\PayoffGapAbsolute` stored the belief gap where the payoff gap was needed.

With the corrected double-sum formula, the updated values are: $V_{\text{iid}} = 0.775$, $V_{\text{Markov}} = 0.569$, effect of persistence = 0.206, overestimation = 36.3%. The text now explicitly distinguishes the belief gap (0.094) from the payoff gap (0.206).

Changes: Section 8.5 corrected. Separate macros introduced: `\PayoffGapPayoff` (payoff gap) and `\BeliefGapBaseline` (belief gap). Terminology changed from “cost of persistence” to “effect of persistence.”

Comment 10: Missing Content in Appendix A.3

Reviewer concern:

“A.3 Monte Carlo Verification” appears to be empty (header only). The main text (Section 5.4) relies on this appendix to verify the KL counting bound.

Disposition: Fully accepted.

Response. We thank the reviewer for identifying this omission. Appendix A.3 contained only a figure reference without explanatory text. We have added a complete description of the Monte Carlo verification procedure:

- **Setup:** Two parallel processes—a Markov chain with parameters ($\alpha = 0.3$, $\beta = 0.5$) and an i.i.d. process with the same stationary distribution $\pi(G) = 0.625$ —are simulated for $T = 5,000$ periods across $N = 500$ independent runs.
- **Methodology:** For each run, Bayesian posteriors are computed and the number of *distinguishing periods*—where the predicted signal distributions under commitment and rational type hypotheses differ by more than η in total variation—is recorded.
- **Finding:** The empirical mean distinguishing-period count is 8.1 (i.i.d.) and 12.7 (Markov), both far below the theoretical bound $\bar{T}(\eta, \mu_0) = 921$, confirming that the KL counting bound requires no mixing-time correction.

Changes: Full methodology text added to Appendix A.3, preceding Figure 6.

Comment 11: Incorrect Inference from Filter Stability

Reviewer concern:

Section 10.2 states “filter stability (SA4) suggests that beliefs may converge to the stationary distribution”—this conflates filter stability (forgetting of initial prior) with convergence of the posterior to π . For small ε , signals remain very informative about θ_t and the belief process continues to track the state.

Disposition: Fully accepted.

Response. This comment identifies the same conceptual error as Comment 7, appearing at a different location in the text. As discussed there, filter stability (SA4) guarantees that the initial prior π_0 is forgotten exponentially fast; it does *not* imply convergence of the filtering distribution to π . For small ε , the posterior continues to fluctuate with the state.

We have corrected the language in Section 10.2 to specify that what filter stability provides is prior-forgetting, and that the relevant open question concerns the *stationary distribution of beliefs* (as a stochastic process indexed by ε), not convergence to a point mass at π .

Changes: Section 10.2 rewritten (jointly with Comment 7). Added explicit distinction between prior-forgetting and posterior convergence.

Comment 12: Contradiction Between Figure 6 and Its Caption

Reviewer concern:

The caption describes empirical distributions as “nearly identical” across Markov and i.i.d. simulations, while Figure 6 shows sample means of 8.1 (i.i.d.) and 12.7 (Markov) with visibly shifted histograms.

Disposition: Fully accepted.

Response. The 57% difference in means (8.1 vs. 12.7) is not “nearly identical.” The important point is that both means are far below the analytical bound of 921 ($< 1.4\%$ of the bound), confirming that the KL counting bound holds with large margin in both settings.

Changes: Figure 6 caption revised to: “both well below the analytical bound (i.i.d.: 8.1; Markov: 12.7).” The phrase “nearly identical” is replaced with accurate comparative language noting the moderate difference while emphasizing that the bound holds in both settings.

Comment 13: Confusion of Reputation Dynamics with State-Belief Dynamics

Reviewer concern:

Section 9.2 writes $B(s_1^, \mu_0(h_t))$ and says the one-shot deviation argument fails “since μ_0 changes.” But μ_0 is the fixed prior over types; posteriors are μ_t . The genuinely new phenomenon is that B becomes state-dependent via $F(\cdot|\theta_t)$, not that type-beliefs evolve (which already occurs in the i.i.d. case).*

Disposition: Fully accepted.

Response. The reviewer correctly identifies two errors:

1. **Notational error:** μ_0 is the fixed prior; the evolving posterior is $\mu_t(\cdot|h_t)$.
2. **Misidentification of the novel phenomenon:** Type-posterior evolution (μ_t changing with history) occurs in the i.i.d. case as well—it *is* the standard reputation dynamics mechanism. What is genuinely new under Markov states is that the Nash correspondence becomes *state-contingent*: $B(s_1^*, F(\cdot|\theta_t))$ varies with the realized state θ_t . This is precisely what the belief-robustness condition (Definition 3.2) addresses.

Quantitatively: for baseline parameters with $\mu^* = 0.60$, the i.i.d. SR always cooperates (since $\pi(G) = 0.625 > 0.60$), while the Markov SR cooperates in state G ($F(G|G) = 0.70 > 0.60$) but defects in state B ($F(G|B) = 0.50 < 0.60$).

Changes: Section 9.2 corrected: $\mu_0(h_t) \rightarrow \mu_t(\cdot|h_t)$; reframed to identify the state-contingent Nash correspondence as the genuinely novel phenomenon. Added cross-reference to Definition 3.2.

Comment 14: Ambiguous Definition of Best-Response Set in Lemma 5.8

Reviewer concern:

$\hat{B}_{\xi(\eta)}(s_1^)$ appears in Lemma 5.8 without redefinition in the Markov extension. The surrounding discussion emphasizes state-dependent correspondences $B(s_1^*, F(\cdot|\theta))$, but Lemma 5.8 uses the static confirmed best-response set. This creates momentary confusion.*

Disposition: Fully accepted.

Response. We agree that the notation requires disambiguation. We have added Remark 5.6 before Lemma 5.8, clarifying:

- $\hat{B}_{\xi}(s_1^*)$ is the ξ -confirmed best-response set from Luo–Wolitzky (2024, Definition 3): the set of SR best-response profiles that remain optimal when the posterior concentrates within ξ of $\{\omega^R, \omega_{s_1^*}\}$. This is a static, type-based object.
- $B(s_1^*, F(\cdot|\theta))$ is the *state-contingent Nash correspondence*: SR behavior conditional on the revealed state θ . This enters only in the payoff bound step (Section 5.7).
- Under belief-robustness, $B(s_1^*, F(\cdot|\theta))$ is constant across θ and the distinction is immaterial.

We have also made explicit in Section 5.7 how the confirmed best-response property from Lemma 5.8 underpins the lower bounds formulated in terms of the state-contingent sets.

Changes: New Remark 5.6 added. Section 5.7 expanded with explicit logical link.

Comment 15: Incorrect Claim About Confound-Defeating Conditions

Reviewer concern:

Section 8.6 states “Persistence thus strengthens identification, making confound-defeating conditions easier to satisfy in the supermodular case.” But Section 6 shows the confound-defeating characterization is unchanged. In what sense is it “easier to satisfy”?

Disposition: Fully accepted.

Response. The mathematical characterization (Proposition 6.1) is indeed unchanged: confound-defeating is equivalent to monotonicity in the supermodular case, regardless of whether the state is i.i.d. or Markov. What persistence improves is the *empirical verifiability*: autocorrelated action sequences are more statistically distinguishable, facilitating verification of the condition from data. The mathematical condition itself neither strengthens nor weakens.

Changes: Section 8.6 revised: “easier to satisfy” replaced with “easier to verify empirically,” with explicit reference to Proposition 6.1 confirming the formal characterization is unchanged.

Comment 16: Ambiguity in “Not Behaviorally Confounded” Application

Reviewer concern:

“Since s_1^ is not behaviorally confounded, any type with the same asymptotic signal distribution must be s_1^* itself” (Part A of Lemma 5.7) relies on an implicit identification between asymptotic conditional signal distributions and the stationary $p(\alpha_0, s_1, \alpha_2)$. This needs to be made explicit under Markov dynamics.*

Disposition: Fully accepted.

Response. The reviewer correctly identifies an implicit step that is automatic in the i.i.d. case but requires justification under Markov dynamics. We have expanded Part A of Lemma 5.7 to make the logical chain explicit:

1. **KL bound** (Lemma 5.6) implies per-period signal convergence (process-independent).
2. **Filter stability** (Proposition A.2) ensures the filtering distribution converges exponentially fast to a limit determined by the strategy and transition kernel, not the initial prior.

3. **Ergodicity** identifies this limit as the unique stationary signal distribution $p_\infty(\alpha_0, s_1, \alpha_2)$.
4. The **not-behaviorally-confounded condition** (Definition 4.3), applied to these stationary distributions, yields $\mu_\infty(\{\omega^R, \omega_{s_1^*}\} \mid h) = 1$.

Changes: Part A of Lemma 5.7 (Section 5.3) expanded with the four-step argument.

Comment 17: Confusing Motivation for Lifted State

Reviewer concern:

Section 1.1 says the lifted state $\tilde{\theta}_t = (\theta_t, \theta_{t-1})$ “provides a stationary distribution $\tilde{\rho}$ on the expanded space and allows the optimal transport framework to apply directly.” This reads as if the original chain lacked a stationary distribution, whereas θ_t already has π by Assumption 2.1.

Disposition: Fully accepted.

Response. The phrasing was misleading. The original chain θ_t possesses a unique stationary distribution π by Assumption 1(b). The true role of the lifting is to construct a “type space” $\tilde{\Theta} = \Theta \times \Theta$ that:

- encodes Markov private information (the transition $\theta_{t-1} \rightarrow \theta_t$) into the signal structure,
- plays the role of the exogenous type distribution in the Luo–Wolitzky optimal transport formulation, with marginal $\tilde{\rho}$.

Changes: Section 1.1 revised to explain the lifting as constructing a suitable type space for Markov private information, not as “providing stationarity.”

Comment 18: Failure of Strict Supermodularity on Lifted Space

Reviewer concern:

$u_1(\tilde{\theta}, a_1, \alpha_2) = u_1(\theta_t, a_1, \alpha_2)$ is constant in θ_{t-1} , suggesting tension with the strict supermodularity invoked in Proposition 6.1 for uniqueness of the OT solution.

Disposition: Fully accepted.

Response. As the reviewer notes upon closer reading, this is resolved by the first-coordinate order on $\tilde{\Theta}$:

$$(\theta_t, \theta_{t-1}) \succeq (\theta'_t, \theta'_{t-1}) \iff \theta_t \succeq \theta'_t.$$

States differing only in θ_{t-1} are not strictly ordered, so the strict increasing-differences condition is never evaluated between such pairs. Strict supermodularity in $(\tilde{\theta}, a_1)$ is

therefore equivalent to strict supermodularity in (θ_t, a_1) , provided the underlying payoff has strictly increasing differences in (θ_t, a_1) .

Since the computational survey in Section 6.2 reports that 4 out of 24 orderings of $\tilde{\Theta}$ preserve supermodularity, and all four correspond to the first-coordinate order, this is consistent.

Changes: Added an explicit sentence to Section 6.2 stating that for θ_t -only payoffs, strict supermodularity on $\tilde{\Theta} \times A_1$ is preserved under the first-coordinate order and coincides with the usual condition on (θ_t, a_1) .

Comment 19: Imprecise Terminology Regarding Filter Stability

Reviewer concern:

“This ensures that the initial condition of the Markov chain is ‘forgotten’ exponentially fast” reads as if referring to θ_0 (chain mixing), whereas Proposition A.2 is about the filter’s dependence on the initial prior π_0 .

Disposition: Fully accepted.

Response. The phrasing is indeed ambiguous. Proposition A.2 concerns filter stability—the filter’s exponential forgetting of its initial prior π_0 —not chain mixing (the Markov chain’s forgetting of its initial state θ_0). Both properties hold but are distinct:

- **Chain mixing:** $P(\theta_t = G \mid \theta_0) \rightarrow \pi(G) = 0.625$ regardless of θ_0 , at rate $|1 - \alpha - \beta|^t = 0.2^t$.
- **Filter stability:** Two filters with extreme priors ($\pi_0(G) = 0.99$ vs. 0.01) converge to agreement exponentially fast, regardless of the true initial state.

Changes: Appendix A.2: “initial condition of the Markov chain is ‘forgotten’” replaced with “initial prior of the filter is forgotten.” Added parenthetical distinguishing filter stability from chain mixing.

Comment 20: Missing Discussion of Stackelberg Well-Definedness

Reviewer concern:

Section 9.5 states that Stackelberg well-definedness for persuasion games is “acknowledged as an open question in Section 10,” but Section 10.2 does not mention it.

Disposition: Fully accepted.

Response. The reviewer has identified a genuine cross-reference error. Section 9.5 claims the topic is discussed in Section 10, but Section 10.2 omitted it. We have added a new paragraph to Section 10.2:

Stackelberg well-definedness in persuasion games. For persuasion games, the Stackelberg strategy is defined via concavification of the sender's value function. Under Markov dynamics, the receiver's prior varies state-by-state through $F(\cdot|\theta)$, and the optimal persuasion scheme may differ for each prior. Whether a single state-independent Stackelberg strategy exists that is simultaneously optimal across all filtering priors remains open.

Changes: New paragraph added to Section 10.2. Cross-reference in Section 9.5 now correctly points to this paragraph.

Comment 21: Inconsistent Payoffs and Stackelberg Strategy

Reviewer concern:

Only the D-row payoffs are displayed ($u_1(G, A) = 1$, $u_1(G, F) = x$, $u_1(B, A) = y$, $u_1(B, F) = 0$). Looking at these alone, A dominates F in state B , making $s_1^(B) = F$ appear suboptimal. The full deterrence specification (including the C-row) is needed.*

Disposition: Fully accepted.

Response. The reviewer is correct that the D -conditional payoffs alone make $s_1^*(B) = F$ appear suboptimal (since $u_1(B, A, D) = y > 0 = u_1(B, F, D)$). The Stackelberg strategy is optimal in the full game because F in state B enables reputation effects: it satisfies the confound-defeating condition and induces SR cooperation in equilibrium. The missing C -column payoffs are essential for understanding this.

We have added the full payoff matrix to Section 7.1, displaying both the C and D columns:

	$a_2 = C$		$a_2 = D$	
	$\theta = G$	$\theta = B$	$\theta = G$	$\theta = B$
$a_1 = A$	g	g	1	y
$a_1 = F$	g	l	x	0

with $g > 1 > x > 0 > l$ and $y \in (0, 1)$, clarifying that s_1^* maximizes the reputation payoff (incorporating SR's equilibrium response), not the per-period payoff against defection.

Changes: Full payoff matrix (both C and D columns) added to Section 7.1 with a clarifying note.

Summary Table

#	Issue	Status	Location of Change
C1	Incorrect scaling of continuation values	Accepted	Remark 5.4 rewritten
C2	$V_{\text{Markov}} \leq V$ claimed without proof	Accepted	Theorem 4.8; Abstract; Section 1
C3	Contradictory i.i.d. benchmark values	Accepted	Sections 7.6–7.7; comparison table
C4	Incorrect V_{Markov} formula (timing)	Accepted	Definition 4.5 (double-sum); Section 2.3
C5	Invalid supermodularity condition	Accepted	Remark 5.4 corrected
C6	Contradiction in one-shot deviation argument	Accepted	Section 5.2 restructured
C7	Conceptual error re filter stability	Accepted	Section 10.2 rewritten
C8	Arithmetic error: $\mu^* = 0.60 < 0.5$	Accepted	Section 7.4; Remark 3.4; stats.tex
C9	Arithmetic error & variable confusion	Accepted	Section 8.5; new macros
C10	Missing Appendix A.3 content	Accepted	Appendix A.3 methodology added
C11	Incorrect inference from filter stability	Accepted	Section 10.2 (joint with C7)
C12	“Nearly identical” with different means	Accepted	Figure 6 caption revised
C13	Reputation vs. state-belief confusion	Accepted	Section 9.2 corrected
C14	Ambiguous \hat{B}_ξ in Lemma 5.8	Accepted	New Remark 5.6; Section 5.7
C15	“Easier to satisfy” confound-defeating	Accepted	Section 8.6 revised
C16	Implicit step in Lemma 5.7 Part A	Accepted	Lemma 5.7 expanded (4-step argument)
C17	Lifted state “provides stationarity”	Accepted	Section 1.1 rephrased
C18	Strict supermodularity on $\tilde{\Theta}$	Accepted	Section 6.2 explicit sentence added
C19	Filter stability terminology	Accepted	Appendix A.2 corrected
C20	Missing Stackelberg/persuasion open question	Accepted	Section 10.2 new paragraph
C21	Incomplete payoff display	Accepted	Section 7.1 full matrix added

Corrected Key Values:

Quantity	Corrected Value
Stationary/i.i.d. payoff V_{iid}	0.775
Markov payoff V_{Markov}	0.569
Effect of persistence $V_{\text{iid}} - V_{\text{Markov}}$	0.206
Overestimation	36.3%
Belief-robust threshold μ^*	0.40
Belief-robust payoff $V(s_1^*)$	0.625
SR cooperation threshold	0.60
Belief gap	0.094