1. Introduction

1.1. Introduction

This book is about the theory of learning in games. Most of non-cooperative game theory has focused on equilibrium in games, especially Nash equilibrium, and its refinements such as perfection. This raises the question of when and why we might expect that observed play in a game will correspond to one of these equilibria. One traditional explanation of equilibrium is that it results from analysis and introspection by the players in a situation where the rules of the game, the rationality of the players, and the players' payoff functions are all common knowledge. Both conceptually and empirically, these theories have many problems.[1]

This book develops the alternative explanation that equilibrium arises as the long-run outcome of a process in which less than fully rational players grope for optimality over time. The models we will discuss serve to provide a foundation for equilibrium theory. This is not to say that learning models provide foundations for all of the equilibrium concepts in the literature, nor does it argue for the use of Nash equilibrium in every situation; indeed, in some cases most learning models do not lead to any equilibrium concept beyond the very weak notion of rationalizability. Nevertheless, learning models

[1] First, a major conceptual problem occurs when there are multiple equilibria, for in the absence of an explanation of how players come to expect the same equilibrium, their play need not correspond to any equilibrium at all. While it is possible that players coordinate their expectations using a common selection procedure such as Harsanyi and Selten's [1988] tracing procedure, left unexplained is how such a procedure comes to be common knowledge. Second, we doubt that the hypothesis of exact common knowledge of payoffs and rationality apply to many games, and relaxing this to an assumption of almost common knowledge yields much weaker conclusions. (See for example. Dekel and Fudenberg [1990], Borgers [1994].) Third, equilibrium theory does a poor job explaining play in early rounds of most experiments, although it does much better in later rounds.. This shift from non-equilibrium to equilibrium play is difficult to reconcile with a purely introspective theory.

can suggest useful ways to evaluate and modify the traditional equilibrium concepts. Learning models lead to refinements of Nash equilibrium: for example, considerations of the long-run stochastic properties of the learning process suggest that risk dominant equilibria will be observed in some games. They lead also to descriptions of long-run behavior weaker than Nash equilibrium: for example considerations of the inability of players in extensive form games to observe how opponents would have responded to events that did not occur suggests that self-confirming equilibria that are not Nash may be observed as the long-run behavior in some games.

We should acknowledge that the learning processes we analyze need not converge, and even when they do converge the time needed for convergence is in some cases quite long. One branch of the literature uses these facts to argue that it may be difficult to reach equilibrium, especially in the short run. We downplay this anti-equilibrium argument for several reasons. First, our impression is that there are some interesting economic situations in which most of the participants seem to have a pretty good idea of what to expect from day to day, perhaps because the social arrangements and social norms that we observe reflect a process of thousands of years of learning from the experiences of past generations. Second, although there are interesting periods in which social norms change so suddenly that they break down, as for example during the transition from a controlled economy to a market one, the dynamic learning models that have been developed so far seem unlikely to provide much insight about the medium-term behavior that will occur in these circumstances.[2] Third, learning theories often have little to say in the short run, making predictions that are highly dependent on details of the learning process and prior beliefs; the long-run predictions are generally more robust to the specification of the

[2] However, Boylan and El-Gamal [1993], Crawford [1995], Roth and Er’ev [1995], Er’ev and Roth [1996], Nagel [1993], and Stahl [1994] use theoretical learning models to try to explain data on short-term and medium-term play in game theory experiments.

model. Finally, from an empirical point of view it is difficult to gather enough data to test predictions about short-term fluctuations along the adjustment path. For this reason we will focus primarily on the long-run properties of the models we study. Learning theory does, however, make some predictions about rates of convergence and behavior in the medium run, and we will discuss these issues as well.

Even given the restriction to long-run analysis, there is a question of the relative weight to be given to cases where behavior converges and cases where it does not. We chose to emphasize the convergence results, in part because they are sharper, but also because we feel that these are the cases where the behavior that is specified for the agents is most likely to be a good description of how the agents will actually behave. Our argument here is that the learning models that have been studied so far do not do full justice to the ability of people to recognize patterns of behavior by others. Consequently, when learning models fail to converge, the behavior of the model's individuals is typically quite naive; for example, the players may ignore the fact that the model is locked in to a persistent cycle. We suspect that if the cycles persisted long enough the agents would eventually use more sophisticated inference rules that detected them; for this reason we are not convinced that models of cycles in learning are useful descriptions of actual behavior. However, this does not entirely justify our focus on convergence results: as we discuss in chapter 8 more sophisticated behavior may simply lead to more complicated cycles.

We find it useful to distinguish between two related but different kinds of models that are used to model the processes by which players change the strategies they are using to play a game. In our terminology, a "learning model" is any model that specifies the learning rules used by individual players, and examines their interaction when the game (or games) is played repeatedly. In particular, while Bayesian learning is certainly a form of learning, and one that we will discuss, learning models can be far less sophisticated, and include for example stimulus-response models of the type first studied by Bush and Mosteller in the 1950's and more recently taken up by economists.[3] As will become clear in the course of this book, our own views about learning models tend to favor those in which the agents, while not necessarily fully rational, are nevertheless somewhat sophisticated; we will frequently criticize learning models for assuming that agents are more naïve than we feel is plausible.

Individual-level models tend to be mathematically complex, especially in models with a large population of players. Consequently, there has also been a great deal of work that makes assumptions directly on the behavior of the aggregate population. The basic assumption here is that some unspecified process at the individual level leads the population as a whole to adopt strategies that yield improved payoffs. The standard practice is to call such models "evolutionary," probably because the first examples of such processes came from the field of evolutionary biology. However, this terminology may be misleading, as the main reason for interest in these processes in economics and the social sciences is not that the behavior in question is thought to be genetically determined, but rather that the specified "evolutionary" process corresponds to the aggregation of plausible learning rules for the individual agents. For example chapter 3 discusses papers that derive the standard replicator dynamics from particular models of learning at the individual level.

Often evolutionary models allow the possibility of mutation, that is, the repeated introduction (either deterministically or stochastically) of new strategies into the population. The causes of these mutations are not explicitly modeled, but as we shall see mutations are related to the notion of experimentation, which plays an important role in the formulation of individual learning rules.

[3] Examples include Cross [1983], and more recently the Borgers and Sarin [1995], Er'ev and Roth [1996], and Roth and Er'ev [1995] papers discussed in chapter 3.

1.2. Large Populations and Matching Models

This book is about learning, and if learning is to take place players must play either the same or related games repeatedly so that they have something to learn about. So far, most of the literature on learning has focused on repetitions of the same game, and not on the more difficult issue of when two games are "similar enough" that the results of one may have implications for the other.[4] We too will avoid this question, even though our presumption that players *do* extrapolate across games they see as similar is an important reason to think that learning models have some relevance to real-world situations.

To focus our thinking, we will begin by limiting attention to two-player games. The natural starting for the study of learning is to imagine two players playing a two person game repeatedly and trying to learn to anticipate each other's play by observation of past play. We refer to this as the *fixed player model.* However, in such an environment, players ought to consider not only how their opponent will play in the future, but also about the possibility that their current play may influence the future play of their opponents. For example, players might think that if they are nice they will be rewarded by their opponent being nice in the future, or that they can "teach" their opponent to play a best response to a particular action by playing that action over and over.

Consider for example the following game:

[4] Exceptions that develop models of learning from similar games are Li Calzi [1993] and Romaldo [1995].

	L	R
U	1,0	3,2
D	2,1	4,0

In almost any learning model, a player 1 who ignores considerations of repeated play will play D, since D is a dominant strategy and thus maximizes 1's current expected payoff for any beliefs about opponents. If as seems plausible, player 2 eventually learns 1 plays D, the system will converge to (D,L), where 1's payoff is 2. But if 1 is patient, and knows that 2 "naively" chooses each period's action to maximize that period's payoff given 2's forecast of 1's action, then player 1 can do better by always playing U, since this eventually leads 2 to play R. Essentially, a "sophisticated" and patient player facing a naive opponent can develop a "reputation" for playing any fixed action, and thus in the long run obtain the payoff of a "Stackelberg leader."

Most of learning theory abstracts from such repeated game considerations by explicitly or implicitly relying on a model in which the incentive to try to alter the future play of opponents is small enough to be negligible. One class of models of this type is one in players are locked in to their choices, and the discount factors are small compared to the maximum speed at which the system can possibly adjust. However, this is not always a sensible assumption. A second class of models that makes repeated play considerations negligible is that of a large number of players, who interact relatively anonymously, with the population size large compared to the discount factor.

We can embed a particular two- (or *N*-) player game in such an environment, by specifying the process by which players in the population are paired together to play the game. There are a variety of models, depending on how players meet, and what information is revealed at the end of each round of play.

Single Pair Model: Each period, a single pair of players is chosen at random to play the game. At the end of the round, their actions are revealed to everyone. Here if the population is large, it is likely that the players who play today will remain inactive for a long time. Even if players are patient, it will not be worth their while to sacrifice current payoff to influence the future play of their opponents if the population size is sufficiently large compared to the discount factor.

Aggregate Statistic Model: Each period, all players are randomly matched. At the end of the round, the population aggregates are announced. If the population is large each player has little influence on the population aggregates, and consequently little influence on future play. Once again players have no reason to depart from myopic play.

Random Matching Model: Each period, all players are randomly matched. At the end of each round each player observes only the play in his own match. The way a player acts today will influence the way his current opponent plays tomorrow, but the player is unlikely to be matched with his current opponent or anyone who has met the current opponent for a long time. Once again myopic play is approximately optimal if the population is finite but large compared the players' discount factors.[5] This is the treatment most frequently used in game theory experiments.

The large population stories provide an alternative explanation of "naive" play; of course they do so at the cost of reducing its applicability to cases where the relevant population might plausibly be thought to be large.[6] We should note that experimentalists

[5] The size of the potential gain depends on the relationship between the population size and the discount factor. For any fixed discount factor, the gain becomes negligible if the population is large enough. However, the required population size may be quite large, as shown by the "contagion" arguments of Ellison [1993].

[6] If we think of players extrapolating their experience from one game to a "similar" one, then there may be more cases where the relevant population is larger than there appear to be at first sight.

often claim to find that a "large" population can consist of as few as 6 players. Some discussion of this issue can be found in Friedman [1996].

From a technical point of view, there are two commonly used models of large populations: *finite populations* and *continuum populations.* The continuum model is generally more tractable.

Another, and important, modeling issue concerns how the populations from which the players are drawn relates to the number of "player roles" in the stage game. Let us distinguish between an *agent* in the game, corresponding to a particular player role, and the actual player taking on the role of the agent in a particular match. If the game is symmetric, we can imagine that there is a single population from which the two agents are drawn. This is referred to as the *homogenous population* model. Alternatively, we could assume that each agent is drawn from a distinct population. This is referred to as the case of an *asymmetric population*. In the case of an aggregate statistic model where the frequency of play in the population is revealed and the population is homogeneous, there are two distinct models, depending on whether individual players are clever enough to remove their own play from the aggregate statistic before responding to it. There seems little reason to believe that they cannot, but in a large population it makes little difference, and it is frequently convenient to assume that all players react to the same statistic.

Finally, in a symmetric game, in addition to the extreme cases of homogeneous and heterogeneous populations, one can also consider intermediate mixtures of the two cases, as in Friedman [1991], in which each player has some chance of being matched with an opponent from a different population, and some chance of being matched with an opponent from the same population. This provides a range of possibilities between the homogeneous and asymmetric cases.

1.3. Three Common Models of Learning and /or Evolution

Three particular dynamic adjustment processes have received the most attention in the theory of learning and evolution. In *fictitious play*, players observe only the results of their own matches and play a best response to the historical frequency of play. This model is most frequently analyzed in the context of the fixed-player (and hence asymmetric population) model, but the motivation for that analysis has been the belief that the same or similar results obtain with a large population. (Chapter 4 will discuss the extent to which that belief is correct.) In the *partial best response* dynamic, a fixed portion of the population switches each period from their current action to a best response to the aggregate statistic from the previous period. Here the agents are assumed to have all the information they need to compute the best response, so the distinctions between the various matching models are unimportant; an example of this is the Cournot adjustment process discussed in the next section. Finally, in the *replicator* dynamic, the share of the population using each strategy grows at a rate proportional to that strategy's current payoff, so that strategies giving the greatest utility against the aggregate statistic from the previous period grow most rapidly, while those with the least utility decline most rapidly. This dynamic is usually thought of in the context of a large population and random matching, though we will see in chapter 4 that a similar process can be derived as the result of boundedly rational learning in a fixed player model.

The first part of this book will examine these three dynamics, the connection between them, and some of their variants, in the setting of one-shot simultaneous-move games. Our focus will be on the long run behavior of the systems in various classes of games, in particular on whether the system will converge to a Nash equilibrium, and, if so, which equilibrium will be selected. The second part of the book will examine similar questions in the setting of general extensive form games. The third and final part of the book will discuss what sorts of learning rules have desirable properties, from both the normative and descriptive points of view.

1.4. Cournot Adjustment

To give the flavor of the type of analyses the book considers, we now develop the example of Cournot adjustment by firms, which is perhaps the oldest and most familiar nonequilibrium adjustment model in game theory. While the Cournot process has many problems as a model of learning, it serves to illustrate a number of the issues and concerns that recur in more sophisticated models. This model does not have a large population, but only one "agent" in the role of each firm. Instead, as we explain below, the model implicitly relies on a combination of "lock-in" or inertia and impatience to explain why players don't try to influence the future play of their opponent.

Consider a simple duopoly, whose players are firms labeled $i = 1,2$. Each player's strategy is to choose a quantity $s^i \in [0, \infty)$ of a homogeneous good to produce.. The vector of both strategies is the strategy profile denoted by s. We let s^{-i} denote the strategy of player *i*'s opponent. The utility (or profit) of player *i* is $u^i(s^i, s^{-i})$, where we assume that $u^i(\cdot, s^{-i})$ is strictly concave. The best response of player *i* to a profile, denoted $BR^i(s^{-i})$, is

$$BR^i(s^{-i}) = \arg\max_{\tilde{s}^i} u^i(\tilde{s}^i, s^{-i}) .$$

Note that since utility is strictly concave in the player's own action, the best response is unique.

In the Cournot adjustment model time periods $t = 1,2,\ldots$ are discrete. There is an initial state profile $\theta_0 \in S$. The adjustment process itself is given by assuming that in each period the player chooses a pure strategy that is a best response to the previous period. In other words the Cournot process is $\theta_{t+1} = f^C(\theta_t)$ where $f_i^C(\theta_t) = BR_i(\theta_t^{-i})$ At each date t player i chooses a pure strategy $s_t^i = BR_i(s_{t-1}^{-i})$. A steady state of this process is a state $\hat{\theta}$ such that $\hat{\theta} = f^C(\hat{\theta})$. Once $\theta_t = \hat{\theta}$ the system will remain in this state forever.

The crucial property of a steady state is that by definition it satisfies $\hat{\theta}^i = BR_i(\hat{\theta}_{-i})$ so that is a Nash equilibrium.

1.5. Analysis of Cournot Dynamics[7]

We can analyze the dynamics of the two player Cournot process by drawing the reaction curves corresponding to the best response function.

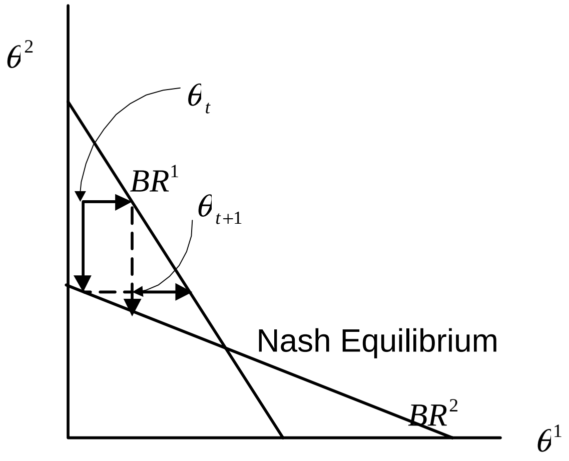


Figure 1.1

As drawn, the process converges to the intersection of reaction curves, which is the unique Nash Equilibrium.

In this example, the firms output levels change each period, so even if they started out thinking that their opponent's output was fixed, they should quickly learn that it is not.

[7] The appendix reviews some basic facts about stability conditions in dynamical systems.

However, we shall see later that there are variations on the Cournot process in which players' beliefs are less obviously wrong.

In Figure 1.1, the process converges to the unique Nash equilibrium from any initial conditions, that is, the steady state is *globally stable*. If there are multiple Nash equilibria, we cannot really hope that where we end up is independent of the initial condition, so we cannot hope that any one equilibria is globally stable. What we can do is ask whether play converges to a particular equilibrium once the state gets sufficiently close to it. The appendix reviews the relevant theory of the stability of dynamical systems for this and other examples.

1.6. Cournot Process with Lock-In

We argued above that interpreting Cournot adjustment as a model of learning supposes that the players are pretty dim-witted: They choose their actions to maximize against the opponent's last period play. It is as if they expect that today's play will be the same as yesterday's. In addition, each player assigns probability one to a single strategy of the opponent so there is no subjective uncertainty. Moreover, although players have a very strong belief that their opponent's play is constant, their opponent's actual play can vary quite a bit. Under these circumstances, it seems likely that players would learn that their opponent's action changes over time; this knowledge might then alter their play.[8]

One response to this criticism is to consider a different dynamic process with alternating moves: Suppose that firms are constrained to take turns with firm 1 moving in periods 1, 3, 5, and firm 2 in periods 2, 4, 6. Each firm's decision is "locked in" for two

[8] Selten's [1988] model of anticipatory learning models this by considering different degrees of sophistication in the construction of forecasts. The least sophisticated is to assume that opponents will not change their actions; next is to assume that opponents believe that *their* opponents will not change their actions, and so forth. However, no matter how far we carry out this procedure, in the end players are always more sophisticated than their opponents imagine

periods: firm 1 is constrained to set its second-period output s_1^1 to be the same as its first-period output s_2^1.

Suppose further that each firm's objective is to maximize the discounted sum of its per-period payoffs $\sum_{t=1}^{\infty} \delta^{t-1} u^i(s_t)$, where $\delta < 1$ is a fixed common discount factor. There are two reasons why a very rational firm 1 would not choose its first-period output to maximize its first-period payoff. First, since the output chosen must also be used in the second period, firm one's optimal choice for a fixed time-path of outputs by firm 2 should maximize the weighted sum of firm 1's first and second period profit, as opposed to maximizing first period profit alone. Second, as in the discussion of Stackelberg leadership in section 1.2, firm 1 may realize that its choice of first-period output will influence firm 2's choice of output in the second period.

However, if firm 1 is very impatient, then neither of these effects matters, as both pertain to future events, and so it is at least approximately optimal for firm 1 to choose at date 1 the output that maximizes its current period payoff. This process, in which firms take turns setting outputs that are the static best response of the opponent's output in the previous period, is called the *alternating-move Cournot dynamic;* it has the qualitatively the same long-run properties as the simultaneous-move adjustment process, and in fact is the process that Cournot actually studied.[9]

There is another variant on the timing of moves that is of interest: instead of firms taking turns, suppose that each period, one firm is chosen at random and given the opportunity to change its output, while the output of the other remains locked in. Then once again if firms are impatient, the equilibrium behavior is to choose the action that maximizes the immediate payoff given the current output of the opponent. There is no

[9] Formally, the two processes have the same steady states, and a steady state is stable under one process if and only if it is stable under the other. .

need to worry about predictions of future because the future does not matter. Note that this model has exactly the same dynamics as the alternating move Cournot model, in the sense that if a player gets to move twice or more in a row, his best response is the same as it was last time, and so he does not move at all. In other words, the only time movement occurs is when players switch roles, in which case the move is the same as it would be under the Cournot alternating move dynamic. While the dating of moves is different, and random to boot, the condition for asymptotic stability is the same.

What do we make of this? Stories that make myopic play optimal require that discount factors be very small, and in particular small compared to the speed that players can change their outputs: the less locked-in the players are, the smaller the discount factor needs to be. So the key is to understand why players might be locked in. One story is that choices are capital goods like computer systems, which are only replaced when they fail. This makes lock-in more comprehensible; but limits the applicability of the models. Another point is that under the perfect foresight interpretation, lock-in models do not sound like a story of learning. Rather they are a story of dynamics in a world where learning is irrelevant because players know just what they need to do to compute their optimal actions.[10]

[10] Maskin and Tirole [1988] study the Markov-perfect *equilibria* of this game with alternating moves and two-period lock in.

1.7. Review of Finite Simultaneous Move Games

1.7.1. Strategic- Form Games

Although we began by analyzing the Cournot game because of its familiarity to economists, this game is complicated by the fact that each player has a continuum of possible output levels. Throughout the rest of the book, we are going to focus on finite games, in which each player has only finitely many available alternatives. Our basic setting will be one in which a group of players $i = 1,\dots, I$ play a *stage game* against one another.

The first half of the book will discuss the simplest kind of stage game, namely *one-shot simultaneous move games.* This section reviews the basic theory of simultaneous-move games, and introduces the notation we use to describe them. The section is not intended as an introduction to game theory; readers who would like a more leisurely or detailed treatment should look elsewhere.[11] Instead, we try to highlight those aspects of "standard" game theory that will be of most importance in this book, and to focus on those problems in game theory for which learning theory has proven helpful in analyzing.

In a one-shot simultaneous-move game, each player *i* simultaneously chooses a strategy $s^i \in S^i$. We refer to the vector of players' strategies as a *strategy profile*, denoted by $s \in S \equiv \times_{i=1}^{I} S^i$. As a result of these choices by the players, each player receives a *utility* (also called a payoff or reward) $u^i(s)$. The combination of the player set, the strategy spaces, and the payoff functions is called the *strategic* or *normal* form of the game. In two-player games, the strategic form is often displayed as a matrix, where rows index

[11] For example, Fudenberg and Tirole [1991] or Myerson [1990].

player 1’s strategies, columns index player 2’s, and the entry corresponding to each strategy profile (s^1, s^2) is the payoff vector

$$(u^1(s^1, s^2), u^2(s^1, s^2)) .$$

In “standard” game theory, that is analysis of Nash equilibrium and its refinements, it does not matter what players observe at the end of the game.[12] When players learn from each play of the stage game how to play in the next one, what the players observe makes a great deal of difference to what they can learn. Except in simultaneous move games, though, it is not terribly natural to assume that players observe their opponents’ strategies, because in general extensive form games a strategy specifies how the player would play at every one of his information sets. For example, if the extensive form is

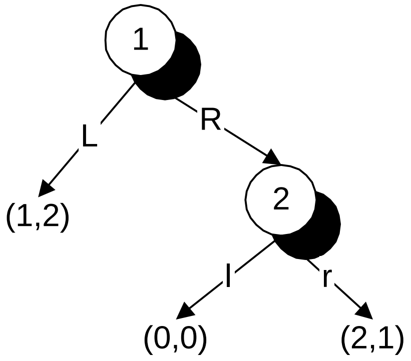


Figure 1.2

and suppose that player 1 plays L. Then player 2 does not actually get to move. In order for player 1 to observe 2's strategy, player 1 must observe how player 2 would have played had 1 played R. We could make this assumption. For example, player 2 may write down his choice on a piece of paper and hand it to a third party, who will implement the choice if 2's information set is reached, and at the end of the period 1 gets to see the piece of paper.

[12] Although what players observe at the end of the stage game in repeated games does play a critical role even without learning. See for example Fudenberg, Levine, and Maskin [1994].

This sounds sort of far-fetched. Consequently, when we work with strategic form games, and suppose that the chosen strategies are revealed at the end of each period, the interpretation is that we are looking at a simultaneous-move game, that is, a game where each player moves only once and all players choose their actions simultaneously This is the case we will consider in the first part of the book

In addition to pure strategies, we also allow the possibility that players use random or "mixed" strategies. The space of probability distributions over a set is denoted by $\Delta(\cdot)$. A randomization by a player over his pure strategies is called a *mixed strategy* and is written $\sigma^i \in \Sigma^i \equiv \Delta(S^i)$. Mixed strategy profiles are denoted $\sigma \in \Sigma = \times_{i=1}^{I} \Sigma^i$. Players are expected utility maximizers, so their payoff to a mixed strategy profile σ is the expected value $u^i(\sigma) \equiv \sum_s u^i(s) \prod_{i=1}^{I} \sigma^i(s^i)$. Notice that the randomization of each player is independent of other players' play.[13]

As in the analysis of the Cournot game, it is useful to distinguish between the play of a player and his opponents. We will write s^{-i}, σ^{-i} for the vector of strategies (pure or mixed) of player *i*'s opponents.

In the game, each player attempts to maximize his own expected utility. How he should go about doing this depends on how he thinks his opponents are playing, and the major issue addressed in the theory of learning is how he should form those expectations. For the moment, though, suppose that player *i* believes that the distribution of his opponents play corresponds to the mixed strategy profile σ^{-i}. Then player *i* should play a *best response,* that is a strategy $\hat{\sigma}^i$ such that $u^i(\hat{\sigma}^i, \sigma^{-i}) \geq u^i(\sigma^i, \sigma^{-i}), \forall \sigma^i$. The set of all best responses to σ^{-i} is denoted by $BR^i(\sigma^{-i})$, so $\hat{\sigma}^i \in BR^i(\sigma^{-i})$. In the Cournot

[13] We will not take time here to motivate the use of mixed strategies, but two motivations will be discussed later on in the book, namely (i) the idea that the randomization corresponds the random draw of a particular opponent from a population each of which is playing a pure strategy, and (ii) the idea that what looks like randomization to an outside observer is the result of unobserved shocks to the player's payoff function.

adjustment process, players expect that their opponent will continue to play as they did last period, and play the corresponding best-response.

In the Cournot process, and many related processes, such as fictitious play, that we will discuss later in the book, the dynamics are determined by best response correspondence $BR^i(\sigma^{-i})$. That is, two games with the same best-response correspondence will give rise to the same dynamic learning process. For this reason, it is important to know when two games have the same best-response correspondence. If two games have the same best-response correspondence for every player, we say that they are *best-response equivalent.*

A simple transformation that leaves preferences, and consequently best-responses unchanged, is a linear transformation of payoffs. The following proposition gives a slight generalization of this idea:

Proposition 1.1: Suppose $\tilde{u}^i(s) = au^i(s) + v^i(s^{-i})$ for all players *i*. Then $\tilde{u}$ and *u* are best-response equivalent.

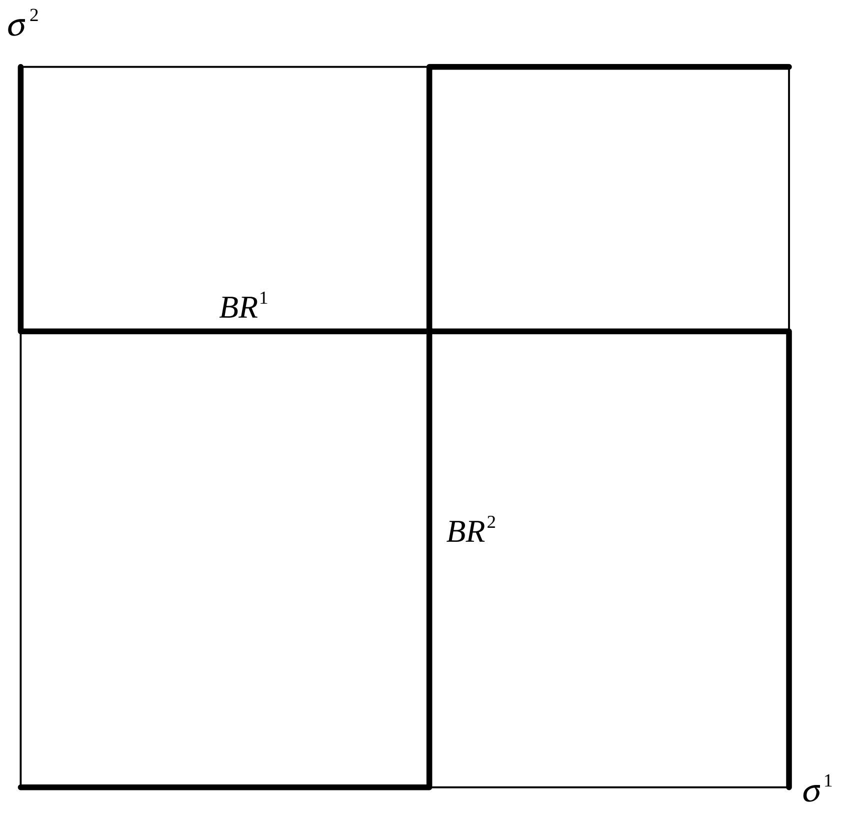
This result is immediate, because adding a constant that depends only on other players actions does not change what is best for the player in question.

An important class of games are zero sum games, which are two player games in which the payoff to one player is the negative of the payoff to the other player.[14] Zero-sum games are particularly simple, and have been extensively studied. A useful result relates best-response correspondences of general games to those of zero-sum games in two-player, two-action games.

[14] Note that the "zero" in "zero-sum" is unimportant; what matters is that the payoffs have a constant sum.

Proposition 1.2 : Every 2x2 game for which the best-response correspondences have a unique intersection that lies in the interior of the strategy space is best-response equivalent to a zero sum game.

Proof: Denote the two strategies A and B respectively. There is no loss of generality in assuming that A is a best-response for player 1 to A, and B is a best response for player 2 to A. (If A was also a best-response to A for 2, then the best-response correspondences intersect at a pure strategy profile, which we have ruled out by assumption.) Let σ^i denote player *i*'s probability of playing A. Then the best-response correspondences of the two players is determined by the intersection point, and is as diagrammed below.



The trick is to show that this intersection point can be realized as the intersection of best-responses of a zero-sum game. Notice that if $1 > a$, then the matrix below are the payoffs to player 1 in a zero sum game in which the best response for player 1 to A is A, and the best response of player 2 to A is B.

$$\begin{bmatrix} 1 & 0 \\ a & b \end{bmatrix}$$

(Recall that player 2's payoffs are the negative of player 1's.) Than player 1 is indifferent between A and B when $\sigma^2 = a\sigma^2 + b(1-\sigma^2)$ while player 2 is indifferent between A and B when $\sigma^1 + a(1-\sigma^1) = b(1-\sigma^1)$. Fixing the intersection point σ^1, σ^2, we may solve these two linear equations in two unknowns to find

$$a = \frac{\sigma^2 - \sigma^1 + \sigma^1\sigma^2}{1+\sigma^1\sigma^2}.$$

Since $\sigma^2 - \sigma^1 < 1$ we see that $a < 1$, as required.

☑

1.7.2. Dominance and Iterated Dominance

The most primitive notion in game theory is that of dominance. Roughly a strategy is dominated if another strategy does better no matter how the player expects his opponents to play. The general idea is that we should not expect dominated strategies to be played.[15] The strongest notion of dominance is that of *strict dominance.*

Definition 1.1: Strategy σ^i is *strictly dominated for player i* if there is some other $\tilde{\sigma}^i \in \Sigma^i$ such that $u^i(\tilde{\sigma}^i, \sigma^{-i}) > u^i(\sigma^i, \sigma^{-i})$ for all $\sigma^{-i} \in \Sigma^{-i}$.

[15] But note that in the non-simultaneous move games, strategies may differ in how much information they generate about opponents play, so that once learning is considered, so a strategy that yields poor payoffs may be played in order to acquire information.

(This condition is equivalent to $u^i(\tilde{\sigma}^i, s^{-i}) > u^i(\sigma^i, s^{-i})$ for all pure strategy profiles s^{-i} of i's opponents, because i*'s* payoff when his opponents play a mixed profile is a convex combination of the corresponding pure strategy payoffs.)

A famous example of a game where dominant strategies play a role is the one-shot prisoner's dilemma game

	A	B
A	3,3	1,5
B	5,1	2,2

In this game the strategy B does better than A no matter what the opponent does. If we eliminate the strategy A, then there is a unique prediction: that both players play B. Note however, that (A,A) Pareto dominates (B,B), which is why there is a dilemma.

In this example, both the dominated strategy and the dominating one are pure strategies. Neither of these are general properties. More precisely, a pure strategy s^i can be strictly dominated by a mixed strategy σ^i and yet not dominated by any pure strategy, as in the next example

	A	B
A	5,0	0,0
B	0,0	5,0
C	2,0	2,0

Here the strategy C is not dominated by either A or B for player 1, but it is dominated by a 50-50 mixture over A and B. Moreover, although any mixed strategy that assigns positive probability to a strictly dominated pure strategy is strictly dominated, a mixed strategy can also be strictly dominated even if it assigns probability 0 to every dominated pure strategy.

If a strategy of player 1's is strictly dominated, then there are several reasons why player 2's beliefs might come to assign that strategy probability 0. The first, traditional, argument is that if player 2 knows player 1's payoff function, and knows that 1 is rational, then player 2 should be able to deduce that player 1 will not use a strictly dominated strategy. Secondly, from the viewpoint of learning theory, if the strategy is strictly dominated then[16] player 1 will have no reason to play it, and so player 2 should eventually learn that the dominated strategy is not being played. Either story leads to the idea of *iterated strict dominance*, in which the deletion of some strategies for one player permits the deletion of some strategies for others, and so on. (It can be shown that the order in which strategies are deleted does not matter so long as the deletion process continues until no more deletions are possible.) We will not give a formal definition of iterated strict dominance here, but the following example should make the idea clear:

	A	B
A	1,-100	1,1
B	2,2	2,1

Here no strategy is dominated for player 2, but the strategy A is strongly dominated for player 1. Eliminating that strategy results in the game

[16] We presume that there are not extensive-form complications of the type mentioned in footnote 15.

	A	B
B	2,2	2,1

In this game, B is strongly dominated for player 2, so the unique survivor of iterated strong dominance is (B,A). Note however, that player 2 must be quite sure that player 1 is not playing A, since (A,A) results in a large loss for him. Consequently, the prediction of (B,A) might be overturned if there were some small probability that player 1 played A. (Perhaps there is some chance that player 1 has different payoffs than those given here, or that player 1 makes a "mistake.")

Related to strict dominance is the notion of *weak dominance.*

Definition 1.2: Strategy σ^i is *weakly dominated for player i* if there is some $\tilde{\sigma}^i \in \Sigma^i$, $\tilde{\sigma}^i \neq \sigma^i$ such that $u^i(\tilde{\sigma}^i, \sigma^{-i}) \geq u^i(\sigma^i, \sigma^{-i})$ for all $\sigma^{-i} \in \Sigma^{-i}$, with strict inequality for at least one σ^{-i}.

Again, there seems no reason to play a weakly dominated strategy, and indeed weakly dominated strategies will not be played in models in which agents "tremble" or, more generally, in models where agents beliefs about their opponents' strategies correspond to a completely mixed strategy. However, the notion of *iterated weak dominance* is problematic, and will not play a role in this book.

1.7.3. Nash Equilibrium

One of the problems with dominance is that in many games of interest, the process of iterated dominance does not lead to strong predictions. This has encouraged the application of equilibrium theory, in which all players simultaneously have correct beliefs about each others play while playing best responses to their own beliefs.

Definition 1.3: A *Nash Equilibrium* is a strategy profile $\hat{\sigma}$ such that $\hat{\sigma}^i \in BR^i(\hat{\sigma}^{-i}), \forall i$.

It follows from the Kakutani fixed point theorem that a Nash equilibrium exists in finite games so long as mixed strategies are permitted. (Many simple games, such as "matching pennies," have no pure-strategy equilibria.) A game may have several Nash equilibria, as in the following example of a *coordination game*:

	A	B
A	2,2	0,0
B	0,0	2,2

Here player 1 picks the row and player 2 the column. Each player has two pure strategies A and B. The numbers in the table denote the utility of player 1 and player 2 respectively. There are two pure strategy Nash equilibria at (A,A) and (B,B). There is also one mixed strategy Nash equilibrium where both player randomize with a 1/2 chance of A and a 1/2 chance of B.

What would we expect to happen in this game? Both players prefer either of the two pure strategy Nash equilibria to the mixed strategy Nash equilibrium, since the expected utility to each player at the pure equilibrium is 2, while the expected utility at the mixed equilibrium is only 1. But in the absence of any coordinating device, it is not obvious how the two players can guess which equilibrium to go to. This might suggest that they will play the mixed equilibrium. But at the mixed equilibrium, each player is indifferent, so while equilibrium requires that they give each strategy exactly the same probability, there is no strong reason for them to do so. Moreover, if player 1, say, believes that player 2 is even slightly more likely to play A than B, then player 1 will want to play A with probability one. From an intuitive point of view, the stability of this mixed strategy equilibrium seems questionable.

In contrast, it seems easier for play to remain at one of the pure equilibria, because here each player strictly prefers to play his part of the Nash equilibrium profile as long as he believes there is a high probability that his opponent is playing according to that equilibrium. Intuitively, this type of equilibrium seems more robust. More generally, this is true whenever equilibria are "strict" equilibria in the following sense:

Definition 1.4: A Nash equilibrium s is *strict* if for each player *i*, s^i is the unique best response to s^{-i}, that is, player i strictly prefers s^i to any other response.

(Note that only pure strategy profiles can be strict equilibria, since if a mixed strategy is a best response to the opponents' play, then so is every pure strategy in the mixed strategy's support.)

This coordination game example, although simple, illustrates the two main questions that the theory of learning in games has tried to address, namely: When and why should we expect play to correspond to a Nash equilibrium? And, if there are several Nash equilibria, which ones should we expect to occur?

Moreover, these questions are closely linked: Absent an explanation of how the players coordinate their expectations on the same Nash equilibrium, we are faced with the possibility that player 1 expects the equilibrium (A,A) and so plays A, while player 2 expects (B,B) and plays B, with the result the non-equilibrium outcome (A.B). Briefly, the idea of learning-based explanations of equilibrium is that the fact that the players share a common history of observations can provide a way for them to coordinate their expectations on one of the two pure-strategy equilibria. Typical learning models predict that this coordination will eventually occur, with the determination of which of the two equilibria arise left either to (unexplained) initial conditions or to random chance.

However, for the history to serve this coordinating role, the sequence of actions played must eventually become constant or at least readily predictable by the players, and there is no presumption that this is always the case. Perhaps rather than going to a Nash equilibrium, the result of learning is that the strategies played, wander around aimlessly, or perhaps play lies in some set of alternative larger than the set of Nash equilibria.

Because of the extreme symmetry of the coordination game above, there is no reason to think that any learning process should tend to favor one of its strict equilibria over the other . The coordination game below is more interesting:

	A	B
A	2,2	-*a*,0
B	0,-*a*	1,1

Here there are two strict Nash equilibria, (A,A) and (B,B); both players would prefer the Nash equilibrium (A,A) with payoffs (2,2), since it Pareto dominates the equilibrium at (1,1). Will players learn to play the Pareto efficient equilibrium? One consideration lies in the risk that they face. That is, if *a* is very large, guessing that your opponent is going to play (2,2) is very risky, because if you are wrong you suffer a large loss. One might expect in a learning setting, that it would be difficult to get to a very risky equilibrium, even if it is Pareto efficient. A notion that captures this idea of risk is the Harsanyi-Selten criterion of *risk dominance.*[17] In 2x2 games, the risk dominant strategy can be found by computing

[17] The use of the word "risk" here is different than the usual meaning in economics. Actual risk aversion by the players is already incorporated into the utility functions.

the minimum probability of A that makes A the best response, and comparing it to the minimum probability of B required to make B the best response. In this example A is optimal if there is probability at least $(a+1)/(3+a)$ that the opposing player plays A, while B is optimal if the probability that the opponent plays B is at least $2/(3+a)$; thus A is risk dominant if $a<1$. Alternatively, and somewhat simpler, risk dominance in 2x2 games is equivalent to a simple concept called 1/2-dominance. An strategy is 1/2-*dominant* if it is optimal for all players to play the strategy whenever their opponents are playing that strategy with probability at least 1/2. Thus A is risk dominant if $2-a>1$, or $a<1$.

In both of the examples above, there is a finite number of Nash equilibria. Although some strategic games can have a continuum of equilibria (for example if each player's payoff function is a constant) generically this is not the case. More precisely, for a fixed strategy space *S,* the set of Nash equilibria is finite (and odd) for an open and dense set of payoff functions (Wilson [1971]).[18] In particular, for generic strategic-form payoffs each Nash equilibrium is locally isolated, a fact that will be very useful in analyzing the stability of learning processes. However, this fact is really only applicable to one-shot simultaneous-move games, since in a general extensive form generic assignments of payoffs to *outcomes* or *terminal nodes* do not generate generic strategic-form payoffs: For example, in the strategic form of the game in Figure 1.3, (L,l) and (L,r) lead to the same outcome and so must give each player the same payoff.

[18] For a fixed strategy space *S*, the payoff functions of the I players correspond to a vector in the Euclidean space of dimension $I \cdot \# S$; a et of payoff functions is "generic" if it is open and dense in this space.

1.7.4. Correlated Equilibrium

There is a second important noncooperative equilibrium concept in simultaneous move games, namely Aumann's [1974] notion of a *correlated equilibrium*. This assumes that players have access to randomization devices that are privately viewed, but perhaps the randomization devices are correlated with each other. In this case, if each player chooses a strategy based upon observations of his own randomization device, the result is a probability distribution over strategy profiles, which we denote by $\mu \in \Delta(S)$. Unlike a profile of mixed strategies, such a probability distribution allows play to be correlated.

As in the theory of Nash equilibrium, suppose that players have figured out how their opponents' play depends on their private randomization device, and know how the randomization device works. Since each player knows what pure strategy he is playing, he can work out the conditional probability of signals received by his opponents, and the conditional probabilities of their play. Let $\mu^{-i}(s^i) \in \Delta(S^{-i})$ denote the probability distribution over opponents play induced by μ conditional on s^i, and let μ^i be the marginal for player *i*. Then if $\mu^i(s^i) > 0$, so that player *i* is willing to play s^i, it must be a best-response to $\mu^{-i}(s^i)$. Formally, if $\mu^{-i} \in \Delta(S^{-i})$ we may calculate the expected utility $u^i(\sigma^i, \mu^{-i})$ and define the best-response $\hat{\sigma}^i \in BR^i(\mu^{-i})$ if $u^i(\hat{\sigma}^i, \mu^{-i}) \geq u^i(\sigma^i, \mu^{-i}), \forall \sigma^i$. A *correlated equilibrium* is then a correlated strategy profile μ such that $\mu^i(s^i) > 0$ implies $s^i \in BR^i(\mu^{-i}(s^i))$.

Jordan's [1993] simple three person matching pennies game illustrates the idea of a correlated equilibrium. This game is a variant on matching pennies, where each player simultaneously chooses "H" or "T", and all entries in the payoff matrix are either +1 (win) or -1 (lose). Player 1 wins if he plays the same action as player 2, player 2 wins if he matches player 3, and player 3 wins by *not* matching player 1. The payoffs are

$$\begin{bmatrix} +1,+1,-1 & -1,-1,-1 \\ -1,+1,+1 & +1,-1,+1 \end{bmatrix} \begin{bmatrix} +1,-1,+1 & -1,+1,+1 \\ -1,-1,-1 & +1,+1,-1 \end{bmatrix}$$

where the row corresponds to player 1 (up for H, down for T), the column to player 2, and the matrix to player three. This game has a unique Nash equilibrium, namely for each player to play (1/2,1/2). However, it also has many correlated equilibria: one is the distribution over outcomes in which each of the profiles (H,H,H), (H,H,T), (H,T,T), (T,T,T), (T,T,H), (T,H,H) have equal weight of 1/6. Notice that in this distribution, each player has a 50% chance of playing H. However, no weight is placed on (H,T,H) so that the play of the players is not independent. (It is "correlated.") Taking player 1, for example, we notice that when he plays H, he faces a 1/3 chance each of his opponents playing (H,H), (H,T) and (T,T). Since his goal is to match player 2, he wins 2/3rds of the time by playing H, and only 1/3 of the time if he plays T. So H is a best response to the distribution of opponents play given H. Similarly when he plays T, his opponents are equally likely to play (T,T), (T,H) and (H,H). Now tails wins 2/3rds the time, as against heads which wins only 1/3rd the time. The idea of correlated play is important in the theory of learning for two reasons. First, the types of learning models players are assumed to use are usually relatively naive, as for example in the Cournot adjustment model. In the Cournot model, it is possible for play to cycle endlessly. One consequence of this is that play viewed over time is correlated. In more sophisticated models, we still have to face the possibility that players incidentally correlate their play using time as a correlation device, and in some instances this results in learning procedures converging to correlated rather than Nash equilibrium. Indeed, this is in a sense what happens if the Cournot adjustment procedure is used in the Jordan game. If we begin with (H,H,H) player 3 will wish to switch, leading to (H,H,T). Then player 2 switches to (H,T,T), then player 1 to (T,T,T). Now 3 wants to switch again, leading to (T,T,H), 2 switches to (T,H,H) and finally 1 to

(H,H,H) completing the cycle. In other words, in this example Cournot best-response dynamics lead to cycling, and if we observe the frequency with which different profiles occur, each of the 6 profiles in the correlated equilibrium is observed 1/6 the time. That is, play in the best-response dynamic resembles a correlated equilibrium.

We should note, however, that the fact that Cournot adjustment leads to correlated equilibrium in this particular game is actually a coincidence. If we modify the payoffs so that when (H,T,T) is played, player 1 gets -100 rather than -1, then the best-response cycle remains unchanged, but it is no longer optimal for player 1 to play H against a 1/3 chance of his opponents playing (H,H),(H,T) and (T,T) with equal probability. It turns out that for some more sophisticated learning procedures, the long run actually will be a correlated equilibrium to a good approximation.[19]

A second reason correlation is important is that during the process of learning, players will have beliefs about the mixed strategies that opponents are using. This takes the form of a probability distribution over opponents mixed profiles. Such a probability distribution is always equivalent to a correlated distribution over opponents pure strategy profiles, but need not be equivalent to a profile of mixed strategies for the opponents. Suppose for example there are two opponents each with two alternative A and B. Player 1believes that there is a 50% chance both opponents are playing A, and a 50% chance both are playing B. If he plays against them for a while he hopes to learn which of these alternatives is correct; that is, he does not think that they are correlating their play. In the meantime, however, he will wish to optimize against the correlated profile 50% (A,A)-50% (B,B).

[19] This is true for consistent procedures discussed in chapters 2 and 4 because the game has only two actions. However, the even more sophisticated calibrated procedures discussed in chapter 8 give rise to correlated equilibrium in all games.

APPENDIX: Dynamical Systems and Local Stability

In general, at any moment of time *t*, certain players are playing particular strategies, and have available certain information on which they base their decisions. We refer to all the variables relevant to determining the future of the system as the *state* and denote it by $\theta_t \in \Theta$. In the Cournot adjustment model, the current state is simply the output levels chosen currently by the two firms. More generally, the variables that are represented in θ_t will depend on the particular model we are studying. In addition to discrete time models where $t = 1,2,\ldots$, such as Cournot adjustment, we will also consider some continuous time models where $t \in [0,\infty)$. In discrete time, the state variable will evolve over time according to the deterministic *law of motion* $\theta_{t+1} = f_t(\theta_t)$ or according to the *stochastic* (Markovian) *law of motion* $pr(\theta_{t+1} = \theta) = \phi_t(\theta|\theta_t)$. In continuous time the deterministic law of motion will be $\dot{\theta}_t = f_t(\theta_t)$. Although we will discuss some results in the case of stochastic continuous time, the notation for these models is complicated, and will be introduced when appropriate.

We begin with some basic definitions and results about stability in dynamic processes; a good reference for this material is Hirsch and Smale [1974]. We let $F_t(\theta_0)$ denote the value assumed by the state variable at time *t* when the initial condition at time 0 is θ_0. I In discrete time $F_{t+1}(\theta_0) = f_t(F_t(\theta_0))$, in continuous time $D_t F_t(\theta_0) = f(F_t(\theta_0))$, and in both cases $F_0(\theta_0) = \theta_0$; the map F is called the *flow* of the system.

Definition 1.5: A *steady state* $\hat{\theta}$ of a flow satisfies $F_t(\hat{\theta}) = \hat{\theta}, t > 0$.

Definition 1.6: A steady state $\hat{\theta}$ of a flow is *stable* if for every neighborhood *U* of $\hat{\theta}$ there is a neighborhood U_1 of $\hat{\theta}$ in *U* such that if $\theta_0 \in U_1$ $F_t(\theta_0) \in U, t > 0$, that is, if the system starts close enough to the steady state, it remains nearby. If a steady state is not stable, we say that it is unstable.

Definition 1.7: A steady state $\hat{\theta}$ of a flow is *asymptotically stable* if it is stable, and in addition if $\theta_0 \in U_1$ then $\lim_{t\to\infty} F_t(\theta_0) = \hat{\theta}$. The *basin (of attraction)* of an asymptotically stable steady state is the set of all points θ_0 such that $\lim_{t\to\infty} F_t(\theta_0) = \hat{\theta}$. If there is a unique steady state with basin equal to the entire state space Θ, it is called *globally stable*.

Definition 1.8: A steady state $\hat{\theta}$ is *locally isolated* if it has an open neighborhood in which there are no other steady states.

Note that an asymptotically stable steady state must be locally isolated, but that a stable steady state need not be.

Definition 1.9: A steady state $\hat{\theta}$ is called *hyperbolic* if $Df(\hat{\theta})$ has no eigenvalues on the unit circle (discrete time) or no eigenvalues with zero real parts (continuous time). If the eigenvalues all lie inside the unit circle (discrete time) or have negative real parts (continuous time) the steady state is called a *sink*; if the eigenvalues all lie outside the unit circle (discrete time) or have positive real parts (continuous time) it is called a *source*. Otherwise a hyperbolic steady state is called a *saddle.*

The critical aspect of a hyperbolic steady state in a non-linear dynamical system is that it behaves locally like the linear system $\theta_{t+1} = \hat{\theta} + Df(\hat{\theta})(\theta_t - \hat{\theta})$ (discrete time) or $\dot{\theta} = Df(\hat{\theta})\theta$ (continuous time). The precise meaning of this can be found in the smooth linearization theorem of Hartmann (see Irwin [1980]), which says that there is a smooth local coordinate system that maps the trajectories of the non-linear system exactly onto the trajectories of the linear system. The most significant case is

Proposition 1.3: A sink is asymptotically stable.

In the two player Cournot process, we may check for asymptotic stability, by computing the appropriate eigenvalues. Denote the slopes of the best response functions $BR_i(s_{-i})$ by $BR_i'(s_{-i})$. We have

$$Df = \begin{pmatrix} 0 & BR_1' \\ BR_2' & 0 \end{pmatrix}$$

with corresponding eigenvalues $\lambda = \pm\sqrt{BR_1' \cdot BR_2'}$. Consequently, the absolute value of λ is smaller than 1 if slope *BR*$_2$ is less than the slope of *BR*$_1$, in which case the process is asymptotically stable.[20]

To the extent that we accept the adjustment process, we can argue that sources will not be observed. However, the case of saddles is more complicated; a flow corresponding to a saddle is illustrated below

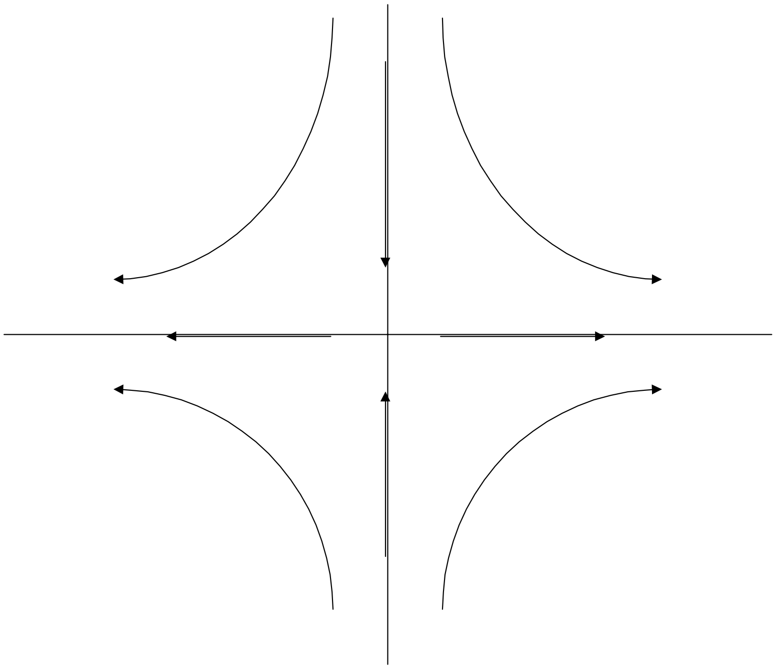


Figure 1.3

[20] Recall that, because s_2 is on the vertical axis, the "slope" of player 1's best response function is $1/BR_1^{'}$.

Here there are paths that approach close to the steady state (at the origin in the figure), but eventually move away. However, once the system moves close to the steady state, the actual speed of movement becomes very slow, so the system will remain near the steady state for a long time before leaving. Consequently, a saddle may be a good model of the "intermediate run," even though it is not a good model of the "long run." This point is emphasized in Binmore and Samuelson [1995], who argue that saddles may in fact be a sensible model of actual behavior over long periods of time.

Even if a game has stable equilibria, it may have more than one of them. Consequently, stability analysis will not in general yield a unique prediction, although it can help reduce the set of possible outcomes. Moreover, the fact that a system has one or more stable equilibria does not imply that the state will approach any of the equilibria.. As a result, we will sometimes have need of the following more general concepts in characterizing the long-run behavior of dynamic systems:

Definition 1.10: The set of ω-limit points of the flow F are the points θ such that for some θ_0, and sequence of times $t_n \to \infty$ the $\lim_{n\to\infty} F_{t_n}(\theta_0) = \theta$. That is, θ is an ω-limit point if there is an initial condition from which θ is approached infinitely often. A set $\Theta' \subseteq \Theta$ is *invariant* if $\theta_0 \in \Theta'$ implies $F_t(\theta_0) \in \Theta'$ for all *t*. An invariant set $\Theta' \subseteq \Theta$ is an *attractor* if it has a compact invariant neighborhood Θ'' such that if $\theta_0 \in \Theta''$, and there is a sequence of times $t_n \to \infty$ for which $\theta = \lim_{n\to\infty} F_{t_n}(\theta_0)$ exists, then $\theta \in \Theta'$.

In addition to containing steady states, the set of ω-limit points can contain *cycles* or even other limit sets known as *strange attractors*. There has been a limited amount of work on strange attractors in the theory of learning, such as that of Skyrms [1992,1993]. So far, however, the existence of strange attractors, and the *chaotic* trajectories that surround them, have not played a central role in the theory of learning in games.

References

Aumann, R. [1974]: “Subjectivity and Correlation in Randomized Strategies,” *Journal of Mathematical Economics*, 1: 67-96.

Binmore, K. and L. Samuelson [1995]: “Evolutionary Drift and Equilibrium Selection,” University College London.

Binmore, K., L. Samuelson and K. Vaughn [1995]: “Musical Chairs: Modelling Noisy Evolution,” *Games and Economic Behavior*, 11: 1-35.

Borgers, T. [1994]: “Weak Dominance and Approximate Common Knowledge,” *Journal of Economic Theory*, 4, 265-276.

Boylan, R. and E. El-Gamal [1993]: “Fictitious Play: A Statistical Study of Multiple Economic Experiments,” *Games and Economic Behavior*, 5, 205-222.

Crawford, V. [1995]: “Adaptive Dynamics in Coordination Games,” *Econometrica*, 63,103-158.

Cross, J. [1983]: *A Theory of Adaptive Economic Behavior*, (Cambridge: Cambridge University Press).

Dekel, E. and D. Fudenberg [1990]: “Rational Behavior with Payoff Uncertainty,” *Journal of Economic Theory*, 52: 243--267.

Ellison, G. [1993]: “Learning, Local Interaction, and Coordination,” *Econometrica*, 61: 1047-1071.

Er’ev, I. and A. Roth [1996]: “On The Need for Low Rationality Cognitive Game Theory: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria,” University of Pittsburgh.

Friedman, D. [1991]: “Evolutionary Games in Economics,” *Econometrica*, 59: 637-666.

Friedman, D. [1996]: “Equilibrium in Evolutionary Games: Some Experimental Results,” *Economic Journal*, forthcoming.

Fudenberg, D. and J. Tirole [1991]: *Game Theory*, (Cambridge: MIT Press).

Fudenberg, D., D. K. Levine and E. Maskin [1994]: “The Folk Theorem with Imperfect Public Information,” *Econometrica*, 62: 997-1039.

Harsanyi, J. and R. Selten [1988]: *A General Theory of Equilibrium Selection in Games*, (Cambridge: MIT Press).

Hirsch, M. and S. Smale [1974]: *Differential Equations, Dynamical Systems, and Linear Algebra*.

Irwin, M. C. [1980]: *Smooth Dynamical Systems*, (New York: Academic Press).

Jordan, J. [1993]: “Three Problems in Learning Mixed-Strategy Equilibria,” *Games and Economic Behavior*, 5: 368-386.

Li Calzi, M. [1993]: “Fictitious Play By Cases,” Istituo di Matemataica \”E. Levi\”, Univeristà di Parma.

Maskin, E. and J. Tirole [1988]: “A Theory of Dynamic Oligopoly, 1: Overview and Quantity Competition with Large Fixed Costs,” *Econometrica*, 56: 549-570.

Myerson, R. [1991]: *Game Theory*, (Cambridge: Harvard University Press).

Nagel, R. [1994]: “Experimental Results on Interactive Competitive Guessing,” D.P. B-236, Universität Bonn.

Romaldo, D. [1995]: “Similarities and Evolution,” mimeo.

Roth, A. and I. Er’ev [1995]: “Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Run,” *Games and Economic Behavior*, 6: 164-212.

Selten, R. [1988]: “Anticipatory learning in two-person games,” University of Bonn.

Skyrms, B. [1992]: “Chaos in Dynamic Games,” *Journal of Logic, Language and Information*, 1: 111-130.

Skyrms, B. [1993]: “Chaos and the Explanatory Significance of Equilibrium: Strange Attractors in Evolutionary Game Theory,” UC Irvine.

Stahl, D. [1994]: “Evolution of Smart n Players,” *Games and Economic Behavior*, 5: 604-617.

Wilson, R. [1971]: “Computing Equilibria of n-person Games,” *SIAM Journal of Applied Mathematics*, 21: 80-87.

2. Fictitious Play

2.1. Introduction

One widely used model of learning is the process of fictitious play and its variants. In this process, agents behave as if they think they are facing a stationary, but unknown, distribution of opponents strategies. In this chapter we examine whether fictitious play is a sensible model of learning, and what happens in games when players all use fictitious play learning rules.

We will begin our discussion of fictitious play in a two player setting. After introducing fictitious play in section 2.2, we will discuss conditions under which fictitious play converges in section 2.3. Although the assumption of stationarity that underlies fictitious play is a reasonable first hypothesis in many situations, we might expect players to eventually reject it given sufficient evidence to the contrary. In particular, if the system in which the agents are learning fails to converge, then the assumption of stationarity does not make much sense. As we will see in section 2.4, most of the problems with the long-run behavior of fictitious play arise in this case of non-convergence.

We then move on to the multi-player case in section 2.5. Here a key issue is the whether players form estimates of the joint distribution of opponents' play by taking the product of independent marginal distributions, or whether they instead form these estimates in ways that allow for either objective or subjective correlation.

One important issue in studying any learning process is whether it succeeds in learning. Section 2.6 discusses the notion of payoff consistency. Since fictitious play only tracks information about the frequency that each strategy is played, it is natural to ask under what conditions fictitious play successfully learns these frequencies, in the sense that the players do as well (as measured by the time average of their payoffs) as if the limiting frequencies were known . We refer to this property as “consistency” and show in section 2.6 that if the course of fictitious play involves less “infrequent” switching between strategies then fictitious play is consistent in this sense. In chapter 4 we show that a randomized version of fictitious play can give rise to universal consistency, meaning that they do as well asymptotically as if the frequencies were known in advance, no matter what strategies are used by opponents.

What can be said about the course of play when players use fictitious play rules? We take up this topic in section 2.8 where we show that fictitious play has dynamics very similar to the partial- best-response dynamic discussed in chapter 1. In chapter 3 , we will consider more closely the behavior of best-response and related dynamics.

For the purpose of modeling short-run behavior in experiments, fictitious play can be improved by allowing for larger weights on more recent observations and player “inertia” in the sense of a probability of repeating the most recently chosen action. These and other variants on fictitious play are discussed in section 2.9; extension to random versions of fictitious play are considered in chapter 4.

One important observation is that the process of fictitious play supposes that players do not try to influence the future play of their opponents. As we discussed in the introduction, there are several models of interactions in a large population in which such “naive” or unsophisticated behavior is sensible. We should also point out that many of the formal results deal with a small finite population, so the conceptually naïve play is problematic. More interesting from an economic point of view is the case of a continuum population, since here it is legitimate to ignore strategic interactions. In the most interesting case, with a large but finite population and anonymous random matching, the matching process adds a source of randomness to the evolution of the system, even if the play of each individual agent is a deterministic function of his observations. However, as we will see in the next chapter, this and other sources of randomness turn out not to have much impact on the qualitative conclusions.

2.2. Two Player Fictitious Play

To keep the formalities reasonably simple, we will start out with the case of a two-player simultaneous-move game, with finite strategy spaces S^1, S^2 and payoff functions u^1, u^2. The model of fictitious play supposes that players choose their actions in each period to maximize that period's expected payoff given their prediction or *assessment* of the distribution of opponent's actions in that period, where this assessment takes the following special form:

> Player *i* has an exogenous initial weight function $\kappa_0^i : S^{-i} \to \Re_+$. This weight is updated by adding 1 to the weight of each opponent strategy each time it is played, so that
>
> $$\kappa_t^i(s^{-i}) = \kappa_{t-1}^i(s^{-i}) + \begin{cases} 1 \text{ if } s_{t-1}^{-i} = s^{-i} \\ 0 \text{ if } s_{t-1}^{-i} \neq s^{-i} \end{cases}.$$
>
> The probability that player *i* assigns to player *-i* playing s^{-i} at date *t* is given by
>
> $$\gamma_t^i(s^{-i}) = \frac{\kappa_t^i(s^{-i})}{\sum_{\tilde{s}^{-i} \in S^{-i}} \kappa_t^i(\tilde{s}^{-i})},$$

Fictitious play itself is then defined as any rule $\rho_t^i(\gamma_t^i)$ that assigns $\rho_t^i(\gamma_t^i) \in BR^i(\gamma_t^i)$. Traditional analyses suppose that the player chooses a pure-strategy best response when indifferent between several pure strategies; since exact indifference will not occur for generic payoffs and priors, the precise specification here is unimportant. Note that there is not a unique fictitious play rule, since there may more than one best response to a particular assessment; note also that the behavior prescribed by fictitious play is a discontinuous function of the player’s assessment, since there does not in general exist a continuous selection from the best-response correspondence.

One way to interpret this method of forming an assessment is to note that it corresponds to Bayesian inference when player *i* believes that his opponents’ play corresponds to a sequence of i.i.d. multinomial random variables with a fixed but unknown distribution, and player *i*’s prior beliefs over that unknown distribution take the form of a Dirichlet distribution.[21] In this case player *i*’s prior and posterior beliefs correspond to a distribution over the set $\Delta(S^{-i})$ of probability distributions on S^{-i}. The distribution over opponent’s strategies γ_t^i is the induced marginal distribution over pure strategies. In particular, if beliefs over $\Delta(S^{-i})$ are denoted by μ^i then we have $\gamma_t^i(s^{-i}) = \int_{\Sigma^{-i}} \sigma^{-i}(s^{-i})\mu_t^i[d\sigma^{-i}]$. The hypothesis to emphasize here is not the Dirichlet functional form, but rather the implicit assumption that the player treats the environment as stationary. In section 2.8 we will discuss the possibility of weighting current observations more heavily than past observations, which is one way that players might respond to the possibility that the environment is nonstationary.

We also define the marginal empirical distribution of *j*’s play as

$$d_t^j(s^j) = \frac{\kappa_t(s^j) - \kappa_0(s^j)}{t}$$

The assessments γ_t^i are not quite equal to the marginal empirical distribution d_t^j (recall that there are only two players, so that *j*=-*i*) because of the influence of player *i*'s prior. This prior has the form of a “fictitious sample” of data that might have observed before play began. However, as observations are received over time, they will eventually outweigh the prior, and the assessments will converge to the marginal empirical distributions.

[21] The Dirichlet distribution and multinomial sampling form a “conjugate family,” meaning that if the prior over the probability distributions belongs to the Dirichlet family, and the likelihood function is multinomial, then posterior is also in the Dirichlet family. See the appendix to this chapter for details.

Compared to the updating rule in the myopic adjustment process, fictitious play has the advantage that as long as all of the initial weights are positive, there is no finite sample to which the beliefs assign probability zero. The beliefs do reflect the conviction that the opponent's strategy is constant and unknown, and this conviction may be wrong along the path of the process, for example, if the process cycles. But any finite string of what looks like cycles is consistent with the belief that the world is constant and the observations are a fluke. If cycles persist, we might expect the player to eventually notice them, but at least his beliefs will not be falsified in the first few periods, as they are in the Cournot process.

2.3. The Asymptotic Behavior of Fictitious Play

One key question about fictitious play is whether or not play converges; if it does, then the stationarity assumption employed by players makes sense, at least asymptotically; if it does not, then it seems implausible that players will maintain that assumption. In this section, we examine some sufficient conditions under which fictitious play converges.

The state of the fictitious play process is the vector of the player's assessments, and not the strategies played at date t, since the latter is not sufficient to determine the future evolution of the system. Nevertheless, in a slight abuse of terminology, we will say that a strategy profile is a steady state if it is played in every period after some finite time T.

Proposition 2.1: (a) If s is a strict Nash equilibrium, and s is played at date t in the process of fictitious play, s is played at all subsequent dates.[22] That is, strict Nash equilibria are absorbing for the process of fictitious play. (b) Any pure strategy steady state of fictitious play must be a Nash equilibrium.

[22] Recall from section 1.6.3 that a strategy profile is a strict Nash equilibrium if each player's strategy is a strict best response to the opponents' play, that is every other choice for the player gives a strictly lower payoff.

Proof: First suppose players' assessments γ_t^i are such that their optimal choices correspond to a strict Nash equilibrium $\hat{s}$. Then when the strict equilibrium is played, each player *i*'s beliefs in period *t+1* are a convex combination of γ_t^i and a point mass on $\hat{s}^{-i}$*:* $\gamma_{t+1}^i = (1-\alpha_t)\gamma_t^i + \alpha_t \delta(\hat{s}^{-i})$. Since expected utilities are linear in probabilities, $u^i(\hat{s}^i, \gamma_{t+1}^i) = \alpha_t u^i(\hat{s}^i, \delta(\hat{s}^{-i})) + (1-\alpha_t) u^i(\hat{s}^i, \gamma_t^i)$, and so if $\hat{s}^i$ is a best response for player *i* for γ_t^i, it is a strict best response for γ_{t+1}^i. Conversely, if play remains at a pure strategy profile, then eventually the assessments will become concentrated at that profile, and so if the profile is not a Nash equilibrium, one of the players would eventually want to deviate.

☑

Since the only pure strategy profiles that fictitious play can converge to are those that are Nash equilibria, fictitious play cannot converge to a pure strategy profile in a game all of whose equilibria are mixed. Consider for example the game "matching pennies":

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

with initial weights (1.5,2) and (2,1.5). Then fictitious play cycles as follows: in the first period, 1 and 2 play *t*, so that the weights the next period are (1.5,3) and (2,2.5). Then 2 plays T and 2 plays H for the next two periods, after which 1's weights are (3.5,3) and 2's are (2,4.5). At this point 1 switches to H, and both players play H for the next 3 periods, at which point 2 switches to T, and so on.

It may not be obvious, but although the actual play in this example cycles, the empirical distribution over player *i*'s strategies d_t^i are converging to (1/2,1/2), so that the product of the two empirical marginal distributions, namely {(1/2,1/2),(1/2,1/2)} is the mixed-strategy equilibrium of the game. The fact that the limiting marginal distributions correspond to a Nash equilibrium is a general property of fictitious play:

Proposition 2.2: Under fictitious play, if the empirical distributions d_t^i over each player's choices converge, the strategy profile corresponding to the product of these distributions is a Nash equilibrium.

Proof: The proof is the same as before: if product of the empirical distributions converges to some profile $\hat{\sigma}$, then the beliefs converge to $\hat{\sigma}$, and hence if $\hat{\sigma}$ were not a Nash equilibrium, some player would eventually want to deviate. In fact, it should be clear that this conclusion does not require that the assessments take the specific form given in fictitious play; it suffices that they are "asymptotically empirical", in the sense that in the limit of a great many observations, they become close to the empirical frequencies.

☑

Proposition 2.3: Under fictitious play, the empirical distributions converge if the stage has generic payoffs[23] and is 2×2 [Robinson, 1951] or zero-sum [Miyasawa, 1961], or is

[23] By "generic" we mean "for a set of payoff vectors that have full Lebesgue measure in the relevant payoff space, which here is $\Re^8$. (See section 1.7.3 for a brief discussion of genericity in strategic form games.) Generic 2x2 games have either one or three Nash equilibria, but for nongeneric payoffs the set of Nash equilibria may be a connected interval of strategy profiles. In this case, whether fictitious play converges can depend on the precise rule used to determine a player's choice when indifferent. See Monderer and Sela [1996] for a discussion of the importance of tie-breaking rules.

solvable by iterated strict dominance [Nachbar, 1990] or has strategic complements and satisfies another technical condition [Krishna and Sjostrom 1995].

The empirical distributions, however, need not converge. The first example of this is due to Shapley [1964], who considered a game equivalent to the following one:

	L	M	R
T	0,0	1,0	0,1
M	0,1	0,0	1,0
D	1,0	0,1	0,0

This game has a unique Nash equilibrium, namely for each player to use the mixed strategy (1/3,1/3,1/3). Shapley showed that if the initial weights lead players to choose (T,M), then the fictitious play follows the cycle (T,M)→(T,R)→(M,R)→(M,L)→(D,L)→(D,M)→ (T,M)..., which is the path of Cournot's alternating-move best-response process. In particular, the three "diagonal profiles" (U,L), (M,M) and (D,R) are never played. Moreover, the number of consecutive periods that each profile in the sequence is played increases sufficiently quickly that the empirical distributions d_t^1, d_t^2 do not converge but instead follow a limit cycle. Shapley's proof of this latter fact explicitly computes the time spent in each stage of the sequence. Monderer, Samet and Sela [1995] have an easier proof of non-convergence that we present in section 2.6.

If the payoffs along the diagonal are increased to (1/2, 1/2), then the example above becomes the zero-sum game "rock-scissors-paper" in which rock smashes scissors, scissors cuts paper and paper wraps the rock. This game has the same, unique, Nash equilibrium, namely for each player to play (1/3,1/3,1/3). Moreover, this is the unique *correlated* equilibrium of the game as well. Note for future reference that since rock-scissors-paper is a zero-sum game, the empirical distributions generated by fictitious play must converge to the Nash equilibrium. However, the outcomes played will follow the same best-reply cycle as in the Shapley example; the empirical distributions converge here because the time spent playing each profile increases sufficiently slowly. The rock-scissors-paper game and its modifications have been very useful in understanding learning and evolution in games, and we will be referring to it again in the next chapter.

2.4. The Interpretation of Cycles in Fictitious Play

The early literature on fictitious play viewed the process as describing pre-play calculations players might use to coordinate their expectations on a particular Nash equilibrium (hence the name 'fictitious' play.) From this viewpoint, or when using fictitious play as a means of calculating Nash equilibria, the identification of a cycle with its time average is not problematic, and the early literature on fictitious play accordingly focused on finding conditions that guaranteed the empirical distributions converge.

However, this notion of convergence has some problems as a criterion for whether players have learned to play the corresponding strategies, as it supposes that the players ignore the persistence of cycles, and suppose that their opponents' play corresponds to i.i.d. draws from a fixed distribution. Moreover, because of these cycles, the empirical *joint* distribution of the two players' play (formed by tracking the empirical frequency of strategy profiles, as opposed to the empirical *marginal* frequencies tracked by the d_t^i) can be correlated. Consider the following example, from Fudenberg and Kreps [1990], [1993]:[24]

	A	B

[24] Young [1993] gives a similar example.

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

Suppose this game is played according to the process of fictitious play, with initial weights (1, sqrt (2)) for each player. In the first period, both players think the other will play B, so both play A. The next period the weights are (2,sqrt (2)) and both play B; the outcome is the alternating sequence ((B.,B),(A,A),(B,B), etc.) The empirical frequencies of each player's choices do converge to 1/2, 1/2, which is the Nash equilibrium, but the realized play is always on the diagonal, and both players receive payoff 0 in every period. Another way of putting this is that the empirical joint distribution on pairs of actions does not equal the product of the two marginal distributions, so that the empirical joint distribution corresponds to correlated as opposed to independent play.

From the standpoint of players learning how their opponents behave, this sort of example, where the joint distribution is correlated, does not seem a very satisfactory notion of "converging to an equilibrium." More generally, even if the empirical joint distribution does converge to the product of the empirical marginals, so that the players will end up getting the payoffs they would expect to get from maximizing against i.i.d. draws from the long-run empirical distribution of their opponents' play, one might still wonder if players would ignore persistent cycles in their opponents' play.

One response to this is to use a more demanding convergence notion, so that a player's behavior only converges to a mixed strategy if his intended actions *in each period* converge to that mixed strategy. The standard fictitious play process cannot converge to a mixed strategy in this sense for generic payoffs, so we postpone our discussion of this response until chapter 4.

An alternative response is to decide that the players ignoring cycles is not problematic because players keep track only of data on opponents' frequency of play.

We explore this response more fully in section 2.7.

2.5. Multi-Player Fictitious Play

We now turn to an important modeling issue that arises in extending fictitious play to games with three of more players: in a 3-player game, say, should player 1's assessment γ^1 about the play of his opponents have range $\Sigma^2 \times \Sigma^3$, with $\gamma^1(s^2,s^3) = \gamma^1(s^1)\gamma^1(s^3)$, so that the assessment always corresponds to a mixed strategy profile, or should the range of the assessment be the space $\Delta(S^2 \times S^3)$ of all the probability distributions over opponents' play, including the correlated ones?

To answer this, we rely on the interpretation of fictitious play as the result of Bayesian updating, with the assessments corresponding to the marginal distribution over the opponents' current period strategies that is derived from the player's current beliefs. From this viewpoint, player 1's current assessment of his opponents' play can correspond to a correlated distribution even if player 1 is certain that the true distribution of his opponents' play in fact corresponds to independent randomizations. Formally, the assumption that 2 and 3 randomize independently implies that

$$\gamma_t^1(s^2,s^3\ ) = \int_{\Sigma^{-i}} \sigma^2(s^2)\sigma^3(s^3)\mu'^{i1}_t[d(\sigma^2,\sigma^3)];$$

the assumption of independent mixing is reflected in the fact that the integrand uses the product of $\sigma^2(s^2)$ and $\sigma^3(s^3)$. Despite this, the assessment γ_t^1 need not be a product measure, and indeed it typically won't be a product measure unless player 1's subjective_ uncertainty about his opponents is also uncorrelated, that is, unless μ_t^1 is a product measure.

To make this more concrete, suppose for example that there is a 1/2 chance both opponents are playing A, and a 1/2 chance both are playing B. Thus the support of player 1's beliefs is concentrated on uncorrelated (in fact, pure) strategy profiles of the opponents, but his current assessment corresponds to the correlated profile 1/2 (A,A)-1/2 (B,B).

Consequently, we see that there is no strong reason to suppose that the players' initial assessments correspond to uncorrelated randomizations. A deeper question is whether the support of the players' *prior beliefs* should be the set of opponents' mixed strategy profiles, as in the calculation of the marginal above, or whether the prior should allow for the possibility that the opponents do manage to consistently correlate their play. If the support is the set of mixed strategies, then over time each player will become convinced that the distribution of opponents' play corresponds to the product of the empirical marginal distribution of each opponent's action, so that persistent correlation will be ignored.

To see the difference this makes, consider a 3-player variant of the rock-scissors-paper game discussed above, where players 1 and 2 have exactly the same actions and payoffs as before, and player 3 has the option of betting on the play of the other two. More precisely, if player 3 chooses Out, 3 gets 0; if 3 chooses In she gets 10 if 1 and 2 play on the diagonal, and -1 if 1 and 2 play on the off-diagonal; player 3's action has no effect on the payoffs of the others. It is easy to see that this game has a unique Nash equilibrium: players 1 and 2 play the mixed strategy (1/3,1/3,1/3), and player 2 chooses In. Moreover, this is the unique correlated equilibrium of the game.

As observed above, the play of players 1 and 2 in this game will cycle through the off-diagonal elements , the empirical distributions over the individual payers' actions will converge to (1/3,1/3,1/3). If player 3 estimates separate distributions for player 2 and player 3, then her beliefs will eventually converge to the empirical distributions, leading her to choose In; yet doing so will result in a payoff of -1 in every period, which is below player 3's reservation utility. If, on the other hand, player 3 instead keeps track of the frequencies of each strategy profile (s^1, s^2) (say by having a Dirichlet prior over the strategy profiles of the opponents) then she will learn to play Out, so that play converges to an outcome that is not a Nash nor even a correlated equilibrium. Moreover, player 3 is recognizing the correlation in the actions of players 1 and 2 that those players themselves ignore.

In our opinion, this example shows that there are problems with either formulation of multi-player fictitious play whenever play *in each period* fails to converge to a fixed strategy profile. This is an additional argument against using the convergence of the empirical distributions as the convergence criterion.

In the introduction, we used Jordan's [1993] simple three person matching pennies game to illustrate the idea of a correlated equilibrium. Here we show how fictitious play leads to a robust cycle in this game, similar to the best-response cycle we discussed in the introduction.

Recall that the game in question is a variant on matching pennies, where each player simultaneously chooses "H" or "T", and all entries in the payoff matrix are either +1 (win) or -1 (lose). Player 1 wins if he plays the same action as player 2, player 2 wins if he matches player 3, and player 3 wins by *not* matching player 1. The payoffs are

$$\begin{bmatrix} +1,+1,-1 & -1,-1,-1 \\ -1,+1,+1 & +1,-1,+1 \end{bmatrix} \quad \begin{bmatrix} +1,-1,+1 & -1,+1,+1 \\ -1,-1,-1 & +1,+1,-1 \end{bmatrix}$$

where the row corresponds to player 1 (up for H, down for T), the column to player 2, and the matrix to player three. This game has a unique Nash equilibrium, namely for each player to play (1/2,1/2), but we observed that the distribution over outcomes in which each of the profiles (H,H,H),(H,H,T),(H,T,T),(T,T,T),(T,T,H),(T,H,H) have equal weight of 1/6 is a correlated equilibrium.

Jordan specifies that players estimate a separate marginal distribution for each opponent as in two-player fictitious play, and that the players' assessments are the product of these marginals; this corresponds to the case where players believe their opponents randomize independently and moreover the subjective prior γ^i of each player is a product measure. However, each player cares only about the play of one of the two opponents, so the particular assumptions about how the opponents' play is correlated is irrelevant. Jordan improves on Shapley's example by providing a game where the empirical distributions fail to converge for all initial conditions outside of a one-dimensional stable manifold, thus showing that the failure to converge is very robust to the initial conditions. The cycle is similar to the best response cycle we discussed above: if players start with assessments that lead them to play (H,H,H), eventually player 3 will want to switch to T. After playing (H,H,T) for a long time, eventually player 2 will want to switch and so forth. As in Shapley's example, the cycle takes longer and longer as time goes on; however, unlike the Shapley example, the joint distribution of player's play converges to the correlated equilibrium putting equal weight on the 6 strategies the cycle passes through.[25] We will see below that it is not a coincidence that the cycle passes through profiles in the same order as the best response dynamic, nor that the joint distribution converges to a correlated equilibrium.

[25] This assertion is not meant to be obvious; we provide a proof in section 2.7.

2.6. Payoffs in Fictitious Play

In fictitious play, players keep track only of data about the frequency of opponents play. In particular, they do not keep track of data on conditional probabilities, and so may not recognize that there are cycles.[26] Given this limitation, , we can still ask whether fictitious play accomplishes its purpose. That is, if fictitious play successfully "learns" the frequency distribution, then it ought, asymptotically at least, yield the same utility that would be achieved if the frequency distribution is known in advance. This section examines the extent to which fictitious play satisfies this property, which we call "consistency."

In this section, we will suppose that if there are more than two players, players assessments track the joint distribution of opponents strategies. Denote the empirical distribution over player *i*'s opponents by D_t^{-i}

Denote the best payoff against the empirical distribution by $\hat{U}_i^t = \max_{\sigma^i} u^i(\sigma^i, D_t^{-i})$. Denote by $U_t^i = (1/t)\sum_{\tau=1}^{t} u^i(s_\tau^i, s_\tau^{-i})$ the time average of player *i*'s realized payoffs.

<u>Definition 2.2 :</u> Fictitious play is ε*-consistent along a history* if there exists a *T* such that for any $t \geq T$, $U_t^i + \varepsilon \geq \hat{U}_t^i$ for all *i*.

Note that this is an "ex-post" form of consistency, in contrast to consistency in the sense of classical statistics. Chapter 4 discusses behavior rules that are "universally ε-consistent" in the sense of being ε-consistent along every possible history.

It is useful to consider not only how well the player actually does against his opponents' empirical play, but also how well he does relative to the utility he expects to

[26] Of course they may keep track of this information; we discuss the consequences of this more fully in chapter 8 below.

get. Because the expected payoffs are linear in the probabilities, a player whose assessment over opponent's date-*t* actions is γ_t^i believes that his expected date-*t* payoff is $U_t^{i*} = \max_{\sigma^i} u^i(\sigma^i, \gamma_t^i)$; . Since the distance between γ_t^i and D_t^{-i} converges to 0 whether or not play converges, and the payoff function is continuous in the opponents' mixed strategies, $\left\| \hat{U}_i^t - U_i^* \right\|$ converges to 0 asymptotically. Consequently, consistency means that not only does a player do as well as he might if he knew the frequencies in advance, but he does as well as he expects to do. For example, if, as in the example of the last section, U_t^i remains less than $\hat{U}_i^t$, player *i* should eventually realize that something is wrong with his model of the environment. This provides an alternative motivation for the notion of consistency.

Our main result concerns the connection between how frequently a player changes strategies, and the consistency of his play. For any time *t* we define the *frequency of switches* η_t^i to be the fraction of periods τ in which $s_\tau^i \neq s_{\tau-1}^i$.

Definition 2.3 Fictitious play exhibits *infrequent switches along a history* if for every ε there exists a *T* and for any $t \geq T$ $\eta_t^i \leq \varepsilon$ for all *i*.

Proposition 2.4 If fictitious play exhibits infrequent switches along a history, then it is it is ε-consistent along that history for every $\varepsilon > 0$.

This result was established independently by Fudenberg and Levine [1994] and by Monderer, Samet, and Sela [1994]; we present the Monderer-Samet-Sela proof since it is shorter and more revealing.[27]

[27] Monderer, Sela and Samet only present the case of fictitious play with a null prior, so that the player's beliefs (at every period after the first one) exactly equal the empirical distribution, so that $\varepsilon = 0$ but their proof extends immediately to general priors.

Intuitively, once there is enough data to swamp the prior, at each date t player i’s action will be a best response to the empirical distribution through date t-1. On the other hand, if player i is not on average doing as well as the best response to the empirical distribution, there must be a nonnegligible fractions of dates t at which the action i chooses at date t is not be a best response to the distribution of opponents’ play through that date. But at such dates, player i will switch and choose a different action at date t+1; conversely, infrequent switches imply that most of the time i’s date t action is a best response to the empirical distribution at the end of date t.

In what follow it is convenient to let $\hat{\sigma}_t^i$ denote the argmax specified by the fictitious play.

Proof of Proposition 2.4:. Observe that

$$U_t^{i*} = u^i(\hat{\sigma}_t^i, \gamma_t^i) \geq u^i(\hat{\sigma}_{t+1}^i, \gamma_t^i) = \frac{\left((t+k+1)u^i(\hat{\sigma}_{t+1}^i, \gamma_{t+1}^i) - u^i(\hat{\sigma}_{t+1}^i, s_t^{-i})\right)}{(t+k)} =$$

$$\frac{\left((t+k+1)U_{t+1}^{i*} - u^i(\hat{\sigma}_{t+1}^i, s_t^{-i})\right)}{(t+k)}$$

so $U_{t+1}^{i\ *} \leq \frac{t+k}{t+k+1} U_t^{i*} + \frac{u^i(\hat{\sigma}_{t+1}^i, s_t^{-i})}{t+k+1}$, and

$$U_{t+1}^{i\ *} \text{-} U_t^{i*} \leq \frac{-1}{t+k+1} U_t^{i*} + \frac{u^i(\hat{\sigma}_{t+1}^i, s_t^{-i})}{t+k+1}.$$

Consequently,

$$U_t^{i*} \leq \frac{\sum_{\tau=0}^{t-1} u^i(\hat{\sigma}_{\tau+1}^i, s_\tau^{-i})}{t+k} + \frac{U_0^{i*}}{t+k}$$

$$= \frac{\sum_{\tau=0}^{t-1} u^i(\hat{\sigma}_\tau^i, s_\tau^{-i})}{t+k} + \frac{\sum_{\tau=0}^{t-1} u^i(\hat{\sigma}_{\tau+1}^i, s_\tau^{-i}) - u^i(\hat{\sigma}_\tau^i, s_\tau^{-i})}{t+k} + \frac{U_0^{i*}}{t+k}.$$

The first quotient in this expression converges to player *i*'s realized average payoff U_t^i. The second sums terms that are zero except when player *i* switches, and so it converges to 0 on any path with infrequent switches, and the third is an effect of the prior beliefs that converges to 0 along any path.

☑

Proposition 2.5: For any initial weights, there is a sequence $\varepsilon_t \to 0$ such that along any infinite horizon history, $U_t^{i^*} \geq U_t^i + \varepsilon_t$. That is, once there is enough data to outweigh the initial weights, players believe that their current period's expected payoff is at least as large as their average payoff to date.

Proof: Let $k = \sum_{s^{-i}} \kappa_0^i(s^{-i})$ be the length of the "fictitious history" implicit in the initial beliefs γ_0^i, and let $\hat{\sigma}_t^i$ denote a best response to γ_t^i. Then

$$U_t^{i^*} = u^i(\hat{\sigma}_t^i, \gamma_t^i) \geq u^i(\hat{\sigma}_{t-1}^i, \gamma_t^i) = \frac{\left(u^i(\hat{\sigma}_{t-1}^i, s_{t-1}^{-i}) + (t+k-1)u^i(\hat{\sigma}_{t-1}^i, \gamma_{t-1}^i)\right)}{(t+k)},$$

where the inequality comes from the fact that $\hat{\sigma}_t^i$ is a best response to γ_t^i. Expanding $u^i(\hat{\sigma}_{t-1}^i, \gamma_{t-1}^i)$ shows that shows that

$$U_t^{i^*} \geq \frac{u^i(\hat{\sigma}_{t-1}^i, s_{t-1}^{-i}) + (t+k-1)\left(u^i(\hat{\sigma}_{t-2}^i, s_{t-2}^{-i}) + (t+k-2)u^i(\hat{\sigma}_{t-2}^i, \gamma_{t-2}^i)\right)/(t+k-1)}{(t+k)} =$$

$$\frac{u^i(\hat{\sigma}_{t-1}^i, s_{t-1}^{-i}) + \left(u^i(\hat{\sigma}_{t-2}^i, s_{t-2}^{-i}) + (t+k-2)u^i(\hat{\sigma}_{t-2}^i, \gamma_{t-2}^i)\right)}{(t+k)};$$

proceeding iteratively shows that

$$U_t^{i^*} \geq \frac{\sum_{\tau=1}^{t-1} u^i(\hat{\sigma}_\tau^i, s_\tau^{-i}) + ku^i(\hat{\sigma}_1^i, \gamma_1^i)}{t+k} = \frac{tU_t^i + ku^i(\hat{\sigma}_1^i, \gamma_1^i)}{t+k}.$$

Taking $\varepsilon_t > (1/t)\max u^i$ completes the proof. Note that this proof does not use the "infrequent switches" property.

☑

The Fudenberg and Kreps example above fails the infrequent switch test because players change their strategies every period. On the other hand, the non-convergent paths in both the Shapley and in the Jordan examples are easily seen to have infrequent switches. Moreover, as noted by Monderer, Samet and Sela, Proposition 2.5 can be used to provide an easy proof that the empirical distributions do not converge in those examples. In the Shapley cycle, for instance, the sum of the realized payoffs is 1 in every period, so that by Proposition 25 the sum of the payoffs U^{*i}_t that the players expect to receive at least 1 for large *t*. On the other hand, if the empirical distributions were to converge they would need to converge to the Nash equilibrium distributions (from Proposition 2.2); thus the players' beliefs would converge to the Nash equilibrium distributions as well, and so their expected payoffs would converge to the Nash equilibrium payoffs, which sum to 2/3.

2.7. Consistency and Correlated Equilibrium in 2 Strategy Games

Because in the Jordan game each player has only two actions, consistency has an interesting consequence: it implies that the long-run average of action profiles resembles a correlated equilibrium.

Specifically, suppose that the outcome of play is ε*-consistent* in the sense of the previous section, with $\varepsilon = 0$. Let $D_t^{-i}[s^i]$ denote the distribution over the play of *i*'s opponents derived from the joint distribution over profiles D_t by conditioning on player *i* playing s^i, and let recall that d_t^i is the marginal distribution of i's play. In particular $D_t^{-i} = \sum_{s^i} d_t^{\,i}(s^i) D_t^{\,-i}[s^i]$. Note also that $\max_{\sigma_i} u^i(\sigma_i, D_t^{-i}) = \max_{s_i} u^i(s_i, D_t^{-i})$.

Consistency, then, is equivalent to the condition that $u^i(D_t) \geq u^i(s_i, D_t^{-i})$ for all s^i. Supposing that player *i* has only two actions, we may write

$$\begin{gathered} d_t^i(s^i)u^i(s^i, D_t^{-i}[s^i]) + d_t^i(r^i)u^i(r^i, D_t^{-i}[r^i]) = u^i(D_t) \geq \\ u^i(s^i, d_t^i(s^i)D_t^{-i}[s^i] + d_t^i(r^i)D_t^{-i}[r^i]) = \\ d_t^i(s^i)u^i(s^i, D_t^{-i}[s^i]) + d_t^i(r^i)u^i(s^i, D_t^{-i}[r^i]) \end{gathered}$$

from which we conclude by subtraction that

$$d_t^i(r^i)u^i(r^i, D_t^{-i}[r^i]) \geq d_t^i(r^i)u^i(s^i, D_t^{-i}[r^i])$$

for all s^i. This says that r^i is a best response to the conditional distribution of opponents actions given r^i whenever it is played with positive probability. We conclude that whenever each player has only two strategies, if D_t is consistent, it is a correlated equilibrium as well.

2.8. Fictitious Play and the Best Response Dynamic

We observed in the Jordan example that the sequence of pure strategy profiles generated by fictitious play (but not the number of times of each profile occurs) is the same as that in the alternating-move best response dynamic. (This is true also in the Shapley example.) It is easy to see that these two processes cannot be the same in general games, for under the alternating-move best response dynamic players only choose strategies that are best replies to some *pure* strategy of their opponents, while under fictitious play a player may choose a strategy that is not a best response to any pure strategy profile, but is a best response to some mixed opponents' strategies. However, the asymptotic behavior of fictitious play is closely related to the asymptotic behavior of the "partial best response dynamic" we discussed in the introduction, if the fraction of the population that adjusts its play each period is sufficiently small.

Let us suppose that there is a continuum population of each type of player, and take the state variable θ_t^i to be the frequency distribution over strategies played by type *i*. That is, $\theta_t^i(s^i)$ is the fraction of type *i*'s that are playing s^i. In discrete time, if a fraction of the population λ is picked at random switch to the best-response to their opponents current play, and the rest of the population continues their current play, the partial best-response dynamic is given by

$$\theta_{t+1}^i = (1-\lambda)\theta_t^i + \lambda BR^i(\theta_t^{-i}) = \theta_t^i + \lambda(BR^i(\theta_t^{-i}) - \theta_t^i),$$

where each BR^i is a (discontinuous) function corresponding to some selection from the best response correspondence for that player. If the time periods are very short, and the fraction of the population adjusting is very small, this may be approximated by the continuous time adjustment process

$$\dot{\theta}_t^i = \beta(BR^i(\theta_t^{-i}) - \theta_t^i).$$

Notice that this dynamic is time homogeneous, as is the discrete-time version of the process, so that the former does not "converge" to the latter as the process evolves; the shift from the discrete-time system to the continuous-time one was justified by considering a change in the system's underlying parameters.

In contrast, the fictitious play process moves more and more slowly over time, because the ratio of new observations to old observations becomes smaller and smaller. Suppose that when there are more than two players, population averages of opponents are viewed as independent, so that asymptotically beliefs γ_t^i are approximately given by the product of marginal empirical distributions $\prod_{j \neq i} d_t^j$. Recalling that d_{t-1}^{-i} is the vector of marginal empirical distributions of the play of players other than player *i*, Then the marginal empirical distributions in fictitious play evolve approximately (ignoring the prior) according to

$$d_t^i = \frac{t-1}{t} d_{t-1}^i + \frac{1}{t} BR^i(d_{t-1}^{-i}) .$$

This is of course very much like the partial best-response dynamic, except that the weights on the past is converging to one, and the weight on the best response to zero.

Moreover, we can make the fictitious play system seem stationary by changing the units in which time is measured. Specifically, let $\tau = \log t$, or $t = \exp \tau$. Suppose that there are infrequency switches, so that play remains more or less constant between τ and $\tau + \Delta$. Observing that

$$\exp(\tau + \Delta) - \exp \tau = (\exp(\Delta) - 1) \exp \tau \approx \Delta \exp \tau = \Delta t ,$$

and letting $\tilde{d}_\tau^i = d_{\exp \tau}^i$ we may write

$$\begin{aligned} \tilde{d}_{\tau+\Delta}^i = d_{t+\Delta t}^i &= \frac{t - \Delta t}{t} d_t^i + \frac{\Delta t}{t} BR^i(d_t^{-i}) \\ &= (1-\Delta) d_t^i + \Delta BR^i(d_t^{-i}) \\ &= (1-\Delta) \tilde{d}_\tau^i + \Delta BR^i(\tilde{d}_\tau^{-i}) \end{aligned} .$$

In the continuous time limit for large *t* and small Δ this may be approximated by

$$\dot{\tilde{d}}_\tau^i = BR^i(\tilde{d}_\tau^{-i}) - d_\tau^i ,$$

which is of course the same as the continuous time partial best-response dynamic.

The conclusion is that with an appropriate time normalization, discrete-time fictitious play asymptotically is approximately the same as the continuous time best-response dynamic. More precisely, the set of limit points of discrete time fictitious play is an invariant subset for the continuous time best- response dynamics, and the path of the discrete-time fictitious play process starting from some large *T* remains close to that of the corresponding continuous time best- response process until some time $T + T'$, where T'

can be made arbitrarily large by taking T arbitrarily large (Hofbauer [1995]). However, starting from the same initial condition the discrete and continuous time processes can tend towards different long-run limits. This sort of relationship between discrete-time and continuous-time solutions is fairly typical, and recurs in the study of the replicator dynamic, as we will see in the next chapter. Subsequently, we will be considering fictitious play type systems with noise. The theory of stochastic approximation studies the exact connection between discrete time stochastic dynamical systems and their continuous time limits, and we will have much more to say on this subject in chapter 4.

2.9. Generalizations of Fictitious Play

As one might expect, many of the results about the asymptotic behavior of fictitious play continue to hold for processes that might prescribe different behavior in the early periods but "asymptotically converge" to fictitious play. Following Fudenberg and Kreps [1993], we say that the player's beliefs are "*asymptotically empirical*" if $\lim_{t\to\infty}\left\|\gamma_t^i - D_t^{-i}\right\| = 0$ along every sequence of infinite histories. It is easy to verify that Proposition 2.1 continues to hold (strict equilibria are absorbing, and pure strategy steady states are Nash) when fictitious play is generalized to allow any asymptotically empirical forecasts. If, in addition, γ_t^i are the product of marginal beliefs, Fudenberg and Kreps show that Propositions 2.2 (convergence of the marginal distributions of play implies Nash) continues hold when fictitious play is generalized to allow any asymptotically empirical beliefs.[28] In a similar vein, Jordan [1993] shows that his example of non-convergent empirical distributions converges to the same limit cycle for any beliefs that (i)

[28] As the example in section 2.5 showed, in the multi-player case, if it is not *a priori* thought that opponents play independently, this result can fail even for fictitious play.

satisfy a "uniform" version of asymptotic empiricism (that is, for any ε the distance between the forecasts and the empirical distribution should become less than ε for *all* histories of length at least some $T(\varepsilon)$) and also (ii) depend on the history only through the empirical distribution.

Milgrom and Roberts [1991] use a still weaker condition on forecasts: they say that forecasts are *adaptive* if the forecasts assign very low probability to any opponent's strategy that has not been played for a long time. Formally, a forecast rule is adaptive if for every ε>0 and for every *t*, there is a there is $T(\varepsilon,t)$ such that for all *t'*> $T(\varepsilon,t)$ and all histories of play up to date t', the forecast γ_t^i assigns probability no more than ε to the set of pure strategies of *i*'s opponent that were not played between times *t* and t'. Since this condition restricts only the support of the forecasts, but not the relative weights of strategies in the support, it is clearly inappropriate for modeling situations where the player's forecasts converge to a mixed distribution. However, as Milgrom and Roberts show, the condition is strong enough to preserve the second part of proposition 2.1: if forecasts are adaptive, and play converges to a pure strategy profile, that profile must be a Nash equilibrium.

One example of a an adaptive forecasting rule is "exponentially weighted fictitious play," under which the forecast probability of strategy s^j at date *t* is $\frac{1-\beta^{\tau-1}}{1-\beta}\sum_{\tau=1}^{\tau-1}\beta^{\tau}I(s_\tau^j = s^j)$, where I is the indicator function and $\beta > 1$. With this rule, the weight given to the most recent observation never vanishes, so that if the opponent's play is a fixed mixed strategy then the assessments do not converge. (If the weight β is allowed to shrink to 0 as t goes to infinity, perhaps to reflect the greater weight given a lengthier past history, then the rule is asymptotically empirical.) We will discuss the properties of this type of exponential weighting scheme below in chapter 4, along with some evidence that exponential weights do a better job than standard fictitious play of describing learning behavior in the experimental play of games. Since these experiments are not run for very large horizons the implications of this for our purposes are not clear.

Finally, many asymptotic properties of fictitious play game are preserved if the assumption that actions are chosen to maximize current payoffs given current forecasts is weakened to only hold asymptotically. Fudenberg and Kreps say that a behavior rule ρ^i (a map from histories to actions) is *strongly asymptotically myopic with respect to forecast rule* γ^i if for some sequence of positive numbers $\{\varepsilon_t)$ converging to 0, for every *t* and every time-*t* history, every pure strategy s^i that has positive probability under ρ^i is an ε_t-optimal response given forecast γ^i.[29] Propositions 2.1 holds for all behavior rules that are strongly asymptotically myopic with respect to some asymptotically empirical forecast rule, and Proposition 2.2 holds if beliefs are the product of independent marginals..

[29] That is, each s^i in the support must satisfy $u^i(s^i,\gamma^i)+\varepsilon_t \geq \max_{s'^i} u^i(s'^i,\gamma^i)$.

Appendix: Dirichlet Priors and Multinomial Sampling

Our summary follows DeGroot [1970].

1) *The Multinomial Distribution:* Consider a sequence of *n* i.i.d. trials, where each period one of *k* outcomes occurs, with p_z denoting the probability of outcome *z*. Denote the outcome of the *n* trials by the vector κ, where κ_z is the number of the outcomes of type *z.* (Think of the outcomes as being the opponent's choice of an action in a simultaneous-move game.) Then the distribution of the κ's, called the *multinomial distribution with parameters n and* $p=(p_1,...,p_k)$, is given by $f(\kappa)=\dfrac{n!}{\kappa_1!\cdots\kappa_k!}p_1^{\kappa_1}\cdots p_k^{\kappa_k}$ for κ such that $\sum_{z=1}^{k}\kappa_z=n$.

2) *The Dirichlet Distribution:* Let Γ denote the gamma function. A random vector *p* has the Dirichlet distribution with parameter vector α ($\alpha_z>0\forall z$) if its density is given by

$$f(p)=\frac{\Gamma(\alpha_1+...+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}p_1^{\alpha_1-1}\cdots p_k^{a_k-1}$$

for all *p*>0 such that $\sum_{z=1}^{k}p_z=1$. This is sometimes called the multivariate beta distribution, because if *p* has a Dirichlet distribution, the marginal distribution of p_z is the beta distribution with parameters α_z and $\sum_{w\neq z}\alpha_z$ In particular, if *p* has the Dirichlet distribution, the expected value of p_z is $\alpha_z/\sum_{w=1}^{k}\alpha_w$.

3) *The Dirichlet Distributions are a Conjugate Family for Multinomial Sampling:* A family of distributions is said to be a conjugate family for a likelihood function if whenever the prior distribution lies in the family, the posterior distribution will lie in the same family for any sample drawn according to the specified form of likelihood function. One classic example is the normal distribution: if samples are drawn according to a normal distribution

with unknown mean, and the prior is itself a normal distribution, then the posterior distribution will also be a normal distribution. Likewise, the Dirichlet distribution is a conjugate family for multinomial sampling.

To see this, suppose that the prior distribution over the probability vector *p* has a Dirichlet distribution with parameter α, so that the density function *f(p)* at each *p* is proportional to

$$\prod_{z=1}^{k} p_z^{\alpha_z - 1}.$$

For each value of *p*, the likelihood of the vector κ of outcomes is proportional to

$$\prod_{z=1}^{k} p_z^{\kappa_z}.$$

To compute the posterior distribution over *p*, we use Bayes rule:

$$f(p|\kappa) = \frac{f(\kappa|p)f(p)}{\int f(\kappa|p)f(p)dp} \propto \prod_{z=1}^{k} p_z^{\alpha_z - 1} \prod_{z=1}^{k} p_z^{\kappa_z} = \prod_{z=1}^{k} p_z^{\alpha_z + \kappa_z - 1},$$

so that the posterior is Dirichlet with parameter α‘, where $\alpha_z' = \alpha_z + \kappa_z$.

If player *i*’s date-t beliefs about -*i’s* mixed strategy have a Dirichlet distribution, player *i’s* assessment of the probability that -*i* plays s^{-i} in period *t* is

$$\gamma_t^i(s^{-i}) = \int_{\Sigma^{-i}} \sigma^{-i}(s^{-i})\mu_t^i[d\sigma^{-i}],$$

which is simply the expected value of the component of σ^{-i} corresponding to s^{-i}; from our remark above if $z = s^{-i}$ this is just $\alpha_z / \sum_{w=1}^{k} \alpha_w$. Therefore, after observing sample κ, player *i*’s assessment of probability that the next observation is strategy *z* is

$$\frac{\alpha_z'}{\sum_{w=1}^{k} \alpha_w'} = \frac{\alpha_z + \kappa_z}{\sum_{w=1}^{k} (\alpha_w + \kappa_w)},$$

which is the formula for fictitious play.

References

Brown, G. W. [1951]: “Iterative Solutions of Games by Fictitious Play,” In *Activity Analysis of Production and Allocation*, Ed. T.C. Koopmans, (New York: Wiley).

Degroot, M. [1970]: *Optimal Statistical Decisions*, (New York: McGraw-Hill).

Ellison, G. [1994]: “Learning with One Rational Player,” MIT.

Fudenberg, D. and D. K. Levine [1995]: “Consistency and Cautious Fictitious Play,” *Journal of Economic Dynamics and Control*, 19 : 1065-1090.

Fudenberg, D. and D. Kreps [1990]: “Lectures on Learning and Equilibrium in Strategic-Form Games,” CORE Lecture Series.

Fudenberg, D. and D. Kreps [1993]: “Learning Mixed Equilibria,” *Games and Economic Behavior*, 5: 320-367.

Hofbauer, J. [1995]: “Stability for the Best Response Dynamic,” University of Vienna.

Jordan, J. [1993]: “Three Problems in Learning Mixed-Strategy Equilibria,” *Games and Economic Behavior*, 5: 368-386.

Krishna, V. and T. Sjostrom [1995]: “On the Convergence of Fictitious Play,” Harvard University.

Milgrom, P. and J. Roberts [1991]: “Adaptive and Sophisticated Learning in Repeated Normal-Form Games,” *Games and Economic Behavior*, 3: 82-100.

Miyasawa, K. [1961]: “On the Convergence of Learning Processes in a 2x2 Non-Zero-Person Game,” Princeton University Research Memo #33.

Monderer, D. and A. Sela [1996]: “A 2 ×2 Game without the Fictitious Play Property,” *Games and Economic Behavior*, 68: 200-211.

Monderer, D., D. Samet and A. Sela [1994]: “Belief Affirming in Learning Processes,” Technion.

Nachbar, J. [1990]: “’Evolutionary’ Selection Dynamics in Games: Convergence and Limit Properties,” *International Journal of Game Theory*, 19: 59-89.

Robinson, J. [1951]: “An Iterative Method of Solving a Game,” *Annals of Mathematics*, 54: 296-301.

Shapley, L. [1964]: “Some Topics in Two-Person Games,” In *Advances in Game Theory*, Ed. M. Drescher, L.S. Shapley, and A.W. Tucker,, (Princeton: Princeton University Press).

3. The Replicator Dynamics and Related Deterministic Models of Evolution

3.1. Introduction

At this point we shift from models that are explicitly based on learning to models based on the idea of evolution. These models originated in the field of evolutionary biology, but such models have become very popular among game theorists in the last few years.[30]

There are three main reasons for this interest. First, although the archetypal evolutionary model, that of the replicator dynamics, was originally motivated by a (simplified version of) biological evolution, the process can also describe the result of some types of "emulation" by economic agents. Second, some of the properties of the replicator dynamic extend to various classes of more general processes that may correspond to other sorts of learning or emulation. Finally, the study of evolutionary models has proved helpful (if controversial) in understanding animal behavior, and while this does not imply that the models have economic interest, it is still an interesting use of the theory of games.

Our discussion begins with the two concepts that have proven central to the study of evolutionary models: the replicator dynamic and the idea of an Evolutionary Stable Strategy or ESS. Section 3.2 begins with the case of a homogeneous population. The replicator dynamic assumes that population playing a particular strategy grows in proportion to how well that strategy is doing relative to the mean population payoff. Every

[30] See for example the symposium issues in the *Journal of Economic Theory* [1992] and *Games and Economic Behavior* [1993].

Nash equilibrium is a steady state in the replicator dynamic, and every stable steady state is a Nash equilibrium. The major question posed in this literature is the extent to which the stability of a steady state leads to a refinement of Nash equilibrium. One major result is that in a homogeneous population a stable steady state must be isolated and trembling hand perfect.

Section 3.4 introduces the notion of an ESS, which is a static concept that was inspired by, but not derived from, considerations of evolutionary dynamics. ESS requires that the strategy be robust when it is invaded by a small population playing a different strategy. Every ESS is Nash, so that ESS is a refinement of Nash equilibrium. One goal of the literature on evolution is to establish more closely the connection between replicator (and related) dynamics and the ESS concept. In the homogeneous population case, an ESS is stable in the replicator dynamic, but not every stable steady state is an ESS.

After examining the homogeneous case, we turn to the case of a heterogeneous population and the asymmetric replicator dynamic in section 3.5. One major result is that in the asymmetric case mixed profiles cannot be asymptotically stable.

Our interest in this book is primarily about the consequences of learning by individual players. Evolutionary models are generally cast in terms of the behavior of an entire population, and are vague about the individual behavior that leads to this population dynamic. However, the work discussed in section 3.6 shows that it is possible to give stories of learning that lead to replicator-like dynamics. One such story we consider is the emulation dynamic in which new player asks an old player what strategy that player used and how well it did. This leads to a model in which it is the deviation from the median rather than mean that determines how rapidly the population playing a strategy grow. There is also a reinforcement model of learning that leads to a dynamic closely related to the replicator. We introduce this model here, but postpone a discussion of its merits to chapter 4, so that we can compare it to variations on fictitious play.

The replicator dynamic is very specific and may not be a good description of many economic situations. Indeed, the learning models that lead to the replicator, can lead also to other “replicator-like” dynamics, in addition to the replicator itself. As a result, much attention has focused on the extent to which results obtained for the replicator dynamic extend to other dynamics with a more concrete economic foundation. Section 3.8 discusses the class of monotonic processes, which incorporate various versions of the idea that strategies that do better grow faster. A weak version of monotonicity is sufficient to assure that strategies that are strictly dominated by pure strategies can be iteratively eliminated; under the stronger condition of convex monotonicity, this conclusion extends to all strictly dominated strategies.

Section 3.8 discusses another generalization of the replicator dynamic called myopic adjustment. This class of processes, which includes the best response dynamic, also is sufficiently strong to yield useful results. In 2x2 symmetric games with a single population if there is a unique mixed strategy equilibrium it is stable. If there is a mixed strategy equilibrium but there are also two pure strategy equilibria, the two pure strategy equilibria are stable.

In addition to point-valued equilibrium notions, it can also be interesting to consider set-valued stability notions such as strategic stability and their relationship to components of steady states in the evolutionary dynamics. One useful result is that attractors in the myopic adjustment process (which may or may not exist) contain a strategically stable set, a rather strong refinement.

Section 3.9 examines the relationship between unmodeled “drift” or “mutation” terms and the possibility that equilibria (or equilibrium components) that do not satisfy strong refinements may nevertheless persist. Section 3.10 examines a set valued concept of stability due to Gilboa and Matsui [1991] that effectively incorporates the idea of drift, and shows how it can be used to eliminate certain mixed strategy equilibria in cheap-talk games.

Most of this chapter, like most of the evolutionary literature, considers continuous-time dynamical systems. Section 3.11 examines discrete-time versions of the replicator dynamic. Unlike the continuous time version, the discrete-time dynamic need not remove dominant strategies. In addition, where the continuous-time dynamic has a center, the discrete-time dynamic instead cycles outward to the boundary.

3.2. The Replicator Dynamics in a Homogeneous Population

Much of the work on evolution has studied the case of a single homogenous population playing a symmetric stage game, so we begin our discussion with this case; we consider models of asymmetric populations later on in this chapter. The most basic evolutionary model is the replicator dynamic. Our goal in this section is to define this dynamic and discuss how it might be interpreted; we will also see how the state states of the dynamic relate to the set of Nash equilibria.

To define the replicator dynamic, suppose that all agents use pure strategies, and specialize to a homogenous population. Let $\phi_t(k)$ be the measures of the set of players using pure strategy s at date *t*; let $\theta_t(s) = \frac{\phi_t(s)}{\sum_{s'} \phi_t(s')}$ be the fraction of players using pure strategy *s* at date *t*, and let the state variable θ_t be the vector of population fractions. Then the expected payoff to using pure strategy *s* at date *t* is $u_t(s) \equiv \sum_{s'} \theta_t(s') u(s, s')$, and the average expected payoff in the population is $\bar{u}_t = \sum_s \theta_t(s) u_t(s)$.

Suppose that each individual is genetically programmed to play some pure strategy, and that this programming is inherited. Finally, suppose that the net reproduction rate of each individual is proportional to its score in the stage game. This leads to the following continuous-time dynamic system:

(3.1) $\dot{\phi}_t(s) = \phi_t(s)u_t(s)$, which implies

(3.2) $$\dot{\theta}_t(s) = \frac{\dot{\phi}_t(s)\sum_{s'}\phi_t(s') - \phi_t(s)\sum_{s'}\dot{\phi}_t(s')}{\left(\sum_{s'}\phi_t(s')\right)^2} = \theta_t(s)[u_t(s) - \overline{u}_t].$$

Equation (3.1) says that strategies with negative scores have negative net growth rates; if all payoffs are negative, the entire population is shrinking. There is no problem with this on the biological interpretation; in economic applications we tend to think of the number of agents playing the game as being constant. But note that even if payoffs are negative, the sum of the population shares is always 1. Note also that if the initial share of strategy *s* is positive, then its share remains positive: the share can shrink towards 0, but 0 is not reached in finite time.

Notice that the population share of strategies that are not best responses to the current state can grow, provided that these strategies do better than the population average. This is a key distinction between the replicator dynamic and the best-response dynamic, and also distinguishes the replicator dynamic from fictitious play. Despite this ability of sub-optimal strategies to increase their share, there is still a close connection between steady states of the replicator dynamic and Nash equilibria. First, every Nash equilibrium is a steady state: in (the state corresponding to a) Nash equilibrium, all strategies being played have the same average payoff, so the population shares are constant. Unfortunately, steady states need not be Nash equilibria: Any state where all agents use the same strategy is a steady state, since the dynamic does not allow the “entry” of strategies that are “extinct”. However, if a profile is not Nash, it cannot be stable: if a small fraction of an improving deviation is introduced, it will grow.

Proposition 3.1: A stable steady state of the replicator dynamics is a Nash equilibrium; more generally, any steady state that is the limit of a path that originates in the interior is a Nash equilibrium. Conversely, for any non-Nash steady state there is a $\delta > 0$ such that all interior paths eventually move out of a δ-neighborhood of the steady state.

Proof: Suppose $\theta*$ is a steady state, but the corresponding strategy profile $\sigma*$ is not a Nash equilibrium. Then, since payoffs are continuous, there exists a pure strategy $s \in \text{support}(\sigma^*)$, a pure strategy s' and an $\varepsilon > 0$ such that $u(s',\sigma^*) > u(s,\sigma^*) + 2\varepsilon$. There is, moreover, a δ such that $u(s',s'') > u(s,s'') + \varepsilon$ for all s'' within δ of σ^*. Hence if there is a path that remains in a δ-neighborhood of σ^*, the growth rate of strategy s' exceeds that of strategy s by an amount that is bounded away from zero. Thus the share of strategy s must converge to 0, which is a contradiction.

☑

Note that this argument does not rely on the special structure of the replicator dynamics; it suffices that the growth rates are a strictly increasing function of the payoff differences. We discuss below another property that the replicator dynamics shares with a broad range of dynamic processes, namely the elimination of dominated and iterated elimination of dominated strategies.

3.3. Stability in the Homogeneous-Population Replicator Dynamic

We have already seen that stable steady states of the replicator must be Nash. We now examine dynamic stability more closely with an eye to answering the following questions: does stability in the replicator dynamic refine Nash equilibrium? This is, can we narrow down the range of Nash equilibria through stability arguments? Does the replicator dynamic necessarily converge to a steady state? That is, are stable steady states the only possible long-run outcomes with the replicator dynamic? We shall see that while it is possible to refine Nash equilibrium through stability arguments, it is also possible that the replicator does not converge to a steady state at all.

We begin with an example of an asymptotically stable steady state.

Example 3.1: Consider the game

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

This game has two asymmetric Nash equilibria, (A,B) and (B,A), and the mixed equilibrium where both players randomize 1/2-1/2. Note that since a homogenous population is assumed, there is no way to converge to the asymmetric equilibria, because there are not separate populations of “player 1's” and “player 2's.” So the only possible steady state is the mixed equilibrium where all players randomize (1/2,1/2). Moreover, this mixed profile is a steady state even though no individual player uses a mixed strategy: when 1/2 of the population uses strategy A, and 1/2 uses strategy B, from an individual player's viewpoint the distribution of opponents' play looks like a mixed strategy. Furthermore, it is easy to check that the mixed strategy equilibrium is asymptotically stable: when fraction $\theta(A)$ of the population plays A, the payoff to A is $\theta(B)=1-\theta(A)$, while the payoff to B is $\theta(A)$. Consequently the average payoff in the population is $2\theta(A)(1-\theta(A))$. Substituting into the replicator equation, we have the one-dimensional system

$$\dot{\theta}_t(A) = \theta_t(A)[(1-\theta_t(A) - 2\theta_t(A)(1-\theta_t(A))] = \theta_t(A)[1-3\theta_t(A) + 2\theta_t(A)^2];$$

this expression is positive for $\theta_t(A) < 1/2,$ exactly 0 at 1/2., and negative at larger values, so that the Nash equilibrium is asymptotically stable. (We will se below that the equilibrium is a saddle when there are distinct populations of player 1's and player 2's.) .

Proposition 3.2 [Bomze 1986]: An asymptotically stable steady state in the homogenous-population replicator dynamic corresponds to a Nash equilibrium that is trembling-hand perfect and isolated.

This result shows that asymptotic stability will be hard to satisfy in games with a non-trivial extensive form, for such games typically have connected sets of equilibria that differ only in their off-path play. For this reason, evolutionary concepts need some modification to be applied to extensive form games: either a set-valued notion of stability must be used, (as for example in Swinkels [1993]) or the model is perturbed with "trembles" so that all information sets have positive probability of being reached.

As with most dynamical systems, there is no guarantee that the replicator dynamics converge, and indeed, there are examples of games with no asymptotically stable steady states. In particular, even a totally mixed equilibrium need not be asymptotically stable. This is significant, because totally mixed equilibria satisfy the standard "equilibrium refinements" based on trembles, including such strong notions as Kohlberg and Mertens [1986] stability. A simple example in which there is no asymptotically stable steady state is the game "rock-scissors-paper:"

Example 3.2:

	R	S	P
R	0,0	1,-1	1,1
S	-1,1	0,0	1,-1
P	1,-1	-1,1	0,0

This may be reduced to a two-dimensional system by substituting $\theta(P) = 1 - \theta(R) - \theta(S)$. Making use of the fact that the average payoff is 0 at every state (because this is a zero-sum game) the resulting replicator dynamics are given by

$$\begin{aligned}\dot{\theta}_t(R) &= \theta_t(R)[2\theta_t(S) + \theta_t(R) - 1] \\ \dot{\theta}_t(S) &= \theta_t(S)[-2\theta_t(R) - \theta_t(S) + 1]\end{aligned}$$

Linearizing at the equilibrium (1/3,1/3) we find that the Jacobian is

$$\begin{bmatrix} 1/3 & 2/3 \\ -2/3 & -1/3 \end{bmatrix}$$

The eigenvalues of this matrix are the solutions of $(1\text{-}3\lambda)(\text{-}1\text{-}3\lambda) + 4 = 0$, or $9\lambda^2 + 3 = 0$, and hence have zero real part. This means that the system is degenerate; it turns out that the steady state is surrounded by closed orbits, so that it is a “center,” and hence is stable but not asymptotically stable.[31] . The phase portrait for this system is illustrated below.

[31] Because the system is degenerate, this cannot be proved by only examining the linearized system

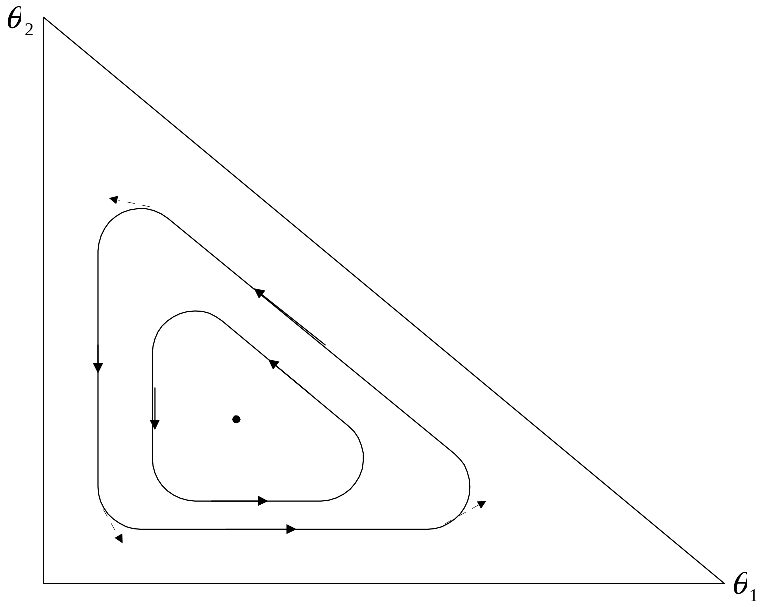


Figure 3.1

Since the 0 real part of the eigenvalue (which means the steady state is not "hyperbolic") is a knife-edge case, we know that there are small changes to the system (that is, small changes to the flow or vector field) that gives the eigenvalues a positive real part. It turns out that such a change can be made simply by changing the payoffs slightly, so that each strategy gets a small $\varepsilon > 0$ when matched against itself.[32] Then the unique Nash equilibrium is still (1/3,1/3,1/3), but now the Nash equilibrium is an unstable steady state, and the trajectories of the replicator dynamics spiral outwards towards the boundary of the simplex without reaching it, as shown by the dashed lines in the figure above.[33] Conversely, the Nash equilibrium is an asymptotically stable steady state for small $\varepsilon < 0$.

[32] The fact that non-hyperbolic steady states are not robust to general perturbations of the dynamics does not imply that they can be destroyed by small changes in the payoffs. Indeed, in asymmetric-population models of 2x2 games, there are centers for a range of payoff values, as we will see in Section 3.5.

[33] This was first shown by Zeeman [1980]. See Hofbauer and Sigmund [1988] and Gaunersdorfer and Hofbauer [1995] for more discussion of variants of this game. If the system is modified by a deterministic flow of "mutants" or new entrants, so that the boundary becomes repelling, then for small $\varepsilon > 0$ there is a cycle (closed orbit) from the Poincare-Bendixson theorem. See Boylan [1994] for a discussion of the properties of steady states that are robust (in the sense of an essential fixed point) to perturbations of the dynamics corresponding to such deterministic mutations.

3.4. Evolutionarily Stable Strategies

In applications, instead of working with explicit evolutionary dynamics, analysts often use the static concept of an *evolutionary stable strategy* or ESS. The idea of ESS is to require that the equilibrium be able to “repel invaders.” After defining the static notion of ESS, and showing that it is a refinement of Nash equilibrium, our goal is to relate it to the evolutionary dynamic. We will find that while every ESS is stable in the replicator dynamic, not every stable steady state needs to be an ESS.

To explain what is meant by “repelling invaders”, suppose that the population is originally at some profile σ, and then a small ε of "mutants" start playing σ'. ESS asks that the existing population gets a higher payoff against the resulting mixture $(1-\varepsilon)\sigma+\varepsilon\sigma'$ than the mutants do. Specifically we require that

(3.3) $$u(\sigma,(1-\varepsilon)\sigma+\varepsilon\sigma') > u(\sigma',(1-\varepsilon)\sigma+\varepsilon\sigma')$$

for all sufficiently small positive ε.

Using linearity of expected utility in probabilities, (3.3) is equivalent to:

$$(1-\varepsilon)u(\sigma,\sigma) + \varepsilon u(\sigma,\sigma') > (1-\varepsilon)u(\sigma',\sigma) + \varepsilon u(\sigma',\sigma').$$

Since this need only hold for ε close to 0, it is equivalent to requiring that for all $\sigma' \neq \sigma$, either

(3.4) $$u(\sigma,\sigma) > u(\sigma',\sigma) \text{ or}$$

(3.5) $$u(\sigma,\sigma) = u(\sigma',\sigma) \text{ and } u(\sigma,\sigma') > u(\sigma',\sigma').$$

There is also a weaker notion of evolutionary stability: Profile σ is a *weak ESS* if every $\sigma' \neq \sigma$ either satisfies (3.4) or satisfies

(3.5') $$u(\sigma,\sigma) = u(\sigma',\sigma) \text{ and } u(\sigma,\sigma') \geq u(\sigma',\sigma').$$

This is a weaker condition because it allows for the case where the invader does just as well against the prevailing population as the population itself; then the invader is not driven out, but it does not grow, either.

Notice first that, an ESS must be a Nash equilibrium; otherwise the first term on the left-hand side of (3.4) is smaller than the first term on the right. Also, any strict Nash equilibrium is an ESS. This follows since strict equilibria by definition satisfy (3.4) for all other strategies. But many games fail to have strict equilibria, because, for example, mixed strategy equilibria can never be strict.

While mixed strategy equilibria can never be strict, they can, however, be ESS. Consider, in particular, example 3.1 above. In this example if both players play the same strategy both get 0; if they play different strategies, both get 1. The unique mixed equilibrium is ½ - - ½ , and if either player is playing this strategy, both players get an expected utility of ½. Examining (3.3), we see that if a player persists in playing ½-½ after an invasion, he gets ½. On the other hand, if he plays the invading strategy σ' he gets

$$u(\sigma',(1-\varepsilon)\sigma+\varepsilon\sigma')=(1-\varepsilon)u(\sigma',\sigma)+\varepsilon u(\sigma',\sigma')=(1-\varepsilon)(1/2)+\varepsilon u(\sigma',\sigma')$$

However, when both players play the same strategy, unless they play ½-½ they always get strictly less than ½. So $u(\sigma',(1-\varepsilon)\sigma+\varepsilon\sigma')$ is strictly less than ½ unless $\sigma'=\sigma$, so the definition of an ESS is satisfied.

A more significant set of examples shows how ESS can reduce the equilibrium set: these are the "cheap-talk" games introduced by Crawford and Sobel [1982]. In these games players are allowed a period of costless and non-binding communication prior to the actual play of the game. Consider for example the coordination game below:

	L	R
L	2,2	-100,0
R	0,-100	1,1

In this game the outcome (L,L) is efficient, but the strategy R is risk dominant, so there are some grounds for expecting the outcome to be (R,R).[34] Suppose next that players can communicate. In the simplest version, this means that there are two stages of play. In the first stage players simultaneously announce their “intended action” L or R; in the second stage they play the game. Talk is cheap in the sense that announcing an action has no direct effect at all on the realized payoffs, which depend only on the second-stage choices. Nevertheless, the ability to communicate can make a difference, since players can now signal their intention to play (L,L) by announcing this in the first stage round. Formally, the two-stage game has a different extensive form than the original one, and so in principle could lead to different conclusions. However, standard equilibrium notions such as subgame perfection, sequential equilibrium, or even Kohlberg-Mertens [1986] strategic stability generate the same predictions with or without cheap talk, in the sense that the sets of equilibrium payoffs and second-stage equilibrium outcomes of the two-stage game are the same as in the one-stage game without communication.

ESS on the other hand does suggest a tendency for meaningful communication to occur. This point was made by Robson [1990], Warneryd [1991], Kim and Sobel [1991], Blume, Kim and Sobel [1993] and Schlag [1993].[35] The outcome (R,R) and a signal that is not sent cannot be evolutionary stable, since mutants could invade the population using the unused signal as a “secret handshake” to indicate to one another their intention to play the

[34] See for example the stochastic perturbation results discussed in Chapter 5.

[35] There are also other models of why cheap talk may have meaning, as in Farrell [1986] or Rabin [1990].

L equilibrium. Such invaders would not suffer against the existing players, and would do even better when matched against one another.[36] Notice however, that no matter how many messages there are, it is an ESS if every message is sent with positive probability and all players play R regardless of the signal. This is a kind of "babbling" equilibrium; since every signal is already being sent with positive probability there is no way for mutants to send their secret handshake.[37] We should also note that ESS arguments (and other arguments we will consider below) are largely limited to pure coordination games; if players disagree about which equilibrium they would like to be at, then an ESS may easily fail to exist as mutants may well enter and wish to move to an equilibrium more favorable to them.

We turn next to the connection between ESS and the replicator dynamic.

Proposition 3.3 (Taylor and Jonker [1978]; Hofbauer et al [1979]; Zeeman [1980]): Every ESS is an asymptotically stable steady state of the replicator dynamics.

The example below shows that the converse need not be true.

Proof: To see that ESS implies asymptotic stability, suppose that σ is an ESS, and let $\sigma(s)$ denote the weight that σ assigns to the pure strategy *s*. Following the proof of Hofbauer and Sigmund [1988], we will show that the "entropy" function $E_\sigma(\theta) = \prod_s \theta_s^{\sigma(s)}$ is a strict local Lyapunov function at σ, that is, that *E* has a local (actually global here) maximum at σ and that it is strictly increasing over time along trajectories in some neighborhood of σ.

To see this, note that

[36] Similar ideas have been used to explain why evolution might tend to select efficient equilibria in repeated games; see for example Binmore and Samuelson [1992] and Fudenberg and Maskin [1990].

[37] Farrell's concept of "neologism-proofness" also supposes that a would-be deviator can always find a previously unsent message by constructing a "neologism," that is, a new message.

$$\frac{\dot{E}_\sigma}{E_\sigma} = \frac{d}{dt}(\log E_\sigma) = \sum_s \sigma(s)\frac{d\log(\theta(s))}{dt} = \sum_s \sigma(s)[u(s,\theta) - u(\theta,\theta)] = u(\sigma,\theta) - u(\theta,\theta).$$

Since σ is an ESS, it satisfies either inequality (3.4) or (3.5) Inequality (3.5) implies directly that the above expression is positive; (3.4) yields the same for all $\theta \neq \sigma$ in some neighborhood of σ by the continuity of *u* in its second argument. Hence *E* is an increasing function of time in this neighborhood as well. Finally, it is well known that *E* is maximized at σ (for example *E* is the likelihood function for multinomial sampling, and the maximum likelihood estimate equals the sample probabilities[38]) so that *E* is a strict local Lyapunov function at σ, and hence σ is asymptotically stable.

☑

The following example from van Damme [1987] shows that not every asymptotically stable steady state is an ESS.

Example 3.3: The payoff matrix is

$$\begin{bmatrix} (0,0) & (1,-2) & (1,1) \\ (-2,1) & (0,0) & (4,1) \\ (1,1) & (1,4) & (0,0) \end{bmatrix}$$

This game has a unique symmetric equilibrium, namely (1/3,1/3,1/3), with equilibrium payoff 2/3.[39] This equilibrium is not an ESS, since it can be invaded by the strategy (0,1/2,1/2), which has payoff 2/3 when matched with the equilibrium strategy, and payoff 5/4 when matched with itself. However, the Jacobian evaluated at the equilibrium is

[38] A direct proof can be given by verifying that *E* is concave and using Jensen’ s inequality.

[39] It also has asymmetric equilibria.

$$\begin{bmatrix} -1/9 & -1/9 & -4/9 \\ -7/9 & -4/9 & 5/9 \\ 2/9 & -1/9 & -7/9 \end{bmatrix}.$$

A computation shows that eigenvalues are -1/3 (twice) and -2/3, so that the equilibrium is asymptotically stable.

The fact that asymptotically stable steady states need not be ESS's is linked to the fact that the replicator dynamics only allow for the inheritance of pure strategies. Bomze shows that if the dynamics are modified so that mixed strategies can be inherited as well, then ESS is equivalent to asymptotic stability under the replicator dynamics. (Since this change in the dynamics does not change the definition of ESS, this statement is equivalent to saying that the change in dynamics renders unstable the non-ESS's that were stable previously.)

This raises the important issue of which replicator model, the pure or mixed strategy model, is more interesting. One of the drawbacks of the evolutionary approach is that because it starts at the aggregate level instead of modeling individual behavior, it cannot answer this question. The answer we would give (see also the discussion in the concluding section of this chapter) is that from an economic perspective, neither replicator model should be taken to be precisely correct. Thus, a primary motivation for our interest in the use of evolutionary models in economics comes from the fact that many of the results discussed later in this chapter, for example about the elimination of dominated strategies, or set-valued notions of stability, hold for a wide range of "replicator-like" dynamics. From this perspective it is troubling that the ESS does not have this robustness property: as noted by Friedman [1991], an ESS need not be asymptotically stable under the sort of monotone dynamics discussed in section 3.6. (We discuss an example that shows this in section 3.10)

3.5. Asymmetric Replicator Models

We now turn to the case where there are distinct populations of player 1's, player 2's and so forth. We consider first how the replicator should be defined in this case, and in particular what to do if the populations of different player types are not the same size. We then show that, in contrast to the symmetric case, mixed equilibria are never asymptotically stable. However, they can satisfy the weaker property of being a center.

How should replicator dynamics be defined if there are two populations that are not the same size? If there are three times as many player 1's, for example, then under a random-matching interpretation of the model each player 1 must on average be involved in only 1/3 as many interactions as each player 2, so that the population of player 1's should evolve more slowly. Instead of examining the complications that stem from differential rates of adjustment, we will follow standard practice and consider only the dynamics

$$\dot{\theta}_t^i(s^i) = \theta_t^i(s^i)[u_t^i(s^i) - \bar{u}_t^i],$$

where the superscript *i*'s refer to the various populations.[40] This dynamics corresponds to the random-matching model provided that the two populations are always the same size. Alternatively, this equation can be viewed as describing a situation where agents know the distribution of opponents' strategies, and the evolution of the state variable reflects the agents' decisions /about revising their choices, as opposed to a hard-wired response to the payoffs obtained from play.

The most striking fact about the asymmetric case (in contrast to the homogenous case) is that interior points, that is, strictly mixed profiles, cannot be asymptotically stable. Hofbauer and Sigmund [1988] gave a proof of this for two-player games based on the fact

[40] Hofbauer and Sigmund [1988, chapter 27] discuss an alternative due to Maynard Smith [1974], in which the relative speeds of adjustment of the two populations are scaled by their average payoffs, which may differ between the two populations, instead of by the population sizes

that the replicator dynamic “preserves volume,” an observation of Akin’s developed in Eshel and Akin [1983]; Ritzberger and Weibull [1995] extended this result to *n*-player games. Rather than give a proof, we will settle for an example that suggests why interior points are less likely to be stable in the asymmetric-population model. The appendix provides a brief summary of volume-preserving maps and Liouville’s formula.

Example 3.1 revisited: Consider again the game in example 3.1, only now with distinct populations of player 1 and player 2. Recall that if both players agree they get 0; if they disagree they get 1. It is easy to see that the two asymmetric equilibrium in which the players disagree are asymptotically stable. The mixed equilibrium, which was asymptotically stable in the homogenous-population model, is now a saddle: If more than 1/2 of the player 1's play A, then the share of player 2’s using B grows, and if more than 1/2 of the 2’s use B, the share of 1’s using A grows, so starting from any point where more than half of the 1’s play A and more than half of the 2’s play B, the system converges to the state where all 1’s play A and all 2’s play B. Likewise, if more than half the 1’s play B and more than half the 2’s play A, the system converges to the other pure strategy equilibrium.

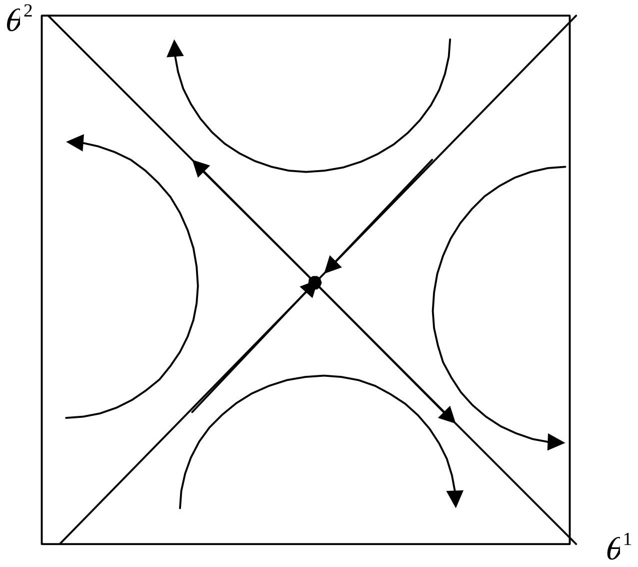


Figure 3.2

Since any open neighborhood of the mixed equilibrium contains points that converge to the two pure strategy, asymmetric equilibria, the mixed equilibria is not stable. Note, though, that the trajectories that starting from any point on the "diagonal", where the share of *A* is the same in each population, do converge to the mixed equilibrium. This is a consequence of the more general fact that, in symmetric games, trajectories that start from symmetric initial positions remain symmetric, and so follow the "same" path as in the one-population model. The contrapositive of this is that a symmetric point that is not stable in the homogenous-population model is not stable with asymmetric populations.

This difference in conclusion with the one-population model should not be surprising: in the first case there is no asymmetry between the players that could allow them to coordinate on one of the pure strategy equilibria; in the second game players can use their labels as a coordinating device.

Although the interior points cannot be asymptotically stable in asymmetric populations, they can satisfy a weaker condition. We say that a steady state is a *center* if it is surrounded by a family of closed orbits and all points that start near the equilibrium remain near it. From the viewpoint of general dynamical systems, being a center is a knife-edge property, meaning that small changes in the dynamics lead to abrupt changes in the asymptotic properties. Examples of small changes that can have this effect are the drift discussed in section 3.8, and the small probability of meeting a player from the same population as discussed in section 3.5. However, as usual with questions of robustness and genericity, a property that is not robust to a broad class of perturbations can be robust to a smaller one. In this case, centers are a robust property of the asymmetric-population replicator dynamics, that is they can arise for a non-negligible set of strategic-form payoffs. In particular, in 2x2 games the steady state is a center whenever the game has no pure-strategy equilibria. (Schuster and Sigmund [1981].)[41] Figure 3.3 depicts the center that arises in the 2x2 game “matching pennies.[42]”

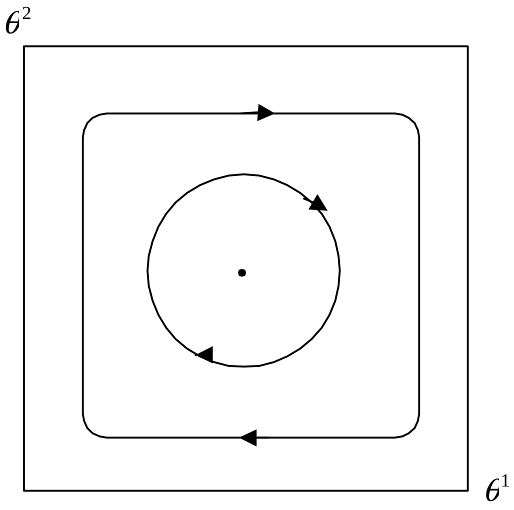


Figure 3.3

[41] This shows that the sensitivity of the center in rock-scissors-paper to payoff perturbations is not completely general–changing the payoff functions need not generate a sufficiently large range of perturbations for structural stability arguments to apply.

[42] To show that this figure is correct, one can note that the function $Q(\theta^1,\theta^2) = \theta^1(H)(1-\theta^1(H))(\theta^2(H)(1-\theta^2(H))$ is a constant along any trajectory, and then plot the level curves of *Q*.

In 3x3 games the asymmetric-population replicator dynamics only has a center for a lower-dimensional set of strategic-form payoffs (Hofbauer [1995]).[43] Hofbauer [1996] conjectures that this extends to larger action spaces.[44]

3.6. Interpretation of the Replicator Equation

3.6.1. Overview

Why should economists, or even game theorists more generally, be interested in the replicator dynamics? After all, we do not think that individuals are genetically programmed to play certain strategies. For that matter, it is not clear that even monkeys are programmed directly for certain behaviors. And even behavior that we do think is inherited probably is not coded by a single gene as this model suggests, but rather results from a complex interaction of genetic factors. Indeed, even a strict biological story does not lead directly to the replicator dynamic in cases where reproduction is not asexual.

How then could we explain this system in economics? Is there an underlying model of learning that gives rise to this dynamic? There are two types of learning stories that have been proposed to explain the replicator dynamic. One is a model of “asking around” or social learning, in which players can learn only from other players in the population. In order for the state of the resulting system to simply be the distribution of strategies currently played, as opposed to some function of the entire history, either the players must not remember their own past experience, or the players must periodically be replaced, so that only new players make choices.

[43] The proof of this is somewhat subtle, since even for these generic payoffs there can be interior steady states where the linearized system has purely imaginary eigenvalues. Hofbauer uses a second-order approximation to determine if the steady state is stable, unstable, or a center.

Alternatively, replicator-like dynamics may be explained by models of aspiration levels, in which players "satisfice" rather than "optimize." There are many ways of formulating such models so that they generate a *payoff monotone* dynamic, that is, a system in which the growth rates of strategies are ordered by their expected payoff against the current population, so that strategies that are "doing better" grow faster. The replicator is the particular form of a payoff monotone dynamic in which the rates of growth are proportional to payoff differences with the mean, and particular specifications of these learning models can give rise to precisely the replicator dynamic. However, since there is typically not a compelling argument for that precise specification, we prefer to focus on the conclusion that a range of learning processes can give rise to payoff monotone dynamics.

3.6.2. Social Learning

Let us first examine the idea that the evolutionary model describes a process of social learning. The simplest such story is one in which each period, some fraction α of the agents leave the system .[45] They are replaced by new agents, each of whom learns something about the prevailing state; to make things concrete let us suppose that each new agent observes the strategy and payoff of one exiting agent and of one other agent drawn randomly from the same population.[46] The new agents then make a once-and-for all choice of strategy, which they do by adopting the strategy with the higher observed payoff, or in case of a tie, the strategy they "inherited" from the exiting agent [47] Moreover, if the agent

[45] This assumption serves to justify the agent's lack of memory. Alternatively, we could suppose that agents do not remember their past experience, or that agents revise their strategies so rarely that they consider their past experience to be irrelevant to the current situation.

[46] In other words the probability of sampling a strategy equals its share in the current population. This sort of "proportional" sampling is standard in the literature, but other sampling rules may be worth considering as well, as noted by Banerjee and Fudenberg [1995] in the related context of "social learning." For purposes of comparison with that paper, note that in process described here, each agent's sample size is 2.

[47] If we think that agents forget, instead of being replaced, then this assumption would be that they stick with their own strategy unless they observe another one that does better.

they sample is using the same strategy that they "inherited," the agents do not switch, even if that strategy is performing poorly.

Intuitively, since a rule like "switch if the other strategy's payoff is higher" depends only on the ordinal rankings of the payoffs, and not on the size of the payoff difference, we would not expect it to lead to a dynamic like the replicator, where the speed of adjustment depends on the size of the payoff differences. Moreover, if agents observe the realized payoff of the agent they sample, as opposed to its average payoff against the current population, the resulting process need not even be payoff monotone . The point is that the rule "switch if the other strategy's observed payoff is higher" favors strategies with a high *median* payoff when matched with the distribution of opponents' play , as opposed to a high *average* payoff.[48]

To see this, consider the following game, where player 2 is a "dummy" whose payoffs are always 0:

$$\begin{bmatrix} (9,0) & (0,0) \\ (2,0) & (2,0) \end{bmatrix}$$

if 1/3 of the player 2's are playing L, then player 1's best response is U, but D has a higher median payoff. Consequently, whenever a player 1 using U samples a player 1 using D, or vice versa, fully 2/3 of them will choose the inferior response D. We call such "median-enhancing" dynamics *emulation dynamics.*

To explore the properties of emulation dynamics, we consider a family of simple two-population models of play in the above game. Suppose that there is a large

[48] This is essentially the same fact as the well-known "probability matching" in the mathematical psychology literature, see for example Norman [1972]. More recently, it is noted in Schlag [1994] in the context of learning in games, and in Ellison and Fudenberg [1993] in the context of agents learning about a move by Nature. Of course, the two cases are the same as far as the response of one population to the play of the "opposing side;" the differences arise from the fact in a game, the distribution of strategies used by the player 2's evolves over time in response to the distribution in the population of player 1's, while the distribution over Nature's moves is usually supposed to be exogenous.

(continuum) population of each player type, and that each period the agents are randomly matched with agents from the other population to play the stage game. Since player 2 is a dummy, we will fix the distribution of player 2 strategies, and to lighten notation we will let $\overline{\theta}_2(L)$ denote the fraction of player 2's playing *L*. Under the simple emulation dynamics described above, new agents who sample someone using the same strategy as their "parent" do not switch; at date *t*, fraction $\theta_1^t(U)$ is using *U*, so a fraction $\theta_1^t(U)^2$ of the active agents is composed of agents with *U* parents who sample another agent using U. Agents with parents using *U* who sample someone using *D* stick with *U* if and only if their own current payoff is 9; that is, if they were matched with strategy *R* last period; the fraction of such agents is $\overline{\theta}_2(L)\theta_1^t(U)\theta_1^t(D)$. Similarly, agents whose parents used *D* switch to *U* if they meet a *U*-user whose last opponent played *L;* this corresponds to fraction $\overline{\theta}_2(L)\theta_1^t(U)\theta_1^t(D)$. Combining terms yields the difference equation

$$\theta_1^{t+1}(U) = (1-\alpha)\theta_1^t(U) + \alpha\left(\theta_1^t(U)^2 + 2\overline{\theta}_2(L)\theta_1^t(U)\theta_1^t(D)\right);$$

substituting $\theta_1^t(D) = 1-\theta_1^t(U)$, simplifying, and passing to continuous time yields the equation

$$\dot{\theta}_1^t = \theta_1^t(1-\theta_1^t)(2\overline{\theta}_2(L)-1).$$

Thus the system converges to $\theta_1^t = 0$ (all player 1's using *D*) whenever the fraction $\theta_1^t(U)$ playing *L* is less than ½, even though U is a best response whenever $\theta_1^t(U) > 2/9$.[49]

In response to this, Schlag [1994] considers a discrete-time system in which agents observe the realized payoff of the agent they sample, and then switch to the better strategy with probability that grows linearly in the payoff difference. In the particular case that Schlag favors, this behavior rule has the following form: If the agent's parent's payoff is u¸

[49] More generally, Ellison and Fudenberg show that if the probability p that *U* is better is i.i.d. instead of constant, the system is ergodic, with the long-run average of $\theta_1^t(U)$ equal to the probability that $\theta_1^t(U)$ >1/2.

and the agent samples an agent with payoff u', the agent switches with probability $\max\{0, b(u-u')\}$.[50] One justification for this behavior is that there is a distribution of switching costs in the population; another, due to Binmore and Samuelson [1993], is that there is a distribution of "aspiration levels," and that agents only become active if their current payoff is below their aspiration level. Schlag then shows that the trajectories of the system converge to those of the continuous-time replicator dynamics as the population grows.

For simplicity, we will present a large-population of his model for the particular game described above. As previously, agents whose parents used *U* and who are matched with *L* will not switch, and neither will agents whose parents used *D* and who sample a *U*-user who is matched with *R*. However, of the agents whose parents used *D* and sample a *U*-user matched with *L*, only some fraction *q* will switch to *U*, while only a fraction *r* of the agents who see *D* get a higher payoff than *U* switch from *D* to *U*. (In Schlag's specification the parameters *q* and *r* are determined by the payoff matrix of the game; we can treat them as fixed so long as we consider a single payoff matrix. Moreover, with a more general payoff matrix where the payoff to *D* depends on 2's strategy, there would be four switching parameters to consider instead of two.) The equation of motion is now

$$\theta_1^{t+1}(U) =$$
$$(1-\alpha)\theta_1^t(U) + \alpha\Big(\theta_1^t(U)^2 + \big[(\bar\theta_2(L) + (1-\bar\theta_2(L))(1-r)\big]\theta_1^t(U)\theta_1^t(D) + \big[\bar\theta_2(L)q\big]\theta_1^t(U)\theta_1^t(D)\Big)$$

and the corresponding continuous-time limit is

$$\dot\theta_1^t = \theta_1^t(1-\theta_1^t)(2z-1),$$

[50] Schlag also considers the more general formulation where agents sometimes switch to a strategy with a lower payoff.

where $z = \overline{\theta}_2(L)(q+r)+(1-r)$.

Thus the system moves in the direction of increasing payoffs (is monotone) if and only if q and r are such that z > 1/2 whenever $\overline{\theta}_2(L)$>2/9. There are many combinations of q and r that satisfy this condition in this particular game; Schlag's result shows that the "proportional" or linear imitation rate guarantees that the corresponding constraint is satisfied for any specification of the payoff matrix, and that a particular proportional scheme produces a discrete-time, stochastic system that converges to the replicator dynamic in the limit of shorter and shorter time periods, where the convergence is in the same (somewhat subtle) sense as in the Borgers-Sarin model described later in this section.

Instead of supposing that agents observe the realized payoffs of other players, Bjornerstedt and Weibull [1995] assume that agents receive possibly noisy statistical information directly about the current expected payoff of the strategy they sample. They show that this assumption, together with the assumption that the support of the noise is sufficiently large, leads to a resulting process that is monotone.

To see this, suppose that the distribution of noise is such that the difference between any two noise terms has c.d.f. Φ_i. Then the probability that a player i who is currently using s^i and who samples a player using $\tilde{s}^i$ will switch is the probability that the noise term is less than the payoff difference $u_t^i(\tilde{s}^i)-u_t^i(s^i)$. This is equal to $\Phi_i(u_t^i(\tilde{s}^i)-u_t^i(s^i))$. Since, under proportional sampling, the fraction that uses s^i and samples $\tilde{s}^i$, namely $\theta^i(s^i)\theta_t^i(\tilde{s}^i)$, equals the fraction that uses $\tilde{s}^i$ and samples s^i; the population evolves according to the dynamic

$$\dot{\theta}_t^i(s^i)=\theta_t^i(s^i)\left[\sum_{\tilde{s}^i}\theta_t^i(\tilde{s}^i)\left(\Phi_i(u_t^i(s^i)-u_t^i(\tilde{s}^i))-\Phi_i(u_t^i(\tilde{s}^i)-u_t^i(s^i))\right)\right],$$

which is payoff monotone whenever the Φ_i are strictly increasing over the range of all payoff differences; this will be the case , whenever the support of the noise is big enough.

If, moreover, the noise has a uniform distribution over a sufficiently large interval, and the distribution is the same for the various players, then $\Phi(u) = a + bu, b > 0$, so the dynamic above simplifies to

$$\begin{aligned}&\dot{\theta}_t^i(s^i) = \theta_t^i(s^i)\left[\sum_{\tilde{s}^i} \theta_t^i(\tilde{s}^i)\left(2b(u_t^i(s^i) - u_t^i(\tilde{s}^i))\right)\right] = \\ &2b\theta_t^i(s^i)(u_t^i(s^i) - \bar{u}_t^i)\end{aligned}$$

which is the replicator dynamics (up to a time rescaling).

Bjornerstedt [1995] develops an alternative derivation of the replicator, based on the idea that only "dissatisfied" agents change their strategy, with the probability of dissatisfaction depending on the agent's own payoff and on some function of the current state such as the current average payoff in the population, or the current lowest payoff. (These functions describe the aggregate play of all agents currently using a given strategy; in some cases they can be built up from behavior rules for individual agents in which each agent only observes the payoff of one other strategy.) Agents who are dissatisfied choose another agent at random (under proportional sampling) and copy that agent's choice regardless of its current payoff.[51] If agents with lower payoffs are more likely to be dissatisfied, the resulting dynamic is monotone; moreover Bjornerstedt shows that the result is exactly the replicator dynamics in the special case where the probability of dissatisfaction is a suitably scaled linear function of the payoffs. (The scaling must ensure that the revision probabilities stay between 0 and 1, and so depends on the payoff function of the particular game.)

One use that has been made of the replicator dynamic is in the study of experimental results, as in the Binmore and Samuelson [1995] paper discussed in section

[51] Similar models of "switch when dissatisfied" have been studied by Binmore and Samuelson [1993] among others. In the Binmore and Samuelson paper, agents become dissatisfied if their payoff is less than some exogenous aspiration level; dissatisfied agents choose what strategy to switch to according to a rule that, as in the example above, leads to greater switching to strategies with higher current payoffs.

3.9. However, the model of agents who forget and ask around does not really apply in this setting, since subjects are not permitted to “ask around” for information about the strategies and realized payoffs of other players. An alternative learning model that gives rise to replicator dynamics is a stimulus-response model, which does not require that agents communicate with or observe one another, or even that there be many agents in each player role.

3.6.3. The Stimulus Response Model

An alternative justification for the replicator dynamic is drawn from the psychological stimulus-response model literature of learning. Basically, it is a model of “rote” learning, in which it is assumed that actions that do well are “reinforced” and so more likely to be used again in the future. Actions that do poorly receive “negative reinforcement,” and are less likely to be used in the future. One example of how a stimulus response type model can lead to replicator-like dynamics can be found in Borgers and Sarin [1995]. In their paper, each agent observes only his own realized action and the payoff that he receives. For simplicity, we specialize to two-player games. Agents at each date use a mixed strategy, and the state of the system at date *t*, denoted (θ_t^1, θ_t^2) is the vector of mixed actions played at time *t* by the two players. Payoffs are normalized to lie between zero and one, so that they may be interpreted as probabilities. The state evolves in the following way: if player *i* plays s_t at date *t*, and the resulting payoff was $\tilde{u}_t^i(s_t)$, then

$$\begin{aligned}
&\theta_{t+1}^i(s) = (1 - \gamma\tilde{u}_t^i(s^t))\theta_t^i(s) + E(s_t, s)\gamma\tilde{u}_t^i(s_t) \\
&E(s_t, s_t) = 1 \\
&E(s_t, s) = 0 \quad s \neq s_t
\end{aligned}$$

Here the reinforcement is proportional to the realized payoff, which is always positive by assumption. This is similar to the "stochastic learning theory" of Bush and Mosteller [1955], in the case where all outcomes provide positive reinforcements; Chapter 4 discusses related models of Borgers and Sarin [1996] and Er'ev and Roth [1996] that do accord more closely with experimental evidence but do not yield the replicator dynamic in the continuous-time limit.

A simple calculation shows that expected increase in the probability that player *i* uses *s* equals the current probability multiplied by the difference between the strategy's expected payoff and the expected payoff of the player's current mixed strategy. In the limit as $\gamma \to 0$ Borgers and Sarin show that the trajectories of this stochastic process converge in probability to the continuous time replicator dynamic. Note, however, that this does not imply that the replicator dynamic has the same asymptotic behavior as the stochastic system: for example in matching pennies, the stochastic reinforcement model will eventually be absorbed by a pure strategy profile, while the continuous time replicator cannot converge to a pure strategy profile. This discontinuity in the asymptotic behavior of discrete and continuous time systems is something we discuss in greater detail below.

3.7. Generalizations of the Replicator Dynamic and Iterated Strict Dominance

The replicator dynamic is very specific. Both the asking around model and the stimulus response model do however lead to dynamics which are payoff monotone, meaning that the number of people playing strategies that are doing well should grow. This leads the question of which properties of the replicator extend to other dynamics that retain this intuitive idea. Here we consider monotonicity and some if its variations, and show

how even relatively weak notions of monotonicity are sufficient to guarantee the iterated elimination of strictly dominated strategies.

Following Samuelson and Zhang [1992], say that an adjustment process (that is, a flow on the state space $\Theta^1 \times \Theta^2 = \Sigma^1 \times \Sigma^2$) is *regular* if (i) it is Lipschitz continuous, (ii) the sum of the flows in each population equals 0, and (iii) strategies with 0 shares have non-negative growth rates. The process is *payoff monotonic* if strategies with higher current payoffs have higher current growth rates: [52]

Definition 3.1: A process is *payoff monotone* if at all interior points,

$$u_t^i(s^i) > (=)\, u_t^i(s^i\text{'}) \Rightarrow \frac{\dot{\theta}_t^i(s^i)}{\theta_t^i(s^i)} > (=) \frac{\dot{\theta}_t^i(s^i\text{'})}{\theta_t^i(s^i\text{'})}\,.$$

Although this condition is quite weak in some respects, the requirement that growth rates are strictly ordered by the corresponding payoffs does rule out the best response dynamics, since under best response all strategies that are not best responses have identical growth rates of -1.

Recall the definition of strict dominance from chapter 1: a strategy σ^i is *strictly dominated* if there is some other (possibly mixed) strategy $\hat{\sigma}^i$ such that

$$u^i(\hat{\sigma}^i, \sigma^{-i}) > u^i(\sigma^i, \sigma^{-i})$$

for all profiles of opponents' strategies σ^{-i}. *Iterated strict dominance* is the process of first removing all strictly dominated strategies for each player, then removing all strategies that become strictly dominated once the dominated strategies are deleted, and so on until no further deletions are possible. Following Samuelson and Zhang, define the process of *iterated pure-strategy strict dominance* to be the analogous iterative process when only

[52] Samuelson and Zhang simply called these processes “monotone.”

dominance by pure strategies is considered. Obviously this process deletes fewer strategies, since a strictly dominated strategy may not be dominated by any pure strategy.

Proposition 3.4 (Samuelson and Zhang): Under any regular, monotone dynamics, if strategy *s* is eliminated by the process of iterated pure-strategy strict dominance, then the share of strategy *s* converges to 0 asymptotically, irrespective of whether the state itself converges.[53]

Sketch of Proof: The easiest case is that of a strategy *s* that is strictly dominated by some other pure strategy $\hat{s}$. Then the growth rate of *s* is always some fixed amount less than the growth rate of $\hat{s}$, and so the share of *s* in the population must go to tend to 0 asymptotically. Once this is seen, it is not surprising that the result extends to iterative deletion. Intuitively, we expect the adjustment process to run through the iterative deletion: once the dominated strategies have shares close to 0, then strategies that are removed at the second round of iterated pure-strategy dominance must have lower payoffs than those of other strategies with non-negligible shares, so their share starts to shrink to 0, and so on. Since the iterative deletion process stops in a finite number of rounds (in stage games with a finite number of actions), the adjustment process should eventually eliminate all of the strategies in question.

To make this intuition more precise, we adapt an argument that Hofbauer and Weibull [1995] used in their proof of proposition 3.5 below. Note first that the share of each strategy that is strictly dominated by a pure strategy is bounded above by a function that converges to 0 at an exponential rate. Since there are only a finite number of such

[53] Nachbar [1990] has a similar result that applies only to “dominance-solvable” games where the iterated deletion process eliminates all but one strategy’s profile. Milgrom and Roberts [1990] obtain a similar result for the class of supermodular games.

dominated strategies, there is, for any positive ε, a finite time T such that at all $t>T$ every one of them has share less than ε.

Let s'^i be a strategy for player i that is not strictly dominated by a pure strategy but is strictly dominated by some $\hat{s}^i$ once the first round of deletions is performed. Since payoff functions are continuous functions of the mixed strategies, $\hat{s}^i$ has a strictly higher payoff than s'^i once the shares of all of the "pure-strategy-strictly dominated" strategies are less than some sufficiently small ε;[54] moreover by taking ε small enough we can ensure that this is true for uniformly over all of the strategies removed at the second round of the iteration. Thus after some finite time T', the shares of all of the strategies that are removed at the second round of iteration are bounded by a function that converges to 0 at an exponential rate. Hence the shares of these strategies become negligible at some finite time T'', and the argument continues on. Since the process of iteration ends in a finite number of rounds in finite games, only a finite number of iterations of the argument are required, and we conclude that there is a finite time T after which the shares of all strategies that are eliminated by iterated pure-strategy strict dominance converge to 0.

☑

An example due to Bjornerstedt [1995] shows that monotone dynamics need not eliminate strategies that are strictly dominated by a mixed strategy. This example uses a version of the "sample-if-dissatisfied" dynamic.

Example 3.3: Consider the following variant of the rock, scissors, paper game due to Dekel and Scotchmer [1992]

[54] Note that this continuity argument does not apply to the concept of weak dominance.

$$\begin{bmatrix} 1.00,1.00 & 2.35,0.00 & 0.00,2.35 & 0.10,1.10 \\ 0.00,2.35 & 1.00,1.00 & 2.35,0.00 & 0.10,1.10 \\ 2.35,0.00 & 0.00,2.35 & 1.00,1.00 & 0.10,1.10 \\ 1.10,0.10 & 1.10,0.10 & 1.10,0.10 & 0.00,0.00 \end{bmatrix}$$

Here, the upper 3x3 matrix is a non-zero sum version of rock, scissors paper, while the fourth strategy is strictly dominated by an equal mixture of the first three strategies, but not by any pure strategy. However, the fourth strategy is a better-than-average response whenever it and one other strategy are scarce. Now consider a "sample-if dissatisfied" dynamic where each player's propensity to sample depends on the current aggregate state as well as on the player's own payoff, and moreover this dependence takes the very special form that players with the lowest possible payoff given the current state are certain to sample. Then since the fourth strategy is typically not the worst-performing one, it can survive even in continuous time unless the system starts at exactly the Nash equilibrium.

Note, incidentally, that in this example the mixed Nash equilibrium is an ESS. Consequently, this example also shows that an ESS does not imply asymptotic stability for general monotone dynamics.

In order to eliminate strategies that are strictly dominated by a mixed strategy, Samuelson and Zhang introduce the condition of "aggregate monotonicity":

Definition 3.2: A system is *aggregate monotonic* if at all interior points,

$$u_t^i(\sigma^i) > u_t^i(\hat{\sigma}^i) \Rightarrow \sum_{s^i} (\sigma^i(s^i) - \hat{\sigma}^i(s^i)) \frac{\dot{\theta}_t^i(s^i)}{\theta_t^i(s^i)} > 0 .$$

This says that, if mixed strategy σ^i has a higher current payoff than mixed strategy $\sigma^{'i}$, then the "growth rate" of σ^i is higher than that of $\hat{\sigma}^i$. It is easy to see that aggregate monotonicity implies monotonicity. Samuelson and Zhang show that the replicator dynamics is aggregate monotonic, and that any aggregate monotonic system deletes all strategies deleted by iterated strict dominance.

Recently Hofbauer and Weibull [1995] have found a weaker sufficient condition that they call *convex monotonicity*.

Definition 3.3: A system is *convex monotonic* if at all interior points,

$$u_t^i(\sigma^i) > u_t^i(s^i) \Rightarrow \sum_s \sigma^i(s^{'i}) \frac{\dot{\theta}_t^i(s^{'i})}{\theta_t^i(s^{'i})} > \frac{\dot{\theta}_t^i(s^i)}{\theta_t^i(s^i)}.$$

In words, this says that if mixed strategy σ^i has a higher current payoff than pure strategy s^i, then the "growth rate" of σ^i is higher than that of s^i.

A convex monotonic system is clearly monotonic, so that convex monotonicity rules out the best-response dynamics. However, there are approximations of the best response dynamic that are convex monotonic. For example, Hofbauer and Weibull note that the following dynamics is convex monotone for any positive λ:

$$\dot{\theta}_t^i(s^i) = \theta_t^i(s^i)\left(\frac{\exp(\lambda u_t^i(s^i))}{\sum_{\hat{s}} \theta_t^i(\hat{s}^i)\exp(\lambda u_t^i(\hat{s}^i))} - 1\right).$$

As λ grows to infinity, this system converges to the best-response dynamics.

Proposition 3.5 (Hofbauer and Weibull): Under any regular, convex monotone dynamics, if pure strategy *s* is eliminated by the process of iterated strict dominance, then the share of strategy *s* converges to 0 asymptotically, irrespective of whether the state converges. Moreover, if mixed strategy σ is removed by iterated strict dominance, then for all $\varepsilon > 0$ there is a time *T* such that for all *t*>*T* the share of at least one of the pure strategies in the support of σ is less than ε.

Sketch of Proof: As with the preceding proposition, the key step is showing that the dominated strategies are removed. To do this, suppose that strategy s^i of payer i is strictly dominated by strategy σ^i, and without loss of generality suppose that σ^i gives strictly positive probability to every strategy but s^i. Now consider the function P defined by $P^i{}_{\sigma^i}(\theta^i) = \theta^i(s^i)\prod_{\tilde{s}^i} \theta^i(\tilde{s}^i)^{-\sigma^i(\tilde{s}^i)}$; thus $P_{\sigma^i}{}^i(\theta^i) = \theta^i(s^i)E_{\sigma^i}(\theta^i)$, where $E_{\sigma^i}(\theta^i)$ is the entropy function used in the proof of Proposition 3.3. Along any interior trajectory,

$$\dot{P}_{\sigma^i}(\theta^i_t) = \dot{\theta}^i_t(s^i)E_{\sigma^i}(\theta^i_t) + \theta^i_t(s^i)\dot{E}_{\sigma^i}(\theta^i_t) =$$

$$\left(\frac{\dot{\theta}^i_t(s^i)}{\theta^i_t(s^i)} + \frac{\dot{E}_{\sigma^i}(\theta^i_t)}{E_{\sigma^i}(\theta^i_t)}\right)P_{\sigma^i}(\theta^i_t) = \left(\frac{\dot{\theta}^i_t(s^i)}{\theta^i_t(s^i)} - \sum_{\tilde{s}^i} \sigma^i(\tilde{s}^i)\frac{\dot{\theta}^i_t(\tilde{s}^i)}{\theta^i_t(\tilde{s}^i)}\right)P_{\sigma^i}(\theta^i_t)$$

which is strictly negative on the interior of the simplex from convex monotonicity and the fact that $E_{\sigma^i}(\theta^i)$ is bounded away from 0. Thus P must converge to 0, and so (again using the fact that $E_{\sigma^i}(\theta^i)$ is bounded away from 0) we conclude that $\theta^i(s^i)$ converges to 0 at an exponential rate.

To show that the process continues to iteratively delete the dominated strategies, we now simply paraphrase the analogous argument from the proof of proposition 3.4 (which was actually taken from Hofbauer and Weibull [1995]) replacing every dominating pure strategy by the dominating mixture.

☑

As a final remark on results about iterative deletion, we should note that even the replicator dynamic need not eliminate a strategy that is weakly dominated, since Nash equilibria in weakly dominated strategies can be stable (but not asymptotically stable.) Section 3.9 gives an example of this, and has an extended discussion of one response to it.

3.8. Myopic Adjustment Dynamics

Besides the aggregate and convex monotonicity discussed in the previous section, there is another useful generalization of monotonicity, called myopic adjustment. This class included not only the replicator dynamic, but also the best response dynamic, and simply requires that utility increase along the adjustment path (holding fixed the play of other players). We consider two applications of this idea. First, we give a complete characterization of myopic adjustment in the case of 2x2 symmetric games with a single population. Second, we consider the set-valued notion of strategic stability, and how it is connected to the property of being an attractor for a myopic adjustment process.

3.8.1. Replicator versus Best-Response

The notion of myopic adjustment is due to Swinkels [1993]. This generalization of monotonicity includes as a special case not only the replicator dynamic, but also the best response dynamic.

Swinkels's condition of myopia is that holding other players' play fixed, utility should be non-decreasing along the adjustment path. Formally

Definition 3.4: A system is a *myopic adjustment dynamic* if

$$\sum\nolimits_{s^i} u_t^i(s^i)\dot{\theta}(s^i) \geq 0$$

We next reconsider the monotonicity condition, that higher utilities imply weakly higher growth rates. One implication of this is that strategies whose share is expanding must have higher utility than strategies whose shares are contracting. Let $\underline{u}^i$ denote the least utility of any strategy whose share is (weakly) expanding, and let $\overline{u}^i$ denote the greatest utility of any strategy whose share is strictly declining. Monotonicity implies $\underline{u}^i \geq \overline{u}^i$. Notice that

$$\sum_{s^i} \dot{\theta}(s^i) = 0 \Rightarrow \sum_{s^i|\dot{\theta}(s^i)\geq 0} \dot{\theta}(s^i) = -\sum_{s^i|\dot{\theta}(s^i)<0} \dot{\theta}(s^i)$$

If we calculate the sum defining the time rate of change of utility separately over those strategies that are expanding, and those that are contracting, we find

$$\begin{aligned}
&\sum_{s^i|\dot{\theta}(s^i)\geq 0} u_t^i(s^i)\dot{\theta}(s^i) + \sum_{s^i|\dot{\theta}(s^i)<0} u_t^i(s^i)\dot{\theta}(s^i) \geq \\
&\underline{u}_t^i \sum_{s^i|\dot{\theta}(s^i)\geq 0} \dot{\theta}(s^i) + \bar{u}_t^i \sum_{s^i|\dot{\theta}(s^i)<0} \dot{\theta}(s^i) = \\
&\underline{u}_t^i \sum_{s^i|\dot{\theta}(s^i)\geq 0} \dot{\theta}(s^i) + \bar{u}_t^i \left(-\sum_{s^i|\dot{\theta}(s^i)\geq 0} \dot{\theta}(s^i)\right) = \\
&(\underline{u}_t^i - \bar{u}_t^i)\sum_{s^i|\dot{\theta}(s^i)\geq 0} \dot{\theta}(s^i) \geq 0.
\end{aligned}$$

This enables us to conclude

Proposition 3.6: Every monotonic regular system is a myopic adjustment dynamic.

As we mentioned above, the best response dynamic is also a myopic adjustment dynamic. That is, reinterpreting the state variable as beliefs rather than a population distribution of play, the system

$$\dot{\theta}^i = BR^i(\theta^j) - \theta^i$$

also increases utility holding the opponents strategy fixed. Indeed in a certain sense, the best-response dynamic increases utility holding the opponents strategy fixed as rapidly as possible. Consequently, results that apply to all myopic adjustment dynamics are applicable to the best response dynamic, and by implication, continuous time fictitious play, as well as to the replicator dynamic.[55]

[55] Remember that the equivalence between the fictitious play and best response dynamic hold only in continuous time, and that the continuous-time model can capture the asymptotic behavior of discrete-time fictitious play, but not that of the discrete-time best response dynamic, for the latter dynamic is time homogenous and does not "slow down" asymptotically.

3.8.2. Two by Two Symmetric Games

The myopic adjustment dynamic is strong enough to yield results in the special case of a symmetric 2x2 game with a symmetric initial condition (or equivalently, one population). In this case the state variable is one dimensional, so that in continuous time models the system cannot cycle and must converge to a steady state. Moreover, the only possible steady states are Nash equilibria of which, generically, there are at most three. The stability properties of steady states is entirely determined by the direction of the flow at each point, the rate of movement makes no difference.

In a symmetric, one-population, 2x2 game, a myopic adjustment dynamic cannot move the system in the direction that corresponds to utility going down holding opponents strategy fixed.[56] If we add to the assumption of myopic adjustment the assumptions that (1) movement must be strictly positive in the utility improving direction (if there is one) , and that (2) every Nash equilibrium is a steady state, then direction of the flow is pinned down everywhere except at non-Nash pure strategy points. (Note that both of these assumptions are satisfied in both the replicator and best-response cases.) Consequently, except for the issue of whether non-Nash pure strategy points are (unstable) steady states (as they are in the replicator dynamics but not under best-response) the global properties of all myopic dynamical systems are exactly the same. In particular, if there is a unique Nash equilibrium (either in the interior or on the boundary), it is an attractor from all interior initial conditions. The other generic possibility is that there is a mixed equilibrium and two strict pure equilibria, as for example in coordination games. Here the two pure equilibria are stable, and the mixed equilibrium is unstable, as illustrated below in Figure 3.4. (If one of the pure equilibria were unstable, then the other strategy would need to be

[56] This is true in all games, but considerably more useful in the 2x2 symmetric case.

the best response to mixed strategies arbitrarily near unstable equilibrium, which would contradict the assumption that the first equilibrium was strict.).

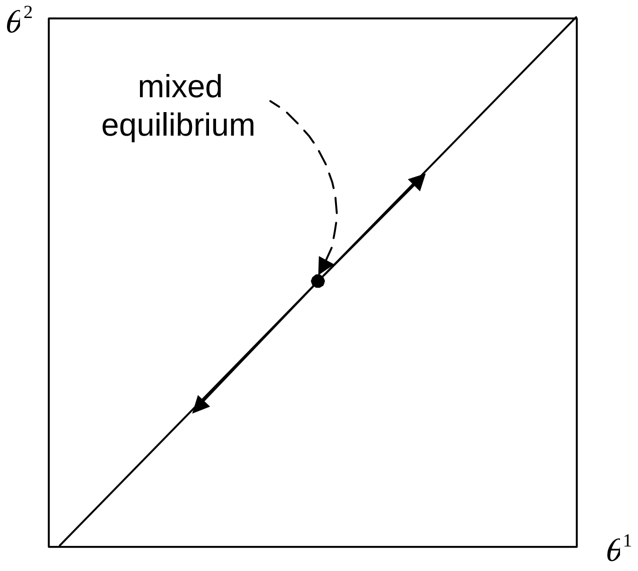


Figure 3.4

3.8.3. Stable Attractors and Strategic Stability

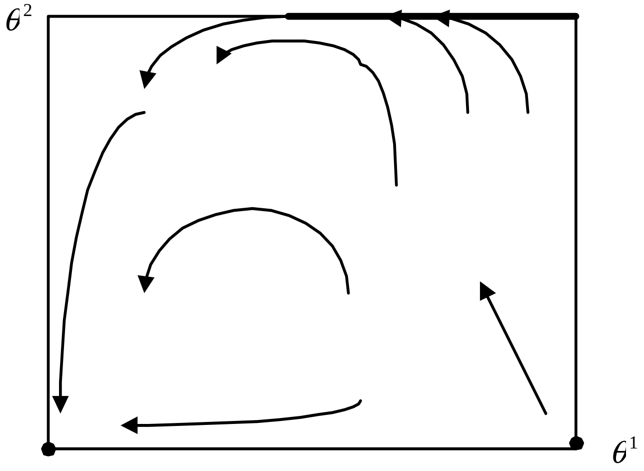
Swinkels [1993] proves a general result about myopic adjustment dynamics: he establishes a connection between stability of the dynamical system, and strategic stability in the game theoretic sense of Kohlberg and Mertens [1986]. A set of mixed strategy Nash equilibria is hyperstable if for every addition of redundant strategies to the game, and every sufficiently small perturbation to payoffs generated by forcing opponents to tremble, the perturbed game has a Nash equilibrium that is close to the original set, and if the set is a minimal set with this property. Kohlberg and Mertens show, for example, that every hyperstable set must contain a subgame perfect equilibrium, and indeed a sequential equilibrium.

Swinkels's result, stated below, restricts attention to asymptotically stable sets that are convex. This effectively rules out limit cycles, focusing attention on sets of steady states, and thus on sets of Nash equilibria. Now recall from chapter 1 that every Nash equilibrium is locally isolated for generic strategic-form payoffs. In fact more is true: For generic payoffs, every Nash equilibrium is strict, and hence hyperstable. This of course is not surprising: The main motivation for looking at refinements of Nash equilibria such as hyperstability comes from the consideration of the sort of non-generic strategic form payoffs that arise when the strategies in the strategic form are complete contingent plans in a non-trivial extensive form game. In effect, Swinkels is following the Kohlberg and Mertens program of adopting a strategic-form approach to a problem that arises from extensive form games. Part 2 of the book will discuss learning in extensive form games using models that explicitly reflect the extensive-form structure.

Proposition 3.7 (Swinkels): If a set is asymptotically stable under a myopic adjustment dynamic for which every Nash equilibrium is a steady state and has a neighborhood contained in the basin of attraction which is homeomorphic to a convex set (and in particular is connected), it contains a hyperstable set. In particular it also contains a sequential equilibrium.

Although the details of Swinkels's proof are quite complex, the basic idea is not. Consider the special case of strictly myopic dynamics, in which the population's utility holding the opponents' strategy fixed strictly increases except at steady states. In this case, steady states coincide with Nash equilibria. The idea of the proof is that since the set in question is asymptotically stable, it must remain so even when the game is perturbed. In other words, we may find a new myopic dynamic in the perturbed game which is close to the old dynamic. Since the original vector field pointed inwards on the boundary of a neighborhood of the set, a small perturbation will not change this. But then the flow maps a set homeomorphic to a convex set to itself, so by the Brower fixed point theorem, the set contains a fixed point of the flow, that is, a steady state, and by construction, this lies near the original set. Since in the strictly myopic case a steady state is a Nash equilibrium, this completes the proof. Swinkels full proof is considerably more complicated because of the need to remove steady states that are not Nash equilibria in the case of weak myopia, and because the missing step of finding a new myopic dynamic in the perturbed game which is close to the old dynamic requires some work.

We should point out however that the property of a set being asymptotically stable is much stronger than that of a point being asymptotically stable. If a steady state fails to be asymptotically stable, then (generically) the set of initial conditions that lead to that steady state in the long run has measure zero: it must be a source or a saddle. However, this need not be true for a set, as the illustration below of a set that is not asymptotically stable illustrates.



Here the solid line on the top represents a set of steady states. Initial conditions on the right of the line converge to the line, those near on the left come close but do not converge to it. None of the steady states are individually asymptotically stable, since a small perturbation can lead to another nearby steady state. Nor is the set asymptotically stable, since a small perturbation away from the set on the left side of the line does not lead back to the line. However, there is a very large (open set) of initial conditions for which there is convergence to the set.

3.9. Set Valued Limit Points and Drift

Expanding on the idea that there may be sets of equilibria that are not stable, but may never-the-less be good asymptotic descriptions of the long-run behavior of the system, Binmore and Samuelson [1995] argue that it is a mistake to take the deterministic learning dynamics of this chapter and the last very seriously in the neighborhood of a set of steady states, and that this has important implications near such sets. The idea is that in addition to the replicator, or other dynamic, there is ordinarily an additional tendency of the system to move, both randomly due to mutations and the like, and deterministically, due to various non-modeled factors. The latter deterministic movement of the system they refer to as drift. If the model without drift is to make sense, the drift should obviously be small. The smallness of drift is enough to guarantee that the asymptotic properties of isolated steady states are not changed by its presence: it is neither strong enough to escape from a stable steady state, nor strong enough to force convergence to an unstable steady state.

Near a set of equilibria, however, Binmore and Samuelson argue, the situation is quite different. This is most easily seen in their example of the ultimatum mini-game

	Y	N
H	2,2	2,2
L	3,1	0,0

The story of this game is that the first player proposes either to split four dollars equally, or to keep three for himself. If an equal split is proposed it is accepted, but if the first player proposes an unequal split, the second player may choose to either accept the split, or reject it, in which case neither player gets anything.

The set of Nash equilibria and the replicator dynamic for this game is sketched in Figure 3.5 below.

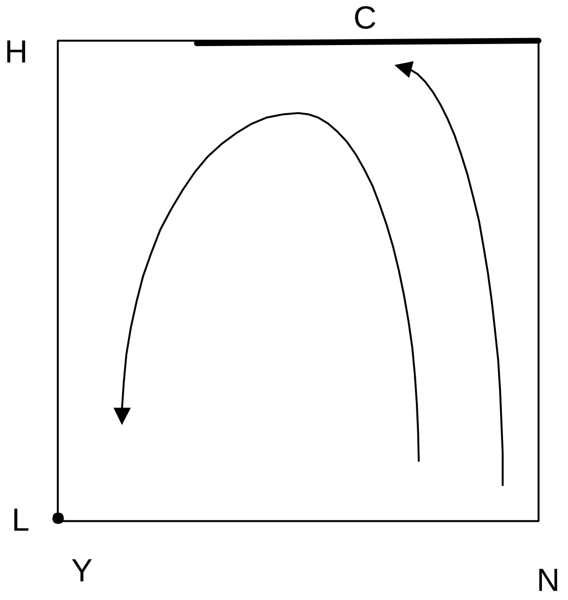


Figure 3.5

In this game there is a strict Nash equilibrium at (L,Y) and a component C of Nash equilibria in which player 1 plays H and player 2 gives probability of 2/3 or more to the weakly dominated strategy N. Since the equilibrium at (L,Y) is strict, it is an attractor of the replicator dynamic. The component of equilibria where 1 plays H is unstable: a small fraction of the population playing L causes the player 2's to gradually move towards Y, so that eventually more than a third of them are playing Y, and the system then moves off towards (L,Y).

The key point of Binmore and Samuelson is that at points near the unstable component C, the "force" or velocity of this drift is very slow, since the player 2's are near indifference. Suppose that the drift is due to some people in the population occasionally choosing at random (50-50) from their two strategies, so that there is a weak tendency *ceteris paribus* to move to the center. Suppose moreover that the player 2's, who have less to lose, are more likely to choose randomly. Then superimposed on the replicator dynamic is a small drift dynamic of the sort illustrated below.

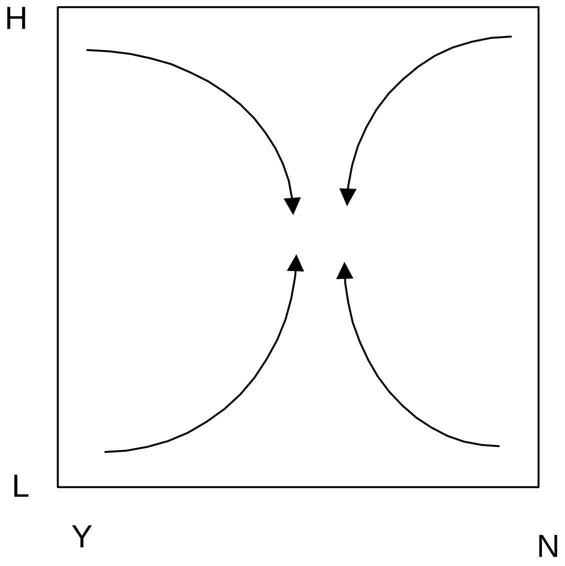


Figure 3.6

When we combine the two dynamics, putting most of the weight on the replicator, and very little weight on the drift, the combined dynamic has the following appearance

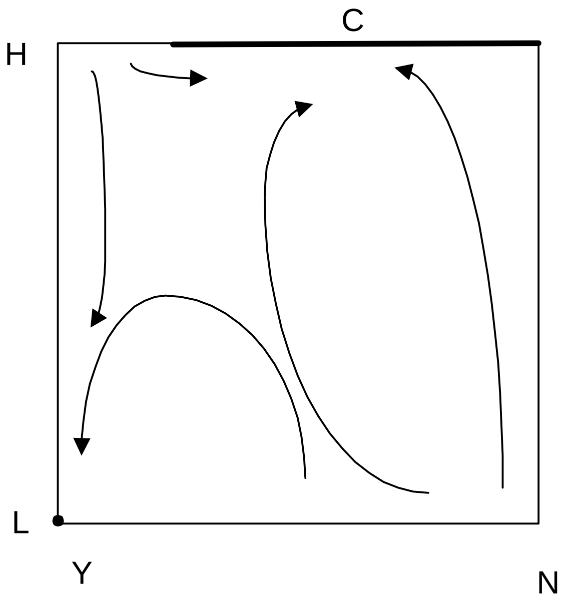


Figure 3.7

Now the tendency of player 2's to randomize between Y and N overcomes their tendency to move towards L when there is a very low probability of L, and the system has stable steady states near the set of equilibria.

Binmore and Samuelson make several other useful observations. First, they note that whether or not drift is important depends both on the nature of the drift and on the nature of the deterministic dynamic. In particular, if both players drift at the same rate, the flow diagram looks much like that without drift, and in particular the set of equilibria is unstable. Moreover, fixing the drift as in figure 3.2, and continuing to suppose that player 2 drifts more rapidly than player 1, we can modify the payoffs of the game, and thus modify the corresponding deterministic dynamics. For example, we can consider

	Y	N
H	2,2	2,2
L	3(1+*a*),1	0,0

Figure 3.8

When $a = 0$, this is the game originally studied. As a increases to 1, the range of mixed strategy equilibria is reduced until the probability of Y drops from 2/3 to 1/3 as illustrated below

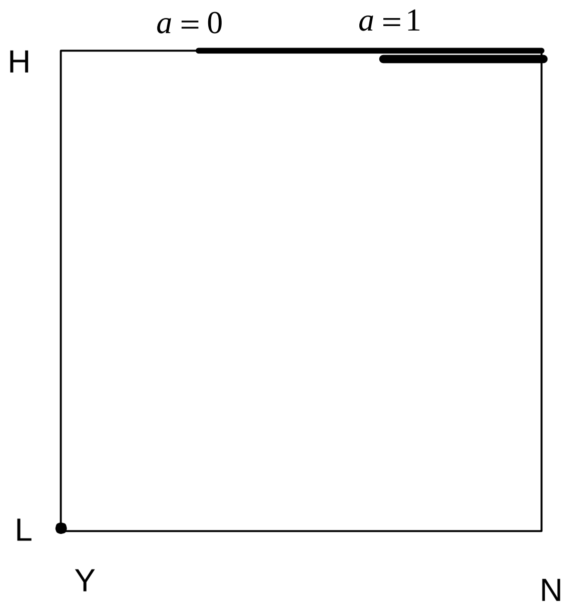


Figure 3.9

Under the proposed model of drift (see the figure 3.2 above) when a=1 the drift does not make any difference, since to the right of player 1 randomizing 50-50 the system is drifting left anyway. Only when a<1/2, so the system drifts to the right near a portion of the equilibrium set, can it begin to stabilize the set of equilibria. Binmore and Samuelson argue that this has important predictive consequences: the bigger is a the more likely we are to see the strict equilibrium at L,Y. Note, however, that this analysis is based on assuming the drift does not change when the payoffs do, while the argument about player 2 drifting faster than player 1 assumes that the drift is in part determined by players' payoffs. In this particular example, the same change in payoff that makes the set of equilibria smaller (bigger a) also makes player 1 relatively less indifferent, and so by the argument should decrease the rate at which he drifts, thereby tending to reinforce the stability of the set of equilibria. However, this does not upset the basic conclusion, because once the segment of equilibria lies entirely to the right of 1/2, regardless of its strength, any drift towards 1/2, reinforces the instability of the segment. This agrees with Binmore and Samuelson’s basic conclusion that the shorter the segment the less likely it is to be stable, but provides a cautionary note about treating the drift as fixed.

These ideas are important because many of the system we will examine, including the smooth version of fictitious play, do exhibit drift. However, as remarked above, all steady states are locally isolated for generic strategic form payoffs, so this observation is not too important in the case of one-shot simultaneous move games. As with Swinkels’s result, Binmore and Samuelson’s analysis is most relevant for strategic forms arising from nontrivial extensive forms. For example, the strategic form mini-ultimatum game is derived the following extensive form mini-ultimatum game:

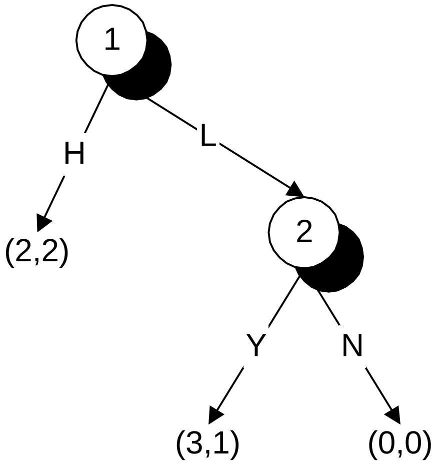


Figure 3.10

Not only do Binmore and Samuelson have this extensive form in mind, they make the stronger argument that learning together with the kind of drift described above can be a theoretical explanation of the empirical phenomenon found in experimental play of extensive-form ultimatum games, namely that the first mover does not get most of the pie.[57].

There are two potential difficulties with this argument. First, Binmore and Samuelson study the replicator dynamic. As we argued above, the stimulus-response rationale for this model is unconvincing, and the asking- around model clearly does not apply in an experimental setting. Of course, the replicator here is used as a convenient way of making things precise; Binmore and Samuelson are motivated by the belief that their result would also obtain in related payoff-monotone dynamics that are more readily justified. More fundamentally, we find it unappealing to attribute the play of N by player 2 to a short-run lack of "knowledge" that player 2 will eventually "learn", since if player 2 understands the rules of the game and her own payoff function, she will understand that playing N is a mistake. One might argue that, despite the experimental instructions, player 2 does not in fact understand (or believe) that the experiment is being conducted anonymously, and so adopts a rule ("do not be taken advantage of") that is optimal in a repeated bargaining setting. However, we find the standard explanation in the experimental literature, namely that the player 2's payoffs depend on other things than the money they receive, to be more convincing.

Note, incidentally, that playing N is weakly dominated in the strategic form game. Indeed one might want to argue that we should consider only rules that put positive weight on strategies that are not weakly dominated, since it makes no sense to play a weakly dominated strategy, and this does not require any learning to find this out (presuming that players know the rules of the game from the outset.) . This idea needs to be qualified, since when we examine extensive form games in more detail later in the book, we will find that

[57] A good discussion of this experimental fact, along with many references can be found in Prasnikar and Roth [1992] .

it does sometimes make sense to play weakly dominated strategies for the information they reveal about opponents' play, but this information-gathering explanation does not apply to a decision at the "end" of the game tree, like strategy N in this game. We should also emphasize that in general extensive form games, segments of equilibria essentially always arise because of the possibility of off-the-path randomization, and they do not always involve the use of weakly dominated strategies. [58]

3.10. Cheap Talk and the Secret Handshake

One interesting set-valued notion of stability closely connected to the best-response dynamic and to the idea of drift is the idea of a "cyclically stable set" introduced by Gilboa and Matsui [1991]. A strategy profile is said to be *accessible* from another profile if there is a continuous time best-response dynamic path from to the other profile. (Note that the relative rates of adjustment of the players may differ or vary over time if necessary.) A set is *cyclically stable* if no point outside the set is accessible from inside it, and every point inside the set is accessible from every other point inside.

The idea that a set that is cyclically stable is relatively stable seem incontrovertible. More questionable is the notion that a set of equilibria that is not cyclically stable is unstable; that is, merely because there is some best-response path that leads out, what reason do we have to believe that this particular path will be followed in practice? This is where the idea of drift is important. Multiple best-response paths occur when there is indifference, and a small amount of drift will tend to move players from one point of indifference to another anyway. So if there are multiple best-response paths, only one of

[58] This is a consequence of Kreps and Wilson' s [1982] theorem that for generic extensive-form payoffs there are finitely many Nash equilibrium "outcomes," that is, probability distributions over terminal nodes.

which leads out of a particular set, we can imagine than random drift will eventually cause that particular path to occur.

One interesting class of games in which cyclically stable sets lead to interesting conclusions are the "cheap-talk" games discussed above in the context of evolutionary stability. Consider again the coordination game example:

	L	R
L	2,2	-100,0
R	0,-100	1,1

Figure 3.11

In addition we allow players a pre-move in which they may send the message L or the message R. We previously saw that ESS could not eliminate "babbling" equilibria of the type "play R no matter what and send all messages with positive probability."

Instead of ESS, Matsui [1991] uses the cyclically stable set idea. Matsui shows that when a round of cheap talk is added to any 2x2 game of common interest there is a unique cyclically stable set[59] and that it has a unique outcome which is the Pareto efficient outcome.

The intuition behind this idea is closely connected to the idea of drift. Suppose that the outcome in the game above is (R,R). If all players use a strategy that ignores first-stage messages and always plays R in the second stage, it is a best response to say anything. Consequently, there is nothing to prevent the system from drifting to a state in which all players are saying (R,R) as well as doing it; that is, there is a path of the continuous- time

[59] This is a two population model; it is not assumed that the two players always play the same way.

best-response dynamics that leads from a “babbling state, to a state in which only a single signal is being sent. Now, however, consider a player 1 who says L , and then plays L if and only if his opponent says L. Since all player 2’s will play R anyway, this strategy is also a best response. Moreover, it is also a best response for some of the player 2’s to drift onto this strategy when all player 1’s play R, and moreover once the process of drift is underway the new strategy becomes a *strict* best response. On the other hand, Matsui shows that when all agents are playing L the equilibrium does not unravel. It is true that there is no disadvantage to saying R and doing R if the opponent says R , and so the system may drift to this state, once it is reached no player will ever wish to make use of the opportunity to induce R to be played.

We will return to this example in chapter 7, in the context of learning in extensive-form games; for the moment we point out only that the continuous-time best response dynamic on *strategies* used to define cyclically stable sets implicitly supposes that players observe the entire strategies of their opponents, and not just the realized action.

3.11. Discrete-Time Replicator Systems

While much of the literature on the replicator dynamics concerns the continuous time system, it is also interesting to consider the extent to which, as in our earlier discussion of the Borgers-Sarin stimulus-response model, discrete time systems give similar or different conclusions. To do so, the first step is of course to define what we mean by the discrete-time replicator system. While several alternatives are possible, perhaps the most obvious discrete-time formulation of the asymmetric-population case (again supposing that each population has the same size) is

$$\phi^i_{t+1}(s) - \phi^i_t(s) = \Delta\phi^i_t(s)u_t(s)\,,$$

where Δ is the length of the time interval[60] and payoffs correspond to the net reproduction rate per unit of time which leads to the population share equation[61]

$$\theta^i_{t+1}(s) - \theta^i_t(s) = \frac{\phi^i_t(s)(1+\Delta u^i_t(s))}{\sum_{s'}\phi^i_{t+1}(s')} - \theta^i_t(s) = \frac{\Delta\theta^i_t(s)\left(u^i_t(s) - \bar{u}^i_t\right)}{1+\Delta\bar{u}^i_t}.$$

Note that as the period length Δ goes to 0, this equation converges to the continuous-time replicator dynamics; note also that the step size in this system depends on the absolute size of the payoffs, so that for example adding 100 to all payoffs shrinks the step size. This is because an additional 1% in absolute growth rate leads to an additional 1% population share if the total population is constant, while it becomes insignificant is all strategies have absolute growth rates that are large.

Alternatively, one can interpret the payoff function as giving the reproduction rate per period, instead of per unit of time; with this interpretation the parameter Δ disappears from the equation of motion, and the way that one models shorter time periods is by lowering all of the payoffs towards 0: in the limit of infinitesimal periods, the population is almost constant from one period to the next.

As is typically the case, the convergence and stability properties of the discrete-time replicator dynamics can be different from the continuous time version, with "long" time periods causing more of a change than small ones do. The most striking example of the effect of long time periods is Dekel and Scotchmer [1992], who show that the discrete-time replicator dynamics need not remove all strictly dominated strategies. (We follow Dekel-Scotchmer in discussing the one-population case; recall that this also describes the evolution of the two-population system from a symmetric initial position.)

[60] If we take the continuous time model more literally, we may wish to view Δ as the exponential of the time interval.

[61] The alternative system mentioned in footnote 5 leads to a corresponding discrete-time alternative.

Dekel and Scotchmer start with a nonzero-sum version of the rock-scissors-paper game as in example 3.3,

$$\begin{bmatrix} 1.00, 1.00 & 2.35, 0.00 & 0.00, 2.35 \\ 0.00, 2.35 & 1.00, 1.00 & 2.35, 0.00 \\ 2.35, 0.00 & 0.00, 2.35 & 1.00, 1.00 \end{bmatrix}$$

In this game, the mixed equilibrium (1/3,1/3,1/3) is an ESS: the equilibrium payoff is $3.35/3$; the payoff of any of the strategies against this mixture is 1. Thus we know that the mixed equilibrium is asymptotically stable in the continuous-time replicator dynamic. However, computation reveals that the discrete-time replicator dynamics (with Δ=1, so that the "large "payoffs here implicitly correspond to a nonnegligible period length) spiral outwards towards the boundary.

Dekel and Scotchmer then add a fourth strategy to the game, resulting in the game described in example 3.3.

$$\begin{bmatrix} 1.00,1.00 & 2.35,0.00 & 0.00,2.35 & 0.10,1.10 \\ 0.00,2.35 & 1.00,1.00 & 2.35,0.00 & 0.10,1.10 \\ 2.35,0.00 & 0.00,2.35 & 1.00,1.00 & 0.10,1.10 \\ 1.10,0.10 & 1.10,0.10 & 1.10,0.10 & 0.00,0.00 \end{bmatrix}$$

Recall that this fourth strategy is strictly dominated by a mixture of the other three strategies, but not by any one of the first three strategies alone, and it does very poorly against itself. However, the fourth strategy is a better-than-average response against states where it and one other strategy are scarce. Since this new strategy is not a best response to the equilibrium (1/3,1/3,13,0), that point is an ESS, and hence is asymptotically stable in the continuous-time replicator dynamic.[62] Moreover, as we

[62] This was first noted by Cabrales and Sobel [1992].

know from section 3.6, the continuous time replicator must eliminate the dominated fourth strategy starting from any interior point. However, a proof by contradiction shows that in discrete time, the share of this fourth strategy does not go to 0: if its share became small, the state would spiral out towards the boundary of the simplex corresponding to the other three strategies, and at most points on this boundary the fourth strategy has a positive growth rate.

The difference between the asymptotic behaviors of the discrete and continuous time systems raises the question of the what can be said about the general relationship between these behaviors. Standard results on the structural stability of dynamical systems (see for example, Hirsch and Smale [1974]) imply that:

1) If a steady state is hyperbolic and asymptotically stable under the continuous-time dynamics then it is asymptotically stable for sufficiently small time periods, and

2) If a steady state is hyperbolic and unstable under the continuous-time dynamics then it is unstable for sufficiently small time periods.

Because the steady state (1/3,1/3,1/3,0) is hyperbolic and asymptotically stable in the continuous-time replicator, it is also asymptotically stable with sufficiently small time periods. Consequently the issue in the Dekel-Scotchmer example is whether long or short time periods are the better description of the situation.

More generally, facts (1) and (2) above show that in many situations the discrepancy between the discrete and continuous time dynamics vanishes in the limit of smaller time periods. A notable exception is the case discussed at the end of section 3.5, where the continuous-time dynamics has a center. If a steady state is a center in the continuous-time dynamics, it is unstable in the discrete-time dynamics even for arbitrarily small period lengths. This is easily seen in a diagram we copied from Borgers and Sarin

[1995] on the replicator dynamics in the zero-sum game “matching pennies” (See also Akin and Losert [1984].)

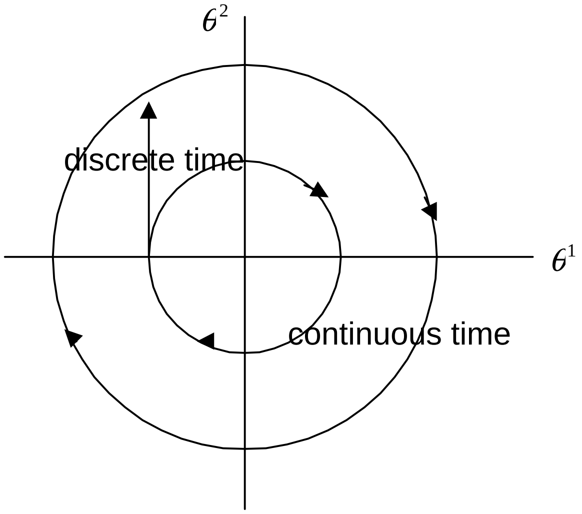


Figure 3.7

In continuous time, the system orbits around the center. To a good approximation, in discrete time, the system moves along tangents to the circle; as can be seen in Figure 3.7. As a result the dynamical system moves points on continuous time orbits to points more distant from the center.

While centers are not structurally stable to general perturbations of the dynamics, they can arise for a “fat” set of payoffs under the replicator dynamics. In particular, as noted in section 3.5 above, centers occur in all 2x2 games with a unique completely mixed strategy equilibrium in the 2-population model. Thus if one considers the narrow question of the relation between discrete and continuous time replicator dynamics, the conclusion is that there are open sets of payoffs for which the two differ even in the limit of period lengths that converge to 0. However, if one allows for other small perturbations of the dynamics, and thus considers the discrete-to-continuous time limit in dynamics “near” the replicator, the answer changes, as such perturbations will (generically) make the centers hyperbolic and either asymptotically stable or unstable, and in either case the asymptotic behavior of the system in the limit of very short discrete time periods will be the same as the asymptotic behavior in continuous time.

Appendix: Liouville's Theorem

The effect of a flow on volume[63] has important consequences for the long-run dynamics of the system. For example, it is easy to see that in a one dimensional system, if the volume of sets is reduced over time, then the system must converge to a unique globally stable steady state. More generally, fix an *n*-dimensional dynamical system, and a measurable set A in the interior of the system's domain of definition. The image of A at time t is $A(t) = \{x(t, x_0) | x_0 \in A\}$.Consider what happens to this image under the flow. The change in the volume of such a set is determined entirely by the *trace* of the matrix $Df(x)$. This is known as the *divergence* of the vector field; that is

$$div(f) = tr(Df) = \sum_i \frac{\partial f_i}{\partial x_i}$$

Specifically, Liouville's theorem says that if the divergence is 0 on an open domain *X*, the image $A(t)$ has the same Euclidean volume as the original set A as long as the system remains in the domain *X*. If the divergence is negative the volume of the image strictly decreases over time and if it is positive, the volume of the image strictly increases.

As noted, in a one dimensional system volume contraction is a strong property since it guarantees convergence to a unique globally stable steady state. Volume contraction is also significant in dimension two, since it implies that the system cannot have a non-trivial closed orbit or "cycle."

A cycle separates the plane into interior and exterior regions, and so the closed set composed of the cycle and its interior must be invariant. Then, however, the flow would be volume-preserving instead of volume-contracting on this set. Volume contraction is less

[63] By volume here we mean the ordinary Euclidean volume $vol(A) = \int_A dx$.

significant in higher dimensions, but it does imply that the system converges to a manifold at least one dimension smaller than the original system.

The case of 0 divergence is of particular use in studying the replicator dynamic. Since asymptotically stable steady states contract volume, a system with 0 divergence cannot have an asymptotically stable steady state. The standard replicator dynamic does not have divergence 0, but it can be transformed into such a system by a (smooth) change of variables. One such transformation is given by Hofbauer [1995]: First, normalize using the size of the population playing one of the strategies as a numeraire, that is set $\zeta^i(s^i) = \theta^i(s^i)/\theta^i(\tilde{s}^i)$ for some strategies $\tilde{s}^i$. Then set $v^i = \log \zeta^i$. Note that this transformation is valid only on the interior of the strategy simplex. The resulting system is *bipartite,* meaning that the terms governing the evolution of v^i depend only on the state variables corresponding to player $j \neq i$. It follows that the diagonal of the Jacobian matrix Df is zero, and so its trace (the divergence) is 0.

What does the fact that this transformed system n has zero divergence reveal about the original system? The paths of the transformed system are transforms of the paths of the original one, so that the steady states and their stability properties are unchanged. In particular, in the interior, the original system cannot have any asymptotically stable steady states.[64] However, the transformation does change the divergence of the map, so the fact that the transformed map preserves volume does not mean that the original one did. Indeed, since boundary steady states that are strict are certainly stable, volume must be contracted near such steady states in the original system. Note also that the transformed map may have a different velocity than the old one, and in particular the sets of point that are reached in finite time need not be preserved under transformations; Weibull [1995]

[64] This conclusion cannot be extended to entire system, as (1) Liouville's theorem only holds on open domains, and (2) the transformation we used is badly behaved at the boundaries of the simplex.

gives an example where a similar transformation of the replicator dynamic results in a system that reached a boundary equilibrium in finite time.

References

Akin, E. and V. Losert [1984]: "Evolutionary Dynamics of zero-sum games," *Journal of Mathematical Biology*, 20: 213-258.

Banerjee, A. and D. Fudenberg [1995]: "Word of Mouth Communication," Harvard.

Binmore, K. and L. Samuelson [1992]: "Evolutionary Stability in Repeated Games Played by Finite Automata," *Journal of Economic Theory*, 57: 278-305.

Binmore, K. and L. Samuelson [1993]: "Muddling Through: Noisy Equilibrium Selection," University College London.

Binmore, K. and L. Samuelson [1995]: "Evolutionary Drift and Equilibrium Selection," University College London.

Bjornerstedt, J. [1995]: "Experimentation, Imitation, and Evolutionary Dynamics," Stockholm University.

Bjornerstedt, J. and J. Weibull [1995]: "Nash Equilibrium and Evolution by Imitation," In *The Rational Foundations of Economic Behavior*, Ed. K. Arrow et al, (London: Macmillan).

Blume, A., Y. G. Kim and J. Sobel [1993]: "Evolutionary Stability in Games of Communication," *Games and Economic Behavior*, 5: 31-57.

Bomze, I. [1986]: "Noncooperative two-person games in Biology: A Classification," *International Journal of Game Theory*, 15: 31-57.

Borgers, T. and R. Sarin [1995]: "Learning Through Reinforcement and Replicator Dynamics," University College London.

Borgers, T. and R. Sarin [1996]: "Naïve Reinforcement Learning With Endogenous Aspirations," University College London.

Boylan, R. [1994]: “Evolutionary Equilibria Resistant to Mutation,” *Games and Economic Behavior*, 7: 10-34.

Bush, R. and R. Mosteller [1955]: *Stochastic Models of Learning*, (New York: Wiley).

Cabrales and J. Sobel [1992]: “On the Limit Points of Discrete Selection Dynamics,” *Journal of Economic Theory*, 57: 473-504.

Crawford, V. and J. Sobel [1982]: “Strategic Information Transmission,” *Econometrica*, 50: 1431-1452.

Dekel, E. and S. Scotchmer [1992]: “On the Evolution of Optimizing Behavior,” *Journal of Economic Theory*, 57: 392-406.

Ellison, G. and D. Fudenberg [1993]: “Rules of Thumb for Social Learning,” *Journal of Political Economy*, 101: 612-643.

Er’ev, I. and A. Roth [1996]: “On The Need for Low Rationality Cognitive Game Theory: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria,” University of Pittsburgh.

Eshel, I. and E. Akin [1983]: “Coevolutionary instability of mixed Nash solutions,” *Journal of Mathematical Biology*, 18: 123-233.

Farrell, J. [1986]: “Meaning and Credibility in Cheap Talk Games,” UC Berkeley.

Friedman, D. [1991]: “Evolutionary Games in Economics,” *Econometrica*, 59: 637-666.

Fudenberg, D. and E. Maskin [1990]: “Evolution and Cooperation in Noisy Repeated Games,” *American Economic Review*, 80: 274-279.

Gaunersdorfer, A. and J. Hofbauer [1995]: “Fictitious Play, Shapley Polygons and the Replicator Equation,” *Games and Economic Behavior*.

Gilboa, I. and A. Matsui [1991]: “Social Stability and Equilibrium,” *Econometrica*, 58: 859-867.

Hirsch, M. and S. Smale [1974]: *Differential Equations, Dynamical Systems, and Linear Algebra*.

Hofbauer, J. [1995]: "Evolutionary Dynamics for Bimatrix Games: A Hamiltonian system?," *Journal of Mathematical Biology*, forthcoming.

Hofbauer, J. [1996]: private communication from author.

Hofbauer, J. and J. Weibull [1995]: "Evolutionary Selection against dominated strategies," Industriens Utredningsintitut: Stockholm.

Hofbauer, J. and K. Sigmund [1988]: *The Theory of Evolution and Dynamical Systems*, (Cambridge: Cambridge University Press).

Hofbauer, J., P. Schuster, K. Sigmund and K. Sigmund [1979]: "A Note on Evolutionary Stable Strategies and Game Dynamics," *Journal of Theoretical Biology*, 81: 609-612.

Kim, Y. G. and J. Sobel [1991]: "An Evolutionary Approach to Preplay Communication," UC San Diego.

Kohlberg, E. and J. Mertens [1986]: "On the Strategic Stability of Equilibria," *Econometrica*, 54: 1003-1038.

Kreps, D. and B. Wilson [1982]: "Sequential Equilibrium," *Econometrica*, 50: 863-94.

Matsui, A. [1991]: "Cheap-Talk and Cooperation in a Society," *Journal of Economic Theory*, 54: 245-258.

Maynard Smith, J. [1974]: "The Theory of Games and Evolution of Animal Conflicts," *Journal of Theoretical Biology*, 47: 209.

Milgrom, P. and J. Roberts [1990]: "Rationalizability, Learning, and Equilibrium in Games with Strategic Complements," Stanford.

Nachbar, J. [1990]: "'Evolutionary' Selection Dynamics in Games: Convergence and Limit Properties," *International Journal of Game Theory*, 19: 59-89.

Norman, M. F. [1972]: *Markov Processes and Learning Models*, (New York: Academic Press).

Prasnikar, V. and A. Roth [1992]: “Considerations of fairness and strategy: experimental data from sequential games,” *Quarterly Journal of Economics*, 107: 865-888.

Rabin, M. [1990]: “Communication Between Rational Agents,” *Journal of Economic Theory*, 52: 1029-1050.

Ritzberger, K. and J. Weibull [1995]: “Evolutionary Selection in Normal-Form Games,” *Econometrica,* forthcoming.

Robson, A. J. [1990]: “Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake,” *Journal of Theoretical Biology*, 144: 379-396.

Samuelson, L. and J. Zhang [1992]: “Evolutionary Stability in Asymmetric Games,” *Journal of Economic Theory*, 57: 363-391.

Schlag, K. [1993]: “Cheap Talk and Evolutionary Dynamics,” University of Bonn.

Schlag, K. [1994]: “Why Imitate, and if so, How? Exploring a Model of Social Evolution,” Universitat Bonn, D.P. B-296.

Schuster and K. Sigmund [1981]: “Coyness, Philandering and Stable Strategies,” *Animal Behavior*, 29: 186-192.

Swinkels, J. [1993]: “Adjustment Dynamics and Rational Play in Games,” *Games and Economic Behavior*, 5: 455-484.

Taylor, P. and L. Jonker [1978]: “Evolutionarily Stable Strategies and Game Dynamics,” *Mathematical Biosciences*, 16: 76-83.

van Damme, E. [1987]: *Stability and Perfection of Equilibria*, (Berlin: Springer-Verlag).

Warneryd, K. [1991]: “Evolutionary Stability in Unanimity Games with Cheap Talk,” *Economic Letters*, 36: 375-378.

Weibull, J. [1995]: *Evolutionary Game Theory*, (Cambridge: MIT Press).

Zeeman, E. [1980]: “Population Dynamics from Game Theory,” *Global Theory of Dynamical Systems*, (Berlin: Springer), Lecture Notes in Mathematics, 819: 472-497.

4. Stochastic Fictitious Play and Mixed Strategy Equilibria

4.1. Introduction

This chapter examines stochastic models in the spirit of fictitious play, in which players randomize when they are nearly indifferent between several choices. One motivation for the material in this chapter is to provide a more satisfactory explanation for convergence to mixed strategy equilibria in fictitious play-like models. Another motivation for looking at stochastic models is to avoid the discontinuity inherent in standard fictitious play, where a small change in the data can lead to an abrupt change in behavior. Such discontinuous responses may not be descriptively realistic in many situations, as psychological experiments show that choices between alternatives that are perceived as similar tend to be relatively random. Moreover, a discontinuous response creates the possibility that the infrequent switching condition described in the previous chapter is violated, which opens the player to sorts of "mistakes" described in chapter 2, where she persistently makes less than her reservation value. In contrast, players can ensure they will obtain at least their reservation value in time average by using the sorts of stochastic rules we develop in this chapter.

The traditional process of fictitious play method is deterministic, except possibly when the historical average is such that the player is indifferent between several actions. Of course, for generic strategic-form payoffs, and a generic prior, there is no sample that makes any player exactly indifferent, so that typically players will use pure strategies in every period. The variations on fictitious play we discussed at the end of chapter 2 do permit players to randomize. Recall in particular the notion of asymptotically empirical beliefs, which requires that beliefs in the limit converge to the frequencies generated by fictitious play while allowing beliefs at any finite time *t* to be arbitrary. As we will see, such procedures permit players to randomize in every period, so potentially such a procedure could converge to a mixed strategy equilibrium. However, the reason that players randomize in this setup is not very satisfactory.

As we mentioned earlier, another motivation for looking at stochastic models is to avoid the discontinuity inherent in standard fictitious play, which is troubling descriptively and can lead to poor long-run performance. These considerations lead us to consider variations on fictitious play in which players randomize when they are nearly indifferent. In studying these stochastic fictitious play-like procedures in discrete time, we will argue that the asymptotic properties of these systems can be understood by reference to a limiting continuous time deterministic dynamical system. Roughly speaking, in fictitious play-like procedures, the averaging of observations over time causes the noise in the system to decrease relative to the speed with which the system moves. If the noise remains large relative to the deterministic movement of the system, then the continuous time limit is less useful, a situation considered in the next chapter.

4.2. Notions of Convergence

Our discussion of fictitious play in the chapter 2 followed the standard practice of saying that play converged if the empirical frequencies of each player's actions converged. As we noted there, this notion of convergence is very weak: Because it requires convergence only of the marginal distribution of individual players' play, it allows the possibility that joint distribution of play is correlated, and this can lead to payoffs that are very different from Nash equilibrium payoffs, as in the example of the coordination game where players always fail to coordinate. If we strengthen the notion of convergence to require convergence of the joint distribution of play, then from a frequency point of view, play in the game does resemble a Nash equilibrium. However, this response is not completely satisfactory because it allows persistent cycles. For example, in a game of matching pennies, deterministic alternation between (H,H), (H,T),(T,H),(T,T) would be viewed as a sequence that "converges" to a Nash equilibrium.

In this chapter, we will follow the approach adopted by Fudenberg and Kreps [1993] and define convergence of the learning process to mean that players' intended play converges. Note that it is not immediately obvious that this is a stronger condition than convergence of either the marginal distribution of players play, or the joint distribution of their play. However Fudenberg and Kreps use a variation on the strong law of large numbers that leads to the conclusion[65] that when the intended play converges the realized joint empirical distribution over profiles converges almost surely to the product of the intended marginals.

Following Fudenberg and Kreps [1993] we say that a strategy profile is *locally stochastically stable* if for every $\varepsilon > 0$ there is some history of play such that the subsequent probability that intended play converges to that profile is at least $1-\varepsilon$. This stochastic version of local stability does not require that behavior converge almost surely, since when behavior is stochastic there is always a small probability of unrepresentative outcomes which would lead players away from the target strategy. It also uses a very weak notion of "local," as it suffices that there be *some* history for which the convergence has high probability, as opposed to requiring that convergence occur for every history in some (suitably defined) neighborhood of the target profile.

[65] Although they show only that convergence of intended play implies the convergence of marginal distributions of play to the intended play, the same argument easily gives the stronger result mentioned here.

4.3. Asymptotic Myopia and Asymptotic Empiricism

We next investigate the extent to which procedures that are asymptotically like fictitious play are locally stable. Since we have now defined convergence to mean the convergence of intended play, behavior that follows a deterministic cycle does not converge, and in particular cannot converge to a Nash equilibrium, even if the empirical marginal frequencies converge to a Nash equilibrium strategy profile. Rather, the only way that play can converge to a mixed equilibrium is if the distribution of play in each period is mixed. This in turn is possible only if players use some type of explicit randomization, and requires us to ask why should the distribution of intended actions at a given date should be random in the first place.

This question is familiar as a critique of mixed strategy equilibrium, so it is not surprising that it reemerges here. One trivial defense is that players are willing to play a mixed strategy so long as every action in the strategy's support yields the same expected payoff. This is true by definition in a mixed-strategy equilibrium. To turn this into an ad-hoc and unsatisfactory "learning" story one could suppose that players start off with the assessment that their opponent's play exactly corresponds to the mixed equilibrium, and that players maintain this assessment unless they get "overwhelming" statistical evidence against it.[66] Suppose moreover that so long as a player does maintain this belief, he chooses his own actions according to his part of the mixed equilibrium. If each player

[66] To be somewhat more precise, fix a mixed strategy equilibrium σ_* of a 2-player game, and let player *i*'s date-*t* assessment be that $\mu_t^i = \sigma_*^{-i}$ so long as $\left\|\sigma_*^{-i} - d_t^{-i}\right\| < 1/n(t)$, with $\mu_t^i = d_t^{-i}$ if the inequality is violated. By the strong law of large numbers, the sequence $n(t)$ may be chosen to converge to infinity slowly enough that there is probability 1 that the above inequality is always satisfied so long as player *-i* follows σ_*^{-i}. Then the assessments are asymptotically empirical (since $n(t) \to \infty$) and if both players use this assessment rule, and play their part of the mixed strategy when indifferent, there is probability 1 that they play according to σ_* in every period.

follows this rule, then by the law of large numbers, neither player will reject the hypothesis that his opponent is following the mixed equilibrium, and play will indeed converge to the mixed equilibrium we specified. Of course, this "explanation" of the persistent mixing in a mixed strategy equilibrium has the defect of building a weak preference for the mixed equilibrium directly into the players' behavior. But then, the story is based on an explanation of equilibrium learning that has the same defect: players are supposed to follow the mixed equilibrium for no positive reason at all.

More generally, we can consider the stability properties of procedures that are asymptotically similar to fictitious play . A behavior rule ρ_t^i for player *i* specifies a mixed strategy based on the history of play. An *assessment for player i* is a map from histories to distributions over the space Σ^{-i} of the opponent's mixed strategies. As in chapter 2, assessments are *asymptotically empirical* if they converge to the empirical average along every sequence of observations, and a behavior rule is *asymptotically myopic* if the loss from player *i*'s choice of action at every history given his assessment goes to zero as the history grows longer.[67] Finally, a profile is *unstable* if for every positive ε, players' behavior is almost surely more than ε away from the profile infinitely often.

Proposition 4.1: (Fudenberg and Kreps): If σ is not a Nash equilibrium, then it is unstable with respect to any behavior rules that are asymptotically myopic with respect to asymptotically empirical assessments.

[67] Formally, behavior rule ρ_t^i is asymptotically myopic with respect to an assessment rule γ_i^t if for some sequence of positive numbers $\varepsilon_t \to 0$, $u^i(\rho_t^i, \gamma_t^i) + \varepsilon_t \geq \max_{s^i} u^i(s^i, \gamma_t^i)$. Note that this definition is in terms of player *i*'s expected utility, where the expectation is taken over any randomness in the play of any player. In particular a strategy that incurs a large loss with a low probability regardless of the opponents' play is treated as having a small loss. Thus this definition of asymptotic myopia is less restrictive than one that requires that player *i* only assign positive probability to pure strategies that come close to maximizing his expected payoff. Moreover, a strategy that is weakly dominated can still be a best response provided that the player's beliefs assign a sufficiently small probability (converging to 0) to opponents' strategies under which the dominated strategy incurs a loss.

The intuition for this is the same as for the corresponding result in chapter 2: if play converged to σ, players assessment would converge to σ as well, but then since σ is not a Nash equilibrium, some player would choose to deviate.

Conversely, Fudenberg and Kreps show that any Nash equilibrium is locally stochastically stable for some behavior rules that are asymptotically myopic with respect to asymptotically empirical assessments. However, the proof of this relies on the construction, sketched above, in which players start out with a strong prior belief in the particular equilibrium, and maintain that belief unless they receive overwhelming evidence to the contrary. Consequently, the stability result just cited does not do a great deal to lend credence to the idea that mixed distributions actually will arise as the result of learning. [68]

4.4. Randomly Perturbed Payoffs and Smoothed Best Responses

To develop a sensible model of learning to play mixed strategies, one should start with a sensible explanation for mixing in the equilibrium context. One such explanation is Harsanyi's [1973] purification theorem, which explains a mixed distribution over actions as the result of unobserved payoff perturbations that sometimes lead players to have a strict preference for one action, and sometimes a strict preference for another.[69] Fudenberg and Kreps [1993] develop a model of fictitious play along these lines. Before considering the application to fictitious play, it is useful to see how in the static case random preferences can provide a positive story of mixed strategy equilibrium.

[68] Fudenberg and Kreps provide a parallel, and no more satisfying, result showing that any Nash equilibrium is locally stochastically stable for behavior that is asymptotically myopic with respect to the exactly empirical assessment rule. Here the players choose to use precisely the equilibrium mixed strategy so long as their perceived loss from doing so is small.

[69] See for example, Fudenberg and Tirole, [1991] chapter 6, or Myerson [1991] for a discussion of the Harsanyi purification theorem and an explanation of the sorts of payoff perturbations it uses.

Ordinarily, the payoff to player i to the strategy profile s would be $u^i(s)$. Now however, we assume that the payoff to player i is $u^i(s)+\eta^i(s^i)$ where η^i is a random vector continuous with respect to Lesbesgue measure on a finite interval. This simplifies Harsanyi's general formulation in that the realized shock to player i's payoff depends on the action he chose but not on the actions of the other players. The basic assumption is that the random payoff shock to each player η^i is private information to that player. Consequently, the game is a Bayesian game of incomplete information, where each player chooses a rule mapping his type to a strategy.

Since each player's type only influences his own payoff, the equilibria of this game can be described in terms of the marginal distributions σ^i over each player's strategies. For each distribution σ^{-i} over the actions of i's opponents, let player i's *best response distribution* $\overline{BR}^i(\sigma^{-i})$ be given by $\overline{BR}^i(\sigma^{-i})(s^i)=\text{Prob}[\eta^i \text{ s.t. } s^i \text{ is a best response to } \sigma^{-i}]$. Since η^i is assumed to have a distribution that is absolutely continuous with respect to Lesbesgue measure, there is a unique best response for almost every type. Consequently, unlike the usual best-response correspondence, the best- response distribution is actually a function. More strongly, the absolute continuity assumption implies that the best response distribution is a continuous function.[70] Looking ahead to the learning model, this means that if the player's assessment converge his behavior will too, which is not the case with standard fictitious play.

The notion of a Nash equilibrium in games with randomly perturbed payoffs can now be defined in terms of the best- response distribution.

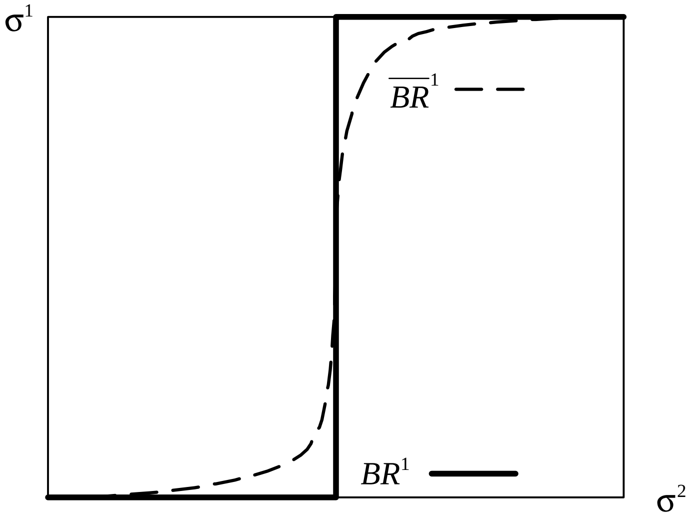
Definition 4.1: The profile σ is a *Nash distribution* if $\overline{BR}^i(\sigma^{-i})=\sigma^i$ for all i.

This distribution may be very different from any Nash equilibria of the original game if the payoff perturbations are large. However, Harsanyi's purification theorem

[70] This is lemma 7.2 in Fudenberg and Kreps [1993].

shows that, for generic payoffs in the original strategic form, the Nash distributions of the perturbed game approach the Nash equilibria of the original game as the support of the payoff perturbations become concentrated about 0. Consequently, for small supports, we can identify a Nash distribution of the perturbed game with the corresponding, possibly mixed, equilibrium of the original game.

The key feature here is that the function $\overline{BR}^i$ is both continuous and close to the actual best response function. For example, in the game of matching pennies, where player 1 wins if matches his opponent the best response correspondence BR^1, and the smooth counterpart $\overline{BR}^i$ are drawn.



Notice that generally even if the opposing player is playing a pure strategy, the smoothed best response $\overline{BR}^i$ will generally still be random, as illustrated in the figure.

At this point we can note that there are other reasons why players may use a smooth best response function such as $\overline{BR}^i$. Here are two of them:

- Random behavior can prevent the player from being "manipulated" by a clever opponent. Thinking of the game of matching pennies, if a player plays a deterministic rule, no matter how complex, an opponent clever enough to deduce what the deterministic rule is can win all the time, despite the fact by a simple 50-50 randomization the player can win nearly half the time. By explicitly randomizing when nearly indifferent, it is possible to prevent this type of manipulation. We will examine the performance of randomized rules in greater depth later in this chapter and also in chapter 8 of the book.

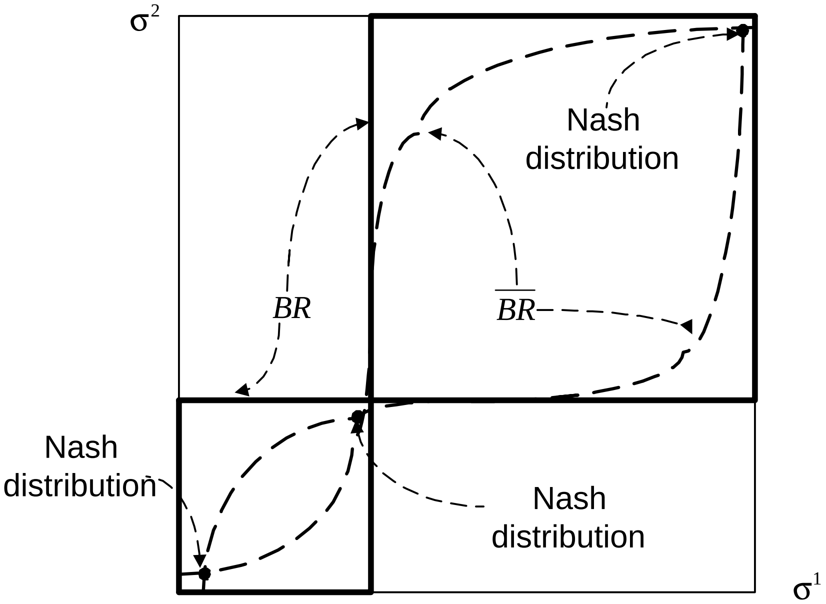
- As mentioned in the discussion of the rote learning model in chapter 3, research in psychology on threshold perception shows that when asked to discriminate between two alternatives, behavior is random, becoming more reliable (that is, deterministic) as the alternatives become more distinct. In choosing between two different strategies, one measure of the distinctness of the strategies is the difference in utility between the strategies. With this interpretation, Thurstone's [1927] law of comparative judgment becomes similar to the random utility model described above. Indeed, the picture of the smooth best response curve drawn above is very similar to behavior curves that have been derived empirically in psychological experiments (including the fact that behavior remains slightly random even far from indifference). A good discussion of psychological models, together with many references and some experimental results can be found in Massaro and Friedman [1990].[71]

The connection between a Nash distribution and Nash equilibrium may perhaps be best seen in an example. Consider the following coordination game

[71] Note, though, that the link between threshold perception and utility maximization is not immediate, because strategies may be perceptibly different even if the utility difference is quite small: We would not expect subjects in a decision problem to randomize when faced with a clear choice between $9.99 and $10.00. On the other hand, there is evidence that some subject do play strictly dominated strategies in games, at least in the early rounds of experiments; we are uncertain how to explain this behavior. In any event, it seems that some choices between strategies yielding nearly the same utility are be relatively ambiguous, while others are not.

	A	B
A	2,2	0,0
B	0,0	1,1

The best responses and smooth best responses are shown in the figure below.



Five points deserve note here. First, there are three Nash distributions corresponding to the three Nash equilibria. Second, none of these three distributions exactly coincides with the corresponding Nash equilibrium of the unperturbed game. The Nash distribution corresponding to the mixed equilibrium lies to the left of it and below, meaning that A is slightly less likely to be played than at the mixed equilibrium. Also, the Nash distributions corresponding to the pure equilibria (A,A) (upper right corner) and (B,B) (lower left corner) both involve some randomization. This is typical: if the best responses involve randomization because, for example, some utility draws make A a best response no matter what the other player does, then there cannot be any “pure” (that is, degenerate) Nash distributions. The third point is that the Nash distribution corresponding to the mixed equilibrium lies to the left and below the actual mixed equilibrium, meaning that A is slightly less likely to be played than at the mixed equilibrium. If the utility shocks are symmetric around zero, this is necessarily the case for this payoff matrix. The reason is easy to see: at the mixed equilibrium, players are indifferent between A and B. Symmetric utility shocks implies in this case that both players have an equal chance of playing A or B. In order to reduce the chance of playing A to approximately 1/3 as it is in the mixed equilibrium, the probabilities must be adjusted to lower the payoff from playing A. This will be the case only if the actual chance of playing A in the Nash distribution is smaller than 1/3.

Fourth, from Harsanyi's purification theorem we know that the distance between the Nash distributions of the perturbed game and the corresponding Nash equilibria of the original game goes to 0 with the magnitude of the payoff shocks. (Think of multiplying the original payoff shocks $\eta_i(s_i)$ by some positive ε and then sending that ε to 0.)

Finally, note that the above diagram would apply to any situation where the player's behavior is described by the smooth best-response distributions pictured, whether or not those distributions arose from unobserved payoff distributions. When these distributions are treated as exogenous and arbitrary functions, there is no obvious reason for their intersection to be called a "Nash distribution," but since doing so will not lead to ambiguity we will use that name anyway.

Corresponding to the static notion of a Nash distribution is the dynamic variation on fictitious play in which, in place of best responses, players respond to an assessment γ^i_{t-1} with a smooth approximation $\overline{BR}^i(\gamma^i_{t-1})$ to the best response. We refer to such a learning rule as *smooth fictitious play*.

4.5. Smooth Fictitious Play and Stochastic Approximation

As we just mentioned, the basic idea of smooth fictitious play is that instead of using the exact best response map BR^i as in fictitious play, players instead take an independent draw from $\overline{BR}^i$ at each date. For a fixed and smooth function $\overline{BR}^i$, we denote the associated smooth fictitious play by $\overline{BR}^i(\gamma_t^i)$, where γ_t^i is a sequence of assessment rules of the kind used in standard fictitious play. One rationale for this type of smooth fictitious play is the random utility model, in which players get an independent draw of η^i every period. A second rationale is that players might explicitly choose to randomize. We examine that explanation in section 4.7 below

The first analysis of smooth fictitious play was done by Fudenberg and Kreps [1993], who showed that the Nash distribution corresponding to the mixed equilibrium is globally stable in 2x2 games with a unique equilibrium that is mixed, provided that the degree of smoothing is sufficiently small. This result shows that smooth fictitious play provides an explanation of how learning can lead agents' play *in each period* to correspond to a mixed strategy equilibrium.[72]

To prove global convergence, Fudenberg and Kreps used the techniques of the theory of stochastic approximation. This is based on the idea that the long-run behavior of discrete-time time-averaging stochastic systems can be determined by analyzing a related deterministic system in continuous time. Benaim and Hirsch [1994] and Kanivkoski and Young [1995] use similar techniques to complete the study of smooth fictitious play in 2x2 games. Specifically, if the 2x2 game has a unique equilibrium that is strict, the unique intersection of smoothed best response functions is a global attractor for smooth fictitious play. In games with two strict equilibria and one mixed equilibrium, with probability one

[72] Their result is only for the random-utility version of smooth fictitious play, but it extends to other interpretations of the model.

the system converges to one of the strict equilibria, with the relative probabilities depending on the initial conditions.[73]

However, smooth fictitious play does not eliminate the possibility of cycling, even in games with a unique Nash equilibrium. Benaim and Hirsch [1994] show that smooth fictitious play converges to a cycle in Jordan's [1993] 3-player matching-pennies game.

Before giving a formal statement of the relevant theory of stochastic approximation, let us first develop a rough intuition. In the case of exact fictitious play, we saw in chapter 2 that along paths where there is infrequent switching, asymptotically the dynamics resemble those of the continuous-time best response dynamics after a rescaling of the measure of time. While smooth fictitious play is random, the time average of many independent random variables has very little randomness, and as a result, a similar result obtains: asymptotically the dynamics resembles that of the continuous time "near best response" dynamics

$$\dot{\theta}^i = \overline{BR}^i(\theta) - \theta^i .$$[74]

Consequently, if the stochastic system eventually converges to a point or a cycle, the point or cycle should be a closed orbit of the continuous-time dynamics. Moreover, if a point or cycle is an unstable orbit of the continuous time dynamics, then we might expect that the noise would eventually "kick" the system off of the corresponding "knife edge," so that the stochastic system can only converge to stable orbits of the continuous-time system.

In the case of 2x2 games, it is easy to see that the mixed equilibrium is unstable under the smooth best response dynamics in games with 2 strict equilibria; it is only

[73] Foster and Kanivkoski [1995] also show that these conclusions extend to the case where players forecast of opponents' play by a randomly-drawn finite-sized sample from the overall history of play, as in Young [1993].

[74] To be somewhat more precise, since the noise has a zero mean, its influence is through its variance, which is of order $1/t^2$, while the deterministic drift corresponding to BR^i is of order $1/t$.

slightly more complicated to show that the mixed equilibrium is globally stable in this dynamic in games like matching pennies with a unique mixed-strategy equilibrium. This explains the basis of the results mentioned above.

Conditions that allow the asymptotic limit points of stochastic discrete-time systems to be inferred from the stability of the steady states of a corresponding deterministic system in continuous time have been developed by Kushner and Clark [1978], Ljung and Soderstrom [1983], Arthur, Ermol'ev and Kanioskii [1983], Pemantle [1990], and Benaim and Hirsch [1995] , among others. Benaim and Hirsch [1995] also show that similar results can be obtained for convergence to closed orbits. We present the basic results without proof, and shows how to use them to characterize the behavior of smooth fictitious play in 2x2 games. The appendix illustrates the techniques involved by giving a proof of a very simple stochastic approximation result.

Consider a discrete-time stochastic processes defined on a compact set in $\Re^n$ by

$$\theta_{t+1} - \theta_t = (F(\theta_t) + \eta_{t+1}) / (t+1),$$

where the function F is smooth (C^2), and the η_t are noise terms satisfying $E[\eta_{t+1}|\theta_t,...\theta_1] = 0$.[75] In the applications to smooth fictitious play, the state space is the empirical distribution of play and the map F is $\overline{BR}(\theta) - \theta$. The noise terms are then the differences between the expected value of $\overline{BR}(\theta_t)$ and its realized value, so that the noise terms have a conditional expectation of zero, but are not in general i.i.d. or even exchangeable. The assumption that $\overline{BR}$ is C^2 is satisfied in the random utility case if the distribution of payoffs is absolutely continuous with respect to the appropriate Lesbesgue measure.[76]

[75] Here the step size at date t is $1/t$, as in fictitious play with no initial weights; the extension to positive and unequal prior weights for different players is immediate but complicates the notation. The important thing is that the steps sizes ω_t be a decreasing series of positive numbers satisfying $\sum_t \omega_t = +\infty, \sum_t (\omega_t)^2 < \infty$

[76] In the case of explicit randomization considered in below, $\overline{BR}$ is C^2 if the v^i are C^3

The idea of this literature is to find conditions under which asymptotic limit points of the sample paths of the $\{\theta_t\}$ are the stable ω-limits of the continuous-time process

$$\dot{\theta} = F(\theta) .$$

A first step is to show that almost surely the sample path lies in some invariant set of the continuous-time process.

Proposition 4.2 : (Benaim and Hirsch [1995,1996]) With probability one, the ω-limit set of a any realization of the discrete- time process is an invariant set of the continuous-time process; the components of this set are compact, connected, and contain no proper subsets that are attractors for the continuous time process.

With this result in hand, we can characterize the long-run behavior of smooth fictitious play in 2x2 games with a unique equilibrium in mixed strategies. If the smooth fictitious play arise from utility perturbations, and the distribution of the perturbations is absolutely continuous with respect to the appropriate Lebesgue measure, then the perturbed game has a unique Nash distribution.[77] Proposition 4.3 shows that smooth fictitious play converges to the Nash distribution in any game where the Nash distribution is a global attractor for the continuous-time dynamics. One way to show this latter fact is by constructing a strict Lyapunov function, and indeed Fudenberg and Kreps construct such a function in the course of providing a direct proof of global convergence.

Benaim and Hirsch [1996] use an alternative and shorter argument: They note first that since the continuous-time smooth best-response process has the form $\dot{\theta}_t^i = \overline{BR}^i(\theta_t^{-i}) - \theta_i$, all of the diagonal entries of its Jacobian are -1, so that the process is

[77] Fudenberg and Kreps [1993].

volume contracting, and second that a volume-contracting process on $\Re^2$ cannot have a limit cycle, and so must converge to a steady state.[78]

Returning to the general stochastic approximation case, the next order of business is to determine which of the ω-limit sets will be selected when there is more than one of them. Say that a steady state is *linearly unstable* if at least one of its associated eigenvalues has positive real part. The next result says that discrete-time system has probability 0 of converging to a steady state that is linearly unstable, provided that there is a non-negligible amount of noise in the evolution of every component of the state.

Proposition 4.3: (Pemantle (1990))[79] Suppose that the distribution of the noise term η_t is such that for every unit vector e_i, $E(\max(0, e_i \circ \eta_t) > c > 0$. Then if θ^* is linearly unstable for the continuous time process, $P\{\lim_{t\to\infty} \theta_t = \theta^*\} = 0$.

This result is almost enough to show that Nash distributions approximating the mixed equilibrium will not occur in 2x2 games like battle of the sexes, with two strict equilibria and one mixed one, since in these game the mixed equilibria is unstable under the exact best response correspondence. Of course, large perturbations of the best response correspondence could introduce interior Nash distributions that are stable, but we would expect that this cannot occur when the smoothed best responses are sufficiently close to the original ones. Benaim and Hirsch show that this is indeed true for smooth fictitious play that arise from sufficiently small payoff perturbations.

To complete the analysis of these 2x2 games with three equilibria, we want to verify that the process will end up at a Nash distribution approximating one of the pure

[78] As we noted in the Appendix to chapter 3, the closed orbit plus its interior must be invariant, which would contradict the flow being volume-contracting. .

[79] There are many earlier and similar results; see the references cited at the beginning of this subsection.

equilibria. That verification has three parts: first, as noted above smooth fictitious play must converge to a steady state in 2x2 games. Second, if the payoff perturbations are small, then there are asymptotically stable Nash distributions in the neighborhood of each pure equilibrium, and these are the only asymptotically stable steady states . Finally, every asymptotically stable steady state has positive probability of being the long-run outcome, again provided that there is "enough noise" in the system. This is a fairly general observation, and not limited to smooth fictitious play, but to keep things simple we will state the version that applies to fictitious play.

Proposition 4.4: (Benaim and Hirsch [1996]) Consider a two-player smooth fictitious play in which every strategy profile has positive probability at any state θ. If θ^* is an asymptotically stable equilibrium of the continuous time process, then regardless of the initial conditions $P[\theta_t \to \theta_*] > 0$.

4.6. Partial Sampling

Fictitious play requires players to track the entire past history. A variant on fictitious play has them randomly sample independently (of one another) from their "recollection" of past observations. There are two models, depending on whether players' recollections go back to the beginning of the game, or only a finite length of time.

The model where observations are draw from the entire past is very much like a smooth fictitious play, and has been studied by Kaniovski and Young [1995]. Here every past period has an equal probability of being sampled. Because all past observations get equal weight, the effect of each successive period on subsequent play diminishes at rate 1/*t*, just as in fictitious play, and once again the long-run behavior of the system can be determined by stochastic approximation techniques, as noted above. Moreover, the limit is exactly the same type of smooth near best-response dynamic discussed above.

When players not only sample randomly but also have finite memory, a situation examined by Canning [1991] and Young [1993], the situation is different, since the system does not "slow down" as data accumulates, and so the effect of the noise terms need not vanish asymptotically. Specifically, each player, plays an exact best response to a randomly chosen sample of k observations drawn without replacement from the outcomes in the $m \geq k$ previous periods. In the case $m = k$, we can interpret the bounded memory as the result of players leaving the game and being replaced; $k < m$ corresponds to a situation where each period new individuals replace the old ones, and the new ones conduct a random poll of recent players.

With this dynamic, strict Nash equilibria are absorbing in the following sense: if a strict equilibrium s^* is played for m periods in a row, it is played forever afterwards, since the only outcome anyone can remember is s^*. Moreover, play cannot converge to a non-Nash equilibrium.

Young [1993] considers the class of *weakly acyclic games*. This means that beginning at any pure strategy profile the alternating-move or Cournot best response dynamic (considering pure best responses only) converges in a finite number of steps L to a strict Nash equilibrium. Young shows that if the sample size k is less than or equal to $\frac{m}{L+2}$, and the draws are without replacement, then in this class of games the dynamics converge almost surely to a strict equilibrium. The method of proof has little to do with stochastic approximation, and instead uses the sort of Markov chain methods discussed in the next chapter.

Rather than give a proof, we will use the following example from Young's paper to illustrate this result.

	A	B

A	0,0	1, $\sqrt{2}$
B	$\sqrt{2}$,1	0,0

If the initial weights in the assessment are (1,1) for both players, then fictitious play cycles between the outcomes (A,A) and (B,B) as in the Fudenberg-Kreps example we examined above.

In this example the path length L = 1, so Young's result shows that if $k \leq m/3$, play eventually stops cycling, and is absorbed at one of the strict equilibria. For simplicity, suppose that k=1 and m = 3. Then at any date t, there is probability 1/81 that both players sample the date-t outcomes at dates t+1 and $t+2$. This implies that every time that play at any date t is either (A,B) or (B,A), there is probability 1/81 that play remains at that profile at all subsequent dates. Thus to prove that there is probability 1 that play converges to a Nash equilibrium, it suffices to show that there will almost surely be infinitely many periods in which either (A,B) or (B,A) is played. But no path can have a run of three or more occurrences of (A,A) or of three or more occurrences of (B,B), and at any date t where both (A,A) and (B,B) have been played in the last three periods, there is a nonnegligible probability that one player will sample (A,A) and the other will sample (B,B), so that there is a nonnegligible probability that the outcome in period t+1 will be (A,B).

This example is special in that the length L of the Cournot adjustment process is 1; in the general case Young shows that each step of the Cournot path has positive probability at every date. As in the example, the intuition is that the noise in the sampling breaks up the “miscoordination” in the cycles. Of course, all of this relies on the restriction to weakly acyclic games, so that the best response process itself does not cycle.

Hurkens [1994] and Sanchirico [1995] use closely related models[80] to study convergence to *CURB* ("Closed Under Best Reply") sets in general games. (CURB sets are sets of strategies E^i for each player such that, for each player *i*, every best response to probability distributions on $E^{-i} \equiv \times_{j \neq i} E^j$ is in E^i.) As in Young's paper, agents in the systems studied by Hurkens and Sanchirico ignore all observations form the sufficiently distant past.[81] Thus if every "recent" observation has belonged to a particular CURB set, the current period's play will lie in that CURB set as well; that is, CURB sets are absorbing. On its own, this result is not very interesting, since the whole strategy space is always a CURB set, but it does help to give a hint of the behavior of these "partially stochastic" systems.

The more interesting result of both Hurkens and Sanchirico gives conditions for the system to be absorbed in a minimal CURB set, that is, one that has no strict subset that is also a CURB set. (Pure-strategy strict Nash equilibria are always singleton and hence minimal CURB sets[82]; in Young's weakly acyclic games, these are the only minimal CURB sets.) This latter conclusion requires two additional assumptions. First of all, since an action might only be a best response if the opponents assign positive probability to all of their actions, the agents must have a long enough memory that the history can generate an assessment that has this full-support property. Secondly, as in Young's model, there must be a source of randomness. Sanchirico arranges both this and the "long enough memory" condition by requiring that there is a positive probability of the agent playing every strategy that is a best response to any distribution of opponents' play whose support is concentrated

[80] Hurkens considers the case where the players' samples from the recent history are drawn with replacement, while Young considers sampling without replacement, but if *k* is large, this should not be an important difference. Sanchirico's model is similar but more general in many respects.

[81] Sanchirico allows for positive but negligible weight even on observations from the distant past; the weight is required to decline to 0 sufficiently quickly.

[82] With Nash equilibria that are not strict it may be necessary to include non-Nash best responses to get a minimal CURB set: recall that CURB requires all best responses be included in the set.

on strategies played in the last *k* periods, where *k* is the number of strategy profiles. An example of a rule with this property would be for each player to first construct an assessment over opponents' play by sampling with replacement from the last *k* periods, and then either playing a best response to that assessment or continuing to play the strategy he used in the previous period.

4.7. Universal Consistency and Smooth Fictitious Play

In this section, we argue that there is a another explanation for smooth fictitious play besides the random utility model, namely that players may choose to randomize even when not indifferent as a sort of protection from mistakes in their model of opponents' play. This randomization in a sense provides a "security level," and is closely related to the use of randomized maximin strategies in the theory of two-player zero-sum games.

In our study of deterministic fictitious play, we saw that fictitious play was approximately consistent for histories that satisfy the infrequent switching condition, but that in the Fudenberg-Kreps example in which infrequent switching was not satisfied, both players got considerably less than the amount they could have guaranteed themselves by randomizing 50-50 in every period. These observations lead to the two *desiderata* for a learning rule studied in Fudenberg and Levine [1995a]. The first is *safety*, meaning that the player's realized average utility is almost surely at least his minmax payoff, regardless of the opponents' play. The second is *universal consistency*, which requires, again regardless of opponents play, that players almost surely get at least as much utility as they could have gotten had they known the frequency but not the order of observations in advance. Since the utility of a best response to the actual frequency distribution must be at least the minmax payoff, it follows that universal consistency implies safety, so we focus on the latter criterion.

Definition 4.2: A rule ρ^i is ε*-universally consistent* if for any ρ^{-i}

$$\lim\sup_{T\to\infty} \max_{\sigma^i} u^i(\sigma^i,\gamma_t^i)-\frac{1}{T}\sum_t u^i(\rho_t^i(h_{t-1}))\le\varepsilon \text{ almost surely.}$$

Notice that specifying universal consistency as an objective differs from the Bayesian approach of specifying a prior beliefs about strategies used by opponents and playing a learning rule that is a best response to those beliefs. However, any Bayesian expects almost surely to be both safe and consistent. These criteria ask for the procedure to be safe and consistent against all alternatives, and not only for those that are regarded *a priori* as having probability one.

It is obvious that no deterministic decision procedure can be safe, or by implication, universally consistent. In the game of matching pennies in which a win counts 1 and a loss -1, any deterministic decision rule can be perfectly defeated by an opponent who knows the rule, resulting in a payoff of -1 for sure. However, by randomizing with equal weights, the minmax payoff of 0 can be guaranteed almost surely. The issue is whether through an explicitly randomized fictitious play procedure, it is possible to (nearly) attain universal consistency.

The affirmative answer was originally given by Hannan [1957], and by Blackwell [1956a] who derived the result from his vector minmax theorem in Blackwell [1956b].
A good exposition of these early results can be found in the Appendix of Luce and Raiffa [1957]. The result has been lost and rediscovered several times since then by many authors, including Banos [1968], Megiddo [1980], Auer et al [1995], Foster and Vohra [1995], and Fudenberg and Levine [1995a], In the computer science literature the basic problem is referred to as the "on-line decision problem," and the result has many applications, including the problem of data compression. Our exposition is based on Fudenberg and Levine [1995b] who, using an argument from Foster and Vohra [1995],

show that universal consistency can be accomplished by a smooth fictitious play procedure in which $\overline{BR}^i$ is derived from maximizing a function of the form $u^i(\sigma)+\lambda v^i(\sigma^i)$. Formally,

Proposition 4.5: Suppose that v^i is a smooth, strictly differentiably concave function satisfying the boundary condition that as σ^i approaches the boundary of the simplex the slope of v^i becomes infinite. Then for every ε there exists a λ such that the smooth fictitious play procedure is ε-universally consistent.

Before proving this result, it is important to note that the function v^i is assumed to be not just continuous, but smooth, and that it satisfies assumptions guaranteeing a unique $\overline{BR}^i$. Moreover, the boundary condition implies strict interiority of the solution to the maximization problem; so that every strategy is played with strictly positive probability, regardless of the frequency of opponents' play.

It may also be useful to have an explicit example of a function that satisfies these assumptions. If we take $v^i(\sigma^i)=\sum_{s^i}-\sigma^i(s^i)\log\sigma^i(s^i)$, we can explicitly solve for $\overline{BR}^i$

$$\overline{BR}^i(\sigma^{-i})[s^i]\equiv\frac{\exp\left((1/\lambda)u^i(s^i,\sigma^{-i})\right)}{\sum_{r^i}\exp\left((1/\lambda)u^i(r^i,\sigma^{-i})\right)},$$

a special case referred to as exponential fictitious play, since each strategy is played in proportion to an exponential function of the utility it has historically yielded. This corresponds to the logit decision model that has been extensively used in empirical work. Notice that as $\lambda\to 0$ the probability that any strategy that is not a best response is played goes to zero. Note also that this function has the property of convex monotonicity that we discussed in chapter 3.

Finally, before proving the theorem, it is useful to define $\vec{u}^i(\sigma^{-i})$ to be the vector of utilities that accrue to different actions for player *i* when opposing players play σ^{-i}. Letting $\vec{u}^i_t = \vec{u}^i(\gamma^i_t)$, the objective function that $\overline{BR}^i$ maximizes may then be written as

$$u^i(\sigma) + \lambda v^i(\sigma^i) = \sigma^i \cdot \vec{u}^i_t + \lambda v^i(\sigma^i).$$

This is important, because it makes clear that to implement a cautious fictitious play, a player need not base his decision on the historical frequencies γ^i_t, but may base his decision solely on the historical utilities $\vec{u}^i_t$ that would have been achieved by different actions.[83]

Proof of Proposition 4.5: Set $V^i(\vec{u}^i_t) = \max_{\sigma^i} \sigma^i \circ \vec{u}^i_t + \lambda v^i(\sigma^i)$ to be the maximized value of the objective function, and denote realized utility by $u^i_t = \sum_{\tau \le t} u^i(s_\tau)$. We then define the cost to be the difference $c^i_t = tV^i(\vec{u}^i_t) - u^i_t$ between the utility that could have been received (according to the approximate function V^i) and the utility actually received. Notice that the loss $\max_{\sigma^i} \sigma^i \cdot \vec{u}^i_t - \dfrac{1}{T}\sum_t u^i(\overline{BR}(\vec{u}^i_t), \rho^{-i}_t(h_{t-1}))$ defining ε-universal consistency is just the expected value of $c^i_t - \lambda v^i(\overline{BR}^i(\vec{u}^i_t))$. Consequently, to demonstrate ε-universal consistency, we can show that for small λ the cost is small.

The increment added to the cost in period *t* if the period-t outcome is *s* is

$$g^i_t(s) = tV^i\left(\frac{(t-1)\vec{u}^i_{t-1} + \vec{u}^i(s^{-i})}{t}\right) - u^i(s) - (t-1)V^i(\vec{u}^i_{t-1}).$$

[83] Suppose that, instead of having the same utility function in each period, player i has a sequence of time-varying utility functions $u^i_\tau(s^i, s^{-i}_t)$ that are uniformly bounded. Define $\vec{u}^i_t(s^i) = (1/T)\sum_{\tau=1}^{t} u^i_\tau(s^i, s^{-i}_t)$, replace $u^i(\sigma^i, \gamma^i_t)$ in the definition of universal consistency with $\sigma^i \cdot \vec{u}^i_t$, and define smooth fictitious play as the solution to maximizing $\sigma^i \cdot \vec{u}^i_t + \lambda v^i(\sigma^i)$. Then Proposition 4.5 still holds, and the proof requires only the obvious notational change of subscripting period *t* utility. This non-stationary version of the Proposition will be used in chapter 9 when we study the choice of experts.

In other words, $c_t^i - c_{t-1}^i = g_t^i(s_t)$. The first step of the proof is to show that if, for all σ^{-i} and h_{t-1}, and all sufficiently large *t*, $g_t^i(\rho^i(h_{t-1}),\sigma^{-i}) \le \varepsilon'$, then ρ^i is $\varepsilon' + \lambda\left\|v^i\right\|$ universally consistent. This is shown by a relatively routine application of the strong law of large numbers: the idea is that the realized increment to the costs g_t^i are, given the history, independent random variables, so their average value must remain close to the average of the conditional expectations $g_t^i(\rho^i(h_{t-1}),\rho^{-i}(h_{t-1}))$. This then implies that the average value of the cost c_t^i is almost surely asymptotically bounded by ε'.

The second step of the proof is to show for any σ^{-i} that $g_t^i(\overline{BR}^i(\vec{u}_{t-1}^i),\sigma^{-i}) \le \lambda\left\|v^i\right\| + \lambda^{-1}B/t$, where B is a constant that depends only on v^i. The first term in the left-hand side of this inequality is the error introduced because $\overline{BR}^i$ maximizes V^i instead of u^i; the second term is the approximation error from replacing the change in V^i with its first derivative times the change in player *i*'s assessment, which is proportional to $1/t$. This upper bound yields the conclusion of the theorem: we choose λ so that $2\lambda\left\|v^i\right\| \le \varepsilon/2$, and observe that for sufficiently large t, $\lambda^{-1}B/t \le \varepsilon/2$ as well.

We now derive this upper bound. Let $\hat{\sigma}_{t-1}^i = \overline{BR}(\vec{u}_{t-1}^i)$ be the mixed strategy that player i will choose at date *t* given assessment γ_{t-1}^i. From the definition of g_t^i, we find that its value at date t when *i*'s opponent plays an arbitrary σ^{-i} is

$$
\begin{aligned}
g_t^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) &= \textstyle\sum_{s^{-i}} tV^i\left(\frac{(t-1)\vec{u}_{t-1}^i + \vec{u}^i(s^{-i})}{t}\right)\sigma^{-i}(s^{-i}) - u^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) - (t-1)V^i(\vec{u}_{t-1}^i) \\
&= \textstyle\sum_{s^{-i}} t\left[V^i\left(\frac{(t-1)\vec{u}_{t-1}^i + \vec{u}^i(s^{-i})}{t}\right) - V^i(\vec{u}_{t-1}^i)\right]\sigma^{-i}(s^{-i}) - u^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) + V^i(\vec{u}_{t-1}^i)
\end{aligned}
\quad .
$$

The term in square brackets is the difference between the maximized payoff given the assessments at dates *t* and *t*-1 respectively. Because we constructed V^i to be smooth, with second derivative proportional to λ^{-1}, we may replace this discrete difference with its linear approximation, and introduce an error of order no more than $\lambda^{-1}(1/t)^2$.[84] Using the envelope theorem to replace the derivative of V^i with realized utility at the optimum, and noting that the change in player *i*'s assessment from *t*-1 to t is of order $1/t$, we find, for some *B* that depends only on v^i, that

$$g_t^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) \le \sum_{s^{-i}}\left[\hat{\sigma}_{t-1}^i\cdot(\vec{u}^i(s^{-i})-\vec{u}_{t-1}^i)\right]\sigma^{-i}(s^{-i}) - u^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) + V^i\left(\vec{u}_{t-1}^i\right) + \frac{\lambda^{-1}B}{t}$$

$$= -\hat{\sigma}_{t-1}^i\cdot\vec{u}_{t-1}^i + V^i\left(\vec{u}_{t-1}^i\right) + \frac{\lambda^{-1}B}{t}$$

where the second step follows from

$$u^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) = \hat{\sigma}_{t-1}^i\cdot\vec{u}^i(\sigma^{-i}) = \sum_{s^i}\hat{\sigma}_{t-1}^i\cdot\vec{u}^i(s^{-i})\sigma^{-i}(s^i)\ .$$

Moreover, in the problem $\max_{\sigma^i}\sigma^i\cdot\vec{u}_{t-1}^i+\lambda v^i(\sigma^i)$, $\hat{\sigma}_{t-1}^i$ is the argmax and $V^i(\vec{u}_{t-1}^i)$ is the maximized value. It follows that

$$g_t^i(\hat{\sigma}_{t-1}^i,\sigma^{-i}) \le \lambda\left\|v^i\right\| + \frac{\lambda^{-1}B}{t}.$$

☑

Notice that there is a tension here between the extent to which V^i approximates u^i and the extent to which it is smooth. The smaller is λ the better the approximation V^i to u^i, and the smaller the approximation error $2\lambda\left\|v^i\right\|$ in the proof. However, smaller λ also increases the second derivative of V^i near point at which player *i* will switch strategies by a factor of λ^{-1}, increasing the loss due to "switching" $\lambda^{-1}B/t$. Consequently, a smaller λ implies that player *i* will have to wait longer before "consistency" becomes relevant.

[84] This is the only step of the proof that uses the fact that V^i is smooth. If $v^i\equiv 0$, the case of ordinary fictitious play, V^i is linear except at points where *i* switches from one strategy to another; consequently, this derivation remains valid except at switch points, which is why ordinary fictitious play is consistent when the no switching condition is satisfied.

4.8. Stimulus-Response and Fictitious Play as Learning Models

Fictitious play or smooth fictitious play are one type of learning model, giving rise through stochastic approximation theory to dynamics that resemble the continuous time best-response dynamic. One interesting class of alternative learning models is based on the idea of "stimulus-response" or "reinforcement" learning. We discussed one such model in chapter 3, the model of Borgers and Sarin [1995], and saw that it converged to the same limit as the discrete-time replicator dynamic in the limit of smaller and smaller time periods. This section presents some related models that are intended to better match the way human agents are thought to behave, and compares the models descriptive performance with that of models in the spirit of fictitious play.

4.8.1. Stimulus Response with Negative Reinforcement

Recall the basic Borgers-Sarin [1995] model: Agents at each date use a mixed strategy, and the state of the system at date *t*, denoted (θ_t^1, θ_t^2) is the vector of mixed actions played at time *t* by the two players. Payoffs are normalized to lie between zero and one, so that they have the same scale as probabilities. The state evolves in the following way: if player *i* plays s_t^i at date *t*, and the resulting payoff was $\tilde{u}_t^i(s_t^i)$, then

$$\begin{aligned}
&\theta_{t+1}^i(s^i) = (1-\gamma\tilde{u}_t^i(s_t^i))\theta_t^i(s^i) + E(s_t^i, s^i)\gamma\tilde{u}_t^i(s_t^i) \\
&E(s_t^i, s_t^i) = 1 \\
&E(s_t^i, s^i) = 0 \quad s^i \neq s_t^i
\end{aligned}$$

A striking and seemingly unrealistic aspect of this model is that if an action is played, it is more likely to be used again in the future than if it had not been played, even if the action resulted in the lowest possible payoff.

In response, Borgers and Sarin [1996] consider a more general stimulus-response model in which reinforcements can be either positive or negative, depending on whether the realized payoff is greater or less than the agent's "aspiration level." Formally, the agent's *aspiration level* in period t is denoted ρ_t^i, and we set $r_t^i(s^i) = \tilde{u}_t^i(s^i) - \rho_t^i$ to be the difference between realized utility in period t and the aspiration level. The system evolves according to

$$\theta_{t+1}^i(s^i) = (1 - r_t^i(s^i))\theta_t^i(s^i) + E(s_t^i, s^i)\max(r_t^i(s^i),0) + (1 - E(s_t^i, s^i))\min(r_t^i(s^i),0)$$

where as above $E(s_t^i, s^i)$ is the indicator function that is 1 if $s_t^i = s^i$. Thus, when the agent is "pleased" with the outcome ($r_t^i(s^i) > 0$) the probability of the corresponding action is increased, while the probability is decreased when the agent is "dissatisfied."

Note that this model reduces to the previous one (with $\gamma = 1$) in the case $\rho_t^i \equiv 0$. Much of Borgers-Sarin [1996] concerns the implications of the way that the aspiration level might vary with the agent' s observations, but the simple case of a constant but nonzero aspiration level is very interesting. Obviously if the aspiration level is greater than 1, so that all outcomes are disappointing, then the agent can never lock on to a pure action. Less obviously, if there are only two strategies H and T, all payoff realizations are either 0 or 1, the probability that H yields 1 is *p,* and the probability that *T* yields 1 is 1-*p,* (as if the agent is playing against an i.i.d. strategy in the game matching pennies) and the aspiration level is constant at ½, and the payoff realizations then $\theta_t^i(s^i)$ converges to *p.*

This strategy of randomizing with probability equal to the probability of success is known as "probability matching." Although such a strategy is not optimal, at one time psychologists believed it was characteristic of human behavior However, subsequent research has shown that moves subjects away from probability matching in the direction of optimality if they are given enough repetitions of the choice (Edwards [1961]) or offered monetary rewards,[85] (Siegel and Goldstein [1959] and consequently the claim that probability matching is prevalent has been discredited. (See, for example, the review of this literature in Lee [1971].) Thus the inability to lock on to the optimal strategy for constant intermediate aspiration levels, even with an arbitrarily long horizon, is a drawback of the stimulus response model.

Er'ev and Roth [1996] develop a closely related variation on the stimulus-response model and use it to study experimental data. The equation of motion they study is

$$\theta^i_{t+1}(s) = \frac{\max\{\nu,(1-\gamma)\theta^i_t(s)+E[s^t,s]r^i_t(s^i)\}}{\sum_{s'}\max\{\nu,(1-\gamma)\theta^i_t(s')+E[s^t,s']r^i_t(s^i)\}}\,.$$

They assume that the aspiration level follows the dynamic equation

$$\rho^i_{t+1} = \begin{cases}(1-w^+)\rho^i_t + w^+\tilde{u}^t_i(s^t)\\(1-w^-)\rho^i_t + w^-\tilde{u}^t_i(s^t)\end{cases}$$

This specification is designed to move the system away from probability matching in the long-run; the parameter ν is designed to keep the probability of strategies bounded away from zero. If we set $\nu = 0$, this reduces to

$$\theta^i_{t+1}(s) = \frac{(1-\gamma)\theta^i_t(s)+E[s^t,s]r^i_t(s^i)}{\sum_{s'}(1-\gamma)\theta^i_t(s')+E[s^t,s']r^i_t(s^i)}$$

which like the Borgers-Sarin [1996] model increases the probability of an action when it is positively reinforced, and decreases when it is negatively reinforced.

[85] The use of monetary rewards is much less common in psychology experiments than in those run by economists.

4.8.2. Experimental Evidence

The stimulus- response model was developed largely in response to observations by psychologists about human behavior and animal behavior. One such observation is the randomness and the smoothness of responses; this is of course also true of smooth fictitious play. There is not a great deal of evidence that enables us to distinguish between the two types of models on empirical grounds: Er'ev and Roth [1996] argue that their variation on the stimulus response model fits the data better than a simple fictitious play. In our view this is largely because they have many free parameters, and most learning models with enough flexibility in functional form are going to fit this data relatively well, because of its high degree of autocorrelation. Er'ev and Roth do show that their model does considerably better than either the best-response dynamic or fictitious play in the two experiments in which they examined individual play. However, the very naïve model of "always at the mixed Nash equilibrium" does marginally better than the Er'ev-Roth model in one experiment, and marginally worse in the other, which suggests that any learning model which converges to Nash equilibrium relatively quickly will do about as well as their model does.

Er'ev and Roth's work does, however, point out an important difficulty with the standard fictitious play model. When agents use the exact best response function the model predicts a deterministic course of play with abrupt switching from one action to another. This counterfactual prediction is not, however, shared by smooth fictitious play. In addition the version of fictitious play studied by Er'ev and Roth has zero prior weights, and so implies that the players will be very responsive to their first few observations, while actual behavior is known to be more "sluggish" than that. Further, the type of exponential weighting of past observations discussed above also seems likely to improve the extent to which fictitious-play like models fit the data while not increasing the number of parameters beyond those used by the Er’ev and Roth model. Cheung and Friedman [1994] have had some success in fitting modified fictitious play models to experimental data, and report that it does better than stimulus- response type models. It is true that Majure [1994] finds that replicator dynamics fit better than fictitious play, but he uses a less satisfactory method of introducing randomness, and does not consider exponentially decreasing weights on past observations.

Finally, Van Huyck, Battalio and Rankin [1996] explicitly consider exponential fictitious play as a learning model and compare it to the replicator and other stimulus response models. They report on an experimental study based on a simple 2x2 coordination game in which players receive nothing if they choose the same action and receive a unit payoff if they choose opposite actions. Four experimental designs are considered: there can be either a single population (that is, the homogenous-population case) , or two of them, and players may or may not have access to a publicly observed coordinating device, namely an assignment of the labels “player 1” and “player 2.” In the two-population case, these labels are chosen once and for all at the start of the experiment; in the one-population case labels are randomly assigned each period.

Without the labels, the situation corresponds to the simple 2x2 game

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

As we discussed in chapter 3, the mixed-strategy Nash equilibrium (1/2,1/2) is globally stable under the replicator dynamic in the homogeneous population treatment of game 1, while with asymmetric populations it is unstable. With one population and labels, the two stable points of the replicator dynamic are the efficient equilibria “A if labeled 1 and B if labeled 2” and “B if labeled 1 and A if labeled 2;” while the inefficient equilibrium where players ignore their labels is not stable.

In the experiments, subjects played between 30 and 75 times. The replicator dynamic explains the basic qualitative features of the data, but in cases in which the symmetric equilibrium is unstable due to labeling, play often remained near the symmetric equilibrium much more than is predicted by the replicator. (This raises the question of what would have happened over a somewhat longer horizon.)

Van Huyck, Battalio and Rankin examine several models of individual learning behavior. They report that they can reject the hypothesis that players are using historical performance of strategies (as they would in stimulus-response type models) in favor of the hypothesis that they are using forecast performance of strategies (as they would in smooth fictitious play type models). Indeed, in their data the model of exponential fictitious play fits the data quite well in all but one of 12 sessions.[86]

4.8.3. Learning Effectiveness

Given the difficulty in distinguishing different learning models from experimental data, and given our prior belief that people are often reasonably good at the sort of learning that is at issue in this book, we think it makes sense to ask which learning models do a reasonably good job of learning.[87] For example, in the case of smooth fictitious play, we

[86] In the anomalous session two populations with labels played each other. In a smooth fictitious play, if there is little smoothing (so that play is nearly like ordinary fictitious play) then when the system converges it should converge (approximately to) a Nash equilibrium. In this particular trial, there is little randomness observed in play, yet convergence is to a point some distance away from the Nash equilibrium. We suspect that the poor fit results from the fact that the lack of randomness in observed play is not consistent with convergence to a point some distance from any Nash equilibrium.

[87] We realize that this belief is controversial among experimentalists.

showed that regardless of opponents' strategies, players do about as well in the time average sense as if they had known the frequencies of opponents play in advance.

The stimulus-response model also is known to have some desirable properties as a learning model. The equations

$$\begin{aligned}
&\theta^i_{t+1}(s^i) = (1-\gamma\tilde{u}^i_t(s^i_t))\theta^i_t(s^i) + E(s^i_t, s^i)\gamma\tilde{u}^i_t(s^i_t) \\
&E(s^i_t, s^i_t) = 1 \\
&E(s^i_t, s^i) = 0 \quad s^i \neq s^i_t
\end{aligned}$$

of the simplest version of the positive reinforcement model studied by Borgers-Sarin correspond to what is called a *learning automaton* in the computer science literature. Narendra and Thatcher [1974] showed that against an i.i.d. opponent, as the reinforcement parameter γ goes to zero, the time average utility converges to the maximum that could be obtained against the distribution of opponents play. However, this relatively weak property is satisfied by many learning rules that are not consistent, including "pure" fictitious play, and on its own does not seem strong enough to indicate that a rule is reasonable. By contrast smooth fictitious play retains its consistency property regardless of how opponents play.

Indeed, while the "learning automaton" does well in the long run against an i.i.d. opponent, it may do very poorly even if in the long-run opponents are approximately i.i.d., as would be the case if the system is converging to an equilibrium. The reason for this is that against an i.i.d. opponent the "learning automaton" eventually gets absorbed by a pure strategy. Consequently, if the distribution of opponent's play is for a long time very different from what it will be asymptotically, the system may be absorbed at the "wrong" pure strategy before the opponent's play shifts to its long run frequency; the probability of this the reinforcement parameter and the length of time before the opponent's play shifts.

To avoid the prediction that play eventually locks on to a pure strategy, the stimulus-reponse model can be modified so that the probability of each action remains bounded away from zero, as in Er'ev and Roth [1996]. In this case Friedman and Shenker [1995] show that if opponents' play is such that eventually one strategy remains optimal for all time, then the "responsive learning automaton" will in the long-run converge to playing the correct strategy with high probability. This covers the case of a system that is converging to equilibrium, but still falls considerably short of universal consistency.

4.8.4. Fictitious Play as a Stimulus-Response Model

One important property of the stimulus-response model is that it only uses information about the learner's realized payoffs in making choices. This may be regarded as a disadvantage or advantage: on the one hand, unlike fictitious play, opponent's play need not actually be observed. On the other hand, if this information is available (as it typically is in experimental settings) it ought not be ignored.

It should be noted that there is a variation on smooth fictitious play that does use only on payoff information. Consider in particular the exponential fictitious play

$$\overline{BR}^i(\sigma^{-i})[s^i] \equiv \frac{\exp\left((1/\lambda)u^i(s^i,\sigma^{-i})\right)}{\sum_{r^i}\exp\left((1/\lambda)u^i(r^i,\sigma^{-i})\right)}.$$

Notice that to compute the probability of playing a strategy it is necessary only to have an estimate of the utility of each action $u^i(r^i,\sigma^{-i})$. Indeed, we can view these utilities as "propensities" to play strategies, much as in the stimulus response model, except that these propensities are computed and used in a different way. This suggests that players keep track of the historic utility of each action; that is, that they compute an estimate

$$\bar{u}_t^i(s^i) = \frac{1}{\kappa_{t-1}(s^i)} E(s_t^i, s^i)\left[\tilde{u}_t^i(s^i) - \bar{u}_{t-1}^i\right] + \bar{u}_{t-1}^i .$$

Where $\kappa_t(s^i)$ is the number of times player *i* has played s^i. We then set the probabilities of playing strategies to

$$\theta_t^i[s^i] \equiv \frac{\exp\left((1/\lambda)\bar{u}_t^i(s^i)\right)}{\sum_{r^i}\exp\left((1/\lambda)\bar{u}_t^i(r^i)\right)}.$$

If opponents play is not converging, this variation on exponential fictitious play is not asymptotically the same as fictitious play. This is because strategies with low probabilities are updated less frequently than those with high probabilities, while in actual exponential fictitious play both are updated equally frequently (since data on opponent's play is used). However, if we set use the alternative weighting rule

$$\bar{u}_t^i(s^i) = \frac{1}{\theta_t^i(s^i)\kappa_{t-1}(s^i)} E(s_t^i, s^i)\left[\tilde{u}_t^i(s^i) - \bar{u}_{t-1}^i\right] + \bar{u}_{t-1}^i$$

then in a large sample, this gives essentially the same result as ordinary exponential fictitious play, so is also universally consistent. Notice that this rule can be interpreted as a kind of stimulus-response model: here an action receives positive reinforcement if it does better than expected, and negative reinforcement if it does worse than expected, where the "aspiration level" is simply average utility to date. By making the probabilities a simple function of the "aspiration level" this rule avoids the need to directly combine probabilities and utilities as in traditional stimulus-response models. This "exponential fictitious play" rule strikes us as being at least as intuitive as way that aspiration levels have been introduced into traditional stimulus-response models.

4.9. Learning About Strategy Spaces

The examples we have considered all involve relatively small strategy spaces. However, in many practical applications, especially those involving repeated play (even in an experimental setting) the space of strategies can be quite large. Both the stimulus-response models and the aspiration level variation on smooth fictitious play we just discussed require a player to track the actual or potential performance of every strategy. This is impractical when the strategy spaces are very large, even in a one person game, so it is natural to ask whether there are methods that involve tracking a smaller subset of strategies. An example of such a method is John Holland's [1975] *genetic algorithm*, a good discussion of which can be found in Goldberg [1989].

A genetic algorithm is somewhat similar to a stimulus-response method, or the aspiration level variation of smooth fictitious play. There are two significant differences. First, only the performance of a small subset of strategies is tracked at any given moment of time, with randomization between these strategies based on their relative performance. Second, there are two methods by which strategies are added to or removed from the subset of strategies actively under consideration. Both of these methods are based upon coding strategies as binary strings; that is each strategy in the strategy space is assigned a unique binary string to identify it. One method of introducing new strategies is through random mutation: existing digits in existing strings are randomly changed to yield new strategies. Since this guarantees that eventually all strategies will be considered, appropriately calibrated it guarantees consistency in a stationary problem. A second method of introducing new strategies is through "crossover" which consists of randomly splitting two existing strings and mating the first half of one string with the second half of the other to create two new strings. The theoretical properties of this procedure are poorly understood, and depend in large part on the way the strategies are "coded," but it is known through practical application that for some methods of coding strategies this results in rapid convergence in relatively difficult combinatoric problems.

There are two problems with using this procedure as a model of learning in games. First, historical performance is used to rate strategies which is fine for stationary problems, but which our analysis of smooth fictitious play suggests is not so good in non-stationary problems such as games. That is, it would be better to use a weighted average of past performance, with weights proportional to the inverse of the frequency with which strategies have been used.

Second, the actual application of genetic algorithms in economic models (primarily in the context of macro-economic price-clearing models, and not in the context of games) has been to assume that an entire population of players jointly implement a genetic algorithm (rather than each individual player implementing a genetic algorithm). This is the case, for example, in Bullard and Duffy [1994], where the strategies currently under consideration are identified with actual players in the game, and players inherit strategies from previous rounds through mutation and cross-over. It is not entirely clear why individually self-interested players would wish to jointly implement a method that learns well in stationary problems, and in addition, individual play of players makes little sense. While we can view this as a model of "asking around" much as in the case of replicator dynamics, so that each player inherits strategies by asking other players what worked for them, this makes sense only if the new players cannot observe the historical performance of strategies directly. However, the algorithm requires players inheriting crossover strategies to implement them based upon performance; that is to choose among the results of crossover based upon their past performance, which contradicts the underlying idea that players to not access to this information.

Appendix: Stochastic Approximation Theory

In this Appendix we illustrate the methods of stochastic approximation by examining when the discrete time system converges almost surely to a particular state θ^*. Sufficient conditions are obtained by the study of the continuous-time system. One case where global convergence obtains is when *F* admits a "quasi-strict" Lyapunov function *V*. This is a function that is strictly decreasing along all non-stationary trajectories of *F*. If, moreover, the minimum of this Lyapunov function is an isolated steady state, then Benaim and Hirsch show that the system converges with probability one to that steady state.

To provide some intuition for the result, we give here a proof for the very simplest case of a one-dimensional state space $[-1,1]$, $F(0) = 0$, $\theta F(\theta) < 0$ for all $\theta \neq 0$. In other words, the point 0 is globally stable in the continuous -time dynamics.

In the special case that $F(\theta) = -\theta$ we have

$$\theta_{t+1} - \theta_t = \frac{-\theta_t + \eta_t}{t+1},$$

or

$$(t+1)\theta_{t+1} - t\theta_t = \frac{t+1}{t+1}(-\theta_t + \eta_t) + \theta_t = \eta_t .$$

This enables us to conclude that $\theta_{t+1} = \sum_{s=1}^{t+1} \eta_s / (t+1)$, and the convergence result reduces to the strong law of large numbers.

For general functions *F*, we can use the Lyapunov function $V(\theta) = \theta^2$ to show that the discrete-time system almost surely converges to 0.

Because the point $\theta = 0$ is a steady state in the continuous-time deterministic dynamics, and the Lyapunov function *V* is positive and strictly decreasing at all other points, it is tempting to think that *V* should be a supermartingale in the stochastic dynamics. However, consideration of the point $\theta = 0$ shows that this is not quite the case, as at this point $E[V(\theta_{t+1})|\theta_t] > V(\theta_t) = 0$. Nevertheless, this intuition is "essentially" correct, in that outside of any fixed neighborhood of 0 we eventually have $E[V(\theta_{t+1})|\theta_t] < V(\theta_t)$ for t sufficiently large. Intuitively, the "deterministic drift" of the system tends to reduce V, but since V is convex the stochastic jumps tend to increase it. However, the size of these jumps diminishes at rate $1/t$, so that the drift term eventually dominates in any region where it is bounded away from zero. For example, in the special case of $F(\theta) = -\theta$ and the η_t have the binomial distribution on {-1,1} with probability 1/ 2, we can take $V(\theta) = \theta^2$, and find that

$$E[V(\theta_{t+1})|\theta_t] - V(\theta_t) = \frac{(\frac{t\theta_t + 1}{t+1})^2 + (\frac{t\theta_t - 1}{t+1})^2 - 2\theta_t^{\ 2}}{2} = \frac{-(2t+1)\theta_t^{\ 2} + 2}{2(t+1)^2}$$

which is negative outside of the interval [-a, a] for all $t > \frac{\sqrt{2}}{a} - 1$. This can be used to show that $\{\theta_t\}$ cannot converge to a limit other than 0.

Instead of pursuing that line, we will offer a direct proof of convergence. Define $M(\theta_t) = E[V(\theta_{t+1}) - V(\theta_t)|\theta_t]$ to be the expected change in the Lyapunov function, which may be either positive or negative, and let $M^+(\theta_t) = \max\{M(\theta_t), 0\}$. Also define

$$V^*(\theta_t) = V(\theta_t) - \sum_{s=1}^{t-1} M^+(\theta_s),$$

so that

$$V^*(\theta_{t+1}) - V^*(\theta_t) = V(\theta_{t+1}) - V(\theta_t) - \max\{M(\theta_t), 0\}.$$

By construction, this is a supermartingale:

$$\begin{gathered}E\left(V^*(\theta_{t+1})|\theta_t\right)-V^*(\theta_t)= \\ E\left(V(\theta_{t+1})|\theta_t\right)-V(\theta_t)-\max\{M(\theta_t),0\}= \\ M(\theta_t)-\max\{M(\theta_t),0\}= \\ \min\{0,M(\theta_t)\}\leq 0\end{gathered}$$

The next step is to check that the supermartingale *V** is bounded below, so that we can conclude it converges almost surely substitute $V(\theta)=\theta^2$ and compute

$$M(\theta_t)=E\left[\left(\frac{(t+1)\theta_t+F(\theta_t)+\eta_{t+1}}{t+1}\right)^2-\theta_t{}^2|\theta_t\right]=\frac{2\theta_tF(\theta_t)}{t+1}+\frac{F(\theta_t)^2+\mathrm{E}\eta_{t+1}{}^2}{(t+1)^2},$$

so that

$$M^+(\theta_t)\leq\frac{F(\theta_t)^2+\mathrm{E}\eta_{t+1}{}^2}{(t+1)^2}$$

In the right-hand side of the equation for M, the first term, which corresponds to the deterministic drift of the system, is non-positive by assumption, and the second has a finite sum, so M is summable. M^+ has only this second term, and so is summable as well, and since *V* is bounded below (by 0) this implies that V^+

is bounded below as well.

Thus V^+ is a supermartingale and is bounded below, so it follows that it converges almost surely. Since $V-V^*$ is a submartingale, bounded above since *V* is bounded above and V^* is bounded below, by the submartingale convergence theorem $V-V^*$ converges almost surely. Since V^* also converges almost surely, it follows that *V* converges almost surely.

The last step is to argue that there cannot be positive probability that *V* has a strictly positive limit. Intuitively, at such a point the “deterministic force” will dominate the noise term and push the system in towards 0. For a formal argument, suppose that there is positive probability of the event that *V* is bounded away from 0; this implies that θ is bounded away from zero as well, and hence for some $\varepsilon > 0$ that $\operatorname{sgn}(\theta_t)F(\theta_t) < -\varepsilon < 0$, and so for some $\delta > 0$, $\operatorname{sgn}(\theta_t)V'(\theta_t) < -\delta < 0$. Since *V* is smooth, we have

$$M(\theta_t) = E[V(\theta_{t+1}) - V(\theta_t)|\theta_t] = E\Big[V\big(\theta_t + (1/(t+1))(F(\theta_t) + \eta_{t+1}\big) - V(\theta_t)|\theta_t\Big]) =$$

$$E\left[V\left(\theta_t + \frac{F(\theta_t) + \eta_{t+1}}{(t+1)}\right) - V(\theta_t)|\theta_t\right] = E\left[V'(\theta_t)\left(\frac{F(\theta_t) + \eta_{t+1}}{(t+1)}\right)\right] + o(\frac{1}{t^2})$$

Since V' and F have opposite signs, and are bounded away from 0 whenever *V* is, we can conclude that along any path where *V* is bounded away from 0

$$M(\theta_t) < \frac{-\lambda}{t+1} < 0$$

for sufficiently large *t* and some $\lambda > 0$.

Define $\overline{V}(\theta_t) = V(\theta_t) - \sum_{s=1}^{t-1} M(\theta_s)$. Note that $\overline{V}$ is a martingale, and that on any path $\overline{V}(\theta_t) \geq V^*(\theta_t)$. Since *V** has a finite limit almost surely, $\overline{V}$ does not have positive probability of converging to -∞. If V remains bounded away from zero, the fact that the upper bound above on *M* is negative and that *V* is bounded below (it is non-negative) implies that $\overline{V}$ converges to +∞. Thus if *V* has positive probability of remaining bounded away from zero, we would have $\lim_{t\to\infty} E\overline{V}(\theta_t) = \infty$ which contradicts the fact that $\overline{V}$ is a martingale. Since we already established that *V* converges, the fact that it cannot remain bounded away from zero with positive probability implies that it converges to zero.

☑

References

Arthur, B., Y. Ermol'ev and Y. Kanioskii [1983]: "A generalized urn problem and applications," *Cybernetica*, 19: 61-71.

Auer, P., N. Cesa-Bianchi, Y. Freund and R. Schapire [1995]: "Gambling in a rigged casino: The adversarial multi-armed bandit problem," *36th Annual IEEE Symposium on Foundations of Computer Science*.

Banos, A. [1968]: "On Pseudo-Games," *Annals of Mathematical Statistics*, 39: 1932-1945.

Benaim, M. and M. Hirsch [1995]: "Asymptotic Pseudo-Trajectories, Chain-Recurrent Flows, and Stochastic Approximation," *Journal of Dynamics and Differential Equations*, forthcoming .

Benaim, M. and M. Hirsch [1996]: "Learning Processes, Mixed Equilibria and Dynamical Systems Arising from Repeated Games," UC Berkeley.

Blackwell, D. [1956b]: "An Analog of the Minmax Theorem for Vector Payoffs," *Pacific Journal of Mathematics*, 6: 1-8.

Blackwell, D. [1956a]: "Controlled Random Walks," *Proceedings International Congress of Mathematicians*, (Amsterdam: North Holland), 1954 III: 336-338.

Borgers, T. and R. Sarin [1995]: "Learning Through Reinforcement and Replicator Dynamics," University College London.

Borgers, T. and R. Sarin [1996]: "Naïve Reinforcement Learning With Endogenous Aspirations," University College London.

Bullard, J. and J. Duffy [1994]: "Using genetic algorithms to model the evolution of heterogenous beliefs," Federal Reserve Bank of St. Louis.

Canning, D. [1991]: "Social Equilibrium," Cambridge University.

Cheung, Y. and D. Friedman [1994]: "Learning in Evolutionary Games: Some Laboratory Results," Santa Cruz.

Edwards, W. [1961]: “Probability Learning in 1000 Trials,” *Journal of Experimental Psychology*, 62: 385-394.

Er’ev, I. and A. Roth [1996]: “On The Need for Low Rationality Cognitive Game Theory: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria,” University of Pittsburgh.

Foster, D. and R. Vohra [1995]: “Asymptotic Calibration,” Wharton School.

Friedman, E. and S. Shenker [1995]: “Synchronous and Asynchronous Learning by Responsive Learning Automata,” Duke University.

Fudenberg, D. and D. K. Levine [1995b]: “Conditional Universal Consistency,” UCLA.

Fudenberg, D. and D. K. Levine [1995a]: “Consistency and Cautious Fictitious Play,” *Journal of Economic Dynamics and Control*, 19 : 1065-1090.

Fudenberg, D. and D. Kreps [1993]: “Learning Mixed Equilibria,” *Games and Economic Behavior*, 5: 320-367.

Fudenberg, D. and J. Tirole [1991]: *Game Theory*, (Cambridge: MIT Press).

Goldberg, D. E. [1989]: *Genetic Algorithms in Search, Optimization and Machine Learning*, (Reading: Addison Wesley).

Hannan, J. [1957]: “Approximation to Bayes Risk in Repeated Plays,” In *Contributions to the Theory of Games*, Ed. M. Dresher, A.W. Tucker and P. Wolfe, (Princeton: Princeton University Press), 3: 97-139.

Harsanyi, J. [1973]: “Games with Randomly Disturbed Payoffs,” *International Journal of Game Theory*, 2: 1-23.

Holland, J. H. [1975]: *Adaptation in Natural and Artificial Systems*, (Ann Arbor: University of Michigan Press).

Hurkens, S. [1994]: “Learning by Forgetful Players: From Primitive Formations to Persistent Retracts,” Tilberg Universtiy.

Jordan, J. [1993]: “Three Problems in Learning Mixed-Strategy Equilibria,” *Games and Economic Behavior*, 5: 368-386.

Kanivokski, Y. and P. Young [1994]: “Learning Dynamics in Games with Stochastic Perturbations,” Johns Hopkins University.

Kushner, H. J. and D. Clark [1978]: *Stochastic Approximation Methods for Constrained Systems*, (New York: Springer).

Lee, W. [1971]: *Decision Theory and Human Behavior*, (New York: Wiley).

Ljung, L. and T. Söoderstrom [1983]: *Theory and Practice of Recursive Identification*, (Cambridge: MIT Press).

Luce, R. and H. Raiffa [1957]: *Games and Decisions*, (London: John Wiley and Sons).

Majure, W. [1994]: “Fitting Learning and Evolutionary Models to Experimental Data,” Harvard.

Massaro, D. and D. Friedman [1990]: “Models of Integration Given Multiple Sources of Information,” *Psychological Review*, 97: 22-252.

Megiddo, N. [1980]: “On Repeated Games with Incomplete Information Played with Non-Bayesian Players,” *International Journal of Game Theory*, 9: 157-167.

Myerson, R. [1991]: *Game Theory*, (Cambridge: Harvard University Press).

Narendra, K. and M. Thatcher [1974]: “Learning Automata: a Survey,” *IEEE Transactions on Systems, Man and Cybernetics*, 4: 889-899.

Pemantle, R. [1990]: “Non-convergence to unstable points in urn models and stochastic approximations,” *Annals of Probability*, 18: 698-712.

Sanchirico, C. [1996]: “A Probabilistic Model of Learning in Games,” *Econometrica*, forthcoming.

Siegel, S. and D. A. Goldstein [1959]: “Decision- making behavior in two-choice uncertain outcome situations,” *Journal of Experimental Psychology*, 57: 37-42.

Siegel, S. and D. A. Goldstein [1959]: “Decision-making behavior in two-choice uncertain outcome situations,” *Journal of Experimental Psychology*, 57: 37-42.

Thurstone, L. [1927]: “Psychophysical Analysis,” *American Journal of Psychology,* 28: 368-389.

Van Huck, J., R. Battalio and F. Rankin [1996]: “On the Evolution of Convention: Evidence from Coordination Games,” Texas A&M.

Young, P. [1993]: “The Evolution of Conventions,” *Econometrica*, 61: 57-83.