Fictitious Play

Game Theory · Module 9 — Learning In Games

Node 9.1

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.1

Fictitious play is the oldest learning algorithm in game theory and the first procedure by which rational agents who know nothing about each other's internal strategies can converge on a Nash equilibrium through repeated play alone. Introduced by George W. Brown in 1951 as a computational method for solving zero-sum games — it was designed as an algorithm, not as a behavioral model — it turned out to have surprising structural properties that made it the starting point for the entire learning-in-games program. The rule is radical in its minimalism: each player maintains a running count of the opponent's historical actions, treats those counts as a probability distribution, and plays a best response against that distribution. No reasoning about the opponent's rationality, no mutual consistency, no equilibrium calculation — just count and respond. Yet in two-player zero-sum games, this local rule provably produces play whose *empirical* frequencies converge to a Nash equilibrium. Fictitious play is the prototype for every learning dynamic that follows: no-regret algorithms, CFR, replicator dynamics, and reinforcement learning in games all inherit its core question — *what happens when boundedly rational agents adapt to each other over time?*

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite normal-form game played repeatedly at times $t = 1, 2, 3, \ldots$. Let $s_i^t \in S_i$ denote the action player i plays at time t.

Empirical frequency. For each player j and action $s_j \in S_j$, define the empirical frequency of s_j through time t:

$$\bar{\sigma}_j^t(s_j) \;=\; \frac{1}{t}\sum_{\tau=1}^{t} \mathbf{1}[s_j^\tau = s_j].$$

This is a probability distribution $\bar{\sigma}_j^t \in \Delta(S_j)$ — the running histogram of player j's plays.

Fictitious play rule. At every time $t+1$, player i chooses a best response to the profile of opponent empirical frequencies $\bar{\sigma}_{-i}^t$:

$$s_i^{t+1} \;\in\; B_i\big(\bar{\sigma}_{-i}^t\big) \;=\; \arg\max_{s_i \in S_i} \; u_i\big(s_i, \bar{\sigma}_{-i}^t\big)\,.$$

Ties are broken by an arbitrary fixed rule. The initial condition $\bar{\sigma}_{-i}^0$ is a specified prior (commonly a unit count on some action for each opponent).

Bayesian interpretation. The rule has a clean probabilistic reading: player i believes the opponents are playing a stationary mixed strategy σ_{-i}, starts with a Dirichlet prior, updates the posterior by counting, and plays the Bayes-optimal response. The empirical frequency is the posterior mean. The definition does not

need this interpretation, but it clarifies the model of the other players that fictitious play implicitly holds: they are stationary random processes rather than adaptive strategic agents.

Convergence concepts. Three distinct senses of "convergence to Nash":

1. **Empirical convergence:** $\bar{\sigma}^t$ converges (in $\prod_i \Delta(S_i)$) to a Nash equilibrium σ^*.
2. **Action convergence:** s^t eventually stabilizes on a Nash equilibrium profile.
3. **Time-averaged payoff convergence:** $\bar{u}_i^t = \frac{1}{t}\sum_\tau u_i(s^\tau)$ converges to the Nash payoff.

For fictitious play, empirical convergence holds in important classes of games, while action convergence typically fails (players keep cycling even when the frequencies settle).

Intuition

Three lenses on fictitious play, each illuminating a different facet.

The statistician. Player i is doing naive Bayesian inference about a stationary but unknown opponent. Each opponent is treated as a random process whose parameter is being estimated from the data. Fictitious play is the Laplace-style point estimate: use the empirical histogram as your best guess for the underlying mixed strategy, then optimize against it. The flaw — obvious but essential — is that the opponent is not actually a stationary process. The opponent is doing the same thing you are, and their histogram evolves in response to your play. The fictitious-play agent ignores this dependence. Remarkably, in zero-sum games this mis-specification does not prevent convergence of frequencies; in general games, it sometimes does.

The optimizer. Fictitious play is a *best-response dynamic* with memory. At each step you act optimally against the *cumulative* distribution of opponent actions, not just the most recent. This averaging across history is the structural feature that distinguishes fictitious play from simple best-response dynamics (the latter can cycle arbitrarily, e.g., in matching pennies, while the former's *averages* settle).

The computer. Brown's original motivation was numerical: use fictitious play as an iterative method for computing equilibria of zero-sum games, as an alternative to solving the LP directly. For a matrix game A, the fictitious-play iterate computes a sequence of row/column plays whose averages converge to the game's value. Brown and Robinson (1951) proved the convergence for zero-sum games, and the rate is polynomial but slow. This slowness is one reason no-regret methods (Node 9.2) and CFR (Node 9.3) eventually displaced fictitious play as practical equilibrium-computation tools.

The deep lesson of fictitious play is that *consistency is not rationality.* Each player is doing something locally reasonable — best-responding to a frequency count — but the joint behavior only aligns with Nash under specific structural conditions on the game (zero-sum, dominance-solvable, certain potential games). In the wrong games (Shapley's 3x3 counterexample), fictitious play *provably* cycles and never converges. The theoretical frontier of learning in games is carved out by asking: *which games force learning dynamics to reach equilibrium, and which do not?*

Structural Role

Fictitious play sits at the root of Module 9 as the ancestor of every learning dynamic:

- **9.2 No-regret learning** generalizes the best-response-to-empirical rule to algorithms with regret guarantees (fictitious play is not no-regret in general, which is a striking fact).
- **9.3 Counterfactual regret minimization** lifts regret-matching to extensive-form games and is the poker-solving workhorse.
- **9.4 Multiplicative weights** replaces pure best response with an exponentiated-weights soft maximum, trading off determinism for no-regret guarantees.
- **9.5 Convergence to equilibrium** catalogs when learning dynamics reach Nash and when they cycle — Shapley's counterexample to fictitious play is the motivating negative result.

- **9.6 Reinforcement learning in games** replaces the "know-the-payoff-function" assumption of fictitious play with payoff *samples*, connecting to Q-learning and actor-critic methods.

Fictitious play is also the simplest nontrivial example of a dynamic on $\prod_i \Delta(S_i)$, and its analysis introduces the differential-inclusion formalism used throughout the literature on continuous-time learning dynamics.

Mathematical Development

Continuous-time formulation

A cleaner analytical setting: pretend the time index is continuous. Let $\bar{\sigma}_i(t) \in \Delta(S_i)$ denote the empirical frequency, evolving by

$$\frac{d\bar{\sigma}_i}{dt} \;\in\; \frac{1}{t}\Big(B_i(\bar{\sigma}_{-i}(t)) - \bar{\sigma}_i(t)\Big),$$

where B_i is the (possibly set-valued) best-response correspondence. This is a *differential inclusion*, because B_i can be multi-valued. After a time rescaling $\tau = \ln t$, the dynamics become autonomous:

$$\frac{d\bar{\sigma}_i}{d\tau} \;\in\; B_i(\bar{\sigma}_{-i}) - \bar{\sigma}_i.$$

This is the **best-response dynamic**, a continuous-time limit of fictitious play. Its fixed points are exactly the Nash equilibria (because at a fixed point $\bar{\sigma}_i \in B_i(\bar{\sigma}_{-i})$ for every i), and its trajectories are the analytical tool for studying convergence.

Brown–Robinson theorem (zero-sum convergence)

Theorem (Brown 1951, Robinson 1951). In any finite two-player zero-sum game, fictitious play converges in empirical frequencies to the set of Nash equilibria. Specifically, the players' time-averaged payoffs converge to the value of the game.

Proof sketch. Let V be the value of the game. Define the minimax gap at time t:

$$\Delta^t \;=\; \max_{s_1} u_1(s_1, \bar{\sigma}_2^t) - \min_{s_2} u_1(\bar{\sigma}_1^t, s_2).$$

This is the "exploitability" of the empirical profile. By the fictitious-play rule, each player at each step picks an action that is optimal against the opponent's empirical frequency. Robinson's argument shows $\Delta^t \to 0$ by a careful induction on the matrix dimensions, using a potential function that tracks the difference between upper and lower envelopes. The convergence rate is polynomial but slow:

$$\Delta^t \;=\; O\Big(t^{-1/(m+n-2)}\Big)$$

for an $m \times n$ matrix game. For even moderately sized games this rate is impractical, which is why no-regret methods eventually replaced fictitious play in computational work.

Zero-sum is special

The Brown–Robinson proof relies heavily on the zero-sum structure. For general-sum games, fictitious play can diverge. The next results carve out the positive classes and illustrate the negative boundary.

Convergent classes beyond zero-sum

- **Dominance-solvable games.** In games solvable by iterated strict dominance, fictitious play converges to the unique surviving profile. Miyasawa (1961) proved convergence for all 2x2 games.
- **Games with a potential.** In a finite game admitting a potential function $\Phi : S \to \mathbb{R}$ such that unilateral deviations change each player's payoff by the same amount as the change in Φ, fictitious play converges to a Nash equilibrium (Monderer and Shapley 1996). This class includes coordination games and congestion games.
- **Supermodular games.** Many classes of games with strategic complementarities also admit convergence results.

Shapley's counterexample (1964)

Consider the 3x3 game with payoff bimatrix

	L	C	R
T	$(0,0)$	$(1,0)$	$(0,1)$
M	$(0,1)$	$(0,0)$	$(1,0)$
B	$(1,0)$	$(0,1)$	$(0,0)$

This is a non–zero-sum variant of Rock-Paper-Scissors. The unique Nash equilibrium is the uniform mixture $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for both players. But **fictitious play does not converge** to this equilibrium from most initial conditions. Shapley showed the empirical frequencies cycle forever, with the sequence of best-response regions traversed in a rotational pattern: each player's best response chases the other's shifting frequency around the simplex, and the frequencies themselves trace out a limit cycle that fails to reach the centroid.

The Shapley counterexample is the canonical negative result of learning in games. It shows that "everybody best-responds to empirical opponent play" is not enough to guarantee convergence outside of special game classes — and it motivates all the structural refinements studied in 9.2–9.5.

The miscoordination problem

Even in games where fictitious play converges in empirical frequencies, *action* convergence typically fails. Consider Matching Pennies: the Nash equilibrium is $(\frac{1}{2}, \frac{1}{2})$ for both. Under fictitious play, each player alternately best-responds to whichever action their opponent has played more often, causing both players to flip their actions in a predictable pattern. The empirical frequencies converge to $\frac{1}{2}$ but the players never settle on a pure action profile. This is fine when all you care about is the distribution of play, but for "predicting what will be played next turn" fictitious play offers no consolation.

Worked Example

Fictitious play in Matching Pennies

Row chooses Heads (H) or Tails (T). Column chooses H or T. Row wins +1 on match, Column wins +1 on mismatch. Initialize Row's belief about Column as (H: 0, T: 1); initialize Column's belief about Row as (H: 1, T: 0).

Each round, each player best-responds to the current empirical belief and the count is updated. Row best-responds to Column = T by playing T; Column best-responds to Row = H by playing T. Then Row updates Column's histogram to (H: 0, T: 2), still best-responds with T; Column updates Row's histogram to (H: 1, T: 1), ties — say plays H. And so on.

Tracing the trajectory, each player follows the opponent's empirical frequency with a lag. Row eventually switches from T to H, Column eventually switches from H to T, and so on. The empirical frequencies $\bar{\sigma} \to (\frac{1}{2}, \frac{1}{2})$ for both — which is the Nash equilibrium — but the actions themselves form an infinite alternating sequence that never stabilizes.

Fictitious play in a dominance-solvable game

Consider the Prisoner's Dilemma with payoffs $(C, C) = (3, 3), (C, D) = (0, 5), (D, C) = (5, 0), (D, D) = (1, 1)$. Defect is strictly dominant for both players. Whatever the initial empirical counts, Row's best response is always D (because $u_1(D, \cdot) > u_1(C, \cdot)$ for every opponent distribution). Same for Column. So from the first step, $s_1^t = s_2^t = D$ for every t, and the empirical frequencies converge instantly to (D, D), the dominance-solvable Nash equilibrium.

Fictitious play in Shapley's 3x3 counterexample

Initialize with unit weight on (T, L). Row sees Column's L and best-responds with B. Column sees Row's T and best-responds with C. Next round: (B, C). Row sees Column's (L: 1, C: 1) and best-responds with M (tie-breaking aside). Column sees Row's (T: 1, B: 1) and best-responds with R. And so on.

Tracing the trajectory, the empirical frequencies move in a rotational pattern around the uniform distribution but never settle there. The limit set is a cycle, not a point. Simulations show the cycle persists indefinitely with amplitude that does not shrink. This is the structural reason fictitious play cannot serve as a universal equilibrium-computation method: for games like Shapley's, the rule is simply wrong.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Empirical frequency $\bar{\sigma}_j^t$	Your HUD stats on an opponent — VPIP, PFR, 3-bet%, c-bet% — running averages of observed actions
Best response $B_i(\bar{\sigma}_{-i}^t)$	The maximally exploitative strategy against the opponent's HUD profile
Fictitious play update	After each hand, update HUD stats and recompute your exploitative plan
Convergence to Nash (zero-sum)	In a heads-up match, repeated exploitation by both players drives play toward GTO
Shapley-style cycle	Two opponents stuck in a rock-paper-scissors loop: each "adjustment" spawns a counter-adjustment that leaves the cycle

Concrete situation: reg-vs-reg HUD-based adjustment

You play long sessions against a pool of regulars at mid stakes. You keep HUD stats on each regular: their preflop frequencies, c-bet frequencies, check-raise frequencies, etc. Over time these become reliable estimates of their mixed strategies at various nodes. Your own strategy is a best response to those running statistics, updated continuously as you see more hands.

This is exactly fictitious play. Your action each hand is the best response to the opponent's empirical frequencies aggregated over all prior observations. The opponents, similarly, are observing your frequencies and adjusting. If the game were two-player zero-sum (it is, modulo rake), fictitious play theory says the empirical frequencies of both players would converge to Nash — the GTO solution — over sufficiently long play. In practice, convergence is slow (Brown–Robinson rate $O(t^{-1/(m+n-2)})$, which for a realistic game tree with hundreds of info sets is glacial), which is why actual poker learning uses CFR (9.3) offline rather than waiting for the mutual-exploitation dynamics to settle at the table.

Concrete situation: the "leveling war" failure mode

In heads-up matches between two aggressive regulars, a common failure of fictitious-play-like reasoning is the "leveling war": I exploit your high c-bet rate by check-raising more, you exploit my high check-raise rate by c-betting thinner, I exploit *that* by check-raising even more, and so on. This is precisely the structural failure mode that Shapley's counterexample demonstrates in the abstract: when the game's best-response dynamics generate cycles rather than attracting to a fixed point, mutual adjustment does not reach equilibrium — it produces an oscillation whose empirical average may or may not be Nash.

The practical upshot: if you are playing someone who adjusts faster than you, pure exploitation will lose because you chase their shifts without ever getting ahead of them. You want a strategy closer to the Nash, which does not require prediction.

What this understanding buys you

Your HUD stats are a fictitious-play prior on your opponent. Recognizing that "I use HUD stats to best-respond" is *literally fictitious play* tells you two things. First, under zero-sum assumptions with a stationary opponent, your empirical exploitation will converge toward optimal — slowly. Second, if your opponents are adapting in parallel, you cannot trust the convergence theorem, because the stationarity assumption is violated. In a reg pool where everyone studies, everyone is updating their strategies on the basis of everyone else's statistics, and the "empirical frequency" you see is a lagged picture of a moving target. Fictitious play's mathematical guarantee is no longer in force.

Static GTO baselines are the answer to non-stationary pools. If fictitious-play dynamics can cycle, what you want is a strategy whose guarantee does not depend on opponent stationarity. That strategy is Nash (GTO): it guarantees the game's value regardless of opponent behavior. This is why the shift from exploit-first to GTO-first poker coincided with the realization, at the theoretical level, that exploit-dynamics can fail to converge in non-zero-sum or non-stationary settings.

Cross-Links

- **1.4 Best response** — the core operation fictitious play iterates.
- **2.1 Nash equilibrium** — the target of convergence (when convergence holds).
- **2.3 Mixed-strategy equilibrium** — the empirical frequency is a mixed strategy.
- **9.2 No-regret learning** — the successor framework with stronger guarantees.
- **9.3 CFR** — the extensive-form descendant that dominates poker solving.
- **9.5 Convergence to equilibrium** — the general theory of which learning dynamics reach Nash.
- **Statistics 3.x Bayesian inference** — the Laplace-point-estimate interpretation.
- **Dynamical systems theory** — differential inclusions, limit cycles.

Structural Compression

Invariant: A player maintaining a running histogram of the opponent's actions and best-responding to the histogram generates an adaptive dynamic whose empirical frequencies converge to Nash equilibrium in two-player zero-sum and certain other structured games, but can cycle indefinitely in general games.

Mechanism: Treat opponent actions as samples from an unknown stationary distribution; estimate the distribution by the empirical histogram; play a best response to the estimate. Iterate. The continuous-time limit is the best-response differential inclusion.

Cross-System Echo: Fictitious play is a stateless version of Bayesian belief updating in a misspecified model: the agent assumes stationarity, estimates via Laplace point estimation, and optimizes greedily. This pattern — local rationality under a mis-specified generative model — recurs throughout statistics (empirical risk minimization), control (certainty-equivalence control), and ecology (optimal foraging under stationary

resource assumptions). In each case, the interesting phenomena happen when the mis-specification bites back, producing limit cycles, cascading failures, or equilibrium selection pathologies.

Exercises

1. Simulate fictitious play on the 2x2 matching-pennies game with initial beliefs $\bar{\sigma}_2^0 = (0.6, 0.4)$ and $\bar{\sigma}_1^0 = (0.3, 0.7)$. Verify that empirical frequencies approach $(\frac{1}{2}, \frac{1}{2})$ for both players while action sequences exhibit persistent alternation.
2. Prove Miyasawa's result: fictitious play converges to a Nash equilibrium in every finite 2x2 game. (Hint: enumerate the structural types — zero-sum, dominance-solvable, coordination, anti-coordination.)
3. Write the continuous-time best-response dynamic for Shapley's 3x3 counterexample and argue that there is a limit cycle around the uniform distribution. Compute the approximate period of the cycle from the best-response structure.
4. Show that in any finite potential game with potential Φ, the fictitious-play empirical frequencies converge to a Nash equilibrium. (Hint: Φ evaluated at the empirical profile is approximately monotone.)
5. Derive the Brown–Robinson convergence rate $O(t^{-1/(m+n-2)})$ for fictitious play on an $m \times n$ zero-sum game, using a bound on the minimax gap Δ^t. Compare to the $O(t^{-1/2})$ rate achievable by multiplicative-weights methods (Node 9.4).
6. Consider the 3x3 Rock-Paper-Scissors game (zero-sum). Prove that fictitious play converges to the uniform equilibrium. Why does the zero-sum structure rescue convergence that Shapley's non-zero-sum variant destroys?
7. **Argue, structurally,** why a learning rule that ignores its own effect on opponent play can converge in zero-sum games but not in general-sum games. What is the structural feature of zero-sum that makes "pretend the opponent is stationary" a viable approximation, and what breaks in general-sum?

Game Theory · Module 9 — Learning In Games · Node 9.1
Concepts: bayesian-updating, convergence, nash-equilibrium
Prerequisites: 1.4, 2.1, 2.3
Enables: 9.2, 9.3, 9.5

hbar.university — own.your.cognition