INFORMATION-PRESERVING EXECUTION AND STABILITY IN BLOCK-STRUCTURED MARKOV GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Simultaneous actions are often sampled in a linear administrative order. BLOCK-SEBIL distinguishes this administrative order from strategic information in finite discounted games with declared block histories and selectors. We prove that hiding sampled actions until commitment preserves the game when the implementation also preserves the prescribed joint policy law, whereas revelation can change it. The resulting protocol-to-evidence contract combines structured sensitivity, invariant-region, and oracle-residual bounds to give conditional contraction and error guarantees. A legal two-state general-sum family attains its comparison bound in span coordinates: with $\gamma = 0.95$ and gain scale 3, the strict local contraction threshold is $L^\star = 4.2935302$ for every $\tau > 0$, and the supremum of certified normalized radii tends to zero there. Controlled game and oracle diagnostics illustrate these conditional guarantees; numerical LP checks are not formal real-arithmetic certificates. In PettingZoo MPE Simple Spread v2, a five-seed fixed-policy comparison gives identical hidden/parallel action traces, while a policy that uses revealed actions changes 77.1% of actions. A long-run aggregate summary records hidden/parallel agreement over $3{,}333{,}325$ cycles per arm, but lacks the per-episode traces required for independent completion verification. These are protocol diagnostics, not learner benchmarks or a general semantic compiler.

Keywords: multi-agent reinforcement learning, block protocols, information structure, small-gain analysis, semantic auditing

1 INTRODUCTION

Sequential and simultaneous decisions are both standard objects in game theory and reinforcement learning. The practical difficulty appears when a system contains both: a stage may contain a simultaneous equilibrium block, while other blocks depend on its committed joint action. Implementations commonly linearize such a block for scheduling or batching. That linearization is harmless only if the sampled actions remain hidden until commitment. If the second component observes the first sampled action, the game becomes a leader–follower protocol.

Administrative sampling order and strategic information order are not interchangeable. The distinction matters beyond game-theoretic notation. A Bellman update built from the wrong information mask can converge to a stable solution of a different game. Conversely, an approximate equilibrium or neural block oracle can be locally accurate while its sensitivity destabilizes a composed recursion. A useful method therefore needs a semantic contract, a stability certificate, and an explicit failure state. Figure 1 summarizes this distinction: hidden administrative order leaves the information set unchanged, whereas revelation creates a strategic edge before commitment.

To make this distinction explicit, BLOCK-SEBIL represents a finite discounted Markov game as an acyclic collection of action blocks. Each block declares its players, actions, predecessor dependencies, observation mask, and selector. Backward evaluation is restricted to public predecessor histories that define continuation subgames: a simultaneous block uses one pre-block history, whereas a sequential block follows the declared order.

The analysis couples this protocol description to a stability interface. Each block exposes its observation mask, selector contract, continuation sensitivities, image region, and semantic residual. For a block B, let S_B be its declared selector and T_B its continuation-value update. We retain

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

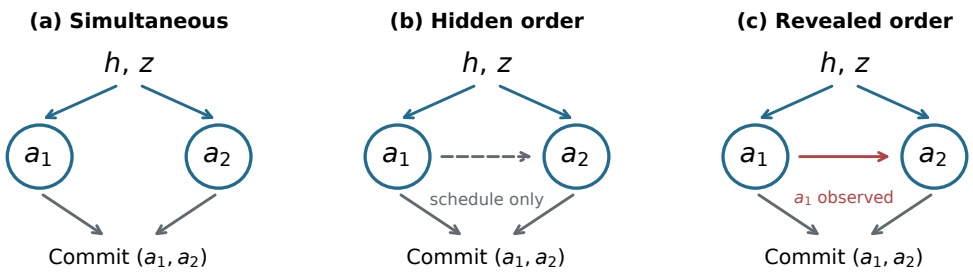

Figure 1: Protocol distinction. In simultaneous play and hidden serialization, both actions use pre-block history h and allowed shared randomness z; equivalence also requires preservation of the prescribed joint policy law. The dashed arrow denotes scheduling, not observation. Revealed serialization also exposes a_1 before a_2 is committed, creating a different information structure.

coordinate-level gains rather than one scalar per block. A valid comparison matrix with $\rho(G) < 1$ on a closed invariant region gives contraction and a fixed-point error tube by standard resolvent weighting. These conclusions are conditional on the declared protocol and analytic evidence; a supplied matrix alone does not establish the required semantic fields.

The resulting comparison is also sharp in the relevant direction. Stochastic action-difference rows annihilate constant value perturbations, so entrywise absolute bounds can be conservative in full coordinates. We construct a legal two-state general-sum game whose span Jacobian equals its comparison matrix entrywise at the centered fixed point. This gives an attainable local threshold, without asserting universal strictness in full coordinates.

The contract therefore distinguishes certification, refutation, and abstention. A loose upper bound above one is not a refutation; a signed point witness is needed. Low output residual alone does not establish contraction. General-sum blocks require a declared branch and regularity assumptions, since exact Nash selection is discontinuous across a simple dominance switch.

Our contributions are:

1. We specify block Bellman–equilibrium semantics for public predecessor histories and prove hidden equivalence under a preserved joint law, with a revealed-order separation witness.
2. We give a protocol-to-evidence contract that combines declared observation masks and selectors with structured small-gain, invariant-region, and semantic-residual checks. Rank-aware matrices and Perron/resolvent weights are certificate primitives inside this contract.
3. We give a multi-block nonnegative span criterion with multiple externality parameters; the symmetric two-block threshold is a closed-form specialization.
4. We construct a legal two-state general-sum family that realizes this criterion and attains the span comparison bound, derive its local threshold $L^\star = 4.2935302$, and prove continuous collapse of its certified regional radius.
5. We use exact games, oracle controls, mixed block compositions, and a fixed-policy PettingZoo MPE protocol stress test as claim-matched diagnostics. These experiments test semantic fidelity and admissible evidence, not learner quality or universal equilibrium computation.

2 Related Work and Positioning

Markov games provide the standard stochastic-game foundation (13; 17). Work on simultaneous-move Markov games addresses learning and value computation under joint actions (20), while turn-based analyses exploit the special structure of action ownership (12). PettingZoo standardizes

multi-agent APIs and the Agent Environment Cycle model (19); we use its MPE environment as an external protocol stress test, not as an environment contribution. These works provide the game and API foundations; our boundary is the explicit contract that keeps administrative sampling order separate from strategic information flow.

Entropy regularization is used for equilibrium computation in normal-form and sequential games (14; 15); in general-sum games it does not by itself ensure a unique smooth QRE. Recent extragradient methods provide linear-rate algorithms for entropy-regularized zero-sum matrix and Markov games (3). We require a specified regular QRE branch and bound its composed sensitivity. The selector regularity frontier is given in Appendix A.2.

General-sum stochastic equilibrium computation has fundamental hardness barriers even in turn-based settings (5). Our guarantees therefore concern declared selectors, not arbitrary stationary Nash computation.

Multi-agent influence diagrams represent information and strategic dependencies (10). Our block notation does not replace that representation; it adds explicit assumptions on the selected value map and its analytic evidence. Perron–Frobenius weighting and spectral small-gain arguments are standard (1; 4). We use them for coordinate-level bounds and resolvent error tubes, without claiming a new small-gain criterion.

The surrounding learning literature also separates coordination, regularization, and evaluation from the semantics of the underlying game. Distributed cooperative learning and counterfactual credit assignment motivate structured coordination mechanisms (6; 11), whereas regularized decision processes and maximum-entropy control analyze a different source of smoothness (7; 8). Our contribution is narrower: the block interface specifies which information may be shared and which continuation value is being composed. The resulting Jacobian condition is therefore a contract on a declared evaluator, not a new equilibrium-selection rule. Standard Markov decision-process notation and matrix-analysis conventions provide the remaining background (9; 16).

For learned block maps, CROWN-style affine relaxation (21), ReLU branch-and-bound (2), and nonlinear BaB (18) provide verification primitives. Our contract requires an image enclosure, a matrix Jacobian envelope, and a semantic residual for a game-level small-gain test. The numerical oracle examples illustrate these requirements; they do not establish an end-to-end semantic verifier.

3 Protocol-Faithful Block Games

3.1 Block-poset Markov games

Let $\mathcal{G} = (\mathcal{S}, \mathcal{B}, \prec, \{\mathcal{A}_i\}, P, \{r_i\}, \gamma)$ be a finite discounted Markov game, with $0 < \gamma < 1$. Blocks form an acyclic precedence graph. Block B contains players I_B and commits $a_B \in \prod_{i \in I_B} \mathcal{A}_i$. We restrict admissible protocols to public predecessor histories h_B: all past actions affecting a continuation subgame are publicly committed, and players have perfect recall. The mask $O_i(B)$ includes h_B and any declared random signal but excludes other current-block actions. Coupled simultaneous decisions belong to one block; the graph schedules already-defined subgame evaluations. Hidden predecessor actions requiring belief-state or imperfect-information continuation solvers are outside this recursion.

For a complete within-stage action history a, form the terminal payoff table

$$Q^V(s,a) = r(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V(s')]. \tag{1}$$

Backward evaluation of descendant blocks at their public histories gives the normal-form table $Q_B^V(h_B, a_B)$. Discounting occurs once per environment transition, not at each internal block. For a declared selector S_B, the history-indexed update is

$$T_B(V) = \mathcal{E}_B\big(Q_B^V, S_B(Q_B^V)\big). \tag{2}$$

Here $\mathcal{E}_{B,i}(Q, \pi) = \mathbb{E}_{a_B \sim \pi}[Q_i(a_B)]$ is the raw-payoff evaluation operator, with no entropy bonus. The global operator T evaluates the root histories by backward induction under the declared selectors. Its semantics are conditional on those selectors, not a claim of Nash selection for arbitrary games.

Whenever a derivative of a general-sum QRE is used, we assume a single-valued continuously differentiable branch on a neighborhood of the region, with invertible response-equation Jacobian through-

out. A positive temperature alone does not supply this assumption. The additive family below has an explicit unique logistic response.

Definition 3.1 (Protocol-faithful implementation)**.** An implementation is protocol-faithful if it preserves the prescribed conditional joint law at every history, respects each policy's mask and shared-randomness coupling, and applies reward and transition to the committed joint action. Hiding samples alone is insufficient if, for example, correlated marginals are resampled independently.

3.2 Hidden order versus revealed order

Theorem 3.2 (Hidden serialization equivalence)**.** *Fix a simultaneous block and a conditional joint policy law given pre-block history and allowed shared randomness. Any internal sampling permutation that preserves this law and all policy inputs, while hiding current actions until commitment, induces the same reward, transition, and continuation-history law.*

Proof. The joint policy is sampled once from the same pre-block sigma-algebra. A permutation changes only the internal sampling trace; marginalizing that trace leaves the same joint action. Since reward and transition consume the joint action, all downstream laws coincide. □

Theorem 3.3 (Revealed serialization separation)**.** *There exists a one-state two-player zero-sum game for which simultaneous play, row-first revealed play, and column-first revealed play have different values.*

Proof. For matching pennies with payoff matrix $\left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$, the simultaneous stage value is 0. Observing the row action lets the column player enforce row payoff -1; reversing the order gives 1. In a one-state discounted repetition the unnormalized returns are 0 and $\pm(1-\gamma)^{-1}$. □

4 Structured Stability and Semantic Error

4.1 Comparison matrices

Let T and its approximation f act on a nonempty closed product region $\mathcal{R}(r) \subset \mathbb{R}^d$. Require a pairwise bound $|T(u) - T(v)| \le G(r)|u - v|$ for all $u, v \in \mathcal{R}(r)$ and a uniform residual $|f(u) - T(u)| \le \beta$. A Jacobian envelope implies the pairwise bound on a convex region when it is valid along every connecting segment. For additive witnesses with equal block coordinate dimensions, local gains M_B and continuation gains C_{BC} yield

$$G(r) = \operatorname{diag}\big(M_B(r_B)\big) + C \otimes I. \tag{3}$$

This additive form requires a derivation from the actual update equations; it is not a general formula for composing a masked DAG. Administrative precedence edges need not be continuation couplings. For exact and approximate iterates in the region, the coordinate error satisfies

$$e^+ \le G(r)e + \beta. \tag{4}$$

Theorem 4.1 (Rank-aware small-gain certificate)**.** *Under the pairwise and residual bounds above, suppose $T(\mathcal{R}) \subseteq \mathcal{R}$, $f(\mathcal{R}) \subseteq \mathcal{R}$, $G \ge 0$, and $\rho(G) < 1$. Then T is contractive in a positive weighted coordinate max norm and has a unique fixed point $v^* \in \mathcal{R}$. Any fixed point $\hat{v}$ of f in this region satisfies $|\hat{v} - v^*| \le (I - G)^{-1}\beta$. Existence or convergence of such an approximate fixed point is not assumed to follow from this bound.*

Proof. Because $G \ge 0$ and $\rho(G) < 1$, the resolvent exists and $w = (I - G)^{-1}\mathbf{1}$ is strictly positive. Moreover, $Gw = w - \mathbf{1} \le qw$ with $q = 1 - 1/\max_i w_i < 1$. Hence the induced weighted coordinate max norm has modulus at most q. Closedness gives completeness, so invariance and Banach's theorem give v^*. For $e = |\hat{v} - v^*|$, $e \le Ge + \beta$ implies the tube by the nonnegative inverse. Irreducible G also admits a positive Perron weight with modulus $\rho(G)$; resolvent weights need not attain that same modulus and also work for reducible G. □

In an asymmetric three-block witness, block scalarization has radius 1.079, while the coordinate comparison has radius 0.596; a scalar row-sum test accepts neither. These saved numerical checks concern the stated stochastic construction and region, not arbitrary supplied matrices.

4.2 Stochastic realization for additive witnesses

For the additive span witnesses, continuation gains can be realized by transferring action-independent probability mass. This construction does not realize arbitrary supplied matrices or observation masks.

Lemma 4.2 (Legal stochastic realization of cross-block gains)**.** *Let a local two-state block b have high-state probabilities $k_b(h,a)$ and low-state probabilities $1-k_b(h,a)$. Set $S_b = \sum_c C_{bc}/\gamma$. Keep k_b unchanged, replace the local low-state mass by $1 - k_b - S_b$, and allocate C_{bc}/γ to the high state of c from the high state of b, and to the low state of c from the low state of b. For these action-independent allocations, the added span term is exactly $C_{bc}[V_c(1) - V_c(0)]$. The kernel is stochastic whenever*

$$\max_{h,a} k_b(h,a) + \sum_c C_{bc}/\gamma \;\le\; 1. \tag{5}$$

Proof. The inequality makes every remaining low-state mass nonnegative and each row still sums to one. The subtracted term $\gamma S_b V_b(0)$ cancels between the high and low evaluations; each new channel contributes the claimed discounted span difference. Action-independent transfers leave own-action payoff gaps unchanged. □

The mixed witness reports slacks 0.058421, 0.473684, and 0.637368 for (5). These checks concern that construction; the separate supplied-matrix diagnostic does not verify stochastic realization.

4.3 Directional cancellation and attainable span bounds

The following distinction separates loss of one cancellation direction from strict slack in a worst-case norm bound.

Proposition 4.3 (Constant-direction cancellation)**.** *Let $P(\cdot \mid h,a)$ be a legal transition kernel on a finite state space and fix two actions with $P(\cdot \mid h,a) \neq P(\cdot \mid h,a')$. Then*

$$\sum_s \big[P(s \mid h,a) - P(s \mid h,a')\big] = 0, \tag{6}$$

so $d = P(\cdot \mid h,a) - P(\cdot \mid h,a')$ has both signs. The bound $|d^\top \delta| \le |d|^\top |\delta|$ is strict for $\delta = \mathbf{1}$, but is attained for $\delta = \operatorname{sign}(d)$. In particular its induced ℓ_∞-to-absolute norm bound $\|d\|_1$ is exact.

Proof. The row sums give $d^\top \mathbf{1} = 0 < \|d\|_1$. For $\delta_s = \operatorname{sign}(d_s)$, both sides equal $\|d\|_1$. □

Thus full-coordinate absolute comparisons need not be strictly conservative. On two states, the span coordinate $x_i = V_i(1) - V_i(0)$ removes the constant direction and permits the following explicit attainability construction.

Theorem 4.4 (Multi-block nonnegative span criterion)**.** *Let F_λ be a continuously differentiable map on a closed invariant region $\mathcal{R} \subset \mathbb{R}^d$, and suppose that on $\mathcal{R}$ it satisfies the componentwise bound*

$$|F_\lambda(x) - F_\lambda(y)| \le G(\lambda)|x-y|, \qquad G(\lambda) = D + \sum_{k=1}^m \lambda_k C_k,$$

where $D, C_k \ge 0$ and $\lambda \in \mathbb{R}^m_+$. Define

$$\mathcal{C} = \{\lambda \in \mathbb{R}^m_+ : \rho(G(\lambda)) < 1\}.$$

Then $\mathcal{C}$ is downward closed, and the map $\lambda \mapsto \rho(G(\lambda))$ is continuous. Every boundary point of $\mathcal{C}$, relative to $\mathbb{R}^m_+$, satisfies $\rho(G(\lambda)) = 1$. Every $\lambda \in \mathcal{C}$ gives a contraction in a positive weighted coordinate max norm and, on $\mathcal{R}$, a unique fixed point together with the usual resolvent error tube. If $G(\lambda)$ is irreducible and $C_k \neq 0$, increasing λ_k strictly increases the spectral radius whenever the resulting matrix remains irreducible. If, in addition, the Jacobian of F_λ at a fixed point equals $G(\lambda)$, then $\rho(G(\lambda)) > 1$ gives a nonzero nonnegative linearized unstable direction.

Proof. For nonnegative matrices, entrywise order implies spectral-radius order, and the Perron–Frobenius theorem gives strict order when the larger matrix is irreducible and the increment is nonzero. Thus $0 \le \mu \le \lambda$ and $\lambda \in \mathcal{C}$ imply $\rho(G(\mu)) \le \rho(G(\lambda)) < 1$, proving downward closure. Continuity of the spectral radius implies that points with radius strictly below or strictly above one have relative neighborhoods inside or outside $\mathcal{C}$, respectively. Every relative boundary point therefore has radius one; the converse is not required. For $\lambda \in \mathcal{C}$, the resolvent $(I - G(\lambda))^{-1}$ exists and provides a positive weight vector; the weighted max-norm contraction and the fixed-point error tube follow from the small-gain argument in Theorem 4.1. Finally, when the Jacobian equals $G(\lambda)$, Perron–Frobenius supplies a nonnegative eigenvector with eigenvalue $\rho(G(\lambda)) > 1$, yielding the stated linearized instability. □

Corollary 4.5 (General symmetric two-block span criterion)**.** *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be continuously differentiable near a centered fixed point $x^\star$, and suppose that its Jacobian and a valid nonnegative comparison matrix at $x^\star$ both have the form*

$$J_F(x^\star) = G(L) = \begin{bmatrix} b & cL \\ cL & b \end{bmatrix}, \qquad 0 \le b < 1, \quad c \ge 0, \quad L \ge 0.$$

If $c > 0$, then the comparison radius is $\rho(G(L)) = b + cL$. Consequently, the local contraction condition is $L < (1-b)/c$, the boundary $L = (1-b)/c$ is not certified by a strict contraction argument, and $L > (1-b)/c$ gives a linearized unstable direction for the stated Jacobian. For $c = 0$, the condition reduces to $b < 1$ independently of L.

Proof. The eigenvalues of the displayed symmetric matrix are $b+cL$ and $b-cL$. Because $G(L) \ge 0$, its spectral radius is the larger eigenvalue $b + cL$. The weighted small-gain result in Theorem 4.1 therefore gives a strict local contraction whenever $b + cL < 1$. At equality the linearization has unit radius, so strict contraction is unavailable and linearization alone is inconclusive. For $b + cL > 1$, the positive eigenvalue of the Jacobian has modulus greater than one, so its Perron eigendirection is linearly unstable. If $c = 0$, the matrix is bI and the stated reduction follows. □

The criterion is attained by the explicit two-state general-sum game in Example A.1. Its span Jacobian equals its comparison matrix; the concrete coefficients and threshold are reported only in that example. The construction and verification appear in Appendix A.1.

Let $R(L)$ be the supremum of positive normalized box radii satisfying the Jacobian and image bounds in Appendix A.1, with value zero if none exists. A box of normalized radius r has physical span radius τr.

Corollary 4.6 (Sharpness as a function of the externality)**.** *For the family of Example A.1, $R(L) > 0$ for $0 \le L < L^\star$, is continuous and strictly decreasing there, and tends to zero as $L \uparrow L^\star$. For $L \ge L^\star$, no positive centered box satisfies this strict contraction certificate.*

Proof. The appendix derives the explicit Jacobian and image-radius inequalities for the centered box. Their strict-contraction margin is positive at $r = 0$ exactly when $b + cL < 1$, which is equivalent to $L < L^\star$. Continuity of the logistic map and of the resulting finite-dimensional bounds gives a positive admissible radius and continuity of the supremal radius throughout the subcritical interval. As L approaches the threshold, the margin at the center tends to zero, so every admissible radius tends to zero. For $L \ge L^\star$, the center already has comparison radius at least one by (7); hence no positive centered box can satisfy the strict certificate. The full expressions are given in Appendix A.1. □

General-sum response sensitivity. Zero-sum minimax values are nonexpansive in the payoff matrix, giving the standard γ-contractive Shapley backup. Here each player's externality leaves its own logit equation unchanged but couples its raw value to the opponent's response. The resulting derivative $0.0456L$ crosses the small-gain threshold at fixed discount. This is a property of the stated general-sum construction, not an obstruction to zero-sum contraction.

Table 1: Revealing an administrative action changes the game, while hidden serialization is numerically exact.

$\lvert\mathcal{A}_i\rvert$	Games	Distorted (%)	Norm. span	P95 span	Hidden resid.
2	1000	32.9 [30.1,35.9]	0.097	1.107	2.2×10^{-16}
3	1000	70.3 [67.4,73.1]	0.159	1.390	4.4×10^{-16}
5	1000	95.8 [94.4,96.9]	0.213	1.649	6.1×10^{-16}

4.4 Selectors are part of the semantics

An unrestricted exact Nash selector is not a smooth black box: a simple dominance switch forces a selected probability to jump across the tie. Thus a general-sum block must restrict to a regular single-valued branch, quantify an approximate selector's regret, use a separate set-valued analysis, or abstain. The sharp Lipschitz frontier and its logit comparison are supporting results in Appendix A.2; selector constants are not used to define the protocol distinction.

4.5 Approximate and neural block oracles

For an approximate block map f_B, the verifier must provide three objects on the declared region: an image enclosure, a nonnegative Jacobian matrix envelope M_B, and a semantic residual β_B relative to the protocol-faithful map. Samples alone are inadmissible. In the one-layer ReLU audit, a 21-by-21 grid misses a width-0.004 activation band and would incorrectly certify the unstable oracle; activation-pattern analysis gives composed radius 1.206. The same three-object interface can be instantiated with multilayer ReLU relaxations, but the deep branch-and-bound and exact-pattern refinements are supporting diagnostics in Appendix B; they do not expand the paper's coverage claim. Enumeration is exact as a combinatorial description; floating-point feasibility, support LPs, and eigensolvers lack outward error certificates. Their verdicts are numerical diagnostics conditional on solver accuracy, and any unresolved LP must cause abstention. The gain must bound the map claimed contractive: a bound on f_B alone does not prove contraction of T_B. If only f is certified, the analogous error-tube argument compares its fixed point to an existing fixed point of T in the same region, without establishing uniqueness for T.

5 Experiments

The experiments use fixed seeds and saved machine-readable outputs. Matrix games test hidden/revealed semantics, PettingZoo probes a third-party API, structured compositions test coordinate retention, and neural controls test the limits of sampled evidence. The appendix gives proofs and numerical details; these diagnostics do not measure learner performance.

Table 12 in the appendix maps each claim to its evidence. The claims concern the specified joint laws, comparison bounds, and numerical oracle examples, rather than arbitrary protocol compilation.

5.1 Protocol distortion

We solve square zero-sum games by primal and dual LPs. For each random matrix, we compare simultaneous play, hidden serialization, and two revealed orders. Table 1 reports 1000 independent games per action size and Wilson 95% intervals for the fraction with nonzero revealed-order distortion.

The distortion fraction rises from 32.9% at two actions to 95.8% at five actions. Hidden-order residuals are below 6.2×10^{-16}; the maximum saved LP duality gap is $8.881784197001252\times10^{-16}$. These are numerical protocol checks, not learner comparisons.

As a separately specified finite stress suite, we also audit five explicit canonical matrices: matching pennies, biased matching pennies, rock–paper–scissors, five-action cyclic dominance, and a rectangular payoff stress case. All five preserve hidden semantics and all five exhibit revealed-order distortion; the full values are reported in Table 9.

Table 2: MPE protocol stress test: a fixed observation-only base policy and an action-dependent revealed variant.

Environment	Episodes	Cycles	Parallel–hidden gap	Revealed action divergence	Return delta	Hidden checks
Simple Spread v2	5	25	0	77.1%	-9.696	true

Table 3: Retaining structured coordinates can certify cases rejected by scalar bounds.

System	Coarse modulus	Retained ρ	Retained structure	Verdict
Mixed QRE/sequential blocks	1.096	0.638	block spectral	certify
Asymmetric QRE blocks	1.079	0.596	coordinate spectral	certify
Shared ReLU recurrence	1.200	0.809	Perron weighted	certify

As a separately specified API check, we additionally extract the payoff and observation behavior of PettingZoo 1.22.2's standard RPS environment using its own environment API, then pass only the extracted JSON to a separate LP auditor. The hidden checks remain exact while the revealed-order span is 2; the Tic-Tac-Toe environment is retained as a negative control because its board and legal-action mask evolve after every move. This is evidence that the protocol contract is observable in a third-party implementation, not a claim of broad benchmark coverage; the extracted values and negative control are listed in Table 10.

We then use PettingZoo MPE Simple Spread v2 only as a fixed-policy protocol stress test, carrying the same distinction into continuous-state multi-agent dynamics without training a learner. Table 2 reports the main comparison: across five fixed seeds and 25 cycles, native parallel execution and hidden AEC serialization have zero mean-return gap and identical action traces, while an action-revealing AEC policy changes 77.1% of actions and shifts mean return by -9.696. Sweeping the leak horizon in the same fixed-policy setting (Table 7) shows increasing action divergence and return loss over these five seeds: letting later agents see one earlier same-cycle action changes 75.2% of actions and shifts return by -5.398, and two actions give 77.1% and -9.696. The revealed policy explicitly uses earlier actions, so this comparison tests that dependence, not the effect of disclosure on an unchanged policy.

A 200-episode evaluation on the same seed schedule preserves hidden/parallel traces; 20 additional resets also pass all-agent pre-commit observation checks. These remain fixed-policy diagnostics. The paired replay return change is positive, so disclosure does not consistently reduce return (Appendix B.1); the long-run summary remains provenance-limited (Table 8).

5.2 Mixed block composition

The mixed witness contains simultaneous logit-QRE blocks and a fixed-policy linear sequential block connected by the stated stochastic construction. Its origin comparison has radius 0.638 but scalar row-sum modulus 1.096; a region along a Perron ray reaches radius target 0.95 with positive transition slack. In the asymmetric witness, block scalarization has radius 1.079, whereas the coordinate matrix has radius 0.596. This tests structured bounds within the declared construction, not general sequential-game solving.

The forcing experiment provides a semantic check rather than only a spectral number: with block forcing $[10^{-4}, 2 \times 10^{-4}, 10^{-4}]$, the observed fixed-point error remains below the componentwise resolvent tube and the forced trajectory stays inside the certified region.

5.3 Certificate tightness

Table 4 instantiates Example A.1 and Corollary 4.6 on the explicit legal two-state general-sum family with $\gamma = 0.95$ and gain scale 3. In panel (a), at the critical externality the span Jacobian radius and the span comparison radius agree to nine decimals and the comparison matrix is entrywise tight at every audited temperature. Central finite differences check the analytic Jacobian; the proof establishes stochasticity and temperature independence for all $\tau > 0$. The saved threshold is identical at $\tau \in \{0.05, 0.2, 1, 2\}$. Panel (b) reports numerical radii from 2.911440 at $L = 0$ to 0.044585 at $L = 4.2$. The search returns zero within 10^{-12} of the local spectral threshold; continuous collapse

Table 4: Certificate tightness. (a) The span comparison bound is attained entrywise at the critical externality, uniformly in temperature. (b) The certified nonlinear region has supremum tending continuously to zero at the strict local threshold.

(a) Entrywise tightness at $L^\star$

τ	$\rho(J_s)$	$\rho(G_s)$	Entrywise tight	FD error	Row-sum error
0.05	1.000000000	1.000000000	yes	2.7×10^{-11}	0
0.2	1.000000000	1.000000000	yes	9.8×10^{-11}	0
1	1.000000000	1.000000000	yes	9.0×10^{-10}	0
2	1.000000000	1.000000000	yes	4.3×10^{-10}	0

(b) Regional collapse

L	Local ρ	Certified radius	Contraction certified
0	0.804215	2.911440	yes
1	0.849815	2.025970	yes
2	0.895415	1.291808	yes
3	0.941015	0.672525	yes
4	0.986615	0.141871	yes
4.2	0.995735	0.044585	yes
4.29353	1.000000	0.000000	no
5	1.032215	0.000000	no

for real L follows from Corollary 4.6, not that numerical cutoff. The supremum radius itself lies on the excluded strict-contraction boundary.

5.4 Evidence failure controls

Activation-pattern analysis gives composed radius 0.810 for the stable one-layer oracle; a coordinate-wise decoupled slope relaxation gives 1.206 and cannot certify it. The unstable control has radius 1.206, while grid sampling reports 0.437 and would incorrectly certify it. These are numerical diagnostics of the evidence requirements. A refutation requires the signed composed Jacobian at a differentiable interior point, not the radius of its absolute comparison. Deep and recurrent examples are in Table 5.

5.5 Role of semantic residuals

On the stored 36-candidate audit, structured small-gain returns 18 certifications, 9 abstentions, and 9 refutations, with no wrong accepts. Interval and affine rules match the same 18/9/9 counts because both use the same witness gate and envelope-radius facts—not because the two rules provide separate evidence. Dropping the semantic residual produces 9 wrong accepts, with an exact rational gap of $2/3$ on the bias coordinates, so the residual is load-bearing for soundness on this set. All passes are numerical predicate checks; the stored formal-certificate flag is false on every row. We therefore do not claim a formal certificate, rule equivalence, or a return ranking. The ablation is shown in Figure 2 in the appendix.

6 Limitations

The semantics require public continuation histories, a preserved joint policy law, and declared selectors; arbitrary imperfect-information masks are not covered. The strongest constructive result is the explicit two-state span family and its regional threshold. It proves attainability, not necessity of small gain for general nonlinear stability, nor strict worst-case slack in full coordinates. The supplied-matrix diagnostic never certifies game semantics. Neural LP and spectral computations are numerical and lack certified rounding bounds. The MPE evidence concerns a fixed policy family, and its long-run completion and version labels have the provenance limitations stated in the appendix. Runtime measurements concern one implementation and machine, not comparative speed.

7 Conclusion

Block-SeBIL separates administrative sampling from strategic information under an explicit joint-law contract. Matrix gains, invariant regions, and uniform residuals then give conditional contraction and error bounds. The legal span family attains the local comparison bound and has a continuously vanishing certified regional radius. The accompanying diagnostics illustrate these assumptions without establishing a general semantic compiler; the long-run cycle count is reported only as provenance.

Reproducibility Statement

All reported diagnostics are backed by archived machine-readable summaries and the generated tables supplied with the paper. The accompanying artifact contains the source code, input records, environment specifications, and instructions for regeneration. Numerical provenance does not validate semantic verdicts, floating-point bounds, or runtime versions. The MPE evidence carries the version labels recorded with each run; these labels are metadata rather than runtime attestations. The short smoke record includes producer-recorded configuration hashes and an episode ledger, but it does not attest the historical long-run completion.

Ethics Statement

This work studies finite game-solving and algorithmic verification without human-subject data. Protocol choices can change strategic outcomes in deployed multi-agent systems; the paper therefore reports the revealed-order failure mode and does not claim universally fair or universally optimal action orderings.

AI Use Statement

Generative AI tools assisted with theoretical analysis, code development, and experiment execution and analysis. The authors take responsibility for the final content.

References

[1] Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, 1994.

[2] Rudy Bunel, Jingyue Lu, Ilker Turkaslan, Philip H. S. Torr, Pushmeet Kohli, and M. Pawan Kumar. Branch and bound for piecewise linear neural network verification. *Journal of Machine Learning Research*, 21(42):1–39, 2020.

[3] Shicong Cen, Yuting Wei, and Yuejie Chi. Fast policy extragradient methods for competitive games with entropy regularization. *Journal of Machine Learning Research*, 25(4):1–48, 2024.

[4] Sergey Dashkovskiy, Björn S. Rüffer, and Fabian R. Wirth. An ISS small-gain theorem for general networks. *arXiv preprint math/0506434*, 2005. URL `https://arxiv.org/abs/math/0506434`.

[5] Constantinos Daskalakis, Noah Golowich, and Kaiqing Zhang. The complexity of markov equilibrium in stochastic games. In *Proceedings of the 36th Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 4180–4234, 2023.

[6] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pp. 2974–2982, 2018.

[7] Matthieu Geist, Bruno Scherrer, and Olivier Pietquin. A theory of regularized markov decision processes. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2160–2169, 2019.

[8] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 1861–1870, 2018.

[9] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2012.

[10] Daphne Koller and Brian Milch. Multi-agent influence diagrams for representing and solving games. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pp. 1027–1034, 2001. URL `https://people.csail.mit.edu/milch/papers/ijcai01maids.html`.

[11] Martin Lauer and Martin Riedmiller. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 535–542, 2000.

[12] Jialian Li, Yichi Zhou, Tongzheng Ren, and Jun Zhu. Exploration analysis in finite-horizon turn-based stochastic games. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pp. 201–210, 2020.

[13] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the Eleventh International Conference on Machine Learning*, pp. 157–163, 1994.

[14] Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995. doi: 10.1006/game.1995.1023.

[15] Julien Perolat, Remi Munos, Jean-Baptiste Lespiau, Shayegan Omidshafiei, Mark Rowland, Pedro Ortega, Neil Burch, Thomas Anthony, David Balduzzi, Bart De Vylder, Georgios Piliouras, Marc Lanctot, and Karl Tuyls. From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8525–8535, 2021.

[16] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 2014.

[17] Lloyd S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39 (10):1095–1100, 1953.

[18] Zhouxing Shi, Qirui Jin, Zico Kolter, Suman Jana, Cho-Jui Hsieh, and Huan Zhang. Neural network verification with branch-and-bound for general nonlinearities. *arXiv preprint arXiv:2405.21063*, 2024.

[19] J K Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, Niall Williams, Yash Lokesh, and Praveen Ravi. Pettingzoo: Gym for multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

[20] Qiaomin Xie, Yudong Chen, Zhaoran Wang, and Zhuoran Yang. Learning zero-sum simultaneous-move markov games using function approximation and correlated equilibrium. In *Proceedings of the 33rd Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 3674–3682, 2020.

[21] Huan Zhang, Tsui-Wei Weng, Pin-Yu Chen, Cho-Jui Hsieh, and Luca Daniel. Efficient neural network robustness certification with general activation functions. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

A Additional Proof Details

A.1 Explicit span map and regional certificate

Example A.1 (A legal Markov-game realization)**.** The general criterion of Theorem 4.4, with the symmetric two-block case given by Corollary 4.5, is attained by the following legal two-state, two-player, two-action general-sum discounted Markov-game construction. Set $\gamma = 0.95$, use gain scale 3, and choose the centered target QRE probabilities $(0.8, 0.8)$. Its transition probabilities to state 1 are 0.01 at state 0 and $0.01 + 3\left[\begin{smallmatrix} 0.10 & 0.20 \\ 0.20 & 0.30 \end{smallmatrix}\right]$ at state 1, with complementary probabilities to state 0. For actions $a_1, a_2 \in \{0, 1\}$, both states have rewards $r_1 = \tau(\ell a_1 + L a_2)$ and $r_2 = \tau(L a_1 + \ell a_2)$, where $\ell = \log 4$. The only varying game parameter is the externality $L \geq 0$; $\tau > 0$ is a common reward scale. Under logit selection and raw-value evaluation, the centered span Jacobian and its comparison matrix coincide entrywise:

$$J_s(L) = G_s(L) = \begin{bmatrix} b & cL \\ cL & b \end{bmatrix} \simeq \begin{bmatrix} 0.8042150228670671 & 0.0456\,L \\ 0.0456\,L & 0.8042150228670671 \end{bmatrix}. \tag{7}$$

Here the fixed numerical coefficients are $b = 0.285(2.6 + 0.16 \log 4) = 0.8042150228670671$ and $c = 0.0456$. Hence $\rho(J_s(L)) = \rho(G_s(L)) = b + cL$. The centered fixed point is locally asymptotically stable for $L < L^\star$ and unstable for $L > L^\star$, where

$$L^\star = (1 - b)/c \simeq 4.293530200283615, \tag{8}$$

independently of the reward scale τ. At equality the linearization has radius one and does not give a strict contraction or, by itself, a nonlinear stability verdict. This is a sharpness result for the displayed construction, not a universal threshold for general-sum Markov games.

Verification. Set $x = \tau y$ and $a = 0.285$. State-1 responses are $p = \sigma(\ell + ay_1)$ and $q = \sigma(\ell + ay_2)$; state-0 responses are both 0.8. Differentiating the normalized raw-value map at zero gives diagonal $2.6a + 0.16a\ell$ and off-diagonal $0.16aL$. Both channels are nonnegative, so their absolute comparison is attained. Appendix A.1 gives the map, derivatives, and regional proof. Write $\ell = \log 4$, $a = 0.95 \cdot 0.3 = 0.285$, and $\sigma(z) = (1 + e^{-z})^{-1}$. Rows and columns below index actions $0, 1$. The state-independent reward matrices and probabilities of next state 1 are

$$r_1 = \tau \begin{bmatrix} 0 & L \\ \ell & \ell + L \end{bmatrix}, \qquad r_2 = \tau \begin{bmatrix} 0 & \ell \\ L & \ell + L \end{bmatrix}, \qquad P_0 = \begin{bmatrix} 0.01 & 0.01 \\ 0.01 & 0.01 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 0.31 & 0.61 \\ 0.61 & 0.91 \end{bmatrix}. \tag{9}$$

All entries lie in $[0, 1]$ and the complementary probability goes to state 0. Let $x_i = V_i(1) - V_i(0) = \tau y_i$. The common continuation constant $\gamma V_i(0)$ cancels from action differences and from the difference between state values. At state 0 both own-action gaps are $\tau\ell$; at state 1 they are $\tau\ell + ax_i$, independent of the opponent's action. Consequently the logit responses are uniquely $p = \sigma(\ell + ay_1)$ and $q = \sigma(\ell + ay_2)$ at state 1, and $p_0 = q_0 = 0.8$ at state 0. Raw expected payoffs give

$$H_L(y) = \frac{T_{s,\tau}(\tau y)}{\tau} = \begin{bmatrix} \ell(p - 0.8) + L(q - 0.8) + ay_1(1 + p + q) \\ L(p - 0.8) + \ell(q - 0.8) + ay_2(1 + p + q) \end{bmatrix}. \tag{10}$$

In particular $H_L(0) = 0$ and the normalized map is independent of temperature. With $p' = ap(1-p)$ and $q' = aq(1 - q)$, differentiation gives

$$DH_L(y) = \begin{bmatrix} a(1 + p + q) + (\ell + ay_1)p' & (L + ay_1)q' \\ (L + ay_2)p' & a(1 + p + q) + (\ell + ay_2)q' \end{bmatrix}. \tag{11}$$

At zero, $p' = q' = 0.16a$, so the direct diagonal is $2.6a$ and the response terms have diagonal $0.16a\ell$ and off-diagonal $0.16aL$. All are nonnegative for $L \geq 0$. Their entrywise-absolute sum therefore equals the signed Jacobian, proving (7). Its eigenvalues are $b \pm cL$, where $b = 2.6a + 0.16a\ell$ and $c = 0.16a = 0.0456$. Thus $\rho = b + cL$ and $L^\star = (1 - b)/c$. Below this threshold, continuity gives a contractive neighborhood; above it the symmetric eigenvector has eigenvalue greater than one, giving instability of the smooth fixed-point iteration. Equality is outside the strict linear stability test.

For $r \geq 0$, let $\mathcal{B}_r = \{y : \|y\|_\infty \leq r\}$ and define

$$\begin{aligned} p_\pm(r) &= \sigma(\ell \pm ar), \qquad d_r = a \max_{u \in [p_-(r), p_+(r)]} u(1 - u), \\ \Delta_r &= \max\{0.8 - p_-(r), p_+(r) - 0.8\}, \\ g_d(r) &= \ell d_r + a(1 + 2p_+(r)) + ard_r, \qquad g_o(r, L) = (L + ar)d_r, \\ F(r, L) &= g_d(r) + g_o(r, L), \qquad I(r, L) = (\ell + L)\Delta_r + ar(1 + 2p_+(r)). \end{aligned} \tag{12}$$

Equation (11) gives the uniform envelope $|DH_L(y)| \leq G_r = \left[\begin{smallmatrix} g_d & g_o \\ g_o & g_d \end{smallmatrix}\right]$ on $\mathcal{B}_r$, with $\rho(G_r) = F(r, L)$. Equation (10) gives $\|H_L(y)\|_\infty \leq I(r, L)$. The regional certificate is

$$F(r, L) < 1, \qquad I(r, L) \leq r, \qquad R(L) = \sup\big(\{r > 0 : F(r, L) < 1,\ I(r, L) \leq r\} \cup \{0\}\big). \tag{13}$$

Every accepted box is closed, invariant, and contractive, so its unique span fixed point is zero and iteration converges from that box. Its radius in the original span coordinates is τr.

Proof of Corollary 4.6. The response derivative bound gives $\Delta_r \leq rd_r$. Hence

$$I(r, L) \leq r\big[(\ell + L)d_r + a(1 + 2p_+(r))\big] = r\big[F(r, L) - 2ard_r\big].$$

The image condition is therefore automatic when $F(r, L) < 1$. The nested intervals $[p_-(r), p_+(r)]$ make d_r continuous and nondecreasing, with $d_r \geq d_0 = 0.16a > 0$. Thus $F(r, L) = (\ell + L + 2ar)d_r + a(1 + 2p_+(r))$ is continuous, strictly increasing in r, unbounded as $r \to \infty$, and strictly increasing in L. Since $F(0, L) = b + cL$, for $L < L^\star$ there is a unique positive root $r_*(L)$ of $F(r, L) = 1$. The accepted radii are exactly $0 < r < r_*(L)$, so $R(L) = r_*(L)$ as a supremum, not an accepted endpoint. Strict monotonicity gives strict decrease in L. Continuity follows by bracketing the unique root between any two nearby radii and using continuity of F in L. For every $\epsilon > 0$, $F(\epsilon, L^\star) > F(0, L^\star) = 1$; continuity then gives $R(L) < \epsilon$ for all sufficiently close subcritical L. For $L \geq L^\star$, $F(r, L) > 1$ at every $r > 0$, completing the proof. □

This statement concerns the analytic real-valued bounds. The numerical search uses a 10^{-12} local-threshold cutoff and returns an interior approximation to the supremum; it cannot resolve all arbitrarily close subcritical cases.

A.2 Selector regularity and regret

Consider two players with actions $\{0, 1\}$ and payoffs $u_i(a_i, a_{-i}; t) = ta_i$, $t \in [-1, 1]$. For either player's action-1 probability $s(t)$, measure input distance by $|t - t'|$ and output distance by $|s(t) - s(t')|$. For a two-player marginal vector use the coordinate maximum; the same scalar result applies to each coordinate. Define deviation regret

$$\epsilon(s) = \sup_{t \in [-1,1]} \begin{cases} t(1 - s(t)), & t \geq 0, \\ (-t)s(t), & t < 0. \end{cases}$$

Proposition A.2 (Exact selector frontier on the dominance-switch family)**.** *No exact single-valued Nash selector on this family is continuous at zero. For $K \geq 0$, the infimum of $\epsilon(s)$ over all K-Lipschitz $s : [-1, 1] \to [0, 1]$ is*

$$\epsilon^*(K) = \begin{cases} \frac{1}{2} - K, & 0 \leq K \leq \frac{1}{4}, \\ \frac{1}{16K}, & K \geq \frac{1}{4}. \end{cases} \tag{14}$$

The permutation-neutral selector $s_K(t) = \min\{1, \max\{0, \frac{1}{2} + Kt\}\}$ attains this frontier. The product bound $K\epsilon \geq 1/16$ consequently requires $K \geq 1/4$ on this domain.

Proof. Exact Nash selection forces $s(t) = 1$ for $t > 0$ and $s(t) = 0$ for $t < 0$. For an approximate selector and $0 < \delta \leq 1$, the two regret inequalities give $s(\delta) \geq 1 - \epsilon/\delta$ and $s(-\delta) \leq \epsilon/\delta$. Lipschitz continuity implies $1 - 2\epsilon/\delta \leq 2K\delta$, or $\epsilon \geq \delta/2 - K\delta^2$. Maximizing over $\delta \in [0, 1]$ gives (14): the maximizer is 1 for $K \leq 1/4$ and $1/(4K)$ otherwise. For s_K, regret at $|t| = \delta$ equals $\max\{0, \delta/2 - K\delta^2\}$, attaining the same maximum. Finally $s_K(-t) = 1 - s_K(t)$ proves neutrality. □

This frontier is exact only for the stated family and norm. It is a lower-bound witness for unrestricted game selectors, not an upper bound for every game. For example $K = 0.1$ gives $\epsilon^* = 0.4$ and product $0.04 < 1/16$.

For comparison, $s_\tau(t) = \sigma(t/\tau)$ has Lipschitz constant $K = 1/(4\tau)$ and regret $t/(1 + e^{t/\tau})$ for $t \geq 0$. Writing $w = W(e^{-1})$, its stationary maximum occurs at $t = \tau(1 + w)$. Thus the maximum is τw when $\tau(1 + w) \leq 1$, and $1/(1 + e^{1/\tau})$ otherwise. In the former regime only, its ratio to (14) at the same K is $4w \simeq 1.113858$. This does not establish uniqueness or smoothness of QRE in general-sum games.

A.3 Comparison recursion

For coordinate radii r and a local perturbation δ, the verified block bounds give

$$|T_B(V + \delta) - T_B(V)| \leq M_B(r_B)|\delta_B| + \sum_C C_{BC}|\delta_C|. \tag{15}$$

Stacking coordinates and adding $|f(V) - T(V)| \leq \beta$ gives (4). If $G \geq 0$, $w > 0$, and $Gw \leq qw$ with $q < 1$, the induced norm of G under $\max_i |x_i|/w_i$ is at most q. The residual series is $\sum_{k \geq 0} G^k \beta = (I - G)^{-1}\beta$. For iterates that remain in the region, $e_k \leq G^k e_0 + \sum_{j=0}^{k-1} G^j \beta$; this gives an asymptotic error tube without requiring convergence of the approximate iteration.

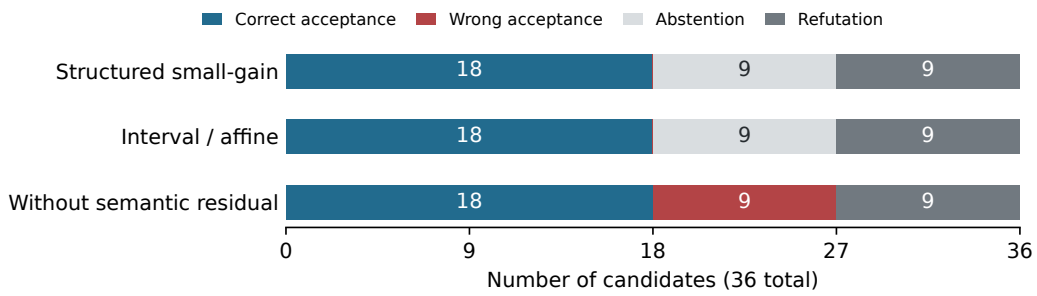

Figure 2: Semantic-residual ablation on the stored 36-candidate audit. The residual prevents nine wrong accepts. Removing it increases total acceptance from 18 to 27 without adding a correct acceptance. Interval and affine rules share the same underlying bounds; their agreement is not independent evidence.

A.4 Exact-pattern leaf verifier

For each activation pattern, preactivation sign constraints define a polyhedron intersected with the input box. The network is affine there, so output supports are LPs; residual supports are LPs when the target is also affine on that region (otherwise further bounds are required). The union over all feasible patterns describes the network exactly in real arithmetic. Only justified infeasibility permits omission: iteration limits, numerical failures, and unknown statuses must propagate as abstention. The current floating-point LP and eigenvalue outputs have no outward-rounding or independently checked dual certificates, so enumeration does not make the numerical bounds exact.

A.5 Finite recurrent unrolling

For $h_{t+1} = \mathrm{ReLU}(Ah_t)$, each one-step derivative has the form $D_t A$, with diagonal entries of D_t in $[0, 1]$. Products, and their generalized-derivative limits, satisfy $|J_H| \le |A|^H$ entrywise. If $\rho(|A|) < 1$, the weight $w = (I - |A|)^{-1}\mathbf{1}$ gives an invariant box $\{|h| \le rw\}$ and geometric decay. A strictly all-active trajectory realizes $J_H = A^H$, which need not satisfy $|A^H| = |A|^H$ for signed A. Equality holds for the nonnegative matrices used in the examples, provided the trajectory stays strictly active; any instability witness must use the actual signed power and an admissible trajectory.

B Additional Experiments

Residual sensitivity on the nonlinear game. For the span map in Appendix A.1, we add a constant signed error to each coordinate, leaving the Jacobian unchanged. We test $L \in \{0, 2, 4\}$, box radii $0.25R(L)$ and $0.75R(L)$, all four error-sign pairs, and error magnitudes 0.01, 0.1, and 0.5 times the available image margin $r - I(r, L)$. All 72 perturbed iterations remain inside their boxes and reach residual below 10^{-13}. Their fixed-point errors use between 0.009 and 0.787 of the comparison tube $\beta/[1 - F(r, L)]$. Every constant-bias perturbation changes the fixed point while preserving the Jacobian at every common input. Thus sensitivity alone cannot replace a semantic-residual bound. These are floating-point checks of the stated nonlinear game, not certified arithmetic.

For the nonnegative matrices and strictly active trajectories in the recurrent examples, the numerical audit gives $|J_H| = |A|^H$ at horizons $1, 2, 4, 8$. The stable example has one-step row-sum modulus 1.2 but spectral radius 0.808545 and horizon-8 radius 0.182656. The unstable control grows from 1.030317 to 1.269895 by horizon 8. This does not extend the equality to signed recurrence or provide a treatment of gated recurrence.

Table 6 reports recomputed implementation medians for zero-sum matrix games up to 64×64, activation-pattern feasibility enumeration up to 4096 patterns, and the relaxed and hybrid deep checks. Each timing uses five repeats; these measurements include the repaired witness and solver-status checks and do not establish comparative speedups.

The supplied-matrix diagnostic checks an administrative DAG and numerical gain and residual arrays. Schema 2 requires explicit residuals and block protocol labels and preserves a topological order. It reports *conditional bound* or abstention for matrix arithmetic; its semantic verdict is always abstention. It does not verify observation masks, selectors, uniform bounds, invariant images, or stochastic realization. Cases up to 32 blocks test spectral/resolvent calculations and scalar row-sum controls under a common tube budget. Some matrices are rescaled to target spectral radii; they are not thereby realized games. Failure of either upper bound is not a refutation of contraction.

B.1 MPE EVIDENCE AND LIMITATIONS

Fresh replay and scheduling controls. A new execution of 200 paired episodes, each of 25 cycles, uses the historical seed schedule starting at 20260730. Its 15,000 aggregate joint steps across three arms reproduce hidden/parallel action traces and mean returns exactly; the revealed arm changes 76.71% of actions. The paired mean-return change per cycle is 0.0510, with a 95% episode-bootstrap interval $[0.0104, 0.0918]$ from 5,000 resamples. This verifies fresh execution on reused seeds, not held-out generalization or an algorithmic improvement.

On 20 additional resets (seeds 92600–92619), all 3,000 checks of every agent's observation after the first and second pre-commit actions agree with the pre-cycle observation. All six orders of computing hidden actions give the same joint action at each of 500 cycles. This permutes action computation, not the simulator's native stepping order. Holding observations fixed, the revealed-action rule changes 53.32% of the 9,000 compared actions; ignoring the transcript changes none. This conditional comparison excludes downstream state drift, unlike the separate-arm rollout statistic. Both studies execute Simple Spread v2 with PettingZoo 1.22.2 and fixed policies; the replay retains per-episode trace hashes, and the additional-reset control retains observations and actions.

An exact shared-randomness control also separates masks from joint-law fidelity. Two binary actions copied from one fair bit have the same individual marginals as independent fair bits, but their joint laws have total-variation distance $1/2$. The reward $\mathbf{1}\{a_1 = a_2\}$ falls from 1 to $1/2$ under independent resampling. Hiding current actions therefore does not suffice when the prescribed coupling is lost.

Historical long-run record. The following longer saved summary is a historical report whose retained provenance does not include a final runtime attestation or per-episode trace ledger. It is therefore used only as an arithmetic consistency check, not as an externally verified completion result. The summary reports $9{,}999{,}975$ aggregate joint steps across parallel, hidden, and revealed arms: 3,333,325 cycles per arm in 133,333 paired episodes, reset every 25 cycles (Table 8). It reports zero mismatched hidden/parallel episode traces, a maximum return gap below 10^{-13}, and 77.0% revealed-action divergence. The long-run return change is positive, 0.0022262543423703 per cycle, so leakage does not consistently reduce return. Action divergence includes downstream state differences. Completion provenance, stored trace coverage, and version-label limitations are detailed in Appendix B.1.

The final long-run JSON reports $133{,}333$ episodes of 25 cycles in each of three arms, $3{,}333{,}325$ cycles per arm, and $9{,}999{,}975$ aggregate joint steps. Its zero mismatch count concerns paired episode traces compared during execution, not an independent audit of stored per-step traces. The retained progress log ends at $1{,}200{,}000$ aggregate steps; no final per-episode trace/hash record accompanies the summary. This limits independent completion verification without showing that the final summary is false.

Observation checks inspect agent 1 after agent 0, not all masks, agents, or sampling permutations. The revealed variant uses *a deterministic transformation of the current action using previously revealed actions*; action divergence therefore combines direct use of revelation with subsequent state differences. Separate-arm marginal return deviations do not determine paired uncertainty. The short AEC output repeats a cycle-average reward in its return vector, so only the reported mean, not those entries as individual-agent returns, is used.

The saved MPE label is PettingZoo 1.22.2; the separate API probe records 1.22.3. These historical version labels were not checked against the installed packages. Counter consistency checks likewise cannot recover missing completion logs, environment hashes, or resume lineage. The five-seed

Table 5: Numerical oracle diagnostics. Pattern analysis and signed point witnesses address different obligations; LP/eigensolver precision is not formally certified.

Instance / evidence	Bound or witness ρ	Decision	Evidence status
Stable, decoupled slope bounds	1.206	abstain	upper bound only
Stable, enumerated patterns	0.810	certify	numerical bound
Expanding, 21x21 samples	0.437	invalid certify	false certification
Expanding, signed point	1.206	refute	point witness
Deep stable, root relaxation	1.206	abstain	upper bound only
Deep stable, BaB (15 cells)	0.437	certify	numerical bound
Deep stable, hybrid (1 cell)	0.437	certify	numerical bound

Table 6: Runtime and scaling. These are implementation medians, not cross-method speedup claims.

Primitive	Scale	Median ms	Unit / outcome
Protocol LP	2×2	1.64	per game
Protocol LP	64×64	5.63	per game
Activation enumeration	$k = 2$ (4 patterns)	2.92	feasibility only
Activation enumeration	$k = 12$ (4096 patterns)	3203.8	feasibility only
Deep BaB	15 cells / 120 LPs	105.6	certified
Hybrid exact leaf	1 cell / 31 LPs	23.7	certified

negative return change and the long-run positive change support protocol sensitivity, not a uniform performance penalty from disclosure.

The retained short smoke record is different: it records PettingZoo 1.22.2, Python 3.10.11, NumPy 2.2.6, dependency versions, producer source/config hashes, 750 aggregate joint steps (250 cycles per arm, 10 episodes per arm, and 2,250 individual actions), and a SHA-256 episode ledger for all ten episodes. This is producer-recorded attestation for the smoke run, not an independent signature and not evidence that the historical long run used the same installed environment. Its observation check covers agent 1 after agent 0 only; it does not cover every agent, mask, or sampling permutation.

C Interpretation of the Evidence

The mathematical contract requires valid pairwise gains on a closed invariant region and a uniform semantic residual. A loose upper bound above one yields abstention. A signed Jacobian of the actual composed map with spectral radius above one at a differentiable interior point refutes uniform contraction on a neighborhood of that point; away from a fixed point it does not establish equilibrium instability. Activation margins and witness coordinates are needed to validate a neural point witness. The supplied-matrix diagnostic establishes none of these semantic obligations. Numerical LP results are conditional on feasibility and optimization accuracy, with unresolved solver outcomes excluded from certification rather than omitted from the domain.

Table 7: Visibility-depth sweep in the same MPE protocol.

Visibility depth	Episodes	Mean return	Delta vs hidden	Action divergence	Trace match	Hidden checks
0	5	-38.856	0.000	0.0%	true	true
1	5	-44.254	-5.398	75.2%	false	true
2	5	-48.552	-9.696	77.1%	false	true

Table 8: MPE saved summary: 9,999,975 aggregate steps across three arms, 3,333,325 cycles per arm, with resets every 25 cycles.

Environment	Aggregate steps	Paired episodes	Gap	Trace mis.	Hidden checks	Revealed div.	Return delta/cycle
Simple Spread v2	9,999,975	133,333	9.9×10^{-14}	0	1.000	77.0%	0.0022

Table 9: Separately specified canonical protocol stress suite.

Canonical game	Size	Simultaneous	Revealed row-first	Revealed column-first	Span
matching pennies	2×2	-0.000	-1.000	1.000	2.000
biased matching pennies	2×2	0.500	-1.000	2.000	3.000
rock paper scissors	3×3	-0.000	-1.000	1.000	2.000
cyclic dominance 5	5×5	-0.000	-1.000	1.000	2.000
rectangular order stress	3×4	-0.000	-2.000	2.000	4.000

Table 10: Third-party protocol extraction and negative control.

External environment	Simult.	Hidden row	Revealed row	Revealed col.	Span	Hidden checks
RPS v2 (PettingZoo 1.22.2)	-0.000	0.000	-1.000	1.000	2.000	9
Tic-Tac-Toe v3 (negative control)	–	–	–	–	–	matrix contract: false

Table 11: Conditional supplied-matrix diagnostic. Matrix arithmetic uses a resolvent-weighted bound; the scalar control uses the maximum row sum. Neither verdict certifies game semantics or refutes contraction.

Supplied-matrix case	Blocks	Edges	$\rho(G)$	Row sum	Matrix	Scalar
rank_sensitive_hybrid_3	3	2	0.274	1.150	CONDITIONAL_BOUND	ABSTAIN
stable_parallel_6	6	5	0.620	0.833	CONDITIONAL_BOUND	CONDITIONAL_BOUND
near_boundary_hybrid_12	12	11	0.940	1.360	CONDITIONAL_BOUND	ABSTAIN
upper_bound_fails_8	8	7	1.040	1.359	ABSTAIN	ABSTAIN
invalid_cycle_4	4	0	–	–	ABSTAIN	ABSTAIN
scale_16	16	15	0.700	0.912	CONDITIONAL_BOUND	CONDITIONAL_BOUND
scale_24	24	23	0.880	1.097	CONDITIONAL_BOUND	ABSTAIN
scale_32	32	31	1.020	1.223	ABSTAIN	ABSTAIN

Table 12: Claim-to-evidence map for the main contract fields.

Contract field	Shortcut ruled out	Primary evidence	Allowed conclusion
Protocol semantics	exposed administrative order	Thms. 3.2, 3.3; Tabs. 1, 8	hidden equivalence; revealed distortion
Structured stability	scalar row-sum certificate	Thm. 4.1; Tab. 3	coordinate/block gains can certify
Certificate tightness	unquantified conservatism	Prop. 4.3, Ex. A.1; Tab. 4	constant-direction cancellation; span attainment
Selector contract	selector as implementation detail	Prop. A.2; selector-frontier JSON	declared selection; $K\epsilon \geq 1/16$ for $K \geq 1/4$
Neural evidence	sampled Jacobians as certificates	Tab. 5	image/Jacobian/residual required
Coverage boundary	universal verifier claim	Tab. 6; Sec. 6	finite, interface-specific, fail-closed