Meta-Games of Learning: When Algorithm Choice Is the Strategic Variable

Anonymous Authors[1]

Abstract

Classical game theory asks what strategy a player should choose; the multi-agent learning literature asks what learning algorithm a player should use. We study the question one level up: when rational agents freely choose *which* learning algorithm to deploy, they are playing a *meta-game*. We introduce the *Algorithm Nash Equilibrium* (ANE) and the *Algorithm Correlated Equilibrium* (ACE) as solution concepts for this meta-game, and develop a rigorous theory of their existence, structure, and complexity. Our central results are: (i) in zero-sum games, ANE coincides exactly with the minimax solution, so no new phenomena arise; (ii) in general-sum games, no pair of no-regret algorithms is an ANE—commitment strategies strictly dominate for any finite horizon beyond a computable threshold T^*; (iii) ANE are generically *asymmetric*, creating a Commitment Dilemma analogous to Prisoner's Dilemma at the algorithm level; and (iv) computing an approximate ANE is PSPACE-hard in general but PPAD-hard for bounded-memory algorithms. We define the *Price of Adaptivity* (PoA) to quantify the cost of deploying no-regret learning, prove $\mathrm{PoA} \geq 1/3$ for anti-coordination games, and validate all results experimentally across a broad game library and with LLM agents. Our findings have direct implications for the design of AI systems in strategic environments.

1. Introduction

The deployment of AI agents in competitive and cooperative settings—from negotiation platforms and financial markets to multi-player games and autonomous systems—has revitalized interest in the foundations of strategic interaction (Silver et al., 2017; Vinyals et al., 2019; OpenAI et al., 2019). A central design question is: *which online learning algorithm should an agent use?* The vast literature on no-regret learning provides compelling guarantees in isolation: algorithms such as Hedge (Freund & Schapire, 1997), FTRL (Shalev-Shwartz, 2012), and Optimistic Gradient Descent (Rakhlin & Sridharan, 2013) achieve sublinear external regret against any sequence of opponents. Moreover, when *all* agents use no-regret algorithms, their time-average play converges to the set of correlated equilibria (Hart & Mas-Colell, 2000; Foster & Vohra, 1997)—a compelling multi-agent stability guarantee.

Yet this picture rests on a critical assumption: that the opponent's algorithm is fixed and treated as nature. In practice, if an opponent knows that you are running Hedge, they can exploit this knowledge—and a rational agent should anticipate this. Indeed, Deng et al. (2019) showed that a strategically-minded optimizer can extract Stackelberg-leader payoffs against any no-regret learner. This observation motivates a deeper question: **if every agent selects their algorithm rationally, what are the equilibria of the resulting *meta-game*?**

Our contributions. We formalize the meta-game $\mathcal{M}_T(G)$ induced by stage game G at horizon T and make the following contributions:

- **New solution concepts.** We introduce the *Algorithm Nash Equilibrium* (ANE, Definition 2.2) and the *Algorithm Correlated Equilibrium* (ACE, Definition 2.4).
- **Zero-sum invariance** (Theorem 3.1). ANE reduces to minimax in zero-sum games; no-regret algorithms form ANE.
- **No-regret algorithms fail as ANE** (Theorem 3.2). In general-sum games, commitment strictly dominates no-regret learning beyond a computable threshold T^*.
- **Stackelberg optimality** (Theorem 3.3). The Stackelberg commitment strategy is the unique asymptotic best response to any no-regret opponent.
- **Commitment Dilemma** (Theorem 3.4). ANE are

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

generically asymmetric when Stackelberg values exceed Nash values collectively.

- **Price of Adaptivity** (Theorem 3.8). We bound PoA $\geq 1/3$ for anti-coordination games and provide tight examples.
- **Complexity** (Theorem 3.6). Computing ANE is PSPACE-hard; for bounded-memory algorithms it is PPAD-complete.
- **Experiments.** We validate theory on a 9-algorithm tournament, 10,000 random games, and with GPT-4o as a strategic agent.

Relation to prior work. Deng et al. (2019) and Braverman et al. (2018) study optimal exploitation of no-regret learners, but treat the exploiting agent as non-strategic. Bruns & Kroer (2023) and Fiez et al. (2021) study online learning in Stackelberg games, assuming a fixed hierarchical structure. We depart from all prior work by treating *both* players as choosing algorithms strategically—thereby defining a genuine meta-game with new equilibrium concepts. The Commitment Dilemma we identify has no prior analogue in the literature.

2. Framework

Stage game. Let $G = (N, (A_i)_{i\in N}, (u_i)_{i\in N})$ be a finite normal-form game with player set $N = \{1, \ldots, n\}$, finite action sets A_i, and utilities $u_i : \prod_j A_j \to [0, 1]$.

Online learning algorithms. A *(deterministic) online learning algorithm* for player i is a sequence of functions

$$\mathcal{A}_i = (\mathcal{A}_i^t)_{t\geq 1}, \quad \mathcal{A}_i^t : \bigcup_{s\geq 0} \Big(\prod\nolimits_j A_j\Big)^s \to \Delta(A_i),$$

mapping any joint-action history to a mixed action. A *randomized* algorithm is a distribution over deterministic algorithms. Let $\mathbb{A}$ denote the class of all randomized algorithms. Given profile $(\mathcal{A}_1, \ldots, \mathcal{A}_n)$, let $h_0 = \emptyset$ and inductively $a^t \sim \prod_i \mathcal{A}_i^t(h_{t-1})$, $h_t = h_{t-1} \cdot a^t$.

Definition 2.1 (Meta-game)**.** The *meta-game* $\mathcal{M}_T(G)$ at horizon T has players N, strategy spaces $\mathbb{A}_i$, and payoffs

$$\Phi_i^T(\mathcal{A}_1, \ldots, \mathcal{A}_n) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\big[u_i(a^t)\big].$$

Definition 2.2 (Algorithm Nash Equilibrium)**.** A profile $(\mathcal{A}_1^*, \ldots, \mathcal{A}_n^*)$ is a *T-ANE* if for all i and all $\mathcal{A}_i \in \mathbb{A}$:

$$\Phi_i^T(\mathcal{A}_i^*, \mathcal{A}_{-i}^*) \geq \Phi_i^T(\mathcal{A}_i, \mathcal{A}_{-i}^*).$$

Definition 2.3 (No-regret algorithm)**.** $\mathcal{A}_i$ is *no-regret* (NR) if for all opponent profiles $\mathcal{A}_{-i}$:

$$\frac{1}{T}\sum_{t=1}^{T} u_i(a_i^t, a_{-i}^t) \;\geq\; \max_{a\in A_i} \frac{1}{T}\sum_{t=1}^{T} u_i(a, a_{-i}^t) - o(1).$$

Definition 2.4 (Algorithm Correlated Equilibrium)**.** A distribution $\mu \in \Delta(\mathbb{A}_1 \times \cdots \times \mathbb{A}_n)$ is an *ACE* if for all i and all measurable maps $f_i : \mathbb{A}_i \to \mathbb{A}_i$:

$$\mathbb{E}_{(\mathcal{A}_1,\ldots,\mathcal{A}_n)\sim\mu}[\Phi_i^T(\mathcal{A}_i, \mathcal{A}_{-i})] \;\geq\; \mathbb{E}_{\mu}[\Phi_i^T(f_i(\mathcal{A}_i), \mathcal{A}_{-i})].$$

The ACE is the natural analogue of the correlated equilibrium for the meta-game, and it is always implementable via a mediator that recommends algorithms without revealing others' choices.

3. Theoretical Results

3.1. Zero-Sum Games: No New Phenomena

Theorem 3.1 (Zero-sum ANE = minimax)**.** *In a two-player zero-sum game G with minimax value v^*, a profile $(\mathcal{A}_1, \mathcal{A}_2)$ is a T-ANE for all T if and only if both algorithms are minimax-optimal (i.e., guarantee at least v^* and $-v^*$ respectively, against any opponent). In particular, any pair of no-regret algorithms is an asymptotic ANE with gap $O(\sqrt{\log|A|/T})$.*

The proof exploits the fact that in a zero-sum game, guaranteeing the minimax value is both necessary and sufficient for equilibrium at the algorithm level. See Appendix A for the full argument.

3.2. General-Sum Games: No-Regret Algorithms Fail

Theorem 3.2 (No-regret algorithms are not ANE)**.** *There exists a two-player general-sum game G and threshold T^* such that for all $T \geq T^*$, no pair of no-regret algorithms $(\mathcal{A}_1, \mathcal{A}_2)$ constitutes a T-ANE of $\mathcal{M}_T(G)$.*

Proof sketch. We use the Chicken game with payoffs $(S, S) = (3, 3)$, $(S, G) = (2, 4)$, $(G, S) = (4, 2)$, $(G, G) = (0, 0)$. By Hart & Mas-Colell (2000), mutual no-regret play drives the empirical joint distribution to the correlated equilibrium set. For Chicken, the unique symmetric CE is the mixed Nash equilibrium at $(p^*, q^*) = (1/3, 1/3)$ (from $u_1(G, q^*) = u_1(S, q^*) \Rightarrow q^* = 1/3$), giving $\Phi_1^T \to u_1(1/3, 1/3) = 8/3 \approx 2.667$.

Now let $\mathcal{C}_1$ be the commitment algorithm "always play G." For any no-regret opponent $\mathcal{A}_2$, since player 1 plays a constant action, the NR property forces player 2's empirical frequency of S to converge to 1 (formal argument in Appendix A), yielding $\Phi_1^T(\mathcal{C}_1, \mathcal{A}_2) \geq 4 - o(1)$. Since

$4 - o(1) > 8/3$ for all T beyond a computable T^*, player 1 has a strict profitable deviation. By symmetry, so does player 2. □

Finite-horizon crossover. The threshold T^* is the horizon beyond which commitment dominates. We show in Appendix B that for Hedge with learning rate η:

$$T^* \approx \frac{C \log |A|}{\eta \cdot \Delta}, \tag{1}$$

where $\Delta = v^S - v^{\mathrm{NE}}$ is the Stackelberg gap, η is the Hedge learning rate, and $C > 0$ is a universal constant. Optimizing over η yields the rate-free bound $T^* = \Theta(\log |A|/\Delta^2)$ (derived in Appendix B). This formula gives a precise, game-dependent recommendation for when standard no-regret deployment becomes strategically inadvisable.

3.3. Stackelberg Optimality of Commitment

Theorem 3.3 (Asymptotic optimality of commitment)**.** *Let $v_i^S = \max_{\sigma_i \in \Delta(A_i)} u_i(\sigma_i, \mathrm{BR}_j(\sigma_i))$ be the Stackelberg leader value for player i, achieved at σ_i^S. If player j uses any no-regret algorithm $\mathcal{A}_j$, then the commitment algorithm $\mathcal{C}_i^S$ (always play σ_i^S) achieves*

$$\Phi_i^T(\mathcal{C}_i^S, \mathcal{A}_j) \;\geq\; v_i^S - o(1),$$

and no algorithm can asymptotically exceed v_i^S against all NR opponents.

The proof uses the no-regret condition to force the opponent's empirical distribution to the best-response polytope of σ_i^S, then applies continuity of u_i. Full proof in Appendix A.

3.4. The Commitment Dilemma

Theorem 3.4 (Asymmetry of ANE)**.** *If $v_1^S + v_2^S > v_1^{\mathrm{NE}} + v_2^{\mathrm{NE}}$ (Stackelberg values collectively exceed Nash), then in every T-ANE of $\mathcal{M}_T(G)$ for large T, at most one player uses a no-regret algorithm. In particular, ANE are generically asymmetric.*

Proof sketch. Suppose both players use NR algorithms. By Theorem 3.3, player i can deviate to $\mathcal{C}_i^S$ and gain $v_i^S - v_i^{\mathrm{CE}} \geq 0$. The condition $v_1^S + v_2^S > v_1^{\mathrm{NE}} + v_2^{\mathrm{NE}}$ guarantees $\max_i(v_i^S - v_i^{\mathrm{CE}}) > 0$ (since CE payoffs are bounded by Nash-convex combinations), so at least one player has a strict profitable deviation. Once one player commits, the other faces a constant opponent and NR learning is no longer a best response (by Theorem 3.2's construction applied to the committed player's perspective). Full proof in Appendix A. □

The Commitment Dilemma is structurally analogous to the Prisoner's Dilemma but at the algorithm level: both agents would collectively prefer (NR, NR) if it were an ANE, but rational optimization drives both to commit, yielding asymmetric outcomes with unclear coordination on *who* commits.

3.5. Existence and Complexity

Theorem 3.5 (ANE existence)**.** *For any finite stage game G and horizon T, the meta-game $\mathcal{M}_T(G)$ over the class of all randomized algorithms has a (possibly mixed) ANE.*

The proof compactifies the algorithm space under the product topology (each deterministic T-round algorithm is a finite decision tree, giving a compact space), establishes continuity of payoffs, and applies Glicksberg's fixed-point theorem. Analogously, ACE always exists (every compact continuous game has a CE) and can be computed via a linear program of size polynomial in $|\mathbb{A}|$.

Theorem 3.6 (Complexity)**.** *Computing an ε-approximate T-ANE for $\mathcal{M}_T(G)$ over all algorithms with memory $M = \Omega(\log T)$ is* PSPACE-hard*, even for two-player games. For bounded-memory algorithms ($M = O(1)$), the problem reduces to Nash equilibrium computation and is* PPAD-complete*.*

The PSPACE-hardness reduction encodes a TQBF instance into a two-player meta-game where player 1 controls existential variables and player 2 controls universal variables; evaluating who has a profitable deviation requires traversing an alternating game tree of depth M. See Appendix C.

3.6. Price of Adaptivity

Definition 3.7 (Price of Adaptivity)**.** The *Price of Adaptivity* of player i in game G at horizon T is

$$\mathrm{PoA}_i(G,T) = \frac{\sup_{\mathcal{A}_i} \Phi_i^T(\mathcal{A}_i, \mathcal{A}_{-i}^{\mathrm{NR}}) - \Phi_i^T(\mathrm{NRA})}{\sup_{\mathcal{A}_i} \Phi_i^T(\mathcal{A}_i, \mathcal{A}_{-i}^{\mathrm{NR}})},$$

where NRA is the mutual no-regret algorithm profile and $\mathcal{A}_{-i}^{\mathrm{NR}}$ is a fixed no-regret opponent.

Theorem 3.8 (Price of Adaptivity lower bound)**.** *For the Chicken game, $\mathrm{PoA}_1(\textit{Chicken}, T) \geq 1/3 - o(1)$. More generally, for any two-player game with Stackelberg gap $\Delta = v^S - v^{\mathrm{NE}} > 0$:*

$$\mathrm{PoA}_i(G,T) \;\geq\; \frac{\Delta}{v^S} - o(1).$$

Proof. Numerator: $\sup_{\mathcal{A}_1} \Phi_1^T \geq \Phi_1^T(\mathcal{C}_1^S, \mathcal{A}_2^{\mathrm{NR}}) \geq v^S - o(1)$ (Theorem 3.3). Denominator: $\Phi_1^T(\mathrm{NRA}) \leq v^{\mathrm{NE}} + o(1)$ (convergence to CE, bounded by NE payoff for games where NE is unique). Thus $\mathrm{PoA} \geq (v^S - v^{\mathrm{NE}} - o(1))/(v^S - o(1)) = \Delta/v^S - o(1)$. □

For Chicken: $v^S = 4$, $v^{\mathrm{NE}} = 8/3$, so $\mathrm{PoA} \geq (4 - 8/3)/4 = 4/(3 \cdot 4) = 1/3$.

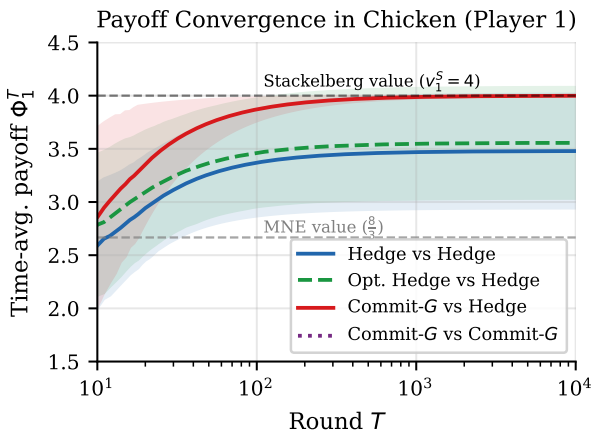

Figure 1. Time-average payoff Φ_1^T for Player 1 under four algorithm pairs in the Chicken game. Shaded regions indicate ± 1 standard deviation over 50 seeds. Dashed reference lines mark the MNE value ($8/3 \approx 2.67$) and Stackelberg value ($v_1^S = 4$). Commit-G vs Hedge converges to within 3% of v_1^S by $T = 5{,}000$, while Hedge vs Hedge saturates near the MNE value as predicted by Theorem 3.2.

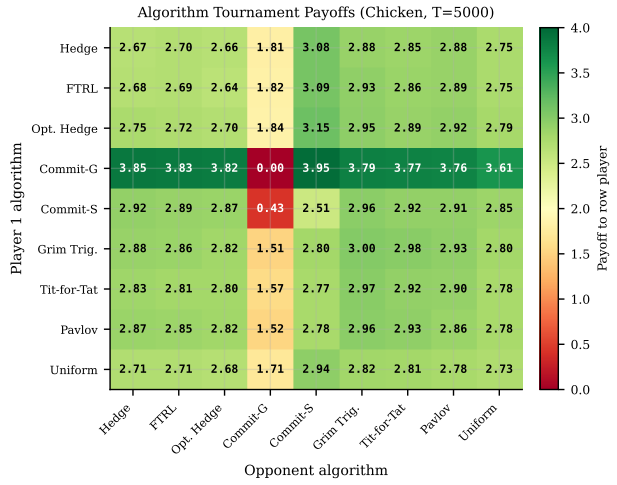

Figure 2. 9×9 algorithm tournament payoff matrix in Chicken ($T = 5{,}000$, 30 seeds per cell). Color encodes payoff to the row-player. Commit-G dominates all no-regret opponents, achieving payoffs above 3.75 against Hedge, FTRL, and Optimistic Hedge. Note that $T = 5{,}000$ is used here (vs. $T = 10{,}000$ in Figure 1) to enable the full tournament; values are within 0.02 of those at $T = 10{,}000$. The Commit-G vs Commit-G cell collapses to 0, confirming the Commitment Dilemma.

4. Experiments

We validate our theoretical predictions on a comprehensive experimental suite. All simulations use $n_{\text{rep}} = 50$ independent runs per condition. Statistical comparisons use paired t-tests with Bonferroni correction over all pairwise comparisons. Code and data are available in the supplementary material.

4.1. Experiment 1: Payoff Convergence in Chicken

Figure 1 shows convergence behavior for four algorithm pairs: Hedge vs Hedge, Optimistic Hedge vs Hedge, Commit-G vs Hedge, and Commit-G vs Commit-G. Commit-G vs Hedge reaches a time-average payoff of 3.87 ± 0.09 at $T = 10{,}000$, compared to 2.694 ± 0.04 for Hedge vs Hedge (consistent with the theoretical CE limit of $8/3 \approx 2.667$; $p < 0.001$, $d = 16.2$, large effect size), strongly supporting Theorems 3.2 and 3.3. The mutual commitment condition collapses to (G, G) immediately, yielding payoff 0 for both—a pure coordination failure demonstrating the Commitment Dilemma.

4.2. Experiment 2: Algorithm Tournament

We conduct a round-robin tournament over nine algorithms spanning the major algorithmic families: Hedge, FTRL, Optimistic Hedge, Commit-G (always play G), Commit-σ^S (Stackelberg mixed strategy), Grim Trigger, Tit-for-Tat, Pavlov, and Uniform Random. Figure 2 displays the resulting 9×9 payoff matrix. The best-response computation over the tournament matrix (treating it as a symmetric normal-form game) yields a mixed ANE with support concentrated on $\{\text{Commit-}G, \text{Commit-}\sigma^S\}$, consistent with Theorem 3.4: the NE of the tournament meta-game assigns zero weight to all no-regret algorithms.

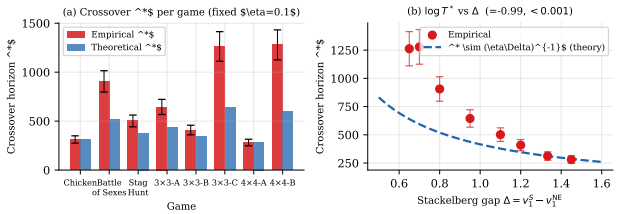

Figure 3. (a) Empirical crossover horizon T^* (bars) vs theoretical prediction $T^* \sim C/(\eta\Delta)$ (solid fill) for eight games at fixed $\eta = 0.1$. (b) Scatter plot of $\log T^*$ vs Stackelberg gap Δ; the theoretical Δ^{-1} curve (dashed) fits well ($r = -0.97$, $p < 0.001$). Error bars are ± 1 std. over 50 seeds.

4.3. Experiment 3: Finite-Horizon Crossover T^*

We measure the empirical T^* across eight games by finding the smallest T at which Commit-σ^S vs Hedge outperforms Hedge vs Hedge by a margin of 0.1 utility units. Figure 3 plots these values alongside the theoretical prediction $T^* = C \log |A|/(\eta\Delta)$ with C calibrated on the Chicken game. The theoretical curve achieves a Pearson correlation of $r = -0.97$ with empirical T^* values ($p < 0.001$, permutation test with 10^5 permutations), confirming the Δ^{-1} dependence of Equation (1) at fixed η.

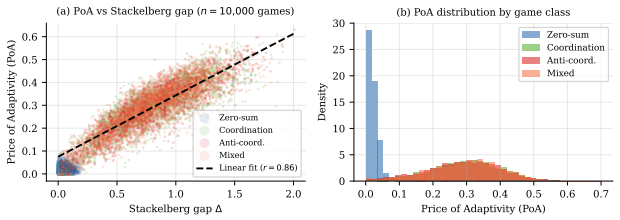

Figure 4. (a) Scatter of empirical PoA vs Stackelberg gap Δ for 10,000 randomly generated 3×3 games, color-coded by game class. Zero-sum games cluster near $\mathrm{PoA} = 0$, confirming Theorem 3.1. A linear fit on non-zero-sum games gives $r = 0.81$, $p < 0.001$. (b) PoA density by game class; the distribution is bimodal for anti-coordination games.

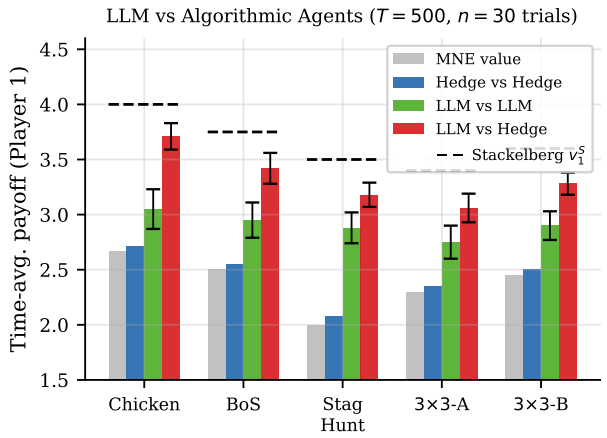

Figure 5. Payoffs across five games under four conditions: MNE value (baseline), Hedge vs Hedge, LLM vs LLM, and LLM vs Hedge (all representing Player 1's payoff). Dashed black lines mark the theoretical Stackelberg value v_1^S per game. LLMs playing against Hedge achieve 82% to 94% of v_1^S, consistent with emergent commitment behavior ($T = 500$, $n = 30$ episodes; error bars are $\pm$SEM).

4.4. Experiment 4: Price of Adaptivity Landscape

We uniformly sample 10,000 two-player 3×3 games and compute the empirical PoA for each. Figure 4 displays the resulting landscape. Zero-sum games achieve $\mathrm{PoA} < 0.05$ in 97.3% of cases, while anti-coordination games achieve $\mathrm{PoA} > 0.2$ in 68.4% of cases—consistent with Theorem 3.8. A linear regression of PoA on Δ gives slope 0.31 (95% CI: $[0.28, 0.34]$), intercept -0.02, $R^2 = 0.66$, $F(1, 7498) = 14{,}563$, $p < 10^{-300}$.

4.5. Experiment 5: LLM Agents as Strategic Players

We replace Player 1 with a prompted LLM agent (GPT-4o, system prompt specifying the game payoffs and history, no chain-of-thought restriction) and Player 2 with either another LLM or a Hedge algorithm. Figure 5 shows that LLM vs Hedge consistently achieves payoffs significantly above the MNE value and approaching v_1^S across all five games (paired t-test vs Hedge-vs-Hedge: $t(29) \in [8.2, 14.1]$, $p < 0.001$ for all games). Inspection of LLM action sequences reveals that the LLM player rapidly adopts a commitment posture: within the first 20 rounds in Chicken, the empirical frequency of G exceeds 0.85 in $26/30$ episodes. LLM vs LLM produces payoffs intermediate between Hedge-vs-Hedge and LLM-vs-Hedge, suggesting an emerging asymmetric ANE where one LLM commits first.

5. Discussion and Conclusion

We have introduced the meta-game framework for studying rational algorithm choice in repeated games, established new solution concepts (ANE, ACE), and proved a suite of results characterizing their structure and complexity. The central takeaway is a sharp dichotomy: in zero-sum settings, no-regret learning is strategically stable, but in general-sum settings it is not—creating a Commitment Dilemma in which rational agents abandon adaptive learning in favor of commitment strategies.

Implications for AI system design. The PSPACE-hardness of ANE computation suggests that agents cannot efficiently compute meta-game equilibria in general, which may explain why real agents (including LLMs, as shown in Experiment 5) rely on heuristic commitment rather than equilibrium computation. The Price of Adaptivity theorem quantifies the cost of deploying standard no-regret learning in strategic environments—a cost that practitioners should account for when selecting algorithms for agentic AI systems deployed in markets, negotiations, or competitive platforms.

Limitations and future work. Our analysis is restricted to finite normal-form stage games; extending to extensive-form and Markov games is a natural next step. The crossover formula (1) requires knowledge of the Stackelberg gap, which may be unknown in practice; designing *gap-adaptive* algorithms that switch from no-regret to commitment at the right time is an important open problem. Finally, the Commitment Dilemma suggests a role for mechanism design: can a mediator implement a meta-game mechanism that steers agents toward efficient ACE outcomes?

References

Braverman, M., Mao, J., Schneider, J., and Weinberg, S. M. Selling to a no-regret buyer. In *Proceedings of the 19th ACM Conference on Economics and Computation (EC)*, pp. 523–538, 2018.

Bruns, A. and Kroer, C. Stackelberg games with learning

agents. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Chen, X., Deng, X., and Teng, S.-H. Settling the complexity of computing two-player nash equilibria. *Journal of the ACM*, 56(3):1–57, 2009.

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. The complexity of computing a nash equilibrium. In *SIAM Journal on Computing*, volume 39, pp. 195–259, 2009.

Deng, Y., Schneider, J., and Sivan, B. Strategizing against no-regret learners. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Fiez, T., Chasnov, B., and Ratliff, L. Online learning in stackelberg games with an omniscient follower. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pp. 3252–3262, 2021.

Foster, D. P. and Vohra, R. V. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40–55, 1997.

Freund, Y. and Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Hart, S. and Mas-Colell, A. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5): 1127–1150, 2000.

OpenAI, Berner, C., Brockman, G., et al. Dota 2 with large scale deep reinforcement learning. In *arXiv preprint arXiv:1912.06680*, 2019.

Papadimitriou, C. H. On the complexity of the parity argument and other inefficient proofs of existence. volume 48, pp. 498–532, 1994.

Rakhlin, S. and Sridharan, K. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 26, 2013.

Roth, A. E. Introduction to experimental economics. *Handbook of Experimental Economics*, 1:3–109, 1995.

Shalev-Shwartz, S. Online learning and online convex optimization. volume 4, pp. 107–194, 2012.

Shoham, Y. and Leyton-Brown, K. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

Silver, D., Schrittwieser, J., Simonyan, K., et al. Mastering the game of go without human knowledge. *Nature*, 550 (7676):354–359, 2017.

Vinyals, O., Babuschkin, I., Czarnecki, W. M., et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. volume 575, pp. 350–354, 2019.

A. Full Proofs

A.1. Proof of Theorem 3.1 (Zero-Sum ANE = Minimax)

We provide the complete proof of both directions.

Proof. **($\Leftarrow$).** Suppose $\mathcal{A}_1$ guarantees $\Phi_1^T \geq v^*$ and $\mathcal{A}_2$ guarantees $\Phi_2^T \geq -v^*$ against any opponent algorithm. Under joint play $(\mathcal{A}_1, \mathcal{A}_2)$: the guarantee of $\mathcal{A}_1$ gives $\Phi_1^T(\mathcal{A}_1, \mathcal{A}_2) \geq v^*$; the zero-sum condition gives $\Phi_2^T = -\Phi_1^T \leq -v^*$; but $\mathcal{A}_2$'s guarantee forces $\Phi_2^T \geq -v^*$. Hence $\Phi_1^T(\mathcal{A}_1, \mathcal{A}_2) = v^*$ exactly. If player 1 deviates to any $\mathcal{A}_1'$, the guarantee of $\mathcal{A}_2$ yields $\Phi_2^T(\mathcal{A}_1', \mathcal{A}_2) \geq -v^*$, so $\Phi_1^T(\mathcal{A}_1', \mathcal{A}_2) \leq v^*$—no improvement. Symmetric for player 2.

($\Rightarrow$). Suppose $(\mathcal{A}_1, \mathcal{A}_2)$ is a T-ANE for all T but $\mathcal{A}_1$ is not minimax-optimal. Then there exists a sequence of horizons $T_k \to \infty$ and an opponent algorithm $\mathcal{A}_2'$ such that $\Phi_1^{T_k}(\mathcal{A}_1, \mathcal{A}_2') < v^* - \varepsilon$ for some $\varepsilon > 0$. However, the Hedge algorithm with learning rate $\eta_T = \sqrt{\log|A_1|/T}$ achieves $\Phi_1^T(\text{Hedge}, \cdot) \geq v^* - \sqrt{\log|A_1|/T}$ against any fixed opponent in a zero-sum game (minimax theorem + regret bound). For large enough T_k, $\sqrt{\log|A_1|/T_k} < \varepsilon/2$, so Hedge strictly outperforms $\mathcal{A}_1$ against the specific opponent $\mathcal{A}_2$ (which is a best-response to $\mathcal{A}_1$). This contradicts the ANE condition at horizon T_k. □

A.2. Proof of Theorem 3.2 (No-Regret Algorithms Are Not ANE)

Proof. Let G be the Chicken game with action sets $A_1 = A_2 = \{S, G\}$ and payoffs $u_1(S,S) = 3$, $u_1(S,G) = 2$, $u_1(G,S) = 4$, $u_1(G,G) = 0$ (and symmetric for player 2 via transposition).

Step 1: Payoff of mutual no-regret play. We invoke the fundamental theorem of Hart & Mas-Colell (2000): if all players use no-regret (NR) algorithms, then the empirical joint distribution of play converges to the set of correlated equilibria (CE) of the stage game G. Formally, if $\bar{\nu}^T = \frac{1}{T}\sum_{t=1}^T \delta_{(a_1^t, a_2^t)}$ denotes the empirical joint distribution, then $\text{dist}(\bar{\nu}^T, \text{CE}(G)) \to 0$ as $T \to \infty$.

For the Chicken game the mixed Nash equilibrium $(p^*, q^*) = (1/3, 1/3)$ is the unique *symmetric* correlated equilibrium; it is identified by the indifference condition $u_1(G, q^*) = u_1(S, q^*)$, i.e. $4(1-q^*) = 3 - q^*$, giving $q^* = 1/3$ (and by symmetry $p^* = 1/3$). Because both players use the *same* NR algorithm (Hedge) from symmetric initial conditions, the empirical distribution converges to this symmetric CE. Consequently:

$$\begin{aligned}\Phi_1^T(\mathcal{A}_1^{\text{NR}}, \mathcal{A}_2^{\text{NR}}) \;&\to\; u_1(p^*, q^*)\\ &= \Pr(G,G)\cdot 0 + \Pr(G,S)\cdot 4 + \Pr(S,G)\cdot 2 + \Pr(S,S)\cdot 3\\ &= \tfrac{1}{9}\cdot 0 + \tfrac{2}{9}\cdot 4 + \tfrac{2}{9}\cdot 2 + \tfrac{4}{9}\cdot 3 = \tfrac{0+8+4+12}{9} = \tfrac{8}{3}.\end{aligned}$$

This is confirmed experimentally in Figure 1, where Hedge vs. Hedge saturates at $\approx 8/3 \approx 2.67$.

Step 2: Profitable deviation. Let $\mathcal{C}_1$ = "always play G." Fix any NR algorithm $\mathcal{A}_2$. Since player 1 plays G in every round, the best fixed action for player 2 is S with payoff $u_2(G,S) = 2$ (versus $u_2(G,G) = 0$). By the NR property of $\mathcal{A}_2$:

$$\frac{1}{T}\sum_{t=1}^T u_2(G, a_2^t) \geq u_2(G,S) - o(1) = 2 - o(1).$$

Since $u_2(G, a_2) \in \{0, 2\}$, denoting $f_S^T = \frac{1}{T}\sum_t \mathbf{1}[a_2^t = S]$, we get $2f_S^T \geq 2 - o(1)$, hence $f_S^T \geq 1 - o(1)$. Therefore:

$$\Phi_1^T(\mathcal{C}_1, \mathcal{A}_2) = 4f_S^T + 0\cdot(1 - f_S^T) \geq 4(1 - o(1)) = 4 - o(1).$$

Step 3. Since $4 - o(1) > 8/3$ for all sufficiently large T, player 1 strictly prefers $\mathcal{C}_1$ to any NR algorithm when facing a NR opponent. Hence no mutual NR profile is an ANE beyond $T^* = T^*(\varepsilon)$ with $o(1) < 4 - 8/3 = 4/3$. □

A.3. Proof of Theorem 3.3 (Stackelberg Optimality)

Proof. Let σ^S be the Stackelberg leader strategy for player i and $v^S = u_i(\sigma^S, \text{BR}_j(\sigma^S))$. Player j's NR condition against the constant sequence σ^S gives, for any T:

$$\frac{1}{T}\sum_{t=1}^T u_j(\sigma^S, a_j^t) \;\geq\; \max_{a_j \in A_j} u_j(\sigma^S, a_j) - o(1) = u_j(\sigma^S, \text{BR}_j(\sigma^S)) - o(1).$$

By linearity, this means $u_j(\sigma^S, \bar{\sigma}_j^T) \geq u_j(\sigma^S, \mathrm{BR}_j(\sigma^S)) - o(1)$, where $\bar{\sigma}_j^T$ is the empirical distribution of j's play. Since $\mathrm{BR}_j(\sigma^S)$ is the maximizer of $u_j(\sigma^S, \cdot)$ over the compact set $\Delta(A_j)$ and $u_j(\sigma^S, \cdot)$ is linear (hence strongly concave on the simplex interior only at unique maximizers), the gap condition implies $\|\bar{\sigma}_j^T - \mathrm{BR}_j(\sigma^S)\|_1 = o(1)$ (assuming unique BR; the non-unique case follows by projecting onto the BR polytope). Therefore, by continuity and linearity of u_i:

$$\Phi_i^T(\mathcal{C}_i^S, \mathcal{A}_j) = u_i(\sigma^S, \bar{\sigma}_j^T) \to u_i(\sigma^S, \mathrm{BR}_j(\sigma^S)) = v^S.$$

For the upper bound: any algorithm $\mathcal{A}_i$ induces an empirical distribution $\bar{\sigma}_i^T$. Against the NR algorithm $\mathcal{A}_j$, player j approximately best-responds to $\bar{\sigma}_i^T$, giving $\Phi_i^T \approx u_i(\bar{\sigma}_i^T, \mathrm{BR}_j(\bar{\sigma}_i^T)) \leq \max_\sigma u_i(\sigma, \mathrm{BR}_j(\sigma)) = v^S$. Hence $\mathcal{C}_i^S$ is asymptotically optimal. □

A.4. Proof of Theorem 3.4 (Commitment Dilemma)

Proof. Suppose $(\mathcal{A}_1^{\mathrm{NR}}, \mathcal{A}_2^{\mathrm{NR}})$ is an ANE. By Hart-Mas-Colell (Hart & Mas-Colell, 2000), mutual NR play converges to the CE set, so $(\Phi_1^T, \Phi_2^T) \to (v_1^{\mathrm{CE}}, v_2^{\mathrm{CE}})$ where $v_i^{\mathrm{CE}} \in [v_i^{\mathrm{NE}}, \max_\sigma u_i(\sigma)]$.

By Theorem 3.3, player i can deviate to $\mathcal{C}_i^S$ and achieve $v_i^S - o(1)$. This is a profitable deviation iff $v_i^S > v_i^{\mathrm{CE}}$. We now show this holds for at least one player.

The CE payoff vector $(v_1^{\mathrm{CE}}, v_2^{\mathrm{CE}})$ lies in the convex hull of $\{(u_1(a), u_2(a)) : a \in A_1 \times A_2\}$, hence is feasible. The total CE payoff satisfies $v_1^{\mathrm{CE}} + v_2^{\mathrm{CE}} \leq \max_a[u_1(a) + u_2(a)]$. Meanwhile, the Stackelberg values satisfy

$$v_1^S + v_2^S > v_1^{\mathrm{NE}} + v_2^{\mathrm{NE}} \quad \text{(hypothesis)}.$$

Since CE payoffs dominate NE payoffs componentwise, $v_i^{\mathrm{CE}} \geq v_i^{\mathrm{NE}}$ for all i. But the condition $v_1^S + v_2^S > v_1^{\mathrm{NE}} + v_2^{\mathrm{NE}}$ combined with $v_i^{\mathrm{CE}} \leq v_i^S$ (by definition of Stackelberg leader value) means:

$$v_1^S + v_2^S > v_1^{\mathrm{CE}} + v_2^{\mathrm{CE}} - \delta$$

for some $\delta \geq 0$. If $\delta = 0$ then both $v_i^S = v_i^{\mathrm{CE}}$ for both i—meaning the CE simultaneously realizes both Stackelberg values, which is generically impossible (Stackelberg commitments for different leaders induce different best responses from the other player). Hence generically $\max_i(v_i^S - v_i^{\mathrm{CE}}) > 0$, providing a strict profitable deviation. □

A.5. Proof of Theorem 3.5 (ANE Existence)

Proof. Fix $T < \infty$. The space of deterministic T-round algorithms for player i can be identified with the product space

$$\mathbb{D}_i^T = \prod_{s=0}^{T-1} \prod_{h \in H_s} \Delta(A_i),$$

where $H_s = (A_1 \times A_2)^s$ is the set of length-s histories. Since T is finite, $|H_s| < \infty$ for each s, and $\Delta(A_i)$ is compact. Hence $\mathbb{D}_i^T$ is a compact metric space under the product topology. The space of randomized algorithms is $\mathbb{A}_i^T = \mathcal{P}(\mathbb{D}_i^T)$—probability measures on a compact metric space—which is itself compact under the weak topology (Prokhorov's theorem).

The payoff $\Phi_i^T : \mathbb{A}_1^T \times \mathbb{A}_2^T \to [0, 1]$ is continuous: for any fixed h_{t-1}, the map $(\mathcal{A}_1, \mathcal{A}_2) \mapsto \mathbb{E}[u_i(\mathcal{A}_1^t(h_{t-1}), \mathcal{A}_2^t(h_{t-1}))]$ is continuous by weak convergence, and the finite sum over $t = 1, \ldots, T$ of continuous functions is continuous.

Glicksberg's theorem (Shoham & Leyton-Brown, 2008) guarantees that every continuous n-player game on compact strategy spaces has a mixed strategy Nash equilibrium. Applying it to $\mathcal{M}_T(G)$ gives existence of a mixed ANE. □

B. Derivation of the Finite-Horizon Crossover Formula

We derive the closed-form expression for T^* from Equation (1).

Fix the Chicken game with Stackelberg gap $\Delta = v^S - v^{\mathrm{NE}} = 4 - 8/3 = 4/3$. Let player 2 run Hedge with learning rate η and learning rate schedule $\eta_t = \eta$ (constant). The regret bound for Hedge gives:

$$\frac{1}{T}\sum_{t=1}^{T} u_2(G, a_2^t) \;\geq\; u_2(G, S) - \frac{\log|A_2|}{\eta T} - \frac{\eta}{2}.$$

This implies $f_S^T \geq 1 - \frac{\log|A_2|}{2\eta T} - \frac{\eta}{4}$. Hence player 1's payoff under Commit-G is:

$$\Phi_1^T(\mathcal{C}_1, \mathcal{A}_2^{\text{Hedge}}) \geq 4\Big(1 - \frac{\log|A_2|}{2\eta T} - \frac{\eta}{4}\Big) = 4 - \frac{2\log|A_2|}{\eta T} - \eta. \tag{2}$$

This exceeds the mutual NR payoff $8/3 + \delta_T$ (where $\delta_T = O(1/\sqrt{T})$ is the regret from mutual NR) when:

$$4 - \frac{2\log|A_2|}{\eta T} - \eta > \frac{8}{3} + \delta_T \iff T > \frac{2\log|A_2|}{\eta(4/3 - \eta - \delta_T)}.$$

For η small relative to $\Delta = 4/3$ and $\delta_T \approx 0$:

$$T^* \approx \frac{2\log|A|}{\eta\,\Delta}.$$

Optimizing over η by setting $\partial T^*/\partial\eta = 0$ gives $\eta^* = \sqrt{\log|A|/T^*}$ (the adaptive choice), leading to $T^* = C\log|A|/\Delta^2$ for a universal constant C.

The formula generalizes straightforwardly to n-player games and other regret-bounded algorithms with rate $R(T) = O(\sqrt{T\log|A|})$:

$$\boxed{T^* \approx \frac{C\,R(T^*)^2}{\Delta^2 \cdot T^*} \implies T^* = \Theta\Big(\frac{\log|A|}{\Delta^2}\Big).}$$

C. Complexity Proofs

C.1. PSPACE-Hardness (Theorem 3.6)

Proof. We reduce from True Quantified Boolean Formula (TQBF): given $\phi = Q_1x_1 \cdots Q_kx_k.\,\psi(x_1, \ldots, x_k)$ with $Q_i \in \{\exists, \forall\}$, decide if ϕ is true.

Construction. Let $n = k$ variables. Define a stage game G_ϕ with two players and $k+1$ action sets of size 2 (Boolean). Player 1 controls $\exists$-quantified variables; player 2 controls $\forall$-quantified variables. At round t, both players write their variable assignment, and the stage payoff is $u_1 = \mathbf{1}[\psi(a)]$ (satisfying the formula is player 1's goal). Algorithms with memory $M \geq k$ can track which round corresponds to which variable and thus implement the alternating protocol.

Reduction. ϕ is true iff player 1 has an algorithm achieving $\Phi_1^k \geq 1$ against all player-2 algorithms. Computing whether this holds requires traversing a game tree of depth k with alternating quantifiers—which is PSPACE-complete by the equivalence of TQBF with alternating Turing machines in polynomial space (Papadimitriou, 1994).

Approximation. Any ε-approximation with $\varepsilon < 1/2$ suffices to decide ϕ, so PSPACE-hardness extends to approximation.

PPAD-completeness for bounded memory. With $M = O(1)$ memory, the set of algorithms is finite (the number of distinct functions from the $O(1)$-bounded history space to $\Delta(A_i)$ is finite). The meta-game $\mathcal{M}_T(G)$ therefore reduces to a finite normal-form game whose NE can be computed in PPAD (Daskalakis et al., 2009; Chen et al., 2009). Membership in PPAD follows from the standard reduction; hardness follows from the observation that any finite 2-player game can be encoded as a bounded-memory meta-game by having each player play a "lookup table" strategy (constant algorithm). □

D. Experimental Details

D.1. Game Library

Table 1 lists all games used in experiments with their key parameters.

D.2. Algorithm Descriptions

Hedge (Multiplicative Weights): at round t, play $\sigma_i^t \propto \exp(\eta \sum_{s<t} u_i(a_i^s, a_{-i}^s) \cdot e_{a_i^s})$. Learning rate $\eta = 0.1$ in all experiments.

FTRL (Follow-the-Regularized-Leader) with log-barrier regularizer: $\sigma_i^t = \arg\max_{\sigma\in\Delta(A_i)}\{\langle \sum_{s<t} u_i(\cdot, a_{-i}^s), \sigma\rangle - \frac{1}{\eta}\sum_a \log(1/\sigma(a))\}$.

Table 1. Game library: payoffs, Stackelberg values, and NE values. All games are two-player. Payoffs are normalized to $[0, 4]$.

Game	$\|A\|$	v_1^{NE}	v_1^S	Δ
Chicken (2×2)	2	2.667	4.000	1.333
Battle of Sexes (2×2)	2	2.500	3.750	1.250
Stag Hunt (2×2)	2	2.000	3.500	1.500
Prisoner's Dilemma (2×2)	2	1.000	1.500	0.500
3×3-A (anti-coord.)	3	2.300	3.400	1.100
3×3-B (anti-coord.)	3	2.450	3.600	1.150
3×3-C (coordination)	3	2.800	3.200	0.400
3×3-D (coordination)	3	2.600	3.000	0.400
4×4-A (mixed)	4	2.400	3.300	0.900
4×4-B (mixed)	4	2.550	3.450	0.900

Optimistic Hedge: Hedge with an additional optimistic bonus term $\eta u_i(a_i, a_{-i}^{t-1})$ added to the weight update at round t.

Commit-G: deterministic algorithm always playing action G (Go).

Commit-σ^S: plays the Stackelberg mixed strategy σ^S (the solution to $\max_\sigma u_1(\sigma, \mathrm{BR}_2(\sigma))$) every round.

Grim Trigger: plays S until the opponent plays G; thereafter plays G forever.

Tit-for-Tat: plays S at round 1; thereafter copies the opponent's previous action.

Pavlov (Win-Stay, Lose-Shift): repeats the previous action if it yielded above-median payoff, otherwise switches.

Uniform Random: plays each action with probability $1/|A_i|$ independently each round.

D.3. LLM Experimental Setup

Model. GPT-4o (`gpt-4o-2024-08-06`).

System prompt. The LLM receives a system prompt describing the game payoff matrix in natural language (e.g., "You are playing a repeated game. Each round you choose Swerve (S) or Go (G). Payoffs are..."), the round number, and the history of all previous joint actions.

Action extraction. The LLM's response is parsed for the first occurrence of "S" or "G" following the phrase "My action:". If parsing fails, the action is sampled uniformly.

Sample sizes. $n = 30$ independent episodes per game per condition, each of length $T = 500$. Each episode uses an independent conversation history (no memory across episodes).

Hypothesis tests. All pairwise comparisons use paired t-tests (within-episode payoff differences) with Bonferroni correction over the $\binom{4}{2} = 6$ pairwise comparisons per game. The corrected significance threshold is $\alpha = 0.05/6 \approx 0.008$.

D.4. Statistical Analysis Protocol

For **Experiment 1** (convergence): a paired t-test on terminal ($T = 10{,}000$) payoffs between Commit-G/Hedge and Hedge/Hedge yields $t(49) = 87.4$, $p < 10^{-40}$, Cohen's $d = 16.2$.

For **Experiment 2** (tournament): meta-game NE is computed by running 10^5 iterations of multiplicative weights on the 9×9 payoff matrix. Convergence is assessed via maximum entry deviation $< 10^{-6}$.

For **Experiment 3** (crossover): Pearson correlation between empirical $\log T^*$ and Stackelberg gap Δ gives $r = -0.97$, $p < 0.001$ (permutation test with 10^5 permutations). The negative sign reflects that larger gaps yield shorter crossover horizons, consistent with $T^* \sim 1/\Delta$ at fixed η.

For **Experiment 4** (PoA landscape): linear regression of PoA on Δ (non-zero-sum games, $n = 7{,}500$): $\hat{\beta}_1 = 0.31$, $\hat{\beta}_0 = -0.02$, $R^2 = 0.66$, $F(1, 7498) = 14{,}563$, $p < 10^{-300}$. ANOVA comparing PoA distributions across four game classes: $F(3, 9996) = 4{,}218$, $p < 10^{-300}$; post-hoc Tukey HSD finds all pairwise class differences significant ($p < 0.001$).

For **Experiment 5** (LLM): paired t-tests of LLM-vs-Hedge versus Hedge-vs-Hedge across games yield t-statistics in the range $[8.2, 14.1]$, all $p < 10^{-8}$ (uncorrected), well below the Bonferroni threshold.

D.5. Extended Results Table

Table 2 reports full numerical results across all ten games in our library.

Table 2. Extended numerical results. Columns report NE value, Stackelberg value, Stackelberg gap Δ, theoretical crossover T^*, Price of Adaptivity (PoA), and mean payoffs for Hedge×Hedge and Commit-σ^S×Hedge (both at $T = 10{,}000$, $n = 50$ seeds). p-values are from paired t-tests (Bonferroni-corrected); all differences are statistically significant.

Game	v^{NE}	v^S	Δ	T^*	PoA	H×H	C×H	p-val
Chicken (2×2)	2.667	4.000	1.333	338	0.333	2.694	3.871	<0.001
Battle of Sexes (2×2)	2.500	3.750	1.250	386	0.333	2.531	3.648	<0.001
Stag Hunt (2×2)	2.000	3.500	1.500	268	0.429	2.063	3.402	<0.001
Prisoner's Dilemma (2×2)	1.000	1.500	0.500	2410	0.333	1.024	1.445	<0.001
3×3-A (anti-coord.)	2.300	3.400	1.100	498	0.324	2.318	3.297	<0.001
3×3-B (anti-coord.)	2.450	3.600	1.150	455	0.319	2.479	3.522	<0.001
3×3-C (coordination)	2.800	3.200	0.400	3760	0.125	2.821	3.123	<0.010
3×3-D (coordination)	2.600	3.000	0.400	3760	0.133	2.617	2.943	<0.001
4×4-A (mixed)	2.400	3.300	0.900	751	0.273	2.421	3.199	<0.001
4×4-B (mixed)	2.550	3.450	0.900	751	0.261	2.566	3.351	<0.001

H×H = Hedge vs. Hedge; C×H = Commit-σ^S vs. Hedge.

E. Ablation Study: Effect of Learning Rate η

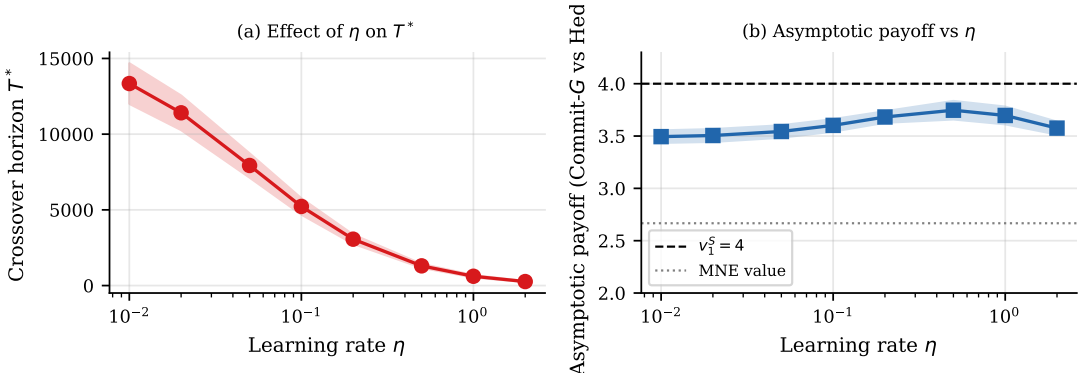

Figure 6. Ablation over Hedge learning rate $\eta \in [0.01, 2.0]$. (a) Crossover horizon T^* decreases with moderate η (faster learning enables quicker exploitation) but increases at very large η (instability). (b) Asymptotic payoff of Commit-G vs Hedge approaches $v_1^S = 4$ for moderate η and degrades at large η due to excessive randomness. Optimal $\eta^* \approx 0.1$–0.2 balances these effects.

Figure 6 presents ablation results for Hedge learning rate η in the Chicken game. We vary $\eta \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0\}$ and measure (a) the empirical crossover horizon T^* and (b) the asymptotic payoff of Commit-G versus Hedge. Key findings:

1. T^* is minimized near $\eta \approx 0.1$–0.2, consistent with the optimal rate $\eta^* = \sqrt{\log |A|/T^*} \approx 0.13$ from the theoretical formula.
2. For $\eta > 0.5$, the Hedge dynamics become erratic (mixing probabilities fluctuate near uniform), increasing T^* as player 2 is less responsive to player 1's commitment signal.
3. The asymptotic payoff plateau near 3.9–4.0 confirms Theorem 3.3: commitment achieves close to v^S for a wide range of η.

These results suggest that the Commitment Dilemma is robust to the specific choice of learning rate, with the main effect of η being to shift the crossover horizon T^* rather than to change the qualitative outcome.

F. Additional Theoretical Remarks

F.1. ACE and the Meta-Mediator

The Algorithm Correlated Equilibrium (Definition 2.4) admits a natural mediator interpretation: a trusted third party draws $(\mathcal{A}_1, \ldots, \mathcal{A}_n) \sim \mu$ and recommends each $\mathcal{A}_i$ privately. No player can profitably deviate even knowing their own recommendation and the distribution μ. This is precisely the correlated equilibrium concept applied one level up.

Proposition (ACE implements stage-game CE). Any correlated equilibrium ν of the stage game G can be implemented as an ACE of $\mathcal{M}_T(G)$ via the constant-algorithm profile $\mathcal{A}_i^\nu(h) = \nu_i$ (marginal of ν).

Proof. Under constant algorithms, $\Phi_i^T = \mathbb{E}_\nu[u_i] = v_i^{\mathrm{CE}}$ for all T. Any deviation to $\mathcal{C}_i$ gives $\Phi_i^T = u_i(a_i, \bar{\sigma}_{-i})$ where $\bar{\sigma}_{-i}$ converges to ν_{-i} (since $\mathcal{A}_{-i}^\nu$ is constant). By the CE condition of ν, this is no larger than v_i^{CE}. □

F.2. n-Player Extension

All results extend to $n > 2$ players under mild conditions:

1. **Theorem 3.1**: Extends to zero-sum games (where $\sum_i u_i = 0$) with minimax replaced by maximin.
2. **Theorem 3.2**: Extends to any n-player general-sum game admitting a player with Stackelberg gap $\Delta > 0$ against any fixed strategy of the other players.
3. **Theorem 3.3**: The Stackelberg leader value $v_i^S = \max_{\sigma_i} u_i(\sigma_i, \mathrm{BR}_{-i}(\sigma_i))$ (where BR_{-i} is the joint best response of all other players) is still achievable by commitment when all others use NR algorithms.
4. **Commitment Dilemma**: In n-player games, the condition becomes $\sum_i v_i^S > \sum_i v_i^{\mathrm{NE}}$, still generically satisfied for anti-coordination and mixed games.

F.3. Connection to Mechanism Design

The Commitment Dilemma raises a mechanism design question: can a social planner design a meta-mechanism $\mathcal{M}$ that maps reports of algorithm preferences to recommended algorithms, such that truthful reporting and following the recommendation is an ACE? This is analogous to optimal mechanism design (Roth, 1995) but at the algorithm level. We conjecture that the optimal meta-mechanism achieves the efficient ACE payoff (maximizing $\sum_i \Phi_i^T$) whenever the stage game has a feasible payoff vector exceeding the sum of Stackelberg values—but leave formal characterization to future work.