Convergence to Equilibrium

Game Theory · Module 9 — Learning In Games

Node 9.5

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.5

This node asks the central question of learning in games: **when do learning dynamics converge to equilibrium, and to which equilibrium?** Nodes 9.1-9.4 developed four learning algorithms – fictitious play, no-regret learning, CFR, and multiplicative weights – each with its own convergence guarantees. Node 9.5 unifies these results, identifies the structural conditions under which convergence occurs, and maps out the landscape of equilibrium concepts that different learning dynamics target. The key insight, and the thesis's contribution to this node, is that **convergence is a property of the dynamics, not of the equilibrium concept**: the same equilibrium set can be approached by dynamics that converge, cycle, or diverge, and the choice of learning dynamics is a design decision analogous to the choice of optimizer in VQA. The separation of "what we are converging to" (the equilibrium concept) from "how we get there" (the learning dynamics) is the exact analogue of the thesis's separation of objective from dynamics in VQA optimization.

Formal Definition

Equilibrium concepts (hierarchy). Consider a finite n-player game $G = (N, \{S_i\}, \{u_i\})$ and a sequence of mixed strategies $(\sigma^1, \sigma^2, \ldots)$ generated by a learning algorithm.

- **Nash equilibrium (NE).** A profile σ^* where no player can improve by unilateral deviation. The strictest standard concept.
- **Correlated equilibrium (CE).** A distribution $\mu \in \Delta(S)$ such that each player's recommended action is optimal given the recommendation: $\mathbb{E}_{s_{-i}|s_i}[u_i(s_i, s_{-i})] \geq \mathbb{E}_{s_{-i}|s_i}[u_i(s_i', s_{-i})]$ for all s_i'. Strictly weaker than NE: NE $\subset$ CE.
- **Coarse correlated equilibrium (CCE).** A distribution μ such that no player can improve by committing to a fixed action *before* seeing their recommendation: $\mathbb{E}_{s\sim\mu}[u_i(s_i, s_{-i})] \geq \mathbb{E}_{s\sim\mu}[u_i(s_i', s_{-i})]$. Strictly weaker than CE: CE $\subset$ CCE.

The hierarchy is NE $\subset$ CE $\subset$ CCE, with strict inclusions in general.

Convergence notions.

- **Time-average convergence.** The empirical distribution of play $\bar{\sigma}^T = \frac{1}{T}\sum_{t=1}^{T} \sigma^t$ converges to an equilibrium set (e.g., to a CE or CCE).
- **Last-iterate convergence.** The actual strategy σ^T (not the average) converges to an equilibrium.
- **Regret convergence.** The total regret R^T grows sublinearly ($R^T = o(T)$), which implies time-average convergence to CCE. Internal regret $R^T_{\text{int}} = o(T)$ implies convergence to CE.

Main convergence results.

Learning algorithm	Convergence guarantee	Target equilibrium
Fictitious play (9.1)	Converges in 2-player zero-sum, 2x2, potential games; may cycle in general	NE (when it converges)
No-regret / MWU (9.2, 9.4)	Time-average converges always; last-iterate may cycle	CCE (external regret), CE (internal regret)
CFR (9.3)	Time-average converges in 2-player zero-sum extensive-form games	NE of the extensive-form game
Optimistic MWU / gradient descent-ascent	Last-iterate converges in 2-player zero-sum	NE

Intuition

The landscape of convergence results is shaped by a fundamental tension: **learning dynamics that are powerful enough to guarantee convergence in general games can only guarantee convergence to weak equilibrium concepts (CCE), while dynamics that converge to strong concepts (NE) only work in restricted game classes.**

This is not a failure of specific algorithms – it is a structural limitation. Hart and Mas-Colell (2003) showed that any "uncoupled" learning dynamics (where each player's strategy update depends only on their own payoffs, not on others' payoff functions) cannot converge to Nash equilibrium in all games. The impossibility is fundamental: Nash equilibrium requires mutual best-response consistency, but uncoupled learning cannot coordinate on this consistency without access to the other players' payoff structure.

The practical consequence is a design tradeoff:

1. **If you want guaranteed convergence in all games**, use no-regret learning (MWU, Hedge). You get CCE convergence. This is enough for many applications but does not pin down a unique outcome.
2. **If you want Nash convergence**, restrict to a well-behaved game class. Two-player zero-sum games are the most important class: CFR and optimistic gradient descent-ascent both converge to NE in last iterate (or time-average) for these games. Potential games and supermodular games also have strong convergence results.
3. **If you want Nash convergence in general games**, you need coupled dynamics (where each player knows or can infer others' payoffs). This is the mechanism-design route: design the game so that the learning dynamics produce the desired outcome.

The thesis's contribution is to frame this tradeoff as a **separation of objective from dynamics**. The equilibrium concept (NE, CE, CCE) is the objective – the target set the dynamics are trying to reach. The learning algorithm is the dynamics – the update rule that generates the trajectory. The convergence theorem specifies which (dynamics, objective) pairs are compatible. This is exactly the VQA framework: the cost function is the objective, the optimizer is the dynamics, and the convergence analysis asks which (optimizer, cost function) pairs yield convergence.

The deeper parallel: in VQA, the thesis argues that the right approach is to design the optimizer (dynamics) to match the structure of the cost landscape (objective). In learning-in-games, the same principle holds: the right approach is to choose the learning algorithm (dynamics) to match the structure of the game (objective). CFR is matched to extensive-form zero-sum games; MWU is matched to general games with external-regret targets; fictitious play is matched to zero-sum and potential games. Mismatch leads to cycling, divergence, or convergence to the wrong concept.

Structural Role

9.5 is the synthesis node of Module 9. It collects the convergence results of 9.1-9.4 into a unified framework and maps out the landscape of what is possible and what is not. It also serves as the bridge to Module 10:

- **9.1-9.4** provide the algorithms and their individual convergence results.
- **9.5** provides the unified view: which algorithms converge to which concepts in which game classes.
- **9.6 (RL in games)** extends to the setting where players do not even know their own payoff functions and must learn from realized rewards.
- **10.5-10.6** use the convergence landscape to discuss institutional design and meta-games: if you want a particular equilibrium, you need to choose (or design) the game and the learning dynamics together.

The thesis voice is strongest here. The framing of convergence as "separation of objective from dynamics" is a direct import from VQA Modules 6-7. The claim is that this framing is not just an analogy – it is a structural isomorphism. The same mathematical objects appear in both settings: gradient-like dynamics on strategy/parameter spaces, state-dependent objectives (payoffs depend on all players' strategies, cost landscapes depend on the quantum state), and convergence conditions that depend on the interaction between the dynamics and the objective's curvature.

Mathematical Development

No-regret implies CCE convergence

Theorem. If all players use no-regret learning (external regret $R_i^T = o(T)$ for each i), then the empirical distribution of play $\bar{\mu}^T = \frac{1}{T}\sum_{t=1}^{T} \delta_{s^t}$ converges to the set of CCE as $T \to \infty$.

Proof sketch. The CCE condition for distribution μ is: for every i and every alternative action s_i':

$$\mathbb{E}_{s\sim\mu}[u_i(s_i, s_{-i})] \geq \mathbb{E}_{s\sim\mu}[u_i(s_i', s_{-i})].$$

This is equivalent to: no player's external regret is positive. If each player's regret is $o(T)$, then the time-average $\bar{\mu}^T$ approximately satisfies the CCE condition with error $O(R^T/T) \to 0$. The empirical distribution converges to the CCE set.

Internal regret implies CE convergence

Theorem. If all players use no-internal-regret learning (internal regret $R_i^{T,\text{int}} = o(T)$ for each i), then $\bar{\mu}^T$ converges to the set of CE.

Internal regret measures: the maximum improvement from swapping every instance of action a to action b in the history, for all pairs (a, b). CE requires that conditional on each recommended action, no switch is profitable – which is precisely the no-internal-regret condition.

Hart-Mas-Colell impossibility

Theorem (Hart-Mas-Colell 2003). There is no uncoupled learning dynamics that converges to Nash equilibrium in all finite games.

Uncoupled means: player i's update depends only on the history of play and their own payoff function u_i, not on others' payoff functions. This is the natural restriction for decentralized learning.

The proof constructs a three-player game where any uncoupled dynamics must cycle or diverge. The structural reason: Nash equilibrium requires mutual consistency of best responses, which is a global property of the game. Uncoupled dynamics can only detect local information (own payoffs), which is insufficient to verify global consistency.

Convergence in two-player zero-sum games

Two-player zero-sum games are the exception to the impossibility. The minimax theorem (von Neumann 1928) ensures that the Nash equilibrium is equivalent to both players minimizing regret, and any pair of no-regret learners converges to NE in time-average.

Stronger results:

- **CFR (9.3)** converges to NE in time-average in extensive-form zero-sum games, with rate $O(1/\sqrt{T})$.
- **Optimistic MWU (Rakhlin-Sridharan 2013, Daskalakis et al. 2015)** achieves last-iterate convergence in matrix zero-sum games.
- **Gradient descent-ascent with extragradient (Korpelevich 1976, Tseng 1995)** converges in continuous zero-sum games.

Convergence in potential games

Theorem. In finite potential games (games where a single potential function $\Phi(s)$ satisfies $u_i(s_i', s_{-i}) - u_i(s_i, s_{-i}) = \Phi(s_i', s_{-i}) - \Phi(s_i, s_{-i})$ for all i, s_i, s_i'), fictitious play converges to NE, best-response dynamics converge to NE, and any no-regret dynamics converges to the set of NE (in time-average).

The reason: the potential function is a Lyapunov function for the dynamics. Every best-response improvement increases Φ, and the dynamics converge to a local maximum of Φ, which is a Nash equilibrium.

Convergence in supermodular games

Theorem (Milgrom-Roberts 1990). In finite supermodular games (8.6), any adaptive learning dynamics converges to the interval $[\underline{s}, \bar{s}]$ between the least and greatest Nash equilibria.

The monotonicity of best responses ensures that the dynamics are sandwiched between two converging sequences (from the top and bottom of the strategy space), both of which converge to the equilibrium set.

The thesis framing: convergence as dynamics-objective compatibility

The convergence landscape can be organized as a compatibility matrix:

Game class	No-regret (MWU)	Fictitious play	CFR	Optimistic GDA
General	CCE ✓	May cycle	CCE ✓	May cycle
2P zero-sum	NE (avg) ✓	NE ✓	NE (avg) ✓	NE (last-iterate) ✓
Potential	NE (avg) ✓	NE ✓	N/A	NE ✓
Supermodular	NE interval ✓	NE ✓	N/A	NE ✓

The thesis claim: the right way to read this table is as a design guide. Given a game class (the "plant"), choose the learning algorithm (the "controller") that is compatible. Do not use fictitious play in general games (it may cycle). Do not use generic no-regret in two-player zero-sum if you want last-iterate NE (use optimistic methods instead). The choice is a control-design problem, not a mathematical inevitability.

Worked Example

Learning dynamics in a 2x2 coordination game

Game:

	A	B
A	$(3,3)$	$(0,0)$
B	$(0,0)$	$(2,2)$

Two pure NE: (A, A) and (B, B). One mixed NE: $(2/5, 2/5)$ (each plays A with probability $2/5$).

Fictitious play. Start with uniform prior. Period 1: both play A (best response to uniform). Period 2: empirical frequency is (A), so both play A again. Convergence to (A, A) immediately. If started with prior slightly favoring B: convergence to (B, B). Fictitious play converges to a pure NE determined by initial conditions. It does *not* converge to the mixed NE.

No-regret (MWU). Both players run MWU with learning rate η. The dynamics cycle near the mixed NE: weights oscillate around $(2/5, 3/5)$ but the time-average converges to the mixed NE (or to a CCE that includes it). Last-iterate does not converge – it cycles.

Interpretation. Different dynamics converge to different equilibria of the same game. The "right" equilibrium depends on the dynamics, not on the game alone. This is the selection-by-dynamics principle.

Learning in a 2-player zero-sum game (Matching Pennies)

Game: $(1, -1), (-1, 1), (-1, 1), (1, -1)$. Unique NE: $(1/2, 1/2)$.

Fictitious play. Converges in time-average to $(1/2, 1/2)$. Last-iterate cycles (Shapley 1964 showed this).

MWU. Time-average converges to NE. Last-iterate cycles (proven by Mertikopoulos et al. 2018).

Optimistic MWU. Last-iterate converges to $(1/2, 1/2)$. This is the state of the art for zero-sum last-iterate convergence.

CFR. Time-average converges to NE. This is the algorithm that solved heads-up limit Texas hold'em (Bowling et al. 2015).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Nash equilibrium	GTO strategy – the mutual-best-response profile
Correlated equilibrium	A "correlated play" recommendation that no player deviates from
CCE	A weaker notion: no player can improve by committing to a fixed strategy regardless of what they're told
No-regret learning	A player who minimizes regret over the long run by adjusting their strategy
CFR convergence	The solver algorithm that computes NE in poker (two-player zero-sum extensive form)
Last-iterate convergence	Whether the solver's *current* strategy (not the average) is close to GTO
Time-average convergence	Whether the solver's *average* strategy over all iterations is close to GTO

Concrete situation: why solvers use CFR and not MWU

Poker solvers (PioSolver, GTO+, Libratus) use CFR or its variants (CFR+, DCFR, LCFR) rather than generic no-regret algorithms like MWU. The reason is structural compatibility: poker is a two-player zero-sum extensive-form game, and CFR is specifically designed for this game class. CFR exploits the extensive-form structure (traversing the game tree, computing counterfactual values at each information set) in a way that MWU cannot. The convergence rate of CFR in poker-sized games is practical ($O(1/\sqrt{T})$ for time-average

NE, with DCFR achieving faster practical convergence); the convergence rate of generic MWU would be impractical for the same game sizes.

This is the thesis's point: the choice of learning algorithm (optimizer) must match the structure of the game (objective). CFR matched to extensive-form zero-sum is the right control design. MWU matched to the same game would technically converge but at impractical rates.

What this understanding buys you

Solver output is a time-average, not a last-iterate. When you load a PioSolver solution, you are seeing the time-averaged strategy over many CFR iterations, not the strategy from the last iteration. This matters because the time-average converges to NE while the last iterate may cycle. If you export a "partially converged" solver solution, the time-average is much closer to NE than the current-iteration strategy.

Convergence tells you what you can compute, not what opponents play. The convergence theory says that CFR finds NE of two-player zero-sum poker. It does not say that your opponents play NE. The gap between "what the solver computes" and "what opponents do" is the gap between equilibrium and learning dynamics. Understanding the convergence landscape helps you understand where solver output is reliable (in well-structured game classes) and where it is not (in multi-player or non-zero-sum settings).

Cross-Links

- **9.1-9.4** – the algorithms whose convergence this node unifies.
- **7.5 Stochastic stability** – evolutionary selection among equilibria.
- **7.6 Adaptive dynamics** – the gradient-flow analogue in evolutionary settings.
- **8.6 Strategic complementarities** – the game class with the strongest convergence results.
- **5.4 Folk theorem** – the multiplicity that convergence theory must address.
- **10.6 Strategic ecosystems** – meta-level design of games and learning dynamics together.
- **VQA 6-7 (Optimization dynamics, control view)** – the thesis's framework for separating objective from dynamics.

Structural Compression

Invariant: No-regret learning converges (in time-average) to coarse correlated equilibrium in all games, to correlated equilibrium with internal-regret minimization, and to Nash equilibrium in two-player zero-sum and potential games. No uncoupled dynamics converges to Nash in all games (Hart-Mas-Colell impossibility). The convergence target depends on the pairing of learning dynamics with game structure.

Mechanism: Regret bounds imply time-average convergence via the equivalence between no-regret and approximate CCE/CE conditions; game-class-specific structure (zero-sum, potential, supermodular) provides additional convergence guarantees by supplying Lyapunov functions, minimax duality, or monotonicity.

Cross-System Echo: The convergence landscape of learning in games is structurally isomorphic to the convergence landscape of optimizers on cost functions in VQA. The game class (zero-sum, potential, supermodular) maps to the cost-landscape class (convex, non-convex, noisy). The learning algorithm (MWU, CFR, fictitious play) maps to the optimizer (Adam, HOPSO, natural gradient). The convergence theorem specifies which (optimizer, landscape) pairs converge. The thesis's claim is that this isomorphism is not accidental: both are instances of the same abstract problem – a controller navigating a state-dependent objective using local feedback – and the design principles for good controllers are the same in both settings.

Exercises

1. Show that if both players in a two-player zero-sum game use no-regret learning, the time-average strategy converges to a Nash equilibrium. (Hint: use the minimax theorem and the definition of regret.)
2. Construct a 3-player game where fictitious play cycles forever. (Hint: use Shapley's example or a modification.)
3. Prove the Hart-Mas-Colell impossibility: no uncoupled dynamics converges to NE in all finite games. (Outline the proof structure and identify the key counterexample.)
4. In a 2x2 coordination game with two pure NE, run MWU for both players for 1000 iterations (numerically). Plot the last-iterate strategies and the time-average strategies. Verify that the time-average converges to a CCE while the last-iterate cycles.
5. Show that in a finite potential game, best-response dynamics converge to a Nash equilibrium in finite time. (Hint: the potential function increases at each step and the strategy space is finite.)
6. For the linear-quadratic network game of 8.6, show that iterated best-response dynamics converge to the unique Nash equilibrium. (Hint: the best-response map is a contraction for small λ.)
7. **Argue, structurally,** why the thesis's "separation of objective from dynamics" principle applies to learning in games. Identify the objective, the dynamics, and the feedback loop. What does the Hart-Mas-Colell impossibility tell us about the limits of this framework, and how does game-class restriction (zero-sum, potential) overcome those limits?

Game Theory · Module 9 — Learning In Games · Node 9.5
Concepts: convergence, nash-equilibrium, dynamical-systems
Prerequisites: 9.1, 9.2, 9.3, 9.4
Enables: 9.6, 10.5, 10.6

hbar.university — own.your.cognition