

Working Paper

On the Nonconvergence of Fictitious Play in Coordination Games

Dean Foster and Peyton Young

WP-95-001
January 1995



International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria

Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

On the Nonconvergence of Fictitious Play in Coordination Games

Dean Foster and Peyton Young

WP-95-001
January 1995

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.



International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria

Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

On the Nonconvergence of Fictitious Play in Coordination Games

Dean Foster and Peyton Young

It is natural to conjecture that fictitious play converges in coordination games, but this is shown by counterexample to be false. Variants of fictitious play in which past actions are eventually forgotten and there are small stochastic perturbations are much better behaved: over the long run players manage to coordinate with high probability.

Key words: fictitious play, coordination game, learning dynamics

Professor Dean Foster
Department of Statistics
University of Pennsylvania
Philadelphia PA 19104-6302

Peyton Young
Department of Economics
Johns Hopkins University
Baltimore MD 21218-2685

On the Nonconvergence of Fictitious Play in Coordination Games

Dean Foster and Peyton Young

Consider a strategic form, two-person game G with finite strategy spaces X_1, X_2 . A *fictitious play sequence* is an infinite sequence $x(t) \in X = X_1 \times X_2$ such that, for $i = 1, 2$ and every $t \geq 1$, $x_i(t + 1)$ is a best response by i to the empirical frequency distribution $f^t(x_{-i})$ of the actions x_{-i} taken by the other side up through time t . (We assume $x(1)$ is arbitrary.) G has the *fictitious play property* if every limit point of the sequence $\{f^t(x_1), f^t(x_2)\}$ is a Nash equilibrium (pure or mixed) of G . Fictitious play was originally proposed as an algorithm for computing equilibria in games (Brown, 1951), and indeed it does have this property for zero-sum games (Robinson, 1951) and 2×2 games (Myasawa, 1961). However, a well-known 3×3 example due to Shapley (1964) shows that games in general do not have the fictitious play property.

Recently there has been a revival of interest in fictitious play and related processes that model how players learn to play a game. (See among others Foster and Young, 1990; Milgrom and Roberts, 1991; Crawford, 1991; Fudenberg and Kreps, 1993; Kandori, Mailath and Rob, 1993; Young, 1993; Kaniovski and Young, 1994; Blume, 1995). Hence it is of interest to know whether fictitious play converges for substantial classes of games, if not for all possible games. Among the classes for which convergence has been established are dominance-solvable games (Milgrom and Roberts, 1991), two-person games with strategic complementarities and diminishing returns (Krishna, 1991), and games with identical interests, that is, games that are best-reply equivalent in mixed strategies to a game in which all players have identical payoff functions (Monderer and Shapley, 1993a).

Many of the games for which fictitious play has been shown to converge have an acyclic best-reply structure. To make this idea precise, let us associate with each strategy profile $x \in X$ a node of a graph. Draw a directed edge from node x to node x' if and only if for some player i , $x'_i \neq x_i$, x'_i is a best reply to x_{-i} , and $x_{-i} = x'_{-i}$. This is called the *best-reply graph* of G . The game G is *acyclic* if it contains no directed cycles. It is *weakly acyclic* if from every node there exists a directed path to a strict Nash equilibrium, that is, to a node

that has no exiting edge (Young, 1993).¹ Every game with identical interests in which no two strategy profiles yield the same payoff for both players is acyclic. (Every directed path in the best-reply graph strictly increases the payoff of both players, so it cannot cycle.)

A particularly natural class of games with an acyclic structure are the coordination games. By a *coordination game* we mean a two-person, $n \times n$ matrix game such that every strategy pair of form (x_j, x_j) is a strict Nash equilibrium, $1 \leq j \leq n$. In the best-reply graph of this game, every edge is directed toward a coordination equilibrium and no edge is directed away from such an equilibrium. Hence there can be no best-reply cycles.

We now exhibit a coordination game that does not have the fictitious play property.

	A'	A''	B'	B''	C'	C''	D'	D''
A'	24, 24	6, 6	0, 18	0, 18	18, 0	18, 0	5, 0	0, 0
A''	6, 6	24, 24	0, 18	0, 18	18, 0	18, 0	4, 0	0, 0
B'	18, 0	18, 0	24, 24	6, 6	0, 18	0, 18	3, 0	0, 0
B''	18, 0	18, 0	6, 6	24, 24	0, 18	0, 18	2, 0	0, 0
C'	0, 18	0, 18	18, 0	18, 0	24, 24	6, 6	1, 0	0, 0
C''	0, 18	0, 18	18, 0	18, 0	6, 6	24, 24	0, 0	0, 0
D'	0, 4	0, 5	0, 2	0, 3	0, 0	0, 1	6, 6	-6, -6
D''	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	-6, -6	6, 6

¹There are several variations of this definition. For example, we could draw a directed edge from x to x' if and only if some player i *strictly prefers* x'_i to x_i given the strategy tuple x_{-i} , and $x'_{-i} = x_{-i}$. (Thus x'_i need not be a best reply to x_{-i} .) The game G is said to have the *finite improvement property* if this graph has no directed cycles (Monderer and Shapley, 1993b). If G has the finite improvement property and there are no ties in payoffs, then G is acyclic, but the converse does not necessarily hold. Another variant is the following: draw an edge from x to x' if x'_i is a *strict* best reply to x_{-i} , $x'_i \neq x_i$, and $x_{-i}' = x_{-i}$. This is a less stringent version of acyclicity. See Monderer and Sela (1994) for a discussion of these ideas and their connections with fictitious play.

Consider a fictitious play sequence in which Row chooses D' and Column chooses D'' in the first period. In the next period the best replies are D'' for Row and D' for Column, and the process unfolds as shown below:

t =	1	2	3	4	5	6	...	17	18	19	20	...	91	92	93	
Row	D'	D''	A'	A''	B'	B''	...	B'	B''	C'	C''	...	C'	C''	A'	...
Column	D''	D'	A''	A'	B''	B'	...	B''	B'	C''	C'	...	C''	C'	A''	...

The role of D' and D'' is to break ties asymmetrically; after the first two rounds they are never chosen again. Given these initial two choices, Row has a slight preference for ' strategies over '' strategies within each category (A, B, C), whereas Column prefers the reverse. This leads the players to coordinate within the same *category* of strategy, but they never actually coordinate. Instead, the process cycles between regimes in which an alternating series of mismatched pairs of A are played, followed by an alternating series of mismatched pairs of B, then of C, then back to A, and so forth. Call each of these alternating series a *run*. Let r_k be the number of periods in run k . The first three runs are of length $r_1 = 2$, $r_2 = 14$, and $r_3 = 74$. In general we have the following recursion

$$\text{for all } k \geq 0, \quad r_{k+3} = 6r_{k+2} - 5r_{k+1}. \tag{1}$$

From this it follows that each run is about five times as long as the previous one. Hence the empirical frequency distribution of strategies never converges, so *a fortiori* it does not converge to an equilibrium.

To prove (1) we proceed by induction on k . For $k = 0$ the result follows by plugging in the values $r_1 = 2$, $r_2 = 14$, and $r_3 = 74$. Suppose now that $k > 0$. Since the game is symmetric in A, B, C, there is no loss of generality in assuming that the $(k + 3)$ rd run is an A-series, that is, $k = 1 \pmod{3}$. Thus the $(k + 2)$ nd run is a C-series, and the $(k + 1)$ st run is a B-series. To find which strategy is a best response by Row at any given time t , it suffices to compute the hypothetical total payoff (to Row) of each strategy assuming it were played against all previous choices by Column up through time $t - 1$. Call this the *score* of the strategy at time t . Fictitious play stipulates that in each period Row choose a strategy with highest score.

Consider the $(k + 3)$ rd run of A's. Each time that Column plays A'A" in succession, both A-strategies for Row increase their score by $24 + 6 = 30$, both B-strategies increase their score by $18 + 18 = 36$, and both C-strategies increase their score by zero. In particular B' gains 6 points relative to A' in every two periods of the current run. Let $S_{A'}$ and $S_{B'}$ be the scores of A' and B' at the *beginning* of the run. Let $[x]$ denote the least integer greater than or equal to x . Then it takes $r_{k+3} = 2[(S_{A'} - S_{B'})/6]$ periods for B' to overtake A' (i.e., for B' to become a better reply than A' by Row), which ends this run and starts the next one.

It remains to compute the difference $S_{A'} - S_{B'}$. Consider the first period of the $(k + 1)$ st run. At this point, B' has just overtaken A'. Moreover if their scores are $S^*_{A'}$ and $S^*_{B'}$, then we have $0 < S^*_{B'} - S^*_{A'} < 6$. (This is because they start period 3 with a difference that is less than 6, and all subsequent actions change the scores by multiples of 6.) During the ensuing B-series, which lasts for r_{k+1} periods, A' increases its score by 0, B' increases its score by $30(r_{k+1})/2$, and C' increases its score by $36(r_{k+1})/2$. After this the C-series commences. This run increases the score of A' by $36(r_{k+2})/2$, the score of B' by 0, and the score of C' by $30(r_{k+2})/2$. Thus we have

$$S_{A'} = 36(r_{k+2})/2 + S^*_{A'} \text{ and } S_{B'} = 30(r_{k+1})/2 + S^*_{B'}. \quad (2)$$

We may assume by induction that r_{k+1} and r_{k+2} are even. From (2) it follows that

$$(S_{A'} - S_{B'})/6 = 6(r_{k+2})/2 - 5r_{k+1}/2 + (S^*_{A'} - S^*_{B'})/6.$$

We also know that $-5 \leq (S^*_{A'} - S^*_{B'}) \leq -1$. Hence

$$[(S_{A'} - S_{B'})/6] = 6(r_{k+2})/2 - 5(r_{k+1})/2,$$

and therefore

$$r_{k+3} = 2 [(S_{A'} - S_{B'})/6] = 6(r_{k+2}) - 5(r_{k+1}).$$

Hence r_{k+3} is even and formula (1) holds for k , from which it follows by induction that (1) holds for all k .

We can think of this game as modelling a squabble among competing doctrines. Imagine two groups of academics (or politicians or religious leaders) who periodically announce a position on some issue. There are three types of positions -- A, B, C -- and each position has two specific variants. It is in the interest of both groups to coordinate on the same position and the same variant of that position. The difficulty is that their preferences differ when they become involved in a doctrinal squabble. By a "squabble" we mean prevarication between two variants of the same policy, say a fifty-fifty probability mixture between A' and A". Once a squabble starts the parties keep shifting position. Both Row and Column prefer either version of B to an A-squabble. The trouble is that their most preferred versions of B differ (because of the initial choice of D-strategies), which leads to a B-squabble. Compared to a B-squabble they would rather choose either version of C, but again they cannot agree on which version of C. Thus one squabble begets another.

Does this counterexample show that agents cannot learn to play coordination equilibria over time? We think not. The reason is that such examples are knife-edge in construction. If there are small stochastic variations and past actions are eventually forgotten (which we think are characteristic of most learning processes), then the process exhibits much better long-run behavior.

To be concrete, suppose that for some large integer m we truncate each fictitious play sequence $\mathbf{x}(t)$ to the most recent m periods. Thus actions more than m periods old are forgotten, and the state at the beginning of time t is a sequence $h(t)$ of form $(\mathbf{x}(t - m), \mathbf{x}(t - m + 1), \dots, \mathbf{x}(t - 1))$ if $t > m$, and of form $(\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t - 1))$ if $1 < t \leq m$. (The process begins at $t = 1$ with the empty sequence.) Suppose further that the players only have incomplete information about what the others have done in the past. In each period $t > 1$, each agent chosen to play draws a random sample of size k without replacement from the sequence $h(t)$. (If $h(t)$ is of length less than k , all entries are sampled.) The draws are independent for the two agents. Each agent then chooses a best reply to the empirical frequency distribution (in his sample) of what the other side has done. It is easy to see that every state in which the same coordination equilibrium is played m times in succession is absorbing. Moreover these are the only absorbing states. It can be shown that, if k/m is sufficiently small (in particular if $k/m \leq 1/2$), the process converges with probability one to an absorbing state. In other words, a coordination equilibrium will eventually be played with probability one (Young, 1993, Theorem 1). the reason this works is that the stochastic variability created by incomplete sampling eventually jostles the process out of uncoordinated cycles. Once the process hits an

absorbing state, however, the sampling variability vanishes and the process stays there forever.

Similar results obtain under other kinds of stochastic perturbation. Suppose, for example, that there is some systematic "error" in the players' responses. Let δ be a small positive number. Suppose that with probability $1 - \delta$ a given agent chooses a best reply to the frequency distribution of the other side's actions in a random sample drawn from $h(t)$, but with probability δ she chooses a strategy at random. The probabilities of these events are independent for the two agents. We then obtain a Markov process P^δ on the finite state space H consisting of all sequences from X of length at most m . The process is ergodic because there is a positive probability of moving from any state to any other in m periods or less. It can be shown that, for all sufficiently small δ , the players play a coordination equilibrium with near certainty over the long run. More precisely, given the process P^δ let π_j^δ be the long-run probability that the j th coordination equilibrium (x_j, x_j) is played in any given period t as $t \rightarrow \infty$. This probability exists because the process is aperiodic and ergodic. It can be shown that, given any $\epsilon > 0$, $\sum_{j=1,n} \pi_j^\delta \geq 1 - \epsilon$ for all sufficiently small δ (Young, 1993). In other words, the probability is at least $1 - \epsilon$ that over the long run the players coordinate at any given time. Indeed, it can be shown that in the absence of ties (i.e., in a generic coordination game) the players coordinate almost all of the time on *exactly one* of the coordination equilibria when the noise δ is small.²

An analogous result holds for weakly acyclic games. Let G be weakly acyclic and let N be the set of all strict Nash equilibria in pure strategies for G . It follows from weak acyclicity that N is nonempty. Let $\pi^\delta(\mathbf{x})$ be the long-run probability that the strategy profile \mathbf{x} is played at any given time t as $t \rightarrow \infty$. Then, for every $\epsilon > 0$, $\sum_{\mathbf{x} \in N} \pi^\delta(\mathbf{x}) \geq 1 - \epsilon$ for all sufficiently small δ . (This follows from the proof of Theorem 4 in Young, 1993). In other words, over the long run a strict Nash equilibrium will be played with high probability when the noise δ is sufficiently small.

In sum, variations of fictitious play that incorporate random perturbations and finite memory have better convergence properties than fictitious play itself for a fairly large class of games that includes coordination games and generic identical interest games.

²See also Kandori, Mailath and Rob (1993) for similar results.

References

Blume, Larry E. (1995): "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, forthcoming.

Brown, G. W. (1951): "Iterative Solution of Games by Fictitious Play," In *Activity Analysis of Production and Allocation*, ed. by Tjalling C. Koopmans. New York: Wiley.

Crawford, Vincent P. (1991): "An Evolutionary Interpretation of Van Huyck, Battalio, and Beil's Experimental Results on Coordination," *Games and Economic Behavior*, 3, 25-59.

Foster, Dean, and H. P. Young (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology*, 38, 219-232.

Fudenberg, Drew, and David Kreps (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320-367.

Kandori, Michihiro, George Mailath, and Rafael Rob (1993): "Learning, Mutation, and Long-Run Equilibria in Games," *Econometrica*, 61, 29-56.

Kaniovski, Yuri, and H. Peyton Young (1994): "Learning to Play Equilibria in Games with Stochastic Perturbations," International Institute for Applied Systems Analysis Working Paper 94-31, Laxenburg, Austria.

Krishna, Vijay (1991): "Learning in Games with Strategic Complementarities," Mimeo.

Milgrom, Paul, and John Roberts (1991): "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior*, 3, 82-100.

Miyasawa, K. (1961): "On the Convergence of the Learning Process in a 2 x 2 Non-Zero-Sum Two-Person Game," Economic Research Program, Research Memorandum No. 33, Princeton University.

Monderer, Dov, and Aner Sela (1993): "Fictitious Play and No-Cycling Conditions," Mimeo.

Monderer, Dov, and Lloyd S. Shapley, (1993a): "Fictitious Play Property for Games with Identical Interests," Mimeo.

Monderer, Dov, and Lloyd S. Shapley (1993b): "Potential Games," Mimeo.

Robinson, Julia (1951): "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54, 296-301.

Shapley, Lloyd S. (1964): "Some Topics in Two-Person Games. " In *Advances in Game Theory*, ed. by M. Dresher, L.S. Shapley, and A.W. Tucker, 1-28, Princeton: Princeton University Press.

Young, H. P., 1993: "The Evolution of Conventions," *Econometrica*, 61, 57-84.