Learning from Personal Experience:
One Rational Guy and the Justification of Myopia

Glenn Ellison*

Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The paper examines a large population analog of fictitious play in which players learn from personal experience, focusing on what happens when a single rational player is added to the population. Because the learning process naturally generates contagion dynamics, the rational player at times has an incentive to act nonmyopically. In 2×2 games the dynamics are asymmetric and favor risk dominant equilibria. A variety of other examples are presented. *Journal of Economic Literature* Classification Number: C7. © 1997 Academic Press

1. INTRODUCTION

One of the most fundamental questions in game theory is the justification of Nash equilibrium or other solution concepts. Recently, there has been a surge of interest in boundedly rational learning processes by which players might come to play an equilibrium, and in particular in models related to the classic model of fictitious play.[1] The standard fictitious play model is quite stylized; in particular, it involves two players who act myopically while facing each other repeatedly. If the model is interpreted literally, it is difficult both to justify why players should do this and to argue that we as economists should devote so much attention to a model which could at most tell us that when two players play each other

[1]This includes Robinson (1951), Miyasawa (1961), Shapley (1964), and the more recent work of Fudenberg and Kreps (1991, 1993), Jordan (1993), Krishna (1991), Krishna and Sjostrom (1995a), Milgrom and Roberts (1991), and Monderer and Shapley (1996), among others. See Fudenberg and Levine (1996) and Krishna and Sjostrom (1995b) for surveys of the literature.

repeatedly they eventually come to play an equilibrium of the static game. The fact that the model has received such great attention is probably a result of people thinking of it instead as a loose representation of the process by which each individual in a society comes to learn its conventions.[2] With such an interpretation myopia may make more sense and we might draw much more interesting conclusions, e.g., that two players who are meeting for the first time might be expected to play an equilibrium if they have grown up in the same society and learned the same conventions. This paper attempts to develop such an interpretation more fully by examining an explicit formal model of a large population of players learning from the personal experience they obtain in of a sequence of encounters with other members of the population.[3]

While the paper is motivated generally by a desire to explore what kinds of things happen in a large population–personal experience model, the discussion focuses on two more concrete questions: when is it reasonable in such a model to assume that players act myopically, and is the outcome of the learning process robust to a slight relaxation of the myopia assumption in the direction of more rationality. The specification of the model follows closely along the lines suggested by Fudenberg and Kreps (1993). A population of N players is repeatedly randomly matched to play the game G, and each player observes only the outcome of matches in which he is personally involved. The one unusual element of the model is that the population consists of $N - 1$ boundedly rational agents who follow specified myopic learning rules along with one patient rational player who knows that the others are boundedly rational and may attempt to manipulate their play.

The primary motivation for this "one rational guy" model is that the question of when a large population–personal experience story justifies the assumption that players are myopic can be addressed by asking when the rational player will play myopically. The first main point which the paper makes is that large populations do not provide as clear cut a justification of myopia as one might have thought. In a 2×2 game with two strict equilibria myopia is justified if one thinks of the population size going to infinity with the discount factor of the rational player held fixed, but for any fixed population size it is also true that for discount factors close enough to one the rational player can attempt to manipulate the play of

---

[2] See Fudenberg and Kreps (1991, 1993) for discussions of the possibility of large population justifications.

[3] See Canning (1990), Fudenberg and Levine (1993), and Young (1993) for discussions of other aspects of large population learning.

the population even after all other players have coordinated on one of the equilibria. The reason why manipulation may be profitable is that contagion dynamics naturally arise in the process of learning from personal experience.[4] While the rational player is a small part of the population, what he does in period t is a nontrivial part of the experience of his period t opponent and thus may affect how that player plays from period $t + 1$ on, how he and others play from period $t + 2$ on, etc. Simulations are used to help assess the practical relevance of the results, with the primary conclusions being that nonmyopic play may be practically important when very strong risk-dominance relationships exist, when players put a large weight on recent experience, and when a population has a local interaction structure.

The one rational guy approach also provides a minimal and fairly natural robustness check in that we can ask whether the predictions of a model change if we merely allow for the possibility that one player might be rational.[5] While the result mentioned above makes it obvious that the stability of equilibria under the learning process will not always be robust, the further observation which may be of additional interest is that the nonrobustness is highly asymmetric—the contagion dynamics allow the rational player to shift play from an equilibrium which is risk dominated to one which is risk dominant, but not vice versa.

The remainder of the paper presents several examples of things that can happen outside the class of symmetric 2×2 coordination games. One of these is an admittedly very special example in which a finite-memory version of the personal experience fictitious play model always converges to a state in which the players coordinate on the unique equilibrium of the game. However, the dynamic process also has a "nearly stable" cycle which is a very strong attractor, and in the intermediate run (meaning in the first few thousand or million periods) cycling will be much more common than equilibrium play. In this environment, a single rational player would be able to keep the state of play permanently away from the equilibrium (or any other stable limit set) and achieve supraequilibrium payoffs. Other examples are used to discuss the scope of nonmyopic play in asymmetric games and how things might be affected by dominated strategies and the presence of other equilibria.

---

[4]See Kandori (1992) and Ellison (1994) for other examples of contagion dynamics. The potentially interesting observation here is that such dynamics are not found only in fanciful bootstrapped equilibria.

[5]See Banerjee and Weibull (1993), Dekel and Scotchmer (1992), and Matsui and Rob (1991) for other discussions of rationality in learning and evolutionary models.

2. MODEL

The basic model of this paper involves a single population of players named $1, 2, \ldots, N$, with N an even number. In periods $t = 1, 2, 3, \ldots$ the players are randomly matched to play the two player game G. Assume that the matching is such that there is a uniform lower bound $p > 0$ on the probability of any two players meeting at any history. The matching is assumed to be anonymous, i.e., players are assumed to be unable to identify or distinguish their previous opponents. I write $s_{i,t}$ for the action chosen by player i in period t, and $s_{-i,t}$ for the action chosen by his opponent. The personal experience assumption is that player i observes only $s_{-i,t}$ and not the actions taken by any of the players with whom he was not matched. In the next three sections the game G will be the 2×2 coordination game shown below whose payoff function is denoted by g. Assume that the payoffs satisfy $a > d$, $b > c$, and $a > b$, so that (A, A) and (B, B) are both equilibria with the former pareto optimal. Let μ be defined by $\mu a + (1 - \mu)c = \mu d + (1 - \mu)b$. If $\mu < 1/2$ then A is said to be risk dominant.

	A	B
A	a, a	c, d
B	d, c	b, b

Player i's information at time t is described by a vector $h_i^t \in H_i^t$ giving the play of player i and his opponent in each previous period. I write (h_i^t, s_i, s_{-i}) for the period $t + 1$ information set in which h_i^t is followed by player i playing s_i and his period t opponent playing s_{-i}. In the models that follow players 1 through $N - 1$ (and sometimes player N) will be assumed to follow simple behavioral rules similar to those discussed in Fudenberg and Kreps (1993) and Milgrom and Roberts (1991). Specifically, each player is assumed to have a belief function $\eta_i: H_i^t \to \Sigma$ giving the player's expectation of the play of his opponent. In period t, each of these players plays a myopic best response to his beliefs, choosing

$$s_{i,t} \in \text{Argmax}_s \, g\big(s, \eta_i(h_i^t)\big).$$

One model of learning which fits into this class is the analog of fictitious play in which each player in each period plays a best response to the empirical frequency distribution of actions he has observed. Another model which reflects the notion that players may give more weight to more recent observations and which is particularly tractable is a finite memory version of fictitious play where players use only observations from the

previous k periods, i.e., the belief functions η_i^k are given by

$$\eta_i^k(h_i^t)(s) = \frac{1}{k} \sum_{r=1}^{k} I(s_{-i,t-r} = s),$$

where $s_{-i,t-r}$ for $r \geq t$ represents an initial fictitious history, and $I(x)$ is the indicator function for condition x being true or false. When discussing this rule, I will assume also that all players use a fixed mixed strategy not depending on the history whenever they are indifferent.

3. PERSONAL EXPERIENCE AND CONTAGION DYNAMICS: A SIMPLE EXAMPLE

The example of this section is meant to illustrate how contagion dynamics may arise when players learn from personal experience and how this may provide an opportunity for a rational player to manipulate the course of play. Consider a population of myopic learners with fictitious play beliefs. In a 2×2 game where each player's pure strategies are A and B this means that each player i begins with weights $(w_{i,0}^A, w_{i,0}^B)$ reflecting the number of times he has "seen" each strategy in some "fictitious history" before play begins. The weights are updated to $(w_{i,t}^A, w_{i,t}^B)$ over time by adding one in each period to the count for whatever strategy player i's opponent used in that period. At time t player i assumes that his opponent will play A with probability $w_{i,t-1}^A/(w_{i,t-1}^A + w_{i,t-1}^B)$. Suppose that the game G is that shown below, and that each of the players initial weights is $(0, 1)$. The players will then all play B in the first period, their beliefs will be reinforced, and they will continue to play B in all future periods.

	A	B
A	10, 10	0, 0
B	0, 0	1, 1

To see how contagion dynamics can arise, suppose that player N were to play A in the first period of the game (and to then revert to following fictitious play). Call player N's first period opponent player 1. In period 2, player 1 will have weights $(1, 1)$. Hence, believing each strategy to be equally likely he has A as a best response. He will continue to play A in periods 3 through 10 (assuming he plays A when indifferent). In period 2 player 1 will likely meet a new opponent, say player 2. By the same

reasoning as above, player 2 will think that A is somewhat likely and will also play A in periods 3 through 10. In the likely event that players 1 and 2 do not meet in period 3, they will "infect" two more players who will play A from period 4 through 10. As the process continues, the number of players playing A may grow exponentially, with up to four players playing A in period 4, eight in period 5, sixteen in period 6, and so on.

The growth of this contagion is sufficiently rapid that in a population of 100 players it may be the case that every player will have switched to playing A by the 10th period, in which case A will be played in all future periods. When this is not the case, each of the players who has seen A only once in the first 10 periods will return to playing B in period 11, and the set of players playing A will start to grow again (provided there is some player who has seen A twice). It is then again possible that everyone will switch to playing A before the 20th period; otherwise another contraction will follow the 20th period.

Imagine now that in the model above player N alone is rational and that he is aware of the fact that the remaining players are fictitious play learners with initial beliefs concentrated on B. Would player N play myopically in such a situation, or is he better off trying to upset the initial equilibrium? It should not be surprising that the answer is that he should try to upset the (B, B) equilibrium if he is somewhat patient. If player N plays myopically, he receives a payoff of 1 in every period. If instead he plays A in the first period and then follows fictitious play, he suffers a short term loss in receiving a payoff of 0 when not coordinating in the first period, and may suffer additional losses from failures to coordinate in some future periods. The potential gain from this strategy, however, is long lived—with some probability the population shifts to the (A, A) equilibrium and player N gets a payoff of 10 in all periods after this occurs. When player N is patient, the possibility of such a long term gain can easily outweigh the short term loss.

Table I contains numerical estimates of the value of the discount factor necessary for the strategy described above to dominate the strategy of myopic play for a range of population sizes. If the model above is meant to capture the way in which people may learn to play games from the experience they obtain in day-to-day interactions, we might imagine that players are involved in tens or hundreds of matches each year. Even

TABLE I
Patience Necessary for Simple Nonmyopic Strategy

Population Size:	4	10	50	100	500	1000	5000	10000
Discount Factor:	0.44	0.55	0.64	0.67	0.74	0.78	0.94	1.00

incentives which exist only for discount factors like 0.999 might then be practically relevant. Given the speed with which play can shift from one equilibrium to the other in this example, far less patience than this is necessary for a wide range of population sizes.

4. A MANIPULATION THEOREM AND
SOME CONVERSES

This section discusses more generally contagion dynamics and incentives to manipulate play in personal experience learning models. The principal insight obtained is that the contagion dynamics are highly asymmetric and favor movement toward risk dominant equilibria. The practical relevance of the results in connection with the justification of myopia is also discussed.

4.1. *Properties of Learning Rules*

The manipulation theorem of this section will apply to all learning rules having three properties outlined here. First, I will say that beliefs are *monotonic* if for all i, t, s_i, and s_{-i}, $\eta_i(h_i^t, s_i, s_{-i})(s_{-i}) \geq \eta_i(h_i^t)(s_{-i})$. This assumption says simply that the strategy which was observed last period is thought to be no less likely than it was previously. The assumption seems innocuous for the problem at hand and is satisfied by each of the examples mentioned above.[6]

Second, I will say that beliefs are *asymptotically majoritarian* if for all i, t, h_i^t, and for all $\epsilon > 0$, there exists T such that if $\tau \geq T$ and $h_i^\tau = (h_i^t, s_i^t, s_{-i}^t, \ldots, s_i^{\tau-1}, s_{-i}^{\tau-1})$ is such that $\eta_i^k(h_i^\tau)(s) \geq \frac{1}{2}$ for all $k \in \{2, 4, \ldots, 2\,\mathrm{Int}((\tau - t)/2)\}$, then $\eta_i^\tau(h_i^\tau)(s) \geq 1/2 - \epsilon$.[7] Informally, the requirement is that when s has been played at least half the time in a large enough number of the most recent periods it is assumed to have a probability which is not too much less than one-half. The requirement is similar in spirit to Fudenberg and Kreps' asymptotic empiricism, although less demanding in that it only applies to strategies played at least half of the time, and only sets a lower bound on beliefs in this case. Note that the standard fictitious play model always has this property, as do finite memory fictitious play beliefs with an even memory length.[8]

---

[6]Aoyagi (1996) argues that monotonicity is inappropriate for players who might recognize cycles or patterns in play. Such cycles, however, will not occur in most of the large population models considered here.

[7]Recall that η_i^k denotes the k-period finite memory fictitious play beliefs.

[8]If the memory length is large and odd the property nearly holds and the results presented below will all go through, provided the game involved has a sufficient degree of risk dominance.

Finally, I will say that a system of beliefs *induces pure strategy convergence* if for any initial history there is probability one that the population will eventually converge to a state in which all players play A or all players play B. The assumption rules out the possibility of play converging to the mixed equilibrium or perpetually cycling in a nonconvergent fashion. It is not so much a property of the beliefs as a joint property of the myopic learning rules which the players are following, the game being played, and the structure of the population interactions.

It is well known that in a standard two player model fictitious play beliefs will not necessarily induce pure strategy convergence in games which have both pure and mixed strategy equilibria. For example, Fudenberg and Kreps (1991) show that in 2×2 coordination games we may get cycles involving persistent miscoordination rather than convergence to either pure equilibrium. So that the reader will not worry that for this reason the set of learning rules satisfying all three of these properties might be empty, I include here a proposition which establishes that finite memory fictitious play learning rules (with an even memory length) induce pure strategy convergence in 2×2 coordination games. Intuitively, the reason why this is true is that because of the randomness in random matching and the contagion dynamics it is always possible for play to take off on a path toward one of the pure strategy equilibria.

PROPOSITION 1. *For k even, k-period finite memory fictitious play belief rules induce pure strategy convergence in 2×2 coordination games.*

The proof of this proposition will be easier to read after seeing Proposition 2 and hence has been relegated to the Appendix.[9]

4.2. *A Manipulation Theorem*

To examine the justification of the assumption of myopic play, consider now a model in which players $1, 2, \ldots, N - 1$ react myopically to beliefs η_i, while player N is rational with discount factor δ and knows the belief functions η_i of the other players. Note that this specification makes the Nth player more informed about the type of irrationality present than is reasonable, but still limits him greatly by not allowing him to recognize

[9]As an aside, it is interesting to note that if players' choices when indifferent are allowed to depend on histories, the proposition is false. Whenever μk is an integer we may make play converge (in time averages) to the mixed strategy equilibrium with an appropriate starting point by assuming particular responses when the players are indifferent. Simply consider a population in which indifferent players play whatever strategy they used k periods ago and an initial history in which all players played A in the first μk periods and B in the next $k - \mu k$ periods. Then, regardless of the realization of the matching, all players remain indifferent and play cycles through periods in which everyone plays A and periods in which everyone plays B.

opponents or observe any aggregate statistics.[10] For myopia to be justified by this large population model, we would want to say that the rational player plays as if he were myopic.

The following proposition identifies a fairly general set of circumstances in which a single rational player will try to manipulate the play of an entire population of opponents. Specifically, such behavior is shown to be optimal whenever the $N - 1$ irrational players all have beliefs initially concentrated on the equilibrium (B, B) which is both pareto inferior and risk dominated provided that the rational player is sufficiently patient. The pareto inferiority of (B, B) makes the rational player want to shift play to the (A, A) equilibrium. The risk dominance of (A, A) ensures that he is capable of doing so (with positive probability).

PROPOSITION 2. *Suppose players* $1, 2, \ldots, N - 1$ *play myopic best responses to belief functions* η_i *which are monotonic, asymptotically majoritarian, and which* (*together with some* η_N *having these properties*) *induce pure strategy convergence. Suppose further that* $\eta_i(h_i^1)(A) < \mu$ *for all i so that all myopic players will play B in period* 1. *If* (A, A) *is risk dominant,* (*i.e., if* $\mu < 1/2$), *then there exists* $\underline{\delta} < 1$ *such that for* $\delta > \underline{\delta}$ *the rational Nth player will not play B in every period.*

Proof. If player N plays myopically, he plays B in every period. By the monotonicity assumption, all other players continue to play B as well, and player N's expected per period payoff is b. To show that the myopic strategy is not optimal, I show that there exists a strategy σ^T for player N which yields a higher expected payoff provided he is sufficiently patient because it causes the population to switch to the equilibrium A with positive probability. In particular, for a value of T to be determined below, the strategy σ^T will consist simply of playing A for the first T periods and then reverting to myopic play given the beliefs $\eta_N(h_N^t)$.

Write s_t for the vector of actions chosen at time t, and **A** and **B** for the vectors in which all players play A and B respectively. With the strategy σ^T, the assumption of induced pure strategy convergence implies that in the limit as $\delta \to 1$, player N's per period payoff approaches

$$a \operatorname{Prob}\{s_t \to \mathbf{A}\} + b(1 - \operatorname{Prob}\{s_t \to \mathbf{A}\}).$$

Hence, it suffices to show that $\exists T$ such that $\operatorname{Prob}\{s_t \to \mathbf{A}\} > 0$ under strategy σ^T.

---

[10] In fact, it is not necessary for the rational player to know the belief functions. It would suffice for Proposition 2 for him to know that the rules satisfy the three properties (with a known T as a function of ϵ).

Using asymptotic majoritarianism and the risk dominance of A, we may choose T_1 such that $\eta_1(h_1^t)(A) > \mu$ whenever $t \geq T_1$ and $s_{-1, t'} = A$ for all $t' < t$ odd. Define T_2 analogously for player 2 when he sees A in even periods and set $T = \text{Max}\{T_1, T_2\}$. Now, consider all realizations of the matching in which player N is alternately matched with players 1 and 2 in the first T periods (an event of nonzero probability). In period $T + 1$ players 1 and 2 will play A, and all players will henceforth follow their myopic rules. It now suffices to establish the following

Claim. In a population of N myopic learners with beliefs satisfying monotonicity and asymptotic majoritarianism, $s_t \to \mathbf{A}$ with positive probability if there exists a period T for which two players have beliefs with $\eta_i(h_i^T) > \mu$.

The claim may be proved by a simple inductive argument starting from the initial fact that when N players simultaneously have such beliefs monotonicity guarantees convergence to $\mathbf{A}$ with probability one. If the result holds when $r + 2$ players have such beliefs for $r \geq 2$ even, suppose players $1, 2, \ldots, r$ have such beliefs and consider realizations of the matching in which those players are matched among themselves in periods $T + 1, \ldots, T + \tau_1$. By monotonicity, each of them plays A in those periods. Asymptotic majoritarianism then guarantees the existence of a τ_1 such that players 1 and 2 play A in all future periods if they see A in all subsequent even periods. Hence, consider matchings in which players 1 and 2 are matched alternately with themselves and with players $r + 1$ and $r + 2$ for the next τ_2 periods while players $3, \ldots, r$ are matched among themselves. (Again this event has positive albeit small probability.) For τ_2 sufficiently large, the first $r + 2$ players will all assign probability greater than μ to A in period $T + \tau_1 + \tau_2$. The claim and hence the proposition follow by induction.

While the inductive argument may not make it clear, the process described above may grow exponentially. If, for example, players 3 through r had been matched alternately with players not playing A, the contagion would have doubled with each interval of time periods rather than growing in size by two players.   QED

Remarks. (1) The assumption that learning is from personal experience alone is critical. If players observed the outcomes of all matches, one player's actions could not significantly affect anyone's beliefs. The effect of learning from personal experience can be thought of as similar to the effect of local interaction discussed in Ellison (1993) in that each player reacts to the actions of a small number of past opponents. One or a few actions can thus have an impact on the process, and "evolution" from the neighborhood of one equilibrium to the other can be both more rapid and possible given a smaller push away from the initial state.

(2) A direct corollary of this proposition is that in a model with r rational players there will be no Nash equilibrium in which all r players play myopically (at all histories). If they did, each of them would find himself in the situation of the one rational player above, and hence have an incentive to manipulate play.

(3) While the statement of the proposition only notes that player N deviates from myopic play at least once, one can deduce from the fact that manipulation is profitable that player N at least occasionally succeeds in dramatically altering the course of play. When players use finite memory learning rules an even stronger conclusion follows. Because there is no buildup of experience to deter him, a rational player in such a model will try to manipulate play again if he ever thinks everyone has returned to playing B, and hence with probability one he will eventually succeed in moving the population to the (A, A) equilibrium. See part 1 of Proposition 5.

(4) The proposition does not establish that a sufficiently patient rational player will behave in a way which is incompatible with Fudenberg and Kreps's (1993) asymptotic myopia. Asymptotic myopia would, however, be violated if for arbitrarily large t there remains a positive probability of reaching a state in which player N believes that the population is sufficiently delicately balanced between tipping toward A and B so that he should again try to push things toward A. I would conjecture that for this reason behavior is not asymptotically myopic, although such violations may not be of much interest because they occur with vanishing probability.

4.3. *Myopia and Patience*

If one wanted to use the large population–personal experience story as a justification for the assumption of myopic play in a standard two player model, one could have written down a model identical to that used here and featured instead a proposition like that given below.

PROPOSITION 3. *Let G be an arbitrary game with (s, s) a strict Nash equilibrium of G. Suppose players $1, 2, \ldots, N-1$ play myopic best responses to monotonic beliefs, and that their initial beliefs are such that they all strictly prefer to play s in the first period. Then for any $\delta < 1$ there exists a population size $\underline{N}$ such that whenever $N > \underline{N}$ a rational player N with discount factor δ will play myopically in every period.*

Proof. Because playing anything other than s incurs a loss in the first period and all payoffs are bounded, it is possible to choose T and ϵ such that a rational player will play s in the first period if with a probability of at least $1 - \epsilon$ his opponents will play s in all of the first T periods

TABLE II
Discount Factors Sufficient for Nonmyopic Play

Payoff, a	Population Size, N							
	4	10	50	100	500	1000	5000	10000
15	0.16	0.35	0.60	0.66	0.73	0.75	0.78	0.79
10	0.23	0.44	0.66	0.70	0.76	0.79	0.87	1.00
8	0.27	0.49	0.69	0.73	0.82	0.86	1.00	1.00
6	0.33	0.55	0.75	0.80	1.00	1.00	1.00	1.00
4	0.44	0.67	0.91	1.00	1.00	1.00	1.00	1.00
2	0.77	0.98	1.00	1.00	1.00	1.00	1.00	1.00

regardless of how he plays in the first period. If one chooses $\underline{N}$ sufficiently large so that there is less than an ϵ chance that a player will meet anyone he has played or anyone who has met anyone he has played, etc., in the first T periods, this is true and thus playing s is optimal. QED

The conflicting conclusions of Propositions 2 and 3 arise from changing the order of the $N \to \infty$ and $\delta \to 1$ limits. Given this conflict, perhaps the best way to try to understand whether or when nonmyopic play may be practically important is simply to check numerically whether one rational player would act myopically for various parameter values.

Table II presents Monte Carlo estimates of how patient players must be for a simple strategy of playing A in the first four periods and then reverting to fictitious play to dominate fictitious play when initial beliefs are very weak.[11] A uniform matching model with $N - 1$ players following fictitious play learning rules is assumed, with the players' initial fictitious histories assumed to consist of having seen B in a single previous period. Payoffs are as described in Section 2, with b fixed at 1, c and d fixed at zero, and a varied across the various experiments.

The most striking regularity in the table is that the degree of risk dominance reflected in the payoff a greatly affects the ease with which play can be manipulated. When a is small and the population size is at all large we are unable to distinguish the required discount factors from one. When the degree of risk dominance is more extreme, manipulation is fairly easy. For a wide range of population sizes the contagion simply grows exponentially until it takes over, and hence the requisite degree of patience increases only very slowly. If one takes the numbers in the table

---

[11] The discount factor is listed as 1.00 if this simple strategy dominates fictitious play only for discount factors which are so close to one as to be indistinguishable from one in the simulations which were carried out. The optimal manipulation strategy would be profitable for players who are somewhat less patient than is indicated in the table.

seriously, one would conclude that nonmyopic play may be relevant when very strong risk-dominance relationships are present or in not so large populations.

On the taking seriously question, it should be noted that the calculations above are based on a model which is in one regard extreme in facilitating manipulation. Because players are assumed to have very little prior experience, seeing A once may lead them to change their play. There are some economic situations where such a sensitivity of behavior to a single observation seems reasonable. For example, we might expect an individual who is mugged while walking down a particular dark alley to change his behavior for a substantial period of time thereafter even if he has walked down that alley with no ill effects a hundred times before. In a business context, it is common also to talk of managers as being careful to avoid making the same mistake twice.

The one experimental paper I am aware of which examines how much weight players put on their most recent observations relative to accumulated experience is that of Cheung and Friedman (1995), which finds players to react even more strongly to new observations than is assumed above. In their random matching/personal experience treatments they find that there is significant heterogeneity but that the typical player dramatically discounts the distant (and even relatively recent) past. In estimating models in which players react to an exponentially weighted average of their previous observations they typically find that the median player discounts t-period ago observations by a factor of about 0.5^{t}.[12] While I have to admit that such discounting seems extreme, if each member of a population were to use such a rule manipulation would be very easy. The number of players playing A could never decrease and if a is at least two the rational player needs only to play A once to ensure that everyone ends up playing A in the long run. Table III reports the discount factors necessary for playing A a few times and then reverting to 0.5-exponentially weighted fictitious play to dominate myopic play when all other players use 0.5-exponentially weighted fictitious play and have initial beliefs which reflect seeing an infinite sequence of B's.[13] The results indicate that when players place such a large weight on recent experience, manipulation no longer requires strong risk dominance relationships and increases in the population size have a very small effect on the requisite degree of patience.[14]

---

[12] They do not always report estimates based on this treatment alone, but find such results both when pooling data from this treatment with data from their other no-history treatment and when pooling with their other random matching treatment.

[13] A few is taken to be 10 for $a = 2$ and 2 for $a \geq 4$.

[14] While computational concerns have led me to stop the table at a population size of 5000, a rough calculation suggests that for $a \geq 4$ manipulation would still be profitable with reasonable discount factors at population sizes in the tens of millions.

TABLE III
Sufficient Discount Factors with 0.5-Exponential Fictitious Play

Payoff, a	Population Size, N						
	4	10	50	100	500	1000	5000
10	0.29	0.51	0.67	0.71	0.76	0.77	0.80
8	0.35	0.55	0.69	0.73	0.77	0.79	0.82
6	0.43	0.62	0.75	0.78	0.82	0.83	0.85
4	0.55	0.70	0.80	0.82	0.85	0.86	0.88
2	0.73	0.86	0.95	0.96	0.98	0.99	0.995

Is manipulation a relevant concern other than in situations where players may react strongly to a single observation? The answer obviously depends on how unlikely a chain of events is necessary to make the contagion take off. To give us some feel for this, Table IV repeats the experiment of Table II but with initial histories consisting of five observations of B. Here, manipulation appears profitable in games without an extreme degree of risk dominance only if the population is very small.

While the table above is more negative on relevance, one can also make things look a little better by noting that while a population of 10 or 50 players is not what models of learning in a social setting are intended to capture, large population–local interaction models where each player mostly interacts with 10 or 50 neighbors may behave similarly. To illustrate this point Table V reports discount factors sufficient to make nonmyopic play profitable despite 5 period initial histories in a local matching model where players interact with their 20 closest neighbors on a circle. In these simulations it appears that manipulation will be possible in larger populations if it is possible in a population of 50 players.[15]

4.4. *Some Nonmanipulation Theorems*

In Proposition 2, it was shown that a single rational player could shift play from an equilibrium which is risk dominated to an equilibrium which is risk dominant. To help clarify how the model's dynamics are dependent

---

[15]To facilitate simulation of the local matching, for this model alone it has been assumed that random matchings occur asynchronously, with each player being matched on average once per period. The discount factors are those which are estimated to make a strategy of deviating from fictitious play to play A in the first 10 periods profitable. Note that this change tends to make manipulation more profitable for players with very small discount factors, as the specification of the random matching allows a player to be matched several times in the first period before any discounting is applied. The standard errors of the large population estimates are approximately 0.01.

TABLE IV
Sufficient Discount Factors with Stronger Initial Beliefs

Payoff, a	Population Size, N						
	4	10	50	100	500	1000	5000
15	0.16	0.32	0.54	0.61	0.71	0.75	0.84
10	0.23	0.40	0.62	0.70	0.90	1.00	1.00
8	0.27	0.45	0.76	0.89	1.00	1.00	1.00
6	0.34	0.65	0.99	1.00	1.00	1.00	1.00
4	0.72	0.96	1.00	1.00	1.00	1.00	1.00
2	0.99	1.00	1.00	1.00	1.00	1.00	1.00

upon risk dominance relationships, Proposition 4 presents a partial converse to this result, showing that manipulation is unprofitable when the initial equilibrium is risk dominant.

PROPOSITION 4.   *For the standard fictitious play beliefs, if* (B, B) *is risk dominant and the initial beliefs are such that all myopic players strictly prefer* B, *then*

1. *For* $N \geq 4$ *player* N'*s optimal strategy is to play myopically.*

2. *For no strategy of player* N *does more than one other player ever play* A *in a given period.*

The proof of Proposition 4 is based on the observation that when (B, B) is risk dominant players will not switch to A unless they have seen A in more than half of the periods to date. Clearly, there will never be two or more players each of whom has played player N more than half the time, and hence player N's manipulations cannot directly convince two or more players to play A. It then remains only to show that he also cannot

TABLE V
Sufficient Discount Factors with Stronger Initial Beliefs
and 20 Neighbor Local Matching

Payoff, a	Population Size, N				
	4	10	50	100	500
15	0.00	0.08	0.27	0.26	0.25
10	0.03	0.21	0.39	0.40	0.39
8	0.04	0.29	0.52	0.52	0.52
6	0.21	0.53	0.85	0.85	0.85
4	0.70	0.93	1.00	1.00	1.00
2	0.97	1.00	1.00	1.00	1.00

somehow indirectly convince two or more players by first convincing one player to play A, etc. Essentially such possibilities do not exist because the timing of these additional manipulations makes them of no help. A formal demonstration may be found in the Appendix.

The second part of the proposition extends the first so that we know not only that manipulation is not profitable, but that it is also impossible. We can thus conclude that the stability of a risk dominant equilibrium under fictitious play is robust not only to the introduction of one rational guy, but also to the introduction of one player of any type.

Note also that the proposition has been stated for fictitious play rather than for general learning rules. The reason for this is that the conditions under which manipulation is not profitable vary from model to model and will not always coincide exactly with risk dominance. In the finite memory fictitious play model, for example, manipulation is somewhat easier and thus turns out to be profitable not only when the initial equilibrium is risk dominated ($\mu < \frac{1}{2}$), but also when the initial equilibrium is only "a little" risk dominant (roughly when $\mu \in (\frac{1}{2}, \frac{2}{3})$).

PROPOSITION 5. *In the model with a single rational player and $N - 1$ players following k-period memory fictitious play rules for k even, $N \geq 4$, and initial beliefs assigning probability one to B*:

1. *For $\lfloor \mu k \rfloor \leq \frac{2}{3}k - \frac{1}{3}$, there exists a $\underline{\delta} < 1$ such that a rational player with a discount factor of at least $\underline{\delta}$ will play nonmyopically and with probability one he will eventually succeed in shifting play to a state where everyone plays A.*

2. *For $\lceil \mu k \rceil > \frac{2}{3}k + \frac{1}{3}$, the rational player will play B in every period and play cannot converge to everyone playing A regardless of the strategy which the Nth player uses.*

Proof. The proof of the first part of the proposition is similar to the proof of Proposition 2 and is left for the Appendix. For the second part it suffices to show that regardless of player N's strategy at most one other player will play A in any given period. To establish this result, suppose not and let t be the first period in which at least two myopic players play A. Call the players 1 and 2. Because each has seen A in at least $\lceil \mu k \rceil$ of the previous k periods, there must be $2\lceil \mu k \rceil - k$ periods in which both saw A simultaneously. In these periods, at most two players play A, so each of players 3 through $N - 1$ face opponents playing B. In the last of these periods, players 3 through $N - 1$ have seen A in at most $k - (2\lceil \mu k \rceil - k - 1)$ of the previous k periods. Under the hypothesis of the proposition, this number is less than $\lceil \mu k \rceil$ so that each of players 3 through $N - 1$ play B in this period. This yields a contradiction, because players 1 and 2 could not have both seen A if they played each other, nor if one of them played a member of the set $\{3, 4, \ldots, N - 1\}$. QED

Intuitively, the extra ability of player N to affect play in this model comes from the possibility that manipulations may be bunched together and hence appear more common in a particular k period history than they are in the entire history.

5. NONEQUILIBRIUM PAYOFFS AND INTERMEDIATE RUN VS LONG RUN BEHAVIOR

In the 2×2 coordination games discussed so far a single rational player may cause a shift in play from one equilibrium to another. This section presents an example involving a 3×3 game in which more complex cyclical contagion dynamics allow a rational player to obtain a payoff which is higher than is possible in any equilibrium. While the example itself is very special, it may be of broader interest as an illustration of the fact that large population learning models may have "nearly stable" behaviors which are more attractive in the intermediate run than are any of the steady states or cycles which a long run analysis would identify.

Consider the 3×3 game shown below, where ϵ is a small positive number (less than 0.01). Note that (B, B) is the unique pure strategy equilibrium of the game and that all mixed strategy equilibria also have a payoff of one. Suppose that N players play the game in a random matching setting and follow finite memory fictitious play rules with a memory length of 10 periods. As was the case in 2×2 coordination games, random matching is sufficient to break up cycles, so while the two player model has a number of stable limit cycles, a four or more player model converges to the unique pure strategy equilibrium from any initial condition.[16] A proof of the proposition is given in the Appendix.

	A	B	C
A	$10 + \epsilon, 10 + \epsilon$	$0, 1$	$0, 11$
B	$1, 0$	$1, 1$	$1, -100$
C	$11, 0$	$-100, 1$	$0, 0$

PROPOSITION 6.   *In the* 10-*period memory fictitious play learning model with $N \geq 4$ players, play converges almost surely to all players playing B given any initial beliefs.*

[16] Seventeen limit cycles are possible in the two player model. The first type of cycle has period 22 and consists of (A, C), $k \times (C, C)$, (B, C), (B, A), $8 - k \times (C, A)$, (C, B), $k \times (A, A)$, $9 - k \times (A, C)$ for some $k \in \{1, 2, \dots, 8\}$ (or the same with the roles of the two players reversed). The second has period 20 and consists of $8 \times (A, C)$, (B, C), (B, A), $8 \times (C, A)$, (C, B), (A, B).

A consequence of the global convergence result is that in a large population any finite set of deviations from the rules will have no effect on the long run behavior of the model. While this might lead one to think that there is no scope for nonmyopic play in this game, the ability to affect limiting behavior with a finite set of manipulations is not a necessary condition for nonmyopic play. What happens in this example is that a single deviation from the equilibrium has a dramatic effect on aggregate behavior in the intermediate run, and (or so I conjecture) the intermediate run benefit can be sufficiently large so that a sufficiently patient rational player will deviate infinitely often and achieve a payoff greater than is possible in any equilibrium.

Let me now describe the unusual cyclical contagion dynamic which the leaning process engenders in the intermediate run. Suppose that all of the players have seen only B played in the past, but that one player trembles and plays A in period 1. His first period opponent will assign a probability of at least one-tenth to each of A and B and hence play A in periods 2 through 10. As in the previous section the set of players playing A begins to grow in a contagious fashion.[17] After A becomes common things start to change. There will be an increasing number of players who have only seen A (and C) in the previous 10 periods and who will therefore switch to C. Later, after most players have switched to C there will be players who no longer remember seeing A and thus return to playing B. This, however, is usually not the end of the effects of the single manipulation. Instead, because different players have different experiences, there will usually be one player who still remembers seeing A after most others have returned to B. This player will play A, initiating another cycle in the population. Hence, while play must converge to B, the convergence may be quite slow and the distribution of play may move far away from the equilibrium.

Figure 1 illustrates the typical path of play following a single deviation. For a randomly selected realization of the matching process, the figure graphs the fraction of the population playing A in each period (and for the 100 player example the fraction playing B as well). Note that instead of smooth gradual convergence to B we see lengthy dramatic cycles where the entire population swings back and forth from one strategy to another before play eventually locks on to B (which is observed here only in the 10 player example). In the 100 player example, the peak fraction of players playing B in each cycle looks random and seems typically to be fairly far from one. We may thus think of convergence to everyone playing B as requiring a very rare random event.

---

[17] Note that while the growth process is still random it is unlike that described in Section 3 in that strategy A cannot simply die out as experience accumulates. Any player who has seen A in the previous 10 periods cannot play B.
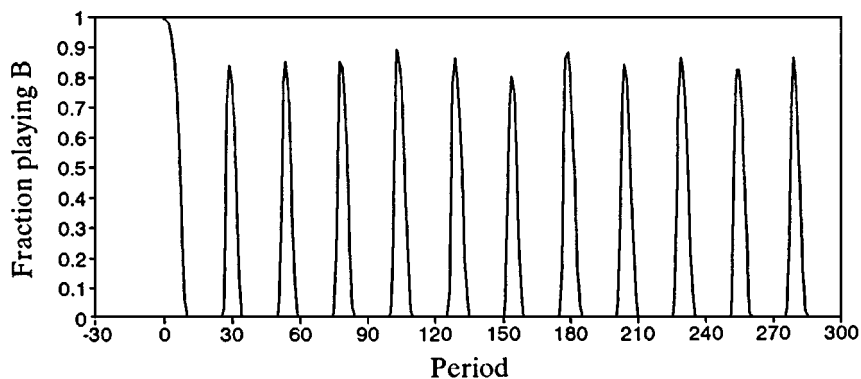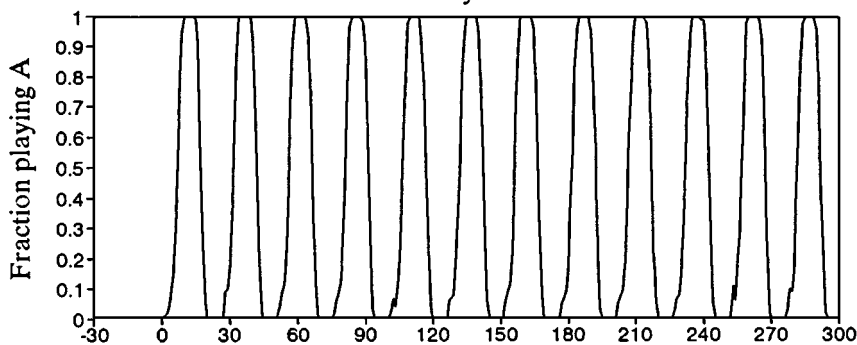
10 Players
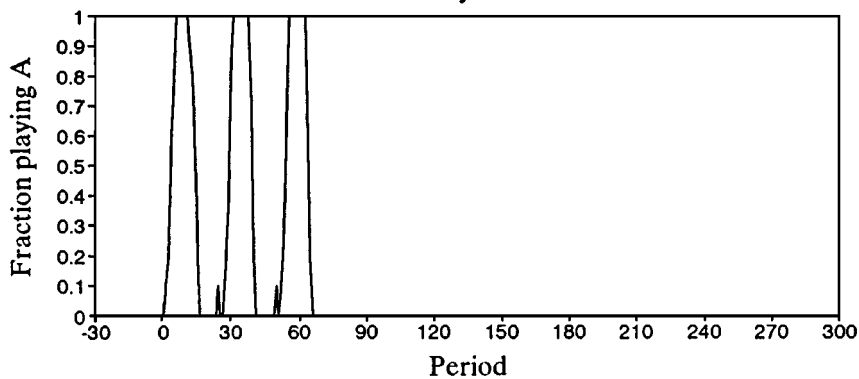
100 Players

FIG. 1. Typical path of play following a perturbation.

Table VI presents (fairly imprecise) Monte Carlo estimates of the length of time until a population of N players return to playing B after a single perturbation. The length of time for which the population cycles increases dramatically when the size of the population is increased. Intuitively, what seems to be going on is that the system has a nearly stable cycle which is broken up only by an unrepresentative realization of the random matching (involving all players' experiences being unusually similar so that they all return to playing B at about the same time). As the population size increases, these unrepresentative realizations of the random matching become more and more unlikely and thus the cycles become more and more nearly stable.

The version of this model without the single rational player illustrates a novel way in which intermediate run and long run behavior can differ. What we see in Fig. 1 (and what seems to happen given virtually any initial history) is that for thousands or millions of periods the players in the model will cycle between the strategies. The standard long run analysis of this model (as contained in Proposition 6) completely misses this in saying only that eventually everyone ends up playing the equilibrium. Practically the dynamics would be much better described by a definition which allowed the cycle to be labeled as an additional "intermediate run limit set" which turns out to be a very strong attractor.

Returning now to the model with one rational player, the fact that aggregate behavior is altered so dramatically by a single manipulation and that manipulations become more effective in larger populations leads one to conjecture that for any population size there exists a discount factor sufficient to make nonmyopic play profitable. Table VII presents additional simulation results which explore this conjecture. What is recorded in the table is the limit as $\delta \to 1$ of the per period payoff obtained by a rational player who follows each of three different strategies when the initial beliefs are concentrated on B. If the rational player plays myopically he receives a payoff of one in every period as (B, B) is played in every match. The second line of the table gives the payoff received by the rational player when he follows the strategy which differs from 10 period memory fictitious play only when the player has seen B 10 times in a row, in which case he is assumed to play A. The third strategy considered is that of always playing A. Each of the latter two strategies is more profitable

TABLE VI
Expected Time to Equilibrium Following a Single Perturbation

Population Size:	2	4	10	20	40	60	80
Approximate Expected Wait:	22	24	58	250	4000	40000	400000

TABLE VII
Estimated Limiting Payoff of Various Strategies

Strategy	Population Size						
	2	4	10	50	100	500	1000
Myopic	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Memory 10 fictitious play + manipulation	3.3	2.7	2.2	2.2	2.2	2.2	2.2
Always A	0.0	0.5	2.9	4.3	4.3	4.3	4.3

than myopic play for the larger population sizes listed in the table, and the pattern of the payoffs in the table suggests that this will be true also in larger populations.

When the rational player in the model above prefers to manipulate play, he will deviate from finite memory fictitious play infinitely often. Otherwise, play will eventually converge back to everyone playing B, and the rational player will assign probability converging to one to this having happened and thus eventually will have an incentive to manipulate play.

Let me say again that the example above is admittedly nonrobust to the learning rules and is meant just to illustrate possible dynamic patterns. The assumption of a finite memory length is critical both because it makes the contagion easy to get started and because it allows the cycles to recur. With standard fictitious play rules there would likely be no discount factor sufficient to induce manipulation in a sufficiently large population. The basic observation that a player may want to manipulate play for a medium run benefit, however, may be regarded as practically robust in that similar cycles would make manipulation profitable if the game was played by a few thousand players following standard fictitious play rules.

6. FURTHER EXAMPLES

This section presents a few more examples illustrating some of the additional considerations which arise outside of the class of symmetric 2×2 games.

6.1. *Asymmetric Games*

The model of this paper can be extended fairly easily to handle asymmetric games, where the primary observation I would like to make is that there is perhaps more scope for nonmyopic play than is made obvious by the previous results. Consider a model where N player 1's and N player 2's are randomly matched to play an asymmetric 2×2 coordination game with equilibria (A, A), (B, B), and $(\mu_1 A + (1 - \mu_1)B, \mu_2 A + (1 - $

$\mu_2)B$). The obvious consequence of the basic manipulation theorem is that nonmyopic play will be optimal for a patient player of either type if $\mu_1 < 1/2$ and $\mu_2 < 1/2$. What may be more interesting is that manipulation will also be optimal in two other sets of circumstances exemplified by the games

	A	B
A	2, 2	$-9, 0$
B	0, 0	1, 1

	A	B
A	2, 2	0, 0
B	$-7, -2$	1, 1

First, suppose beliefs also have the property that the probability of a strategy which is seen from time t on approaches one, and consider a game like that on the left where $\mu_1 < 1/2$, but $\mu_2 > 1/2$. (The example has $\mu_1 = 1/3$ and $\mu_2 = 5/6$.) In such a game, a sufficiently patient rational player 1 will try to manipulate play. The argument is similar to those of the previous section and is only sketched here. Consider initial beliefs for which play converges to B, and suppose the Nth player 1 (N_1 for short) plays A in the first several periods and then reverts to fictitious play. Again **A** and **B** are steady states of the dynamic system and it will suffice to show that a positive probability exists of converging to **A** after the manipulation. If in the first T periods players N_1 and 1_1 are alternately matched with players 1_2 and 2_2, then players 1_2 and 2_2 will eventually switch to A. Because he sees A in every subsequent period, player 1_1 will then adopt A as well (for T large enough). In the next T periods suppose that 1_1 and 2_1 are matched alternatingly with 1_2 and 2_2, while players N_1 and 3_1 meet players 3_2 and 4_2. The first group will sustain themselves for the rest of the game, while player N_1's manipulation adds the third and fourth player to the group playing A. Continuing in this way, we reach a state in which everyone plays A.

Second, suppose beliefs are asymptotically empirical (as in fictitious play) and suppose $\mu_1 + \mu_2 < 1$. The game on the right above is an example ($\mu_1 = 3/5$ and $\mu_2 = 1/10$). In such a game a rational player of *either* type will want to manipulate play. That a rational player 2 wants to manipulate play follows from the first example. To see also that there is scope for a rational player 1 to do so as well, again look at the effect of player N_1 playing A at the start of the game. If player 1_2 is matched with player N_1 in a fraction μ_1 of the periods and with player 1_1 in the remainder, eventually player 1_2 starts playing A, and this in turn causes player 1_1 to switch as well (because $1 - \mu_1 > \mu_2$). Again, the contagion may now spread to the rest of the population through further manipulations by player N_1 or simply on its own as an additional player 1 is caused to play A which in turn adds another player 2, and so on.

6.2. *Dominated Strategies*

Among games with more than two strategies, the simplest ones to analyze are those which reduce to 2×2 games after the removal of dominated or iteratively dominated strategies. For example, consider a 3×3 game in which C is strictly dominated and the payoffs when A and B are played are as before. Clearly, all of the results so far showing that manipulation is possible will remain true with only minor changes. If initial beliefs assign probability close to one to the risk dominated equilibrium (B, B), player N can still manipulate play using the same strategy as before. No strategy other than A or B will ever be played and dynamics proceed exactly as in the 2×2 game. More generally, because no player other than the rational player will play the strictly dominated strategy, adding such strategies never reduces the set of opportunities for the manipulation of play.

On the other hand, the converses in Propositions 4 and 5 do not carry over so easily. Though they are dominated, the additional strategies provide player 1 with an additional tool with which to manipulate play. For example, if we eliminate the dominated strategy in the game below, we obtain a 2×2 game with $\mu = 3/4$, so that manipulation is impossible in a neighborhood of **B** in both the finite and infinite memory fictitious play models. In the full 3×3 game, strategy A is much better than B when there is a significant probability of one's opponent player C. For this reason, player N can now manipulate play by playing C in the first several periods. When he does so, there is some probability that two or more players will assign a probability of at least one-fifth to C and switch to playing A. If these players are subsequently matched, they will continue to play A while player N's actions convince more players to switch to A. Once enough players are playing A (here at least five of them) there is a positive probability that playing A will spread to the rest of the population. The basic point to take away is that dominated strategies cannot be ignored and may affect the selection the model produces.

	A	B	C
A	$2, 2$	$-5, 0$	$0, -20$
B	$0, -5$	$1, 1$	$-20, -20$
C	$-20, 0$	$-20, -20$	$-30, -30$

6.3. *Strategic Risk*

Evaluating myopia in games which do not involve simply the addition of strictly dominated games is much more complicated. It is possible to get a

generalization of the basic manipulation theorem to some games by adding hypotheses which ensure that the best response to a mixture of A and B is always either A or B (see Kandori and Rob (1993) for an example). With such an assumption, the same strategy as before will allow a player to manipulate play away from a strategy which is pairwise risk dominated. In general, however, manipulation may be made more difficult by the presence of additional undominated strategies. The game shown below is an illustration.

	A	B	C
A	12, 12	0, 0	$-100, 5$
B	0, 0	8, 8	$-100, 5$
C	5, -100	5, -100	0, 0

In this game, a rational player who tries to shift play from B to A is faced with the possibility that at some point along the way one of the myopic players may assign a probability in the interval $(3/8, 5/12)$ to A. That player will then play C and may trigger a contagion which leads everyone to play C. In this steady-state, player N receives a lower payoff than at **B**, and because C is a best response to any mixture placing probability at least $1/2$ on C, it would be impossible for player N to manipulate play back to **A** or **B**. Because of this type of strategic risk there may be more stable states once we move beyond 2×2 games.

7. CONCLUSION

In this paper, I have tried to take seriously the idea that learning is a social process in which individuals react to the experience they gain in a series of encounters with others over the course of their lifetime. The primary observation which results is that large population behavior is complex and a variety of contagion and other dynamics arise quite naturally. For this reason, justifying the assumption that players would want to act myopically may require that population sizes be extremely large.

The dynamics of learning in large populations suggest a tendency for play to shift toward risk dominant equilibria. The practical importance of this effect depends on the degree of risk dominance, with selection most probable when the pareto optimum is strongly risk dominant, and perhaps otherwise likely only when players react strongly to their most recent experiences or with local interaction. Further analyses of large population models may yield other insights unavailable in two player models.

A couple of connections with the literature on learning and evolution with noise following Foster and Young (1990), Kandori *et al.* (1993), and Young (1993) are probably also worth noting. First, given the rough similarity between the personal experience model of this paper and Young's model of learning by sampling, one might expect that similar dynamics might make evolution much faster in Young's model than in the KMR model (especially if the samples on which players base their beliefs are small and the degree of risk dominance is strong). Second, the example of Section 5 provides another reason to be cautious when applying limiting analyses to these models. The evolutionary forces created when noise is added to the example of that section would be powerful—we would expect to see cycling almost all the time if each member of a population of 100 players were to tremble as often as once every billion periods—but would not be picked up by the standard limiting analysis.

APPENDIX

Proof of Proposition 1.

The system of myopic responses to the beliefs generates a Markov process on the state space of the previous k observations of the players. The state in which everyone has seen A in each of the last k periods and that in which everyone has seen B are clearly steady states. Call these states **A** and **B**. It then suffices to show that for any other state z_t there exists T such that $\mathrm{Prob}\{z_{t+T} = \mathbf{A}\} > 0$ or $\mathrm{Prob}\{z_{t+T} = \mathbf{B}\} > 0$. Renaming the strategies if necessary, assume that $\mu \leq 1/2$ and that players play A with positive probability when indifferent if $\mu = 1/2$. Write $A(z_t)$ for the number of players with beliefs such that they play A with positive probability in period t.

The proof proceeds by considering three cases depending on the value of $A(z_t)$. First, in the proof of Proposition 2, a claim is established which guarantees (with a minor modification if $\mu = \frac{1}{2}$) that play converges to **A** with positive probability from any state with $A(z_t) \geq 2$. Second, the result is trivial when $A(z_t) = 0$—play is already at **B**. Finally, it remains only to consider the case of z_t with $A(z_t) = 1$.

For this case, suppose player 1 plays A with positive probability and all other players do not. Consider realizations of the matching in which players 1 and 2 are matched in the first k periods. If in any period both play B, then because there is a positive probability they will continue to play B for as long as their beliefs are not adjusted in the direction of A, there is a positive probability that the state **B** will be reached within k periods. On the other hand, if at least one of players 1 and 2 play A in the

first k periods, we have $\eta_1(h_1^{t+k}) + \eta_2(h_2^{t+k}) \geq 1$. Noting that the player who assigns lower probability to A always has his belief increase when at least one player plays A and that k is even, there must have been an intermediate period $t + r$ in which $\eta_1(h_1^{t+r}) \geq 1/2$ and $\eta_2(h_2^{t+r}) \geq 1/2$. Hence, we have a state with $A(z_{t+r}) \geq 2$, and there is again a positive probability of play reaching **A**. QED

Proof of Proposition 4. Define z_{it} to be the number of past observations of player i which would have to be switched from B to A to induce him to play A. Formally, set $z_{it} = (\mu - \eta_i(h_i^t)(A))(w_i + t - 1)$ where w_i is the sum of the weights in player i's initial fictitious history. Let $x_{1t}, x_{2t}, \ldots, x_{N-1t}$ be the first through $N - 1$st order statistics of the z_{it}. To show that at most one irrational player plays A in any given period, I simultaneously establish conditions I and II below by induction on t.

I: $x_{1t} + x_{2t} > 0$.
II: $x_{1t} \leq 0 \Rightarrow x_{2t} + x_{3t} \geq 2\mu(1 - \lceil x_{1t}/\mu \rceil)$.

(The notation $\lceil x \rceil$ indicates the smallest integer greater than or equal to x.)

Given the initial conditions on beliefs, $z_{i1} > 0$ for all i and conditions I and II are satisfied at $t = 1$. To see that they hold at $t + 1$ whenever they hold at t, divide the problem into two cases.

Case 1. $x_{1t} > 0$. Here, no myopic players play A in period t, so that at most one of them has an opponent playing A. Hence,

$$x_{1t+1} + x_{2t+1} \geq x_{1t} + \mu - 1 + x_{2t} + \mu > 2\mu - 1 > 0,$$

and condition I holds at time $t + 1$.

As for condition II, if $x_{1t+1} \leq 0$ it must be that player 1's opponent played A in period t and the other myopic players' opponents did not. (By player 1 is meant the player with the largest z_{it+1}). Hence,

$$x_{2t+1} + x_{3t+1} = x_{2t} + \mu + x_{3t} + \mu > 2\mu = 2\mu(1 - \lceil x_{1t+1}/\mu \rceil),$$

with the final inequality from $x_{1t+1}/\mu \geq \mu - 1/\mu > -1$.

Case 2. $x_{1t} \leq 0$. Here, we distinguish between two subcases where player 1 does or does not meet player N in period t. In the first of these, condition I follows because again at most one myopic player has an opponent who plays A. In the second subcase, condition I follows because player 1 does not see A giving

$$x_{1t+1} + x_{2t+1} \geq \text{Min}\{x_{1t} + x_{2t} + 2\mu - 1, \ x_{2t} + x_{3t} + 2\mu - 2\}$$
$$\geq \text{Min}\{2\mu - 1, \ 2\mu + 2\mu - 2\} > 0.$$

To establish condition II for the first subcase, note that

$$x_{2t+1} + x_{3t+1} = x_{2t} + x_{3t} + 2\mu \geq 2\mu(1 - \lceil x_{1t}/\mu \rceil + 1)$$
$$\geq 2\mu(1 - \lceil x_{1t+1}/\mu \rceil),$$

again because $x_{1t} - x_{1t+1} \leq 1 - \mu < \mu$. In the second, we have

$$x_{2t+1} + x_{3t+1} \geq x_{2t} + x_{3t} + 2\mu - 2 \geq 2\mu(1 - \lceil x_{1t}/\mu \rceil) + 2\mu - 2$$
$$> 2\mu(1 - \lceil x_{1t}/\mu \rceil) - 2\mu = 2\mu(1 - \lceil x_{1t+1}/\mu \rceil),$$

with the last line following from $x_{1t+1} = x_{1t} + \mu$. This completes the proof that conditions I and II hold.

Condition I immediately gives the second conclusion of the proposition. The first conclusion follows from the second, because player N expects to meet an A player in each period with probability at most $1/(N-1)$, and this is less than μ when $N \geq 4$.   QED

Proof of Proposition 5. (1) As a first step, I show that $s_t \to \mathbf{A}$ with probability one if the rational player N plays A in every period. To see this, note that as in Lemma 1 we have a Markov process on the state space of the players' k period histories and that $\mathbf{A}$ is a steady state. It then suffices to show that there is a positive probability path from every state to $\mathbf{A}$. To do so, I simply exhibit a realization of the first several periods of matching which guarantees that play converges to $\mathbf{A}$ regardless of the initial state. Suppose first that player N meets player 1 in the first k periods. In the next $k - \lfloor \mu k \rfloor$ periods suppose that player 1 meets player 2, while player N meets player 3. (Note that players 1 and N play A in these periods.) In the following $\lfloor \mu k \rfloor$ periods, suppose that player N alternately meets players 2 and 3. After $2k$ periods a state has been reached in which players 2 and 3 have seen A in at least $k - \lfloor \mu k \rfloor + \lfloor \lfloor \mu k \rfloor/2 \rfloor$ of the last k periods. This is at least $\lfloor \mu k \rfloor + 1$ whenever $k - \lfloor \mu k \rfloor + (\lfloor \mu k \rfloor - 1)/2 \geq \lfloor \mu k \rfloor + 1$, which is equivalent to the assumption $\lfloor \mu k \rfloor \leq \frac{2}{3}k - \frac{1}{3}$. In period $2k + 1$, players 2 and 3 therefore both play A.

To finish the argument, we simply repeat the argument above with players 2 and 3 meeting for the following $2k$ periods while player N meets player 1 then alternately meets players 4 and 5 until they switch to A. In the next $2k$ periods we can have player N meet players 1, 6, and 7 until 6 and 7 switch to A, and so on.

Given that the expected per period payoff when player N plays A approaches a as $\delta \to 1$, it is immediate that the probability of play converging to $\mathbf{B}$ when player N uses his optimal strategy must vanish as

$\delta \to 1$. To show in fact that play converges to **A** with probability 1 for a neighborhood $[\underline{\delta}, 1)$ of discount factors, however, requires a little more work.

To begin, I show that $\exists \epsilon_1 > 0$ such that for any $\delta \in (0, 1)$ player N's optimal strategy must prescribe playing A whenever he assigns a probability of at least $1 - \epsilon_1$ to the true state z_t being **A**. (The state here being the previous k observations of all N players.) For any other strategy, let t be the first period in which player N plays B despite having the beliefs above. The expected payoff loss in period t is at least $(1 - \epsilon_1)(a - d) - \epsilon_1(b - c)$. The expected gain in the future is bounded above by the gain possible relative to playing A in every future period which is

$$\epsilon_1 \sum_{s=1}^{\infty} \delta^s \text{Prob}\{z_{t+s} \neq \mathbf{A} | z_t, z_t \neq \mathbf{A}\}(a - c),$$

where the probability term concerns the evolution of the state when player N plays A in every period. Because the Markov process describing that evolution converges at an exponential rate, the right-hand sum has a finite limit (bounded independently of the initial state) as $\delta \to 1$. Hence, we may choose ϵ_1 sufficiently small so that any potential gain is less than the short term loss regardless of the discount factor. Because the convergence is exponential, it also follows that for any $\eta > 0$ there exists $\underline{\delta} < 1$ such that $\delta > \underline{\delta}$ implies that player N's average payoff conditional on any private history is at least $a - \eta$.

Next, note that we may choose T such that player N's belief satisfies

$$\text{Prob}\{z_t = \mathbf{A} | s_{-N, s} = A \, s_{N, s} = A \, \forall s \in \{t - 1, \ldots, t - T\}, \, h_{t-T}\} > 1 - \epsilon_1,$$

and such that there is a positive probability that the contagious process leads everyone to play A T periods after the state was **B** and player N started playing A. When player N receives a payoff of a for T periods in a row he plays A for as long as he continues to see A. Hence, if player N receives a payoff of a for T periods in a row infinitely often there is probability one of playing converging to **A**.

Now, choose $\underline{\delta}$ such that player N's average payoff conditional on any history h_t is greater than that obtained when he receives a payoff of a in at most T out of every $T + 1$ periods. Suppose $\delta > \underline{\delta}$. If play does not converge to **A** with probability one when the rational player plays his optimal strategy it must be that for some k there is a positive probability of the event Λ_k in which he and his opponent play A for T periods in a row exactly k times. By Levy's zero or one law $\text{Prob}\{\Lambda_k | h_t\} \to 1_{\Lambda_k}$ almost surely, so there exists a positive probability set of histories H_t for which

player N assigns probability at least $1 - \epsilon$ to his never again receiving a
payoff of a in T consecutive periods. For ϵ sufficiently small this is
impossible given player N's maximization and the choice of δ. Hence, play
converges to **A** almost surely as desired.   QED

Proof of Proposition 6. The dynamics of the model are described by a
Markov process on the finite state space consisting of the 10 previous
observations of each player. Write **B** for the state in which all players have
seen B in each of the previous 10 periods. This is clearly a steady state. It
then suffices to show that given any other state s there exists a time T
such that there is a positive probability of the system moving from s to **B**
in T periods. Note further that it suffices to prove the result for the case of
$N = 4$ players. If the result holds for $N = 4$ we may use an inductive
argument to show that it holds for all larger even N, because there is a
positive probability of a realization of the matching in which the first four
players are matched among themselves for a time sufficiently long to allow
them to all reach B followed by a long period in which players 1 and 2 are
matched with each other (continuing to play B) while players 3 through N
are matched among themselves. If the second long period is long enough
for a group of $N - 2$ players to have a positive probability of all reaching
B from any initial condition, then with this matching there is a positive
probability of all N players having reached B.
    To prove the result for $N = 4$, I consider a large set of possible first
period action profiles. For each I exhibit a realization of the first several
periods of matching which leads to the state **B**.
    It may make the proof clearer to note the order in which I will show that
first period profiles give a positive probability of a transition to **B**: (1) all
four players playing B; (2) two or more playing B; (3) one or more B and
one or more C; (4) two or more C; (5) two or more A.

    1. First, suppose that all four players play B in period 1. This profile
is repeated forever and the state **B** is reached.

    2. Suppose that two or more players play B in period 1, say players 1
and 2. Consider a realization where players 1 and 2 are matched with each
other for a long time. They continue to play B while players 3 and 4
eventually enter one of the limit sets described above. If players 3 and 4
reach the equilibrium we are done. If they reach the first type of limit
cycle, we break up the cycle by supposing that after the 12th period of the
cycle (with the first period being the profile listed first in the description in
Lemma 2) player 1 meets player 3 and player 2 meets player 4 in periods
13 through 34. It is easy to verify that players 1 and 3 play (B, A),
$9 \times (A, A)$, (C, A), $9 \times (C, C)$, (C, B), and (B, B) in those periods, and
that players 2 and 4 both play B in all subsequent periods. If instead they

reach the second type of limit cycle, we obtain the identical path to **B** if after the 19th period of the cycle player 1 meets player 3 and player 2 meets player 4 in periods 20 through 41. Players 2 and 4 both play B in all subsequent periods. Players 1 and 3 play (B, A), $9 \times (A, A)$, (C, A), $9 \times (C, C)$, (C, B), and (B, B).

3. Next, consider all profiles where one player plays B and one player plays C. If two or more play B the result follows from the last paragraph. If the first period profile is $BCAA$, then if players 1 and 2 are matched in the first period, the second period profile (if it does not have two Bs) can be $BAAA$, $BAAC$, or $BACC$. In the first case match players 1 and 3 in period 2. Players 2 and 3 have memories B then A and A then B of the first two periods, so if they are subsequently matched we get the same path as in the previous paragraph starting from 9 (A, A)'s, leading both to play B (which gives a positive probability of getting to the equilibrium by the previous paragraph). In the second case matching players 2 and 3 in period 2 implies players 2 and 4 have memories B then A and A then B and we again get a positive probability of reaching **B** from the state which occurs after they meet in the next 21 periods. In the third case matching players 3 and 4 they will play (C, C) for 10 periods and then switch to (B, B).

If the first period profile is $BCCA$, then if players 1 and 2 are matched we either get a state with two Bs or $BACA$. Either of these gives a positive transition probability.

If the first period profile is $BCCC$, then we again get a profile with two Bs or $BACC$. This again gives a positive transition probability.

4. Now, suppose two or more players play C in period 1. If those two players are matched in period 1 and in the next several periods they will eventually switch to (B, C) or (B, B). Both of these give a positive transition probability.

5. Finally, suppose that two or more players play A in period 1. Matching those two players in period 1 and the next several periods, they eventually switch to (C, C) or to (A, C) (which leads subsequently to (B, C) or (C, C)). In any case the result follows from the paragraphs above.   QED

REFERENCES

Aoyagi, M. (1996). ''Evolution of Beliefs and Nash Equilibria in Normal Form Games,'' *J. Econ. Theory* **70,** 444–469.

Banerjee, A., and Weibull, J. (1993). ''Evolutionary Selection with Discriminating Players,'' mimeo. Cambridge, MA: MIT.

Canning, D. (1990). ''Social Equilibrium,'' Cambridge University Economic Theory Discussion Paper 150.

Cheung, Y., and Friedman, D. (1995). "Individual Learning in Normal Form Games: Some Laboratory Results," mimeo. Santa Cruz: University of California.

Dekel, E., and Scotchmer, S. (1992). "On the Evolution of Optimizing Behavior," *J. Econ. Theory* **57,** 392–406.

Ellison, G. (1993). "Learning, Local Interaction, and Coordination," *Econometrica* **61,** 1047–1071.

Ellison, G. (1994). "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Rev. Econ. Studies* **61,** 567–588.

Foster, D., and Young, H. P. (1990). "Stochastic Evolutionary Game Dynamics," *Theoret. Popul. Biol.* **38,** 219–232.

Fudenberg, D., and Kreps, D. (1991). "Learning and Equilibrium in Games," mimeo. Cambridge, MA: MIT.

Fudenberg, D., and Kreps, D. (1993). "Learning Mixed Equilibria," *Games Econ. Behav.* **5,** 320–367.

Fudenberg, D., and Levine, D. K. (1993). "Steady State Learning and Nash Equilibrium," *Econometrica* **61,** 547–573.

Fudenberg, D., and Levine, D. K. (1996). *Theory of Learning in Games*, unpublished manuscript.

Jordan, J. (1993). "Three Problems in Learning Mixed-Strategy Nash Equilibria," *Games Econ. Behav.* **5,** 368–386.

Kandori, M. (1992). "Social Norms and Community Enforcement," *Rev. Econ. Studies* **59,** 63–80.

Kandori, M., Mailath, G., and Rob, R. (1993). "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica* **61,** 29–56.

Krishna, V. (1991). "Learning in Games with Strategic Complementarity," Harvard Business School Division of Research Working Paper 92-073.

Krishna, V., and Sjostrom, T. (1995a). "On the Convergence of Fictitious Play," Harvard Institute of Economic Research Discussion Paper 1717.

Krishna, V., and Sjostrom, T. (1995b). "Learning in Games: Fictitious Play Dynamics," mimeo. University Park, PA: Pennsylvania State University/Cambridge, MA: Harvard University.

Matsui, A., and Rob, R. (1991). "Evolution, Rationality, and Equilibrium Selection in Societal Games," CARESS Working Paper 91-31. Philadelphia, PA: University of Pennsylvania.

Milgrom, P., and Roberts, R. (1991). "Adaptive and Sophisticated Learning in Normal Form Games," *Games Econ. Behav.* **3,** 82–100.

Miyasawa, K. (1961). "On the Convergence of the Learning Process in a 2×2 Nonzero-Sum Two Person Game," Princeton University Economic Research Program Research Memorandum 33.

Monderer, D., and Shapley, L. (1996). "Potential Games," *Games Econ. Behav.* **14,** 124–143.

Robinson, J. (1951). "An Iterative Method of Solving a Game," *Ann. Math.* **54,** 296–301.

Shapley, L. (1964). "Some Topics in Two-Person Games," in *Advances in Game Theory* (M. Dresher, L. Shapley, and A. Tucker, Eds.). Princeton, NJ: Princeton University Press.

Young, H. P. (1993), "The Evolution of Conventions," *Econometrica* **61,** 57–84.