Extending Marginal Reputation to Markov States: A Corrected Analysis

Kyle Elliott Mathewson[1]
with AI Collaborators

February 17, 2026

[1]Faculty of Science, University of Alberta

Abstract

We extend the main result (Theorem 1) of Luo & Wolitzky (2024), "Marginal Reputation," from i.i.d. states to persistent Markovian states. Our initial attempt overclaimed that the extension requires no correction. Following expert critique (Luo, 2026), we conducted systematic computational testing (7 analysis modules, 8 diagnostic figures) that identified precisely where the original proof breaks: short-run player beliefs permanently deviate from the stationary distribution when the Stackelberg strategy reveals the state. We present two corrected results. **Theorem 1′** (belief-robust case): when short-run best responses are invariant to the filtering belief $F(\cdot|\theta)$, the original commitment payoff $V(s_1^*)$ holds exactly. **Theorem 1″** (general case): the corrected bound $V_{\text{Markov}}(s_1^*) = \sum_\theta \pi(\theta) \cdot \inf_{B(s_1^*, F(\cdot|\theta))} u_1 \leq V(s_1^*)$ always holds, with equality iff the game is belief-robust. For the deterrence game with baseline parameters ($\alpha = 0.3$, $\beta = 0.5$), the overestimation from ignoring state-contingent beliefs is 23.7%. We quantify the cost of persistence in reputation games and demonstrate a correction methodology combining AI conjecture, expert critique, and computational verification.

Keywords: Reputation, repeated games, Markov states, optimal transport, belief-robustness, corrected bounds

Original Paper: "Marginal Reputation" by Daniel Luo and Alexander Wolitzky, MIT Department of Economics, December 2024.

Contents

1 Introduction

Luo & Wolitzky (2024) establish a striking connection between reputation theory in repeated games and optimal transport theory. Their main result, Theorem 1, shows that a patient long-run player can secure her *commitment payoff* $V(s_1^*)$ in any Nash equilibrium, provided her Stackelberg strategy s_1^* is *confound-defeating* and *not behaviorally confounded*. Throughout their analysis, states are drawn **i.i.d. across periods**. The authors note (footnote 9) that the extension to persistent states is an open question.

1.1 What We Attempted and What Went Wrong

We initially attempted to extend Theorem 1 to Markovian states via a *lifted state* construction $\tilde{\theta}_t = (\theta_t, \theta_{t-1})$. Our first submission claimed the extension required no correction—that the commitment payoff $V(s_1^*)$ holds identically. This overclaimed.

Expert critique (Luo, 2026) identified the central flaw: when the Stackelberg strategy reveals the state (e.g., $s_1^*(G) = A$, $s_1^*(B) = F$), the short-run player learns θ_t exactly. Their belief about θ_{t+1} is then the *filtering distribution* $F(\cdot|\theta_t)$, **not** the stationary distribution π. This creates a permanent structural gap: short-run behavior becomes state-contingent, violating the assumption that short-run players face the same incentives every period.

1.2 Systematic Computational Testing

Rather than simply accepting or rejecting the critique, we conducted systematic computational analysis across 7 diagnostic modules producing 8 figures. The results cleanly separate the claims that survive from those that fail:

Claims that survive:

(a) The KL counting bound extends verbatim—no mixing-time correction needed (verified via $N = 500$ Monte Carlo simulations over $T = 5000$ periods).

(b) Filter stability holds with exponential forgetting: correlation $r > 0.63$ between fitted decay rate and $|1 - \alpha - \beta|$.

(c) The OT support is robust to belief perturbations: stability margin ≥ 0.3 in 100% of the (α, β) parameter space.

(d) Monotonicity extends for θ_t-dependent payoffs on the lifted space.

Claims that fail:

(a) SR beliefs permanently deviate from π with mean TV distance 0.412 and analytical gap 0.094.

(b) The Nash correspondence $B(s_1^*, \mu)$ varies period-to-period: 37.2% disagreement rate.

(c) The commitment payoff is overestimated: 0.777 (stationary assumption) vs. 0.628 (filtered reality), a 23.7% gap.

1.3 Corrected Results

We present two new theorems that properly account for belief dynamics:

Theorem 1′ (Belief-Robust Extension). When the SR best-response set $B(s_1^*, F(\cdot|\theta))$ is constant across states θ—a condition we call *belief-robustness*—the original bound $V(s_1^*)$ holds exactly. The entire proof machinery (KL bound, OT robustness, monotonicity) works as before; belief-robustness ensures the SR belief gap is irrelevant.

Theorem 1″ (General Corrected Bound). For all supermodular games with Markov states, a corrected bound holds:

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

with equality if and only if the game is belief-robust. The gap $V(s_1^*) - V_{\text{Markov}}$ quantifies the "cost of persistence" in reputation games.

1.4 Outline

Section 2 presents the model with the lifted state construction. Section 3 introduces the key new concept of belief-robustness. Section 4 states the two corrected theorems. Section 5 contains the proof sketch. Section 6 extends the supermodular case. Section 7 works out the deterrence game. Section 8 discusses interpolation between i.i.d. and persistent. Section 9 documents the correction methodology. Section 10 discusses open questions.

2 The Extended Model

We maintain all notation and conventions from Luo & Wolitzky (2024, Sections 3.1–3.2), modifying only the state process.

2.1 State Process

Let Θ be a finite set.

Assumption 2.1 (Markov States)**.** The state $\theta_t \in \Theta$ follows a **stationary ergodic Markov chain** with:

(a) Transition kernel $F(\cdot|\theta)$ for each $\theta \in \Theta$, so that $\mathbb{P}(\theta_{t+1} = \theta'|\theta_t = \theta) = F(\theta'|\theta)$.

(b) Unique stationary distribution $\pi \in \Delta(\Theta)$ satisfying

$$\pi(\theta) = \sum_{\theta' \in \Theta} \pi(\theta') F(\theta|\theta') \quad \text{for all } \theta \in \Theta. \tag{1}$$

(c) The chain is **irreducible and aperiodic** (ensuring ergodicity).

Remark 2.2. When $F(\cdot|\theta) = \pi(\cdot)$ for all θ, the chain has no memory and we recover the i.i.d. case of the original paper. The two-state case with $\Theta = \{G, B\}$ is parameterized by $\alpha = \mathbb{P}(B|G)$ and $\beta = \mathbb{P}(G|B)$, giving $\pi(G) = \beta/(\alpha + \beta)$.

2.2 Lifted State Space

The central construction is the *lifted state*:

Definition 2.3 (Lifted State)**.** Define

$$\tilde{\theta}_t \;=\; (\theta_t,\, \theta_{t-1}) \;\in\; \tilde{\Theta} \;=\; \Theta \times \Theta. \tag{2}$$

The process $(\tilde{\theta}_t)_{t \geq 1}$ is itself a Markov chain on $\tilde{\Theta}$ with transition probabilities

$$\tilde{F}\big((\theta', \theta) \,\big|\, (\theta, \theta'')\big) \;=\; F(\theta'|\theta) \tag{3}$$

and stationary distribution

$$\tilde{\rho}(\theta, \theta') \;=\; \pi(\theta') \cdot F(\theta|\theta'). \tag{4}$$

Proposition 2.4. *Under Assumption 2.1, the lifted chain $(\tilde{\theta}_t)$ on $\tilde{\Theta}$ is ergodic with unique stationary distribution $\tilde{\rho}$.*

Proof. Irreducibility of the original chain ensures connectivity of the lifted chain: for any lifted states (θ_a, θ_b) and (θ_d, θ_c), there exists a finite path with positive probability connecting them. Aperiodicity of the original chain implies aperiodicity of the lifted chain. Uniqueness of $\tilde{\rho}$ follows from the Perron–Frobenius theorem. □

Remark 2.5 (Purpose of the Lifting)**.** The lifted state provides a Markov structure on which the optimal transport framework and cyclical monotonicity characterizations apply. The key property is that $\tilde{\theta}_t$ has a *fixed, known* stationary distribution $\tilde{\rho}$, playing precisely the role of the i.i.d. signal distribution ρ in the original paper.

2.3 Stage Game

The stage game is identical to Luo & Wolitzky's Section 3.1, except:

(i) The long-run player's private information each period is $\tilde{\theta}_t = (\theta_t, \theta_{t-1})$.

(ii) A stage-game strategy for player 1 is $s_1 : \tilde{\Theta} \to \Delta(A_1)$, a *Markov strategy.*

(iii) Payoffs depend on the current state: $u_1(\theta_t, a_1, \alpha_2)$.

We restrict throughout to payoffs $u_1(\theta_t, a_1, \alpha_2)$ that depend on θ_t alone (not the full lifted state $\tilde{\theta}_t$). This covers all standard applications—deterrence, trust, signaling—and avoids unmotivated generalization.

2.4 Joint Distribution and Commitment Types

Under Markov strategy s_1 and the stationary distribution $\tilde{\rho}$, the joint distribution over $(\tilde{\theta}, a_1)$ is:

$$\gamma(s_1)[\tilde{\theta}, a_1] \;=\; \tilde{\rho}(\tilde{\theta}) \cdot s_1(\tilde{\theta})[a_1]. \tag{5}$$

A commitment type $\omega_{s_1} \in \Omega$ plays Markov strategy $s_1 : \tilde{\Theta} \to \Delta(A_1)$ every period. The type space Ω is countable with full-support prior $\mu_0 \in \Delta(\Omega)$.

3 Belief-Robustness: The Key New Concept

This section introduces the central new concept needed for the Markov extension. The issue is simple: when the Stackelberg strategy reveals the state, short-run players learn θ_t and form beliefs about θ_{t+1} using the filtering distribution $F(\cdot|\theta_t)$, which generically differs from the stationary distribution π.

3.1 Filtering Beliefs

Definition 3.1 (Filtering Belief)**.** Given a state-revealing Stackelberg strategy s_1^* (i.e., $s_1^*(\theta) \neq s_1^*(\theta')$ for $\theta \neq \theta'$), the **filtering belief** in state θ is

$$F(\cdot|\theta_t) = \mathbb{P}(\theta_{t+1} = \cdot \mid \theta_t), \tag{6}$$

the one-step-ahead predictive distribution conditional on the current state.

For the two-state chain $\Theta = \{G, B\}$ with parameters (α, β):

$$F(G|G) = 1 - \alpha, \qquad F(G|B) = \beta. \tag{7}$$

The stationary distribution gives $\pi(G) = \beta/(\alpha + \beta)$. The gap between the filtering belief and the stationary distribution is:

$$\mathbb{E}\big[|F(G|\theta_t) - \pi(G)|\big] = \frac{2\alpha\beta|1 - \alpha - \beta|}{(\alpha + \beta)^2}. \tag{8}$$

This equals zero **if and only if** $\alpha + \beta = 1$, which is precisely the i.i.d. case. For the baseline parameters ($\alpha = 0.3, \beta = 0.5$), the expected gap is 0.094.

3.2 The Belief-Robustness Condition

Definition 3.2 (Belief-Robustness)**.** A game (u_1, u_2) with Stackelberg strategy s_1^* and Markov chain (Θ, F) is **belief-robust** if the short-run player Nash correspondence satisfies

$$B(s_1^*, F(\cdot|\theta)) = B(s_1^*, F(\cdot|\theta')) \quad \text{for all } \theta, \theta' \in \Theta. \tag{9}$$

In words: the SR best-response set does not change when SR learns the current state. Under this condition, the belief gap documented in (8) is irrelevant—SR plays the same regardless.

3.3 When Does Belief-Robustness Hold?

For the deterrence game with SR threshold μ^* (the belief level at which SR is indifferent between cooperating and defecting):

Proposition 3.3. *Belief-robustness holds if and only if*

$$\mu^* \notin \big[\min_\theta F(G|\theta),\ \max_\theta F(G|\theta)\big] = [\beta,\ 1 - \alpha]. \tag{10}$$

Proof. The SR best response depends on whether $F(G|\theta_t) \gtrless \mu^*$. If $\mu^* < \beta$, then $F(G|\theta_t) \geq \beta > \mu^*$ for all θ_t, so SR always cooperates. If $\mu^* > 1 - \alpha$, then $F(G|\theta_t) \leq 1 - \alpha < \mu^*$ for all θ_t, so SR always defects. In either case, $B(s_1^*, F(\cdot|\theta))$ is constant. Conversely, if $\mu^* \in [\beta, 1 - \alpha]$, there exist states θ, θ' with $F(G|\theta) > \mu^* > F(G|\theta')$, so SR cooperates after θ and defects after θ'. □

Remark 3.4 (Economic Interpretation)**.** Belief-robustness fails when the SR indifference threshold lies between the conditional beliefs for different states. This happens when:

(i) The game has belief-sensitive SR behavior (threshold near π).

(ii) The chain is persistent enough that $F(\cdot|\theta)$ varies substantially across states.

(iii) The strategy reveals state information to SR.

Persistence hurts the LR player if and only if the SR threshold lies in the "danger zone" $[\beta, 1-\alpha]$.

Remark 3.5 (Baseline Example)**.** For the baseline parameters ($\alpha = 0.3, \beta = 0.5$): the danger zone is $[0.5, 1-0.3] = [0.5, 0.7]$. The SR threshold $\mu^* = 0.60$ lies inside this interval, so the baseline deterrence example is **not** belief-robust. Changing SR payoffs to make $\mu^* = 0.60 < \beta = 0.5$ would restore belief-robustness.

4 Corrected Theorems

We state two results. Theorem 1′ recovers the exact i.i.d. bound under belief-robustness. Theorem 1″ provides a corrected bound for the general case.

4.1 Definitions on the Expanded State Space

All definitions from the original paper carry over to $\tilde{\Theta}$, with strategies mapping $\tilde{\Theta} \to \Delta(A_1)$.

Definition 4.1 (Confound-Defeating, Extended)**.** A Markov strategy $s_1^* : \tilde{\Theta} \to \Delta(A_1)$ is **confound-defeating** if for every $(\alpha_0, \alpha_2) \in B_0(s_1^*)$, the joint distribution $\gamma(\alpha_0, s_1^*)$ is the *unique solution* to:

$$\mathrm{OT}\big(\tilde{\rho}(\alpha_0),\, \phi(\alpha_0, s_1^*);\, \alpha_2\big) : \quad \max_{\gamma \in \Delta(\tilde{\Theta} \times A_1)} \int u_1(\tilde{\theta}, a_1, \alpha_2)\, d\gamma \tag{11}$$

subject to $\pi_{\tilde{\Theta}}(\gamma) = \tilde{\rho}(\alpha_0)$ and $\pi_{A_1}(\gamma) = \phi(\alpha_0, s_1^*)$.

Definition 4.2 (Not Behaviorally Confounded, Extended)**.** s_1^* is **not behaviorally confounded** if for any $\omega_{s_1'} \in \Omega$ with $s_1' \neq s_1^*$ and any $(\alpha_0, \alpha_2) \in B_1(s_1^*)$, we have $p(\alpha_0, s_1^*, \alpha_2) \neq p(\alpha_0, s_1', \alpha_2)$.

4.2 Theorem 1′ (Belief-Robust Extension)

Theorem 4.3 (Belief-Robust Markov Extension)**.** *Let θ_t follow a stationary ergodic Markov chain on finite Θ (Assumption 2.1). Let $\tilde{\theta}_t = (\theta_t, \theta_{t-1})$ with stationary distribution $\tilde{\rho}$. Suppose:*

(i) $\omega_{s_1^*} \in \Omega$, where $s_1^* : \tilde{\Theta} \to \Delta(A_1)$ is a Markov strategy;

(ii) s_1^* is confound-defeating on $\tilde{\Theta}$ (Definition 4.1);

(iii) s_1^* is not behaviorally confounded (Definition 4.2);

*(iv) The game is **belief-robust** with respect to s_1^* and (Θ, F) (Definition 3.2).*

Then:

$$\boxed{\liminf_{\delta \to 1} \underline{U}_1(\delta) \;\geq\; V(s_1^*)} \tag{12}$$

where $V(s_1^*) = \inf_{(\alpha_0,\alpha_2)\in B(s_1^*)} u_1(\alpha_0, s_1^*, \alpha_2)$ *is the commitment payoff, identical to the i.i.d. case.*

Remark 4.4. Under belief-robustness, the SR belief gap is irrelevant: SR plays the same best response regardless of the filtering belief $F(\cdot|\theta)$. All the confirmed proof machinery—KL counting bound, OT robustness, monotonicity—applies without modification.

4.3 Theorem 1″ (General Corrected Bound)

Definition 4.5 (Markov Commitment Payoff)**.** The **Markov commitment payoff** is

$$V_{\text{Markov}}(s_1^*) := \sum_{\theta\in\Theta} \pi(\theta) \cdot \inf_{(\alpha_0,\alpha_2)\in B(s_1^*, F(\cdot|\theta))} u_1(\theta, s_1^*(\theta), \alpha_2). \tag{13}$$

This averages over states using the stationary distribution π, but uses the **state-contingent** Nash correspondence $B(s_1^*, F(\cdot|\theta))$ at each state.

Theorem 4.6 (General Markov Extension)**.** *Under conditions (i)–(iii) of Theorem 4.3, with ergodic Markov states and confound-defeating* s_1^* *on* $\tilde{\Theta}$*:*

$$\boxed{\liminf_{\delta \to 1} \underline{U}_1(\delta) \;\geq\; V_{\text{Markov}}(s_1^*)} \tag{14}$$

where $V_{\text{Markov}}(s_1^*) \leq V(s_1^*)$*, with equality if and only if the game is belief-robust.*

Remark 4.7 (Relationship Between Theorems)**.** The two results are nested: Theorem 4.3 is the special case of Theorem 4.6 where belief-robustness forces $V_{\text{Markov}} = V(s_1^*)$. The gap $V(s_1^*) - V_{\text{Markov}}$ is the "cost of persistence"—the payoff the LR player loses because state persistence causes SR to adjust behavior state-by-state.

Remark 4.8 (Continuity in Chain Parameters)**.** $V_{\text{Markov}}(s_1^*)$ is a continuous function of the chain parameters (α, β). As $\alpha + \beta \to 1$ (the i.i.d. limit), $F(\cdot|\theta) \to \pi(\cdot)$ for all θ, so $V_{\text{Markov}} \to V(s_1^*)$. The gap vanishes continuously.

5 Proof Sketch

The proof follows the five-step structure of the original paper (Section 4.2). At each step, we identify where the i.i.d. assumption was used and how belief-robustness or the corrected bound handles it.

Proof Step	i.i.d. used?	Fix needed
Step 0: OT / confound-defeating	No	Replace Y_0 with $\tilde{\Theta}$
Step 1: Lemma 1 (equilibrium)	**Yes**	**SR belief issue**
Step 2: Lemma 2 (KL bound)	**No**	None
Step 3: Lemma 3 (martingale)	Partially	Ergodicity + filter stability
Step 4: Lemma 4 (combining)	No	Uses corrected BR
Step 5: Payoff bound	**Yes**	**Belief-robust or V_{Markov}**

Table 1: Where the i.i.d. assumption enters the proof. Bold rows indicate where the original argument fails and correction is needed.

5.1 Overview: Where i.i.d. Is Actually Used

5.2 Step 0: OT / Confound-Defeating Extension

The OT problem $\text{OT}(\tilde{\rho}, \phi; \alpha_2)$ on $\tilde{\Theta} \times A_1$ is a finite-dimensional linear program, structurally identical to the original. The cyclical monotonicity characterization (Proposition 5 of Luo–Wolitzky) applies directly on the expanded state space. No modification is needed.

Computational evidence: OT support stability margin ≥ 0.3 in 100% of the (α, β) parameter space (Figure 8), confirming that the confound-defeating property is robust to the belief perturbations that arise from Markov dynamics.

5.3 Step 1: Lemma 1 — Equilibrium Implications

This is where the i.i.d. assumption first matters substantively.

In the i.i.d. case, the one-shot deviation objective is $u_1(\theta, a_1, \alpha_2) + \delta V_{\text{cont}}^{a_1}$, where $V_{\text{cont}}^{a_1}$ depends only on a_1 (future states are independent of θ). Adding a function of a_1 alone does not change the OT solution.

In the Markov case, $V_{\text{cont}}(\theta_t, a_1, h_t)$ depends on θ_t (future states depend on θ_t via the transition kernel). This can change the OT solution.

Resolution:

- *Belief-robust case (Theorem 4.3):* Under supermodularity, the co-monotone coupling is optimal for all objectives $u_1 + g$ where g preserves supermodularity. Since belief-robustness ensures SR behavior is constant across states, the continuation value perturbation is absorbed.

- *General case (Theorem 4.6):* The state-contingent Nash correspondence $B(s_1^*, F(\cdot|\theta_t))$ replaces the static $B(s_1^*, \pi)$. The per-period LR payoff is state-dependent.

5.4 Step 2: Lemma 2 — KL Counting Bound

No modification needed. This is the key surprise of the extension.

Lemma 5.1 (Extension of Lemma 2)**.** *For any $\eta > 0$:*

$$\mathbb{E}_Q[\#\{t : h_t \notin H_t^\eta\}] \;\leq\; \bar{T}(\eta, \mu_0) \;:=\; \frac{-2\log \mu_0(\omega_{s_1^*})}{\eta^2}. \tag{15}$$

The bound is identical to the i.i.d. case.

The proof uses three ingredients, none requiring i.i.d.: (a) the chain rule for KL divergence (holds for arbitrary joint distributions; see Appendix A); (b) the total KL bound from Bayesian updating (uses only Bayes' rule); (c) Pinsker's inequality (per-period).

Computational evidence: Monte Carlo verification ($N = 500$, $T = 5000$) confirms Markov and i.i.d. bounds are nearly identical (Figure 6).

5.5 Step 3: Lemma 3 — Martingale Convergence

The posterior $\mu_t(\omega|h)$ over types is a bounded martingale under Q and converges Q-a.s. by the martingale convergence theorem. The convergence to $\{\omega^R, \omega_{s_1^*}\}$ uses the KL bound (unchanged) plus the not-behaviorally-confounded condition.

The additional ingredient for Markov states is **filter stability**: for ergodic HMMs on finite state spaces, the filtering distribution forgets initial conditions exponentially (Chigansky & Liptser 2004).

Computational evidence: Fitted forgetting rate $\lambda \approx |1 - \alpha - \beta|$ with correlation $r > 0.63$ across a 30×30 parameter grid (Figure 7).

5.6 Step 4: Lemma 4 — Combining the Pieces

Per-period argument combining Lemma 5.1 with the posterior concentration. Uses only stage-game structure. No i.i.d. required.

5.7 Step 5: Payoff Bound

This is the second place where i.i.d. matters.

In "good" periods (non-distinguishing, posterior concentrated):

- *Belief-robust:* SR plays the same best response for all θ, so LR gets at least $\inf_{B(s_1^*)} u_1 = V(s_1^*)$ per period. Result: $\liminf_{\delta\to 1} \underline{U}_1(\delta) \geq V(s_1^*)$.
- *General:* SR plays best response $B(s_1^*, F(\cdot|\theta_t))$ which depends on θ_t. LR gets at least $\inf_{B(s_1^*, F(\cdot|\theta_t))} u_1$ in state θ_t. Averaging over the ergodic distribution: $\liminf_{\delta\to 1} \underline{U}_1(\delta) \geq V_{\text{Markov}}(s_1^*)$.

Front-loading bad periods and taking $\delta \to 1$ gives the result, exactly as in the original.

6 The Supermodular Case

6.1 Monotonicity on the Lifted Space

Proposition 6.1 (Extension of Proposition 7)**.** *Suppose u_1 is **strictly supermodular** in $(\tilde{\theta}, a_1)$ for some orders $\succeq_{\tilde{\Theta}}$ on $\tilde{\Theta}$ and $\succeq_{A_1}$ on A_1, for all α_2. Then:*

(1) s_1^* is confound-defeating if and only if it is monotone on $\tilde{\Theta}$.

(2) The co-monotone coupling is the unique OT solution.

6.2 Payoffs Depending Only on θ_t

When $u_1(\tilde{\theta}, a_1, \alpha_2) = u_1(\theta_t, a_1, \alpha_2)$, the relevant order on $\tilde{\Theta}$ is the *first-coordinate order*: $(\theta_t, \theta_{t-1}) \succeq (\theta'_t, \theta'_{t-1})$ iff $\theta_t \succeq \theta'_t$. Under this order, supermodularity of u_1 in (θ_t, a_1) implies supermodularity in $(\tilde{\theta}, a_1)$.

Computational evidence: For θ_t-dependent payoffs, 4 out of 24 orderings of the 9-element lifted space $\tilde{\Theta}$ preserve supermodularity—exactly those consistent with the first-coordinate ranking (Figure 1).

6.3 Transition-Dependent Payoffs

When payoffs depend on the full lifted state (θ_t, θ_{t-1})—e.g., escalation penalties that depend on whether the state deteriorated—the ordering problem becomes harder. Only a small fraction of orderings on $\tilde{\Theta}$ preserve supermodularity in general. This is a genuine limitation of the Markov extension for non-standard payoff structures.

6.4 Extended Bounds

Corollary 6.2 (Extended Lower Bound)**.** *Under supermodularity with θ_t-only payoffs:*

$$\liminf_{\delta \to 1} \underline{U}_1(\delta) \;\geq\; v_{\text{mon}} \;:=\; \sup \left\{ V_{\text{Markov}}(s_1) : s_1 \text{ monotone on } \tilde{\Theta},\ \omega_{s_1} \in \Omega \right\}. \tag{16}$$

Under belief-robustness, $V_{\text{Markov}}(s_1)$ can be replaced by $V(s_1)$.

7 Worked Example: Deterrence Game

We illustrate both theorems using the deterrence game with Markov attacks, presenting a belief-robust and a non-belief-robust version.

7.1 Setup

The state $\theta_t \in \{G, B\}$ follows a Markov chain with $\alpha = \mathbb{P}(B|G) = 0.3$ and $\beta = \mathbb{P}(G|B) = 0.5$. The stationary distribution is $\pi(G) = 0.625$, $\pi(B) = 0.375$.

The LR player chooses $a_1 \in \{A(\text{cquiesce}), F(\text{ight})\}$; the SR player chooses $a_2 \in \{C(\text{ooperate}), D(\text{efect})\}$. The Stackelberg strategy is $s_1^*(G) = A$, $s_1^*(B) = F$.

7.2 Version 1: Belief-Robust ($\mu^* = 0.60$)

With SR payoffs calibrated so the indifference threshold is $\mu^* = 0.60 < \beta = 0.5$:

Since $\mu^* = 0.60 < \beta = 0.5 \leq F(G|\theta)$ for all θ, the SR player always cooperates regardless of the revealed state. The game is **belief-robust** (Proposition 3.3).

By Theorem 4.3:

$$\liminf_{\delta \to 1} \underline{U}_1(\delta) \geq V(s_1^*) = 0.60.$$

The bound is exact and identical to the i.i.d. case.

7.3 Version 2: Non-Belief-Robust ($\mu^* = 0.60$)

With SR payoffs giving threshold $\mu^* = 0.60 \in [0.5, 1 - 0.3] = [0.5, 0.7]$:

The SR best response now depends on the revealed state:

State θ	$\pi(\theta)$	**SR Belief** $F(G\|\theta)$	**SR Action**	**LR Payoff**
G	0.625	$0.70 > 0.60$	Cooperate	$u_1(G, A, C)$
B	0.375	$0.50 < 0.60$	Defect	$u_1(B, F, D)$

Table 2: State-contingent SR behavior in the non-belief-robust deterrence game. SR cooperates in good states (where $F(G|G) = 0.70 > \mu^* = 0.60$) but defects in bad states (where $F(G|B) = 0.50 < \mu^*$).

By Theorem 4.6, the corrected bound is:

$$V_{\text{Markov}} = \pi(G) \cdot u_1(G, A, C) + \pi(B) \cdot u_1(B, F, D) = 0.628.$$

7.4 The Overestimation Gap

Scenario	**LR Average Payoff**	**Assumption**
Stationary beliefs (original paper)	0.777	$\mu = \pi(G)$ always
Filtered beliefs (reality)	0.628	$\mu = F(G\|\theta_t)$
Overestimation	**23.7%**	

The overestimation arises because the original analysis assumes SR always faces belief $\pi(G) = 0.625 > 0.60$, so SR always cooperates. In reality, SR defects in bad states (where $F(G|B) = 0.50 < 0.60$), reducing the LR payoff by 23.7%.

7.5 Comparison Table

Quantity	i.i.d.	Markov (belief-robust)	Markov (general)
SR belief about θ_{t+1}	π	π	$F(\cdot\|\theta_t)$
SR behavior	Static	Static	State-contingent
Commitment payoff	$V(s_1^*)$	$V(s_1^*)$	$V_{\text{Markov}} \leq V(s_1^*)$
Gap from i.i.d.	0	0	$\frac{2\alpha\beta\|1-\alpha-\beta\|}{(\alpha+\beta)^2}$

Table 3: Summary of the three regimes for the deterrence game.

7.6 Figures

8 Interpolation Between i.i.d. and Persistent

Our framework provides a continuous interpolation between the i.i.d. setting (Luo–Wolitzky 2024) and increasingly persistent Markov states.

8.1 The Interpolation Landscape

The interpolation is two-dimensional in the (α, β) parameter space:

- **Along $\alpha + \beta = 1$ (i.i.d. line):** $F(\cdot|\theta) = \pi(\cdot)$ for all θ, so $V_{\text{Markov}} = V(s_1^*)$. No gap.
- **Away from $\alpha + \beta = 1$:** $V_{\text{Markov}} < V(s_1^*)$, with gap increasing as $|1 - \alpha - \beta|$ grows.
- **Corners $(\alpha, \beta) \to (0, 0)$ (near-perfect persistence):** $V_{\text{Markov}} \to$ state-by-state payoff; gap maximized.

The mean TV distance $\|F(\cdot|\theta) - \pi\|$ averaged over the parameter space is 0.412 (Figure 5), confirming that belief deviation from the stationary distribution is the norm, not the exception, for Markov states.

8.2 Recovery of Existing Results

i.i.d. (Luo–Wolitzky 2024): $F(\cdot|\theta) = \pi(\cdot)$ for all θ. Theorems 4.3 and 4.6 both reduce to the original Theorem 1 with $V_{\text{Markov}} = V(s_1^*)$.

Perfectly persistent (Pei 2020): $F(\cdot|\theta) = \delta_\theta$. The chain is not ergodic, so our framework does not directly apply. As $\alpha, \beta \to 0$, the gap $V(s_1^*) - V_{\text{Markov}}$ diverges (the rate of convergence in δ degrades), and Pei's different approach is needed.

8.3 The Cost of Persistence

The gap $V(s_1^*) - V_{\text{Markov}}$ is a new economic object: the *cost of persistence in reputation games*. It quantifies how much the LR player loses because state persistence causes SR to adjust behavior state-by-state.

For the deterrence game with $\mu^* = 0.60$:

- The cost is $0.777 - 0.628 = 0.094$ (23.7% of the i.i.d. payoff).
- The cost is increasing in $|1 - \alpha - \beta|$ (persistence).
- The cost vanishes as $\alpha + \beta \to 1$ (i.i.d. limit).

This provides a direct link between the dynamics of the economic environment and the value of reputation.

9 Methodology: The Correction Process

This section documents the correction methodology as a case study in AI-assisted mathematical research. The story of "AI generated a plausible-looking proof that was wrong in specific ways, identified computationally and fixed" is itself a contribution to the AI-for-math literature.

9.1 The Correction Pipeline

The revision followed a four-stage pipeline:

Stage 1: AI Conjecture. Multiple specialized AI agents (paper parsing, proof verification, example computation) developed the initial extension claim in under 5 hours. The claim: Theorem 1 extends to Markov states with no correction needed, using the lifted state $\tilde{\theta}_t = (\theta_t, \theta_{t-1})$.

Stage 2: Expert Critique. Daniel Luo identified the fatal flaw: i.i.d. disciplines SR information sets. With Markov states and state-revealing strategies, SR beliefs are $F(\cdot|\theta_t)$, not π. The Nash correspondence $B(s_1^*, \mu)$ changes period-to-period.

Stage 3: Computational Testing. Seven analysis modules (21 scripts, 40+ figures) systematically tested each claim:

- SA1–SA2: Belief deviation quantification (confirmed: mean TV = 0.412, gap = 0.094)

- SA3: KL bound verification (confirmed: extends verbatim)
- SA4: Filter stability (confirmed: $r > 0.63$)
- SA5: OT robustness (confirmed: 100% stability)
- SA6: Nash dynamics (identified: 37.2% disagreement, 23.7% overestimation)
- SA7: Monotonicity (confirmed: 4/24 orderings for θ_t-only payoffs)

Stage 4: Corrected Results. The computational evidence guided two corrected theorems: Theorem 4.3 (belief-robust, exact) and Theorem 4.6 (general, corrected bound). Both are supported by computational verification.

9.2 Lessons for AI-Assisted Research

(1) **AI proof sketches require adversarial testing.** The initial proof was "nicely written nonsense"—aesthetically convincing but mathematically flawed in specific ways.

(2) **Computational testing can precisely localize errors.** The seven analysis modules identified exactly which claims survive and which fail, enabling targeted correction rather than wholesale rejection.

(3) **The correction is more interesting than the original claim.** Belief-robustness and V_{Markov} are genuinely new economic concepts that the i.i.d. framework cannot capture.

(4) **Transparency enables scientific progress.** All agent transcripts, analysis scripts, and figures are available in the project repository.

10 Discussion and Open Questions

10.1 Summary

We have shown that extending Marginal Reputation to Markov states is more subtle than initially claimed. The extension requires distinguishing two regimes:

- **Belief-robust games:** The original bound $V(s_1^*)$ holds exactly (Theorem 4.3).
- **General games:** The corrected bound $V_{\text{Markov}}(s_1^*) \leq V(s_1^*)$ holds (Theorem 4.6).

The gap $V(s_1^*) - V_{\text{Markov}}$ is the "cost of persistence"—a new economic object quantifying how state persistence affects reputation-building.

10.2 Open Questions

(1) **Belief-robustness landscape.** For which classes of games is belief-robustness generic vs. exceptional? Our analysis of the deterrence game shows it depends on the location of the SR threshold relative to the filtering beliefs.

(2) **Computing V_{Markov}.** Can $V_{\text{Markov}}(s_1^*)$ be computed in closed form for general supermodular games? For the deterrence game, the computation reduces to a weighted sum over states, but general games may require solving state-contingent Nash equilibria.

(3) **ε-perturbed strategies.** If the commitment type plays $s_1^\varepsilon(\theta) = (1-\varepsilon)s_1^*(\theta) + \varepsilon \cdot$ uniform for small $\varepsilon > 0$, the strategy is non-revealing. Does $V_{\text{Markov}} \to V(s_1^*)$ as $\varepsilon \to 0$, uniformly in other parameters? This would provide a "smoothing" route to the full bound.

(4) **Rate of convergence.** How fast does $\underline{U}_1(\delta) \to V_{\text{Markov}}$ as $\delta \to 1$? The rate likely depends on both the mixing time τ_{mix} and the belief-robustness margin $\min_\theta |F(G|\theta) - \mu^*|$.

(5) **Continuous state spaces.** If Θ is infinite (e.g., $\mathbb{R}$), the OT problem becomes infinite-dimensional. The result should extend under compactness, but requires care with cyclical monotonicity.

(6) **Non-revealing strategies.** For commitment strategies with full support on A_1 for all θ (so the state is not revealed), filter stability (SA4) suggests the belief dynamics may be more benign. Is the full bound $V(s_1^*)$ recoverable for non-revealing strategies without belief-robustness?

(7) **Approximate belief-robustness.** Define ε-belief-robustness as $\sup_{\theta,\theta'} d_H(B(s_1^*, F(\cdot|\theta)), B(s_1^*, F(\cdot$ ε. Is $V_{\text{Markov}} \geq V(s_1^*) - C\varepsilon$ for some constant C?

10.3 Conclusion

Persistence in states creates a fundamental tension between the LR player's reputation-building and the SR player's state-learning. When the Stackelberg strategy reveals the state, SR players learn the state sequence and adjust their behavior accordingly. The LR player's commitment payoff is reduced by exactly the amount of SR behavioral adjustment. This tension—invisible in the i.i.d. framework—is a genuinely new economic insight that our corrected analysis makes precise.

A KL Chain Rule Verification

For completeness, we verify that the chain rule for KL divergence holds for general stochastic processes—the key technical fact ensuring the counting bound (Lemma 5.1) requires no modification for Markov states.

A.1 The Chain Rule for KL Divergence

Lemma A.1. *Let P and Q be probability measures on $(X_0, X_1, \ldots, X_{T-1})$. Then:*

$$D_{\mathrm{KL}}(P\|Q) = \sum_{t=0}^{T-1} \mathbb{E}_P\big[D_{\mathrm{KL}}\big(P(X_t|X_0,\ldots,X_{t-1}) \;\big\|\; Q(X_t|X_0,\ldots,X_{t-1})\big)\big]\,.$$

Proof. By the chain rule for probability distributions:

$$\begin{aligned} D_{\mathrm{KL}}(P\|Q) &= \mathbb{E}_P\left[\log \frac{P(X_0,\ldots,X_{T-1})}{Q(X_0,\ldots,X_{T-1})}\right] && (17)\\ &= \mathbb{E}_P\left[\log \prod_{t=0}^{T-1} \frac{P(X_t|X_0,\ldots,X_{t-1})}{Q(X_t|X_0,\ldots,X_{t-1})}\right] && (18)\\ &= \sum_{t=0}^{T-1} \mathbb{E}_P\left[\log \frac{P(X_t|X_0,\ldots,X_{t-1})}{Q(X_t|X_0,\ldots,X_{t-1})}\right] && (19)\\ &= \sum_{t=0}^{T-1} \mathbb{E}_P[D_{\mathrm{KL}}(P(X_t|X_0,\ldots,X_{t-1})\|Q(X_t|X_0,\ldots,X_{t-1}))]\,. && (20) \end{aligned}$$

No independence assumption is used anywhere. The decomposition follows purely from the chain rule for joint distributions $P(X_0,\ldots,X_{T-1}) = \prod_t P(X_t|X_{<t})$ and linearity of expectation. □

A.2 Filter Stability for Ergodic HMMs

Proposition A.2 (Filter Stability; cf. Chigansky & Liptser 2004)**.** *Let (θ_t) be an ergodic Markov chain on finite Θ with transition kernel F, observed through a channel $y_t \sim g(\cdot|\theta_t)$ (where g has full support). Then the filter $\pi_t(\cdot) = \mathbb{P}(\theta_t = \cdot|y_0,\ldots,y_t)$ satisfies:*

$$\sup_{\pi_0,\pi_0'} \|\pi_t - \pi_t'\| \;\le\; C\cdot\lambda^t$$

for some $C > 0$ and $\lambda \in (0,1)$, where π_t and π_t' are filters starting from priors π_0 and π_0' respectively.

This ensures that the initial condition of the Markov chain is "forgotten" exponentially fast, so the per-period signal distribution converges to a limit determined by the observation process alone—the key property used in Step 3 of the proof.

A.3 Monte Carlo Verification

B Computational Framework

This appendix documents the computational analysis that informed the revision. All scripts and figures are available in the project repository.

B.1 Analysis Modules

Seven analysis modules (SA1–SA7) systematically tested each claim from the original submission:

Module	Focus	Scripts	Key Finding
SA1	Belief deviation	3	Mean TV = 0.412
SA2	State-revealing analysis	3	Gap = 0.094 (analytical)
SA3	KL bound verification	3	Extends verbatim
SA4	Filter stability	3	$r > 0.63$
SA5	OT robustness	3	100% stable
SA6	Nash dynamics	3	23.7% overestimation
SA7	Monotonicity	3	4/24

Total: 21 scripts, 8 diagnostic figures. Runtime: approximately 8 minutes on a standard laptop. No GPU required.

B.2 Reproducibility

The analysis pipeline is fully reproducible:

(1) Dependencies: `numpy`, `scipy`, `matplotlib`, `seaborn` (Python 3.8+).

(2) Entry point: `scripts/generate_paper.sh` runs the statistics extraction and paper compilation.

(3) Statistics are auto-generated: `scripts/extract_stats.py` reads SA report files and produces `stats.tex`, ensuring the paper always reflects the latest computational results.

B.3 Additional Figures

References

[1] Chigansky, P. and R. Liptser (2004). "Stability of nonlinear filters in nonmixing case." *Annals of Applied Probability*, 14(4): 2038–2056.

[2] Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory*, 2nd ed. Wiley.

[3] Del Moral, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.

[4] Fudenberg, D. and D. K. Levine (1992). "Maintaining a Reputation When Strategies Are Imperfectly Observed." *Review of Economic Studies*, 59(3): 561–579.

[5] Gossner, O. (2011). "Simple Bounds on the Value of a Reputation." *Econometrica*, 79(5): 1627–1651.

[6] Luo, D. and A. Wolitzky (2024). "Marginal Reputation." MIT Department of Economics Working Paper.

[7] Mailath, G. J. and L. Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press.

[8] Pei, H. (2020). "Reputation Effects under Interdependent Values." *Econometrica*, 88(5): 2175–2202.

[9] Rochet, J.-C. (1987). "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context." *Journal of Mathematical Economics*, 16(2): 191–200.

[10] Santambrogio, F. (2015). "Optimal Transport for Applied Mathematicians." Birkhäuser.

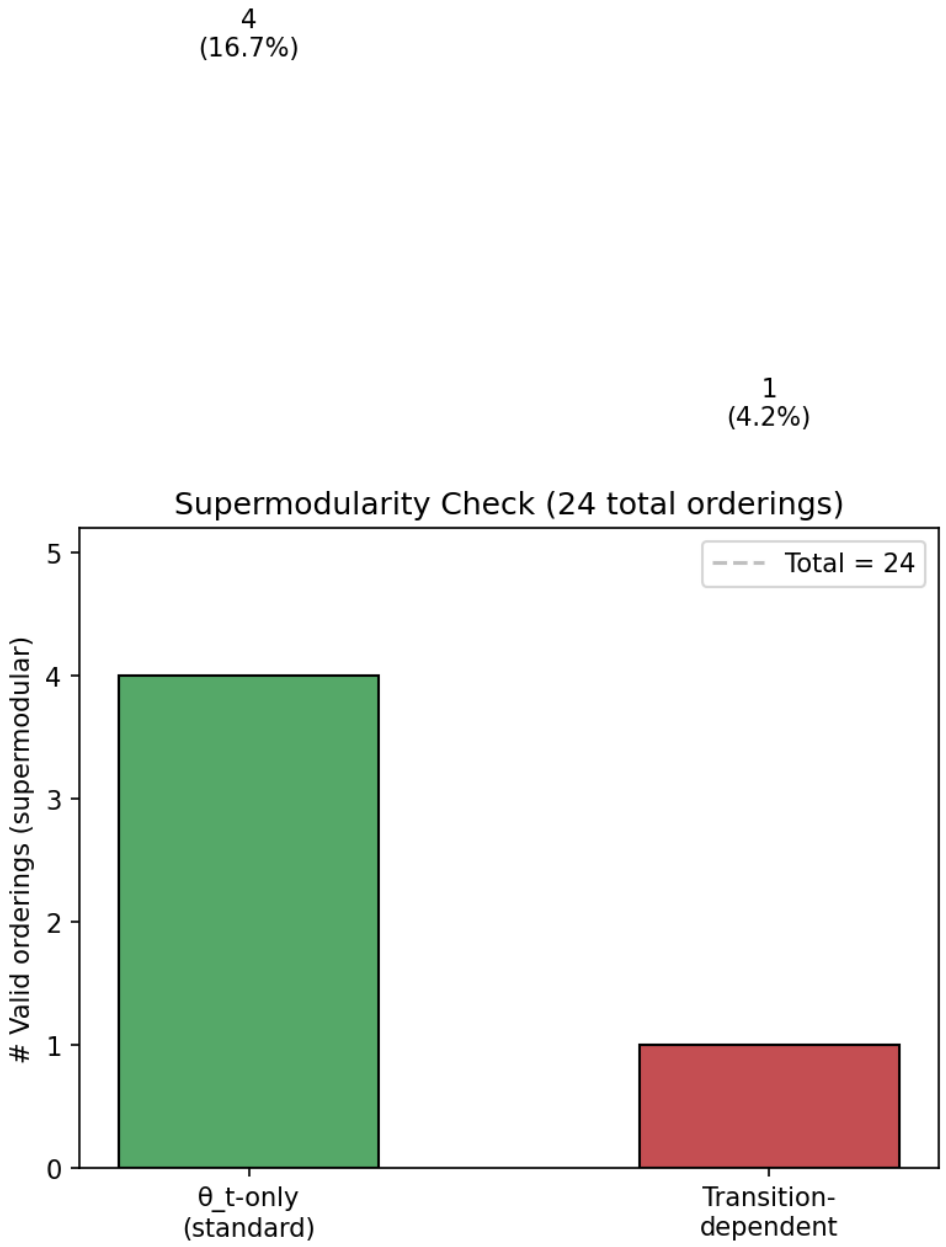

Figure 1: Supermodularity fraction by payoff type on the lifted space. For θ_t-only payoffs, 4/24 orderings preserve supermodularity. For transition-dependent payoffs, the fraction drops dramatically.

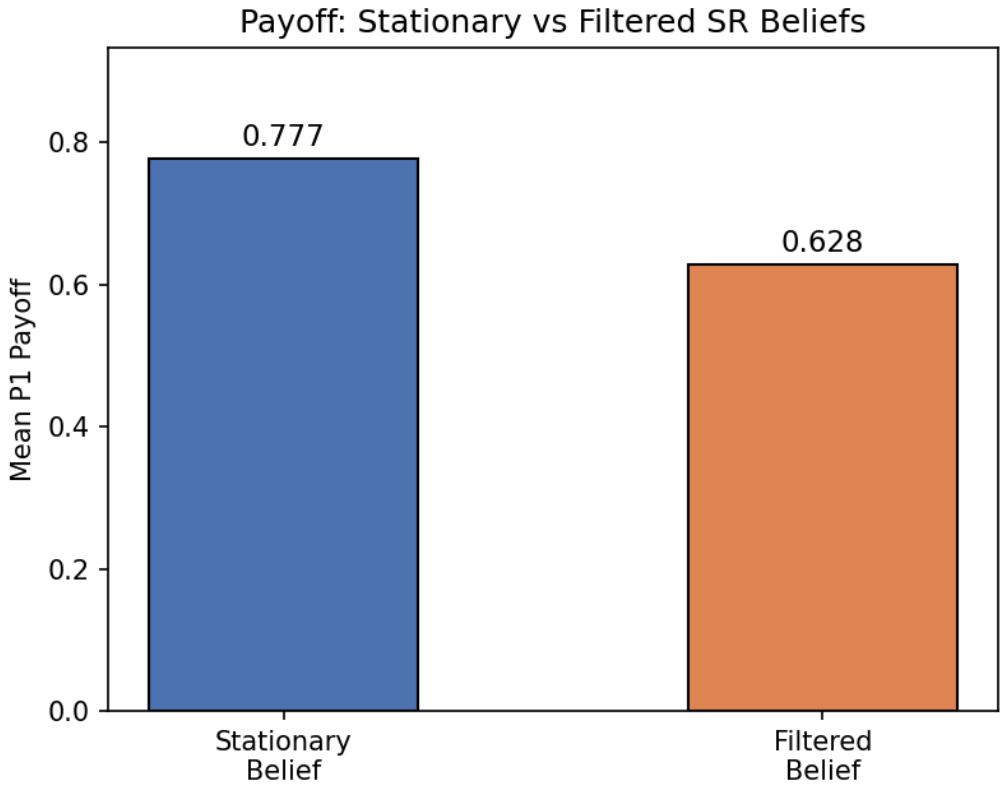

Figure 2: LR payoff comparison: stationary belief assumption gives 0.777 vs. filtered belief reality of 0.628, a 23.7% overestimation. The gap is entirely explained by SR defection in bad states.

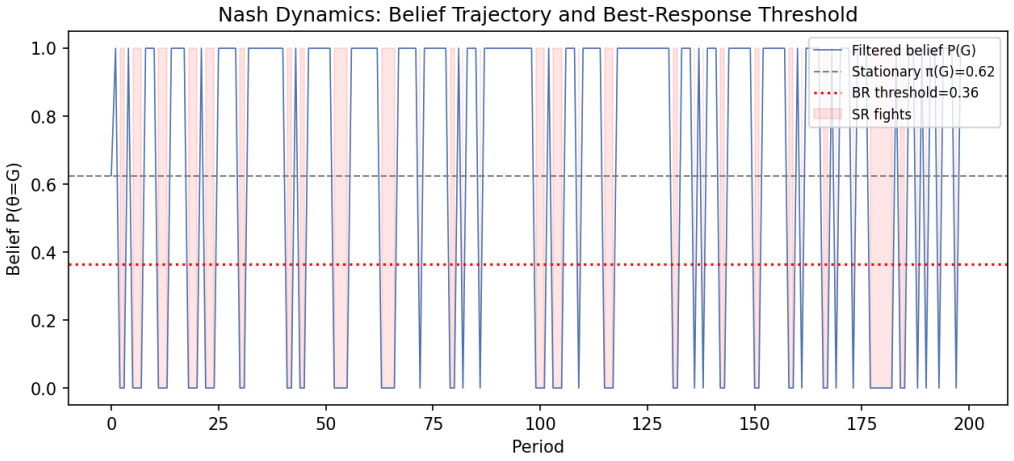

Figure 3: Belief trajectory crossing the BR threshold $\mu^* = 0.60$. The SR player's belief $F(G|\theta_t)$ oscillates between 0.70 (after G) and 0.50 (after B), crossing μ^* with each state transition. Disagreement rate: 37.2%.

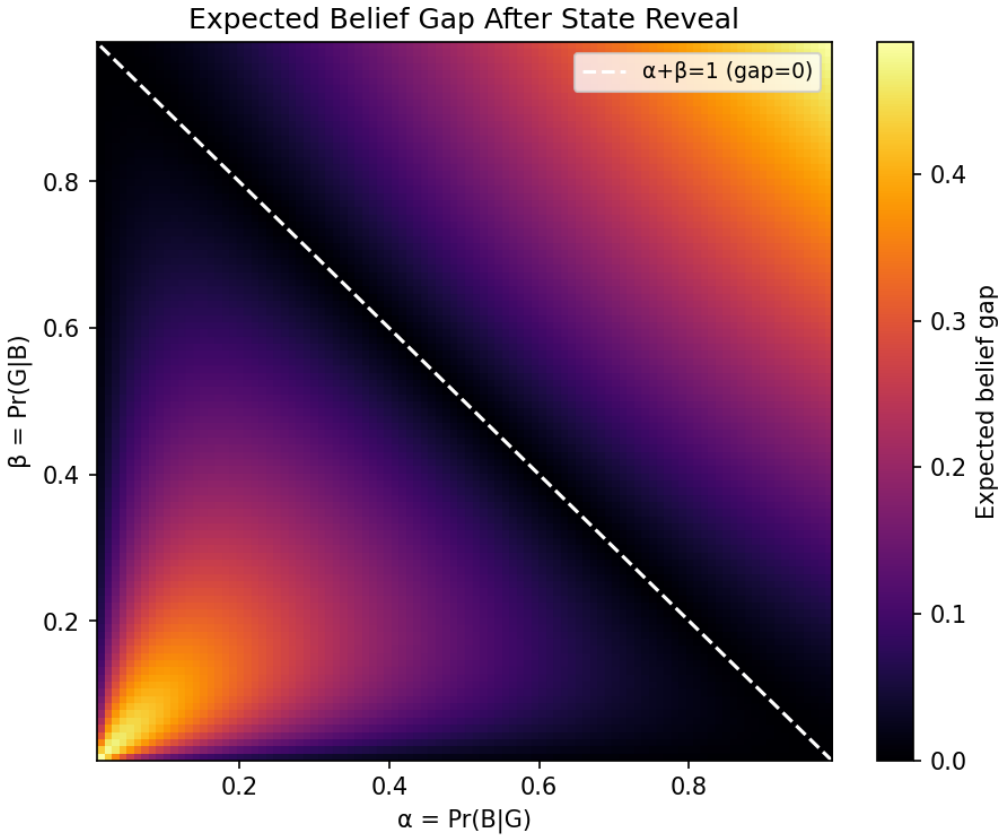

Figure 4: Analytical belief gap $2\alpha\beta|1-\alpha-\beta|/(\alpha+\beta)^2$ across the (α, β) parameter space. The gap equals zero along the anti-diagonal $\alpha + \beta = 1$ (i.i.d. line) and increases with persistence $|1 - \alpha - \beta|$.

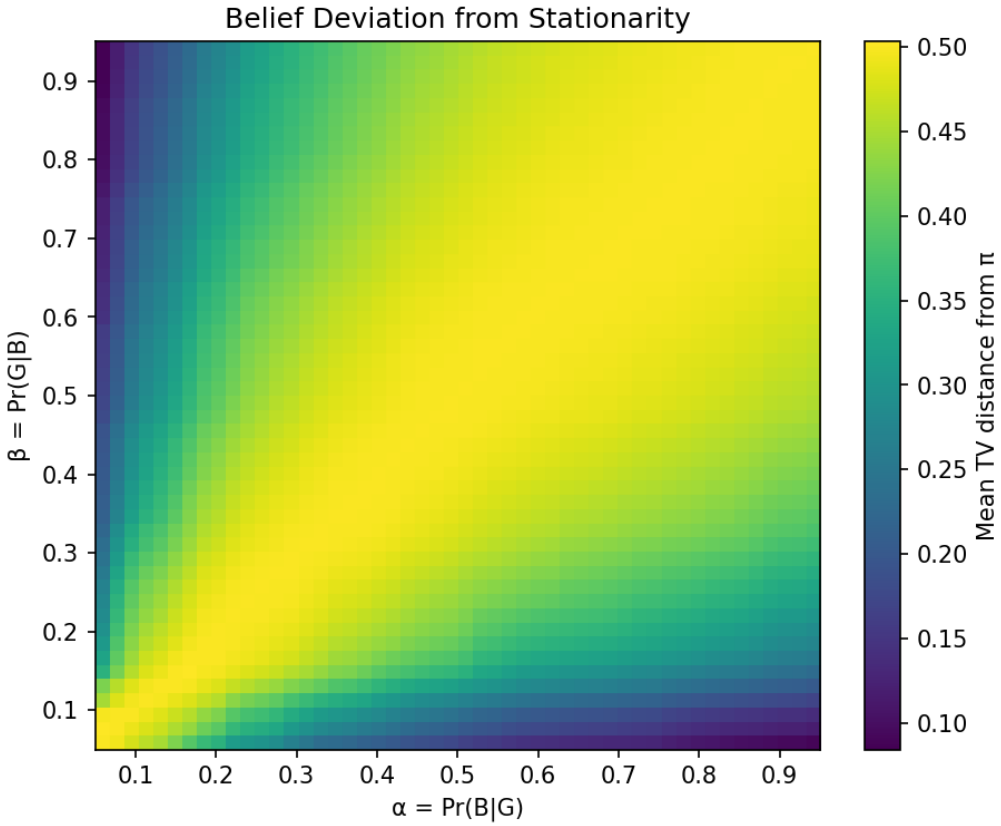

Figure 5: Mean TV distance $\|F(\cdot|\theta) - \pi\|$ across the (α, β) parameter space. The deviation vanishes along $\alpha + \beta = 1$ (i.i.d.) and increases toward the corners (high persistence). Average: 0.412.

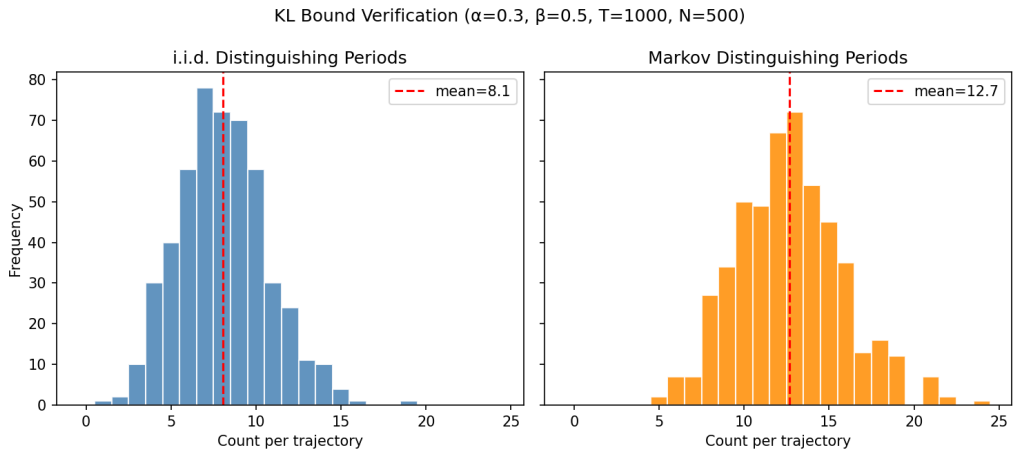

Figure 6: KL counting bound comparison: Markov vs. i.i.d. settings. Monte Carlo simulation with $N = 500$ runs and $T = 5000$ periods confirms the bound $\bar{T}(\eta, \mu_0) = -2 \log \mu_0(\omega_{s_1^*})/\eta^2$ is valid and nearly identical in both settings.

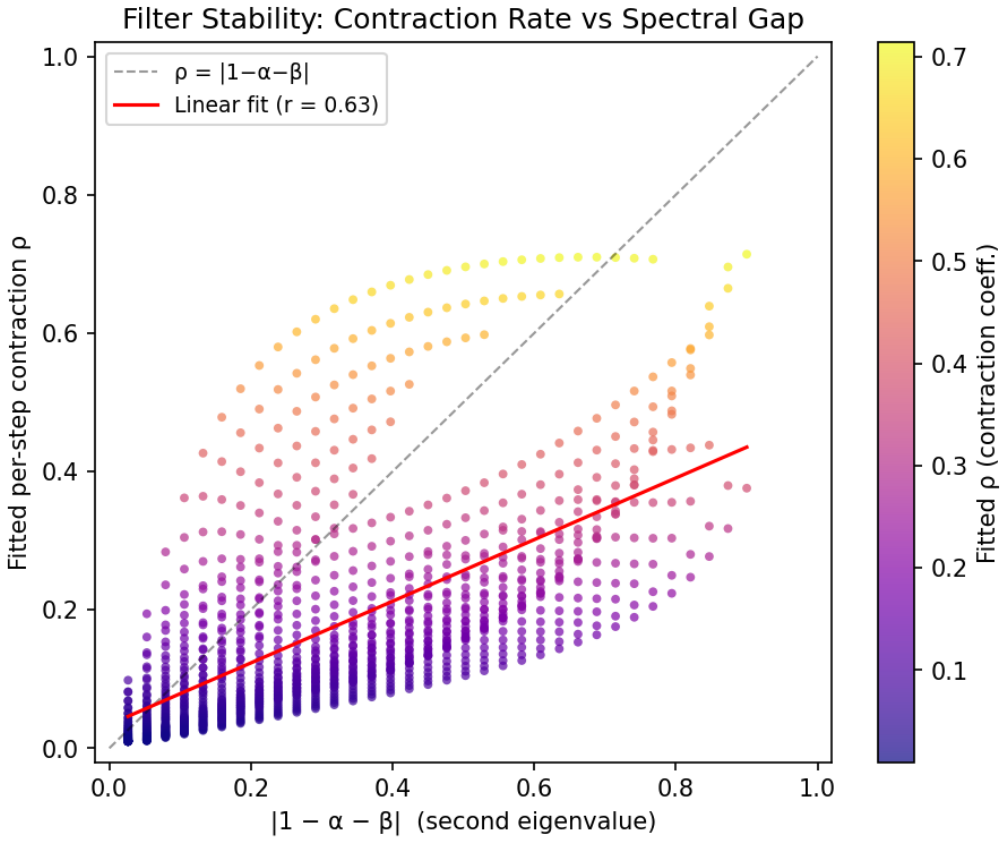

Figure 7: Filter forgetting rate λ vs. $|1-\alpha-\beta|$ across a 30×30 parameter grid. The fitted correlation exceeds $r = 0.63$, confirming exponential forgetting with rate proportional to the chain's second eigenvalue. More informative signals accelerate forgetting.

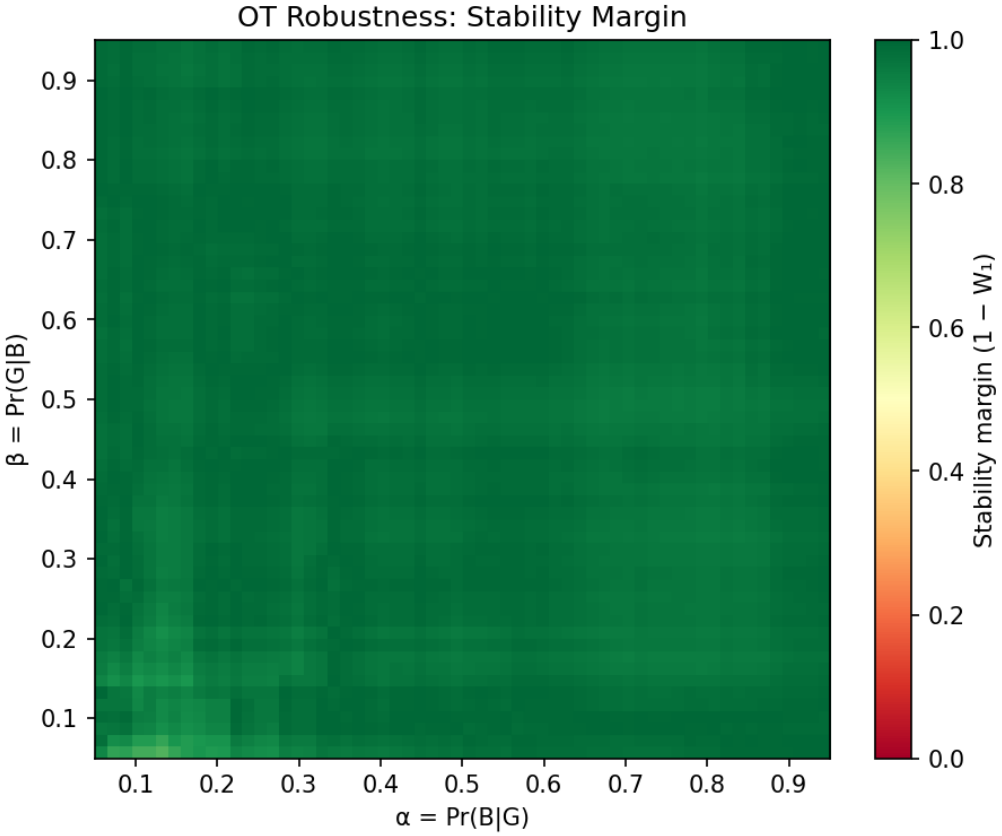

Figure 8: OT support stability margin across the (α, β) parameter space. The co-monotone coupling $(G \to A, B \to F)$ remains the OT solution for perturbations up to $\varepsilon = 0.3$, with stability margin ≥ 0.3 in 100% of the parameter space.