Appendix B

Draft 0.11 — July, 1988
Preliminary and Incomplete

A Theory of LEARNING, EXPERIMENTATION, AND EQUILIBRIUM IN GAMES

Drew Fudenberg and David M. Kreps
Department of Economics, MIT, and Graduate School of Business, Stanford

We are grateful to Robert Anderson, David Levine, and Ramon Marimon for helpful comments. The financial assistance of the National Science Foundation is gratefully acknowledged.

INTRODUCTION

In recent years, noncooperative game theory, and especially Nash equilibrium analysis, has been used in the study of many economic situations. Along with the many and varied applications has come persistent criticism: Why (or when) is equilibrium analysis appropriate? Where do equilibria come from? How does one choose among many equilibria in making predictions? In this monograph, we begin to develop one sort of answer to these questions. We develop a model where equilibrium analysis is appropriate because players hold fairly strong and approximately consistent beliefs about each others' actions. We suppose that those beliefs come from past experiences in similar situations. And we see how the ways in which players will "experiment", to test their beliefs, can influence equilibrium selection.

The formal literature on the foundations of Nash equilibrium (Aumann, 1987; Brandenburger and Dekel, 1987) interprets equilibrium as a situation in which the players have identical beliefs about how each will play. This literature, taken literally, may not justify applying equilibrium analysis to the study of economic problems, becuase the requirement of exactly identical beliefs is very strong.

This raises two questions. First, in what ways can the assumption of identical beliefs be relaxed? Second, how might the players come to have identical or almost identical beliefs? Satisfactory answers to both of these questions seem necessary in order to trust in the use of equilibrium analysis. And since we have both contributed to the flood of papers that use equilibrium analysis to study economic phenomena, we are particularly interested in looking for answers.

Of course, the best reason for trust, or the clearest reason to dismiss this form of analysis, must come from empirical and experimental evidence. Experimental economists have not been slow to test some of the game theoretic models that populate the journals, and we would assert that the evidence, while mixed, is not completely discouraging. (To be entirely self-serving, we direct the reader to Camerer and Weigelt, 1988, and to Banks, Camerer and Porter, 1988.) Empirical evidence is coming in more slowly, but what there is is also not damning. (See,

for example, Bresnahan, 1988.)

But, to supplement these empirical/experimental studies, one would like some theoretical reason to think that Nash equilibrium analysis is appropriate even when the common knowledge structures that formally justify it are absent, and even when players behave in less than perfectly and completely rational fashion.

There has long been a verbal tradition that gives such a reason. Suppose that, in a game with finitely many strategies for each player, we are studying a particular strict equilibrium. [1] Then, of course, it is not necessary that every player is certain of the strategies of his opponents, in order to find that his part is optimal. For each player to be choosing according to the equilibrium, it will be sufficient that each player attaches high probability to his opponents following their parts of the equilibrium. There is no need that beliefs are identical; simply that each player, viewing the problem as a single-person decision problem, has come to have beliefs about his fellow players that are close to the equilibrium. Following this observation, one can tell a story about why such "roughly consistent" beliefs might persist — if the players have engaged in this game in the past, and if they have any empirical sense at all, then they will use the data derived from past experiences to predict how their opponents will act in the future. If beliefs somehow settle down "close to" the equilibrium, a reinforcement feedback occurs — players play their part in the equilibrium, which will lead others to expect with even greater certainty that the equilibrium predicts what their opponents will do. [2]

Moreover, this suggests a further story as to how beliefs might come to be approximately consistent, namely through the same learning process. To take a very simple example, imagine that the game being played is a simple two person, two-by-two coordination game. If the players play up-right or down-left, each gets $1. If they play up-left or down-right, each gets nothing. If the players play this game once, we would be surprised to find them coordinating

[1] By a *strict* equilibrium, we mean one in which each player's strategy is uniquely a best response to the strategies of others.

[2] A formal criterion for this sort of *local stability* will be given in section 4. We will, however, take this opportunity to note that we do not regard it as sufficient to show that if beliefs are "close" to a strict equilibrium, then behavior next period will probably conform to the equilibrium. We ask for a bit more, namely that once beliefs are "close" to an equilibrium, then with probability approaching one they stay close forever after.

their actions. But if they play it over and over, we expect that they will learn how to coordinate. Note well, in this case it may well be as likely that they will "learn" to coordinate in one way as in another. But we do expect that, eventually, they will figure out how to coordinate. Of course, this is a very simple example, and it is far from clear that in every game a simple process of learning will lead to an equilibrium. (That is to say, in our analysis, we will not be able to provide the reader with a general global convergence result.) But learning models give us a reason to think that roughly consistent beliefs might emerge; and our formal analysis of learning models will suggest situations in which this can be guaranteed. [3] While we neither show nor believe that equilibrium analysis is appropriate for all situations, we think that are results suggest when it might be.

Note that these stories about learning, even told this informally, suggest some problems in cases where the equilibrium is not strict. Mixed strategies would seem to cause problems. Perhaps more importantly from the point of view of many of the applications that we see is that problems would arise with equilibria that are not strict because they arise from an extensive form game. Put somewhat differently, repeated play of an extensive form game might not tell players much about how each other would act "out of equilibrium," and so there is less reason to suppose that beliefs about out of equilibrium actions are roughly consistent.

Note also that this story is very different in spirit from the introspective analysis justification for Nash equilibrium that one sometimes finds in the literature. That is, game theory is traditionally distinguished from choice theory by the notion that each player is able to put himself in the situation of his opponents, to see what their motivations are and what actions they might take. It would seem that no allowance is made for this type of analysis in the informal story. We shall return to this point, but for now we offer the following elaboration of the story: Introspective analysis may well be used by players in forming their "prior" beliefs as to what will happen when the game is played. This, of course, gives us a first answer to the question: Where do these almost-consistent beliefs come from? But players are not so sure about their powers of analysis that they will be certain of their conclusions, nor will they hold onto beliefs based on introspection in the face of overwhelming evidence to the contrary. The informal story, then, suggests why such beliefs derived of introspective analysis might be reinforced

[3] For a different approach to learning in pure coordination games, an approach which uses standard equilibrium analysis, see Crawford and Haller (1987).

in repeated play. Finally, note that this informal story depends on the situation in question being repeated. There are many contexts to which Nash equilibrium analysis has been applied in which the game played is quite different from any other game players have encountered. In such contexts, the story above gives little comfort. [4]

There have been a few formal analyses of this verbal story. Huang (1985) shows how, in such a story applied to extensive form games (and non-strict equilibria), beliefs might never move from a non-Nash point. This analysis suggests an important point of departure for this study: In extensive form games, if Nash equilibrium (or some refinement of Nash equilibrium) is to be the solution concept, then it becomes necessary that players receive evidence about what will happen out of equilibrium. Otherwise, radically divergent out-of-equilibrium conjectures might persist, giving rise to non-Nash equilibria as stable points. [5] Game theorists who stress evolutionary approaches also analyze this sort of story — see, for example, Canning (1987) and Friedman (1988).

In this monograph, we will add to the formal treatment of this sort of story, stressing two particular and related points: bounded rationality and experimentation. Our analysis makes use of players who are boundedly or limitedly rational. In our models, players are somewhat rational in that they learn by their experiences and, given what they think they know, they nearly always optimize in the near term. When they don't choose what looks like near-term optimal actions, they do so in order to experiment with other options they may have; moreover these experiments are chosen somewhat sensibly. But players are not completely rational: They do not work out a grand, dynamically optimal strategy. In other words, within a small scale, they act (nearly) optimally. Within the larger scale, they act "sensibly". [6]

The notion that players experiment with (seemingly) suboptimal actions is crucial to the models that we build. We should stress that "suboptimality" in the previous sentence means suboptimal in the short-run — the motive for these

---

[4] When we speak later of "similar games and inference across information sets", we will attempt to brighten this rather bleak picture.

[5] This point may be a bit opaque without an example. We will return to it, with an example, in section 5.

[6] The distinction here is familiar from work on learning and rational expectations literature. In that literature, there are models of rational learning, such as Bray and Kreps (1987), and models of boundedly rational, sensible learning, as in Bray (1982). This paper, then, is in the tradition of the second sort of model.

experiments is to see if one's current conjectures about the actions of opponents out-of-equilibrium are in fact correct. But, on grounds of bounded rationality, these experiments are not truly optimal in the long-run — players do not solve the fantastically complex problem (more complex than a multi-armed bandit problem, since the arms of the bandit are simultaneously solving a similar problem) of finding the best experimentation procedure. Indeed, if players discounted payoffs between periods, the theory of the multi-armed bandit would suggest that, at some point, experiments would cease. In our models of behavior, we assume that this does not happen. The rate of experimentation is presumed to vanish if there is a strategy which seems clearly optimal, based on the evidence from play. But the rate of experimentation does not vanish too quickly: If given the opportunity to do so, each player will try every possible action infinitely often over the course of infinitely many repetitions of play. These experiments then generate enough information about out-of-equilibrium play so that non-Nash profiles are not stable.

Besides enabling us to justify, at least in part, the informal story supporting equilibrium analysis, our model provides a few further insights that the pure theory of Nash equilibrium does not provide. For one thing, we will see that a certain level of dispersion in beliefs about what players are doing is not inconsistent with Nash equilbirium. For another thing, our study of local stability will suggest that certain classes of Nash equilibria are, from our perspective, not very likely to persist. At one level, we are led to question the reasonableness of mixed strategy equilibria, at least insofar as our story is the motive story for equilibrium.[7] More interestingly, we will be able to relate broad classes of experimentation procedures to various equilibrium refinements.

Our analysis is particularly germane to refinements of the sort given in Banks and Sobel (1987) and Cho and Kreps (1987). The stories that surround those refinements concern deviations that are accompanied by unmodelled "speeches." These stories about speeches suffer from the weakness that they are not fully consistent or complete: Players make out-of-equilibrium inferences, but there is no formal story about why there would be any out-of-equilibrium actions to observe. Without such a formal story, it is difficult to evaluate an equilibrium refinement. One would like to know if there is a plausible, internally consistent

[7] While mixed equilibria are locally stable for very specially selected behavior, it seems unlikely to us that the requirements on behavior would ever be met. But, as we shall see, the purification theorems suggest a way around this difficulty.

story of (infrequent) deviations that justifies only the allowed inferences.

An example of a complete theory is one in which deviations are the result of mistakes or trembles. Here allowed inferences are those which can be justified by mistakes, with the further condition (perhaps) that each player is assumed to tremble independently of others. Another complete theory is implicit in Fudenberg, Kreps and Levine (1988), which concerns games with a small amount of payoff uncertainty. Here deviations occur when a player's payoffs are different than had been supposed to be likely.

The experiments aspect of our model yields a third kind of complete theory of equilibrium refinements: Out-of-equilibrium actions are interpreted as experiments, and by imposing conditions on the likelihood of various experiments, we can obtain different sorts of refinements. More specifically, when players are learning from past experience, they will have more observations about, and thus be more certain of, play along the prevailing "equilibrium path" than about play that is only (rarely) observed when someone experiments. These differences in degree-of-certainty suggest that certain types of experiments are more likely than others: Players have relatively little incentive to experiment with actions that cannot yield a higher payoff than the equilibrium for any response by their opponents. This leads to the definition of "equilibrium domination" and "conditional domination." Briefly, a strategy is equilibrium dominated if it yields strictly less than the equilibrium payoff for all strategies of the opponents; and conditional dominance extends this idea to players who are off of the equilibrium path. Our learning-and-experimentation model provides a justification for restricting attention to those sequential equilibria where beliefs assign probability zero to actions that are conditionally dominated. In signaling games, these equilibria are closely related to those that satisfy the intuitive criterion of Cho and Kreps (1987). But when applying our model to signaling games, it seems natural to go beyond the intuitive criterion to a more restrictive refinement that we call "co-divinity." Co-divinity is very close in spirit to Banks and Sobel's (1987) notion of divinity. As we will see, however, the discipline imposed by working with a complete theory leads to a refinement that is a bit different from divinity.

A fourth dividend from our approach is that it suggests a somewhat natural framework for thinking about inferences across different games. That is to say, in the basic story we will tell, players infer what their opponents will do at a particular information set from what has happened at that *exact* information set in previous play of the game. If the game in question is indeed repeated over

and over, then this might well provide players with the sort of rich data base needed to have substantial confidence in their predictions about their opponents. But it seems unreasonable to expect the exact same game to be repeated over and over; put another way, if we could only justify the use of Nash analysis in such situations, we would not have provided much reason to have faith in the many widespread applications that are found in the literature. Faith can be greater if, as seems reasonable, players infer about how their opponents will act in one situation from how opponents acted in other, similar situations. One is led naturally to model cases in which players make such cross-information set and cross-game inferences, and one can then look formally both at questions of stability and effectiveness of a supposed "similarity" that players think they see.

Indeed, it is here that we think we can connect our approach to the story of introspective analysis. What is going on in such analysis, it seems clear, is that players are using their experience from other games to infer what will happen in the current situation. The grounds for calling situations "similar" are extremely complex and are buttressed by involved logical analysis, but no one would trust to such analysis or to the "similarities" that this analysis suggests without some reinforcement from observations.[8] We will not, in our simple models, provide anything are complex as the similarities that are suggested by the logic of game theory. But, in principle, we believe that the logical application of game theory is of a kind with the simple similarities that we do model.

Our approach has some substantial drawbacks. Because it is based on models of limitedly rational behavior, it is open to the criticism of being *ad hoc*. In our view this criticism is entirely valid. We will, as much as possible, attempt to give results for large classes of behavior, moving beyond a single parametric specification for inference or for experimentation. Even so, all members of our "broad classes" will have some very restrictive qualitative features built in; features which, if changed, could change dramatically the results obtained. We think that the class of behaviors that we study is interesting, but our interest in this class should not be taken to mean that we believe all other classes to be uninteresting. In fact, quite the reverse is true. We have picked this class of behavior because it is rich enough to give us a number of preliminary insights, and also to allow us to suggest which of these insights might not survive in different looking specifications. But it does not give us everything we might want. (For example,

[8] Another way to say the same thing is to repeat that the real test of whether game theory is useful as a tool of analysis is empirical.

our classes of behavior are not very good on justifying mixed equilibria. Compare with the independent work of Canning (1987), in which mixed equilibria can arise entirely naturally, because Canning builds a source of "within-equilibrium" variability into his model.) We will try throughout to mention some alternative specifications and how they might change our results. But, in the end, our analysis is almost entirely based on a particular sort of model, which is only one of the many that might be contemplated.

In fact, in many places our specifications of rules for behavior are driven by a desire to place standard equilibrium notions within our framework of experimentation and learning. Because of this, the assumptions we make about behavior sometimes have unappealing aspects. We hope that the reader will draw from this the same conclusions that we do: Insofar as we must take forms of behavior with unappealing aspects in order to "justify" standard equilibrium concepts, the appeal of the standard concepts is called into question. We conclude this study convinced that, at least within our general framework, standard equilibrium concepts are less than the best reduced form solution concepts. The reader will, we expect, be similarly convinced.

Secondly, our formal analysis is restricted to a small class of games, in order to keep the analysis from getting in the way of any insights that might be derived. We excuse ourselves from the task of giving a theory for general games because, it seems to us, a theory that is general in terms of the games it encompasses but rather special in the class of behavior it allows would not be worth the cost of the generality.

To conclude this introduction, we now outline what we will do. Part I gives our basic analysis for the concept of Nash equilibrium. A basic model is presented, and general classes of "reasonable" behavior are identified. We then connect those general classes of behavior with the concept of Nash equilibrium, showing that outcomes will be locally stable for these classes of behavior (in a sense to be made precise) if and only if the outcomes correspond to Nash equilibrium. We say what little we know about global stability and about models where players are randomly matched. Part II investigates the consequences of assuming that players are sensitive to small *ex ante* differences in payoff if those differences are large *ex post*. We are led to redo much of the analysis of Part I at a higher level — in place of outcomes, the basic object of study becomes full behavior strategies, and in place of Nash equilibrium we find that sequential equilibria emerge as the candidate "outcomes" of our behavioral models. Part III treats two sorts of

rfinements based on extensions of equilibrium dominance: P-perfection, which we develop for general games; and codivinity, which is concerned with the class of signaling games studied in Cho and Kreps (1987) and Banks and Sobel (1987). In Part IV, we move on to the notion of inference across "similar" situations in different games... In a brief epilogue, we offer some concluding remarks.

PART I — NASH EQUILIBRIUM

I.1. FORMULATION

Consider a finite N-player extensive form game. The game tree (collection of nodes) will be denoted by T (assumed to be a finite set), $\succ$ will denote precedence in the game tree, X the subset of nodes where a player is called upon to take an action, Z the subset of terminal nodes, $n(x)$ (for $x \in X$) the player whose turn it is to move at node x, H the set of all information sets (a partition of X), $n(h)$ the player who moves at information set $h \in H$, H^n the set of information sets belonging to player n, $A(h)$ the set of actions available to player $n(h)$ at information set h, and $u^n(z)$ the payoff to player n at terminal node z. We will follow the practice of putting all of nature's moves at the start of the tree, with W denoting the possible initial nodes (moves by nature) and ρ the probability distribution over W that gives the distribution of nature's moves. (For more detail about this type of extensive form representation, see Kreps and Wilson, 1982.)

We make one important and nontrivial assumption about the extensive form: For each player $n = 1,...,N$, no node $x \in h \in H^n$ precedes another node $x' \in h' \in H^n$, where this condition includes the case $h = h'$. This guarantees that the game has perfect recall, of course, but it does an enormous amount more — it guarantees that, in any course of play, no player can ever be called upon to move twice. As we shall see, this assumption simplifies our analysis in very substantial ways; we will say later a few things about doing away with this assumption. Given this assumption, the strategy space of player i is fairly trivial: A pure strategy is an element s^n of $S^n = \Pi_{h \in H^n} A(h)$. For mixed strategies, all that is important to the computation of outcomes are the marginal distributions over actions at each information set (that is, the behaviorally mixed strategies), and we write $\Delta(A(h))$ for the set of probability mixtures on $A(h)$, so that the set of (behaviorally) mixed strategies for player n is $\Sigma^n = \Pi_{h \in H^n} \Delta(A(h))$. We

will assume that players' randomizations are independent — correlated mixed strategies are ruled out. [9]

We imagine that our N players play this particular game over and over again, against each other. (We will consider in a later section the more complex situation where there is a large population of players who are randomly matched each round.) A "play" of the game results in a particular terminal node Z, so that a history of k plays of the game is an element of Z^k. We write ζ_k for such a k-length history. We assume that each player, at the end of each round, knows what happened (what was the outcome z), so that players begin stage $k+1$ knowing ζ_k.

We assume that our N players, in round $k+1$ of the game, play according to some behavioral rules that may depend on the history ζ_k of past play. [10] We write $\phi(k+1,\zeta_k)$ for the N vector $(\phi^1(k+1,\zeta_k),...\phi^N(k+1,\zeta_k))$ of behavior strategies for the N players in round $k+1$. We write $\phi^n(k+1,\zeta_k,a)$ for the probability that player n takes action $a \in h \in H^n$ in round $k+1$ at history ζ_k, if h is reached. And we write $\phi = (\phi^1,...,\phi^N)$ for a specification of behavior rules of play for our N players for every round $k = 1,2,...$ of the game.

We will assume that our N players, in deciding what to do in any given round, do so on the basis of conjectures they have about the actions of the other players. We write $\sigma^{-n}(k+1,\zeta_k)$ for the conjectures of player n about the actions of the other players at date $k+1$ as a function of past history ζ_k; formally $\sigma^{-n}(k+1,\zeta_k)$ is an element of $\Pi_{h\in H^{-n}}\Delta(A(h))$, where $H^{-n} = \cup_{n'\neq n}H^{n'}$. We assume that no player ever assigns zero probability to some action by his opponents, so that $\sigma^{-n}(k+1,\zeta_k)$ is strictly positive. We also assume that each player maintains the hypothesis that his opponents choose their actions independently, so that the joint probability assessed by player n of any sequence of actions by his opponents is the product of the probabilities of each individual action.

Every specification of behavior rules ϕ determines in obvious fashion a "law of (stochastic) motion" for the repeated game: $\phi(1)$ determines a probability distribution over the first outcome of the game ζ_1, and then $\phi(2,\zeta_1)$ determines transition probabilities to ζ_2, etc. (Note that the independence of different play-

[9] If one wants a theory for correlated equilibria to emerge, one should build into the extensive form the correlating device. That is, rely on the observation that a correlated equilibrium is a Nash equilibrium for the game with the correlating device made explicit.

[10] For $k=0$, the usual convention that Z^0 is some convenient singleton set is followed.

ers' mixed strategies will play a role in computing these transition probabilities.) At a very formal level, the program for this first part of the monograph is: By imposing restrictions on the behavior rules ϕ that players use, what can be said about the dynamics of play in the repeated game? Will play eventually settle down to a Nash equilibrium? If play gets close to looking like a Nash equilibrium, will it stay there? Can play settle down to something that doesn't look like a Nash equilibrium? [11]

[11] We should stress: Because the game is repeated, there can be many equilibria of the repeated game. We use the term "Nash equilibrium" here to mean an equilibrium of the single-stage game. It will become apparent shortly why it is that we are able to restrict attention to static equilibria.

I.2. AN EXAMPLE

Part I concerns behavior rules that lead to a theory of Nash equilibrium. To obtain this theory, we study rules that have three properties: (i) Players end up with "enough" observations of their opponents' play to learn enough about their opponents' strategies. (ii) Players draw the "appropriate" inferences from their observations. (iii) Players, in their choose of action, eventually relect the information they gather, in the sense that they choose best responses to what they assess as their opponents' play. In order to understand some of the issues that are involved, we turn to an example. Consider the game depicted in figure I.1.

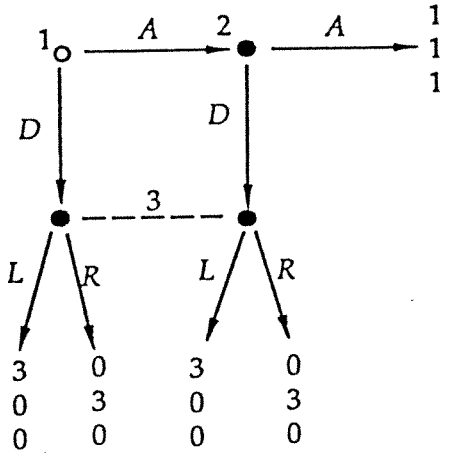

Figure I.1.

In this game, player 1 begins, and gives the move either to player 2 or to 3. If 2 gets the move, then 2 can either end the game or give the move to 3. If called on to move, 3 is unaware of whether 1 gave him the move directly, or if 1 moved to 2, who then gave 3 the move.

The reader will note that in no Nash equilibrium of the game does 1 give the move to 2 who then ends the game. For whatever player 3 does (either as a pure or a mixed strategy), either 1 or 2 (or both) strictly prefer that 3 be given

the move.

We will now describe dynamic behavior rules for the three players. For player 3, things are simple — he always chooses to move to the left, given the chance. For 1 and 2 things are a bit more complex. Player 1 has beliefs about what both 2 and 3 will do if put on the move. At any point in time, those beliefs are given by a Dirichlet posterior distribution. That is, player 1, at any time t, assigns positive integers to the two actions (across and down) of player 2 and to the two actions (left and right) of player 3, and the respective ratios of those pairs of integers give the ratios of the probabilities that 1 assesses about 2 and 3, respectively. Suppose that player 1 assigns, in round k, the numbers 100 to A and 1 to D, and 200 to R and 1 to L. Then player 1, at time t, assesses probability $100/(100+1)$ that 2 will move A, $1/101$ that 2 will move D, $200/201$ that 3 will move R, and $1/201$ that 3 will move L. Player 1 then chooses whichever action (across to 2 or down to 3) gives him the higher expected value, given those beliefs: Choosing D will net $3 \times (1/201)$ and A will net $1 \times (100/101) + 3 \times (1/101)(1/201)$ (or so 1 believes, given his beliefs about the behavior of players 2 and 3 described above). Thus player 1 chooses A in this instance.

For player 2 the same is true, although he only really needs beliefs about player 3. (Still, we will give 2 beliefs about 1, for purposes of later discussion.) Suppose that in round k 2 assigns integers 300 to A for 1 and 1 to D, and 400 to L for 3 and 1 for R. Then 2's *ex ante* expected payoff if he elects to choose A is $(1/301)(1/401) \times 3 + (300/301) \times 1$, while D nets $(1/301)(1/401) \times 3 + (300/301)(1/401) \times 3$. Note well that we did *ex ante* calculations for 2's payoff, so we had a need of 2's assessment concerning 1's action, those assessments are irrelevant to a relative comparison of the two expected payoffs. [12] In any event, 2 nets more moving A than D given his beliefs about 1 and 3, and so we can imagine that 2 will choose A.

This gives the behavior in round k of players 1 and 2 as a function of their beliefs. We will now describe the evolution of those beliefs, and (therefore) the dynamic law of motion of behavior in repeated play of this game with these three players. Player 1 began the game (at time zero) assigning integers 10 to A for 2

[12] Although it may be a trifle mysterious at this point, we cannot resist foreshadowing later developments: The difference between Part I and Part II is that in Part I we will use this sort of *ex ante* calculation and in Part II we will use *ex post* or conditional calculations. In the final paragraph of this section, we will say why this can make a difference.

and 1 to D, and 200 to R for 3 and 1 to L. After each round of play, player 1 increases by one the integer assigned to any action that he sees chosen. So if, in the first round, player 1 chooses A as does player 2, then in his computations for the second round, player 1 assigns integers 11 to A for 2 and 1 to D, and (since he got no information in the first round about player 3) the same 200 to R and 1 to L as before.

Player 2 updates his beliefs similarly; and he starts with 20 for A for 1 and 1 for D, and 400 to L for 3 and 1 for R.

It isn't hard to see where this behavior will lead. Players 1 and 2 each choose A in the first round, because each believes (given their Dirichlet priors — those initial assignments of integers) that 3 is going to take an action they won't like. Accordingly, they get no evidence about what 3 would actually do (which, recall, is to choose L), and in the second round each will again choose A. And so on. Neither chooses to move D, because each fears what 3 will do. Of course, it is crucial here (to get both to choose A) that their fears are based on inconsistent priors about 3. But neither ever gets any evidence that their fear is misguided, and so each continues with A. Of course, each becomes more convinced, as evidence accumulates, that the other will move A. But this evidence doesn't cause either to change from A. Only evidence about 3 could do that, and no such evidence is being accumulated.

So we see that if the players' behavior is to be such that non-Nash outcomes are not "stable," then their play must involve experimentation with actions that are not currently perceived as optimal.

The motivation for this experimentation is that the players are not certain that their predictions are correct. Accordingly, if evidence accumulates that confirms their predictions, they will experiment less and less. However, if we are to avoid getting stuck at non-Nash outcomes, we must assume that players experiment "enough" so that, given their inference procedures, they come in the end to sufficiently consistent beliefs.

A major focus of our analysis is in making precise the qualifier "enough" in the previous sentence. To show why some precision is necessary, consider the following modification to behavior in the example. In round k, player 1 chooses (based on his beliefs) whichever action maximizes his expected payoff with probability $(k-1)/k$, and he chooses the other action with probability $1/k$. Note that the rate of experimentation is assumed to fall off. But we have picked a rate of falling off which guarantees that, with probability one, player

1 will each of his options infinitely often. (We imagine that player 2 does a similar amount of experimenting, although we won't need to worry about that for this example.) This means that player 3 will be given the opportunity to move infinitely often. And if, as we have supposed, player 3 always moves L, then as player 1 digests this information (that is, updates his Dirichlet assessments), player 1 will eventually decide that D is better than A after all, at which point he will mostly move D, and we will be at a Nash equilibrium. Note well, if player 3 played some other strategy than always L it wouldn't matter — whatever 3 does, after 3 has been given a large number of opportunities to move, the evidence accumulated by 1 and 2 will overwhelm their initial, inconsistent priors. They will come to have almost identical beliefs about what 3 will do, and if their beliefs are close together, then one or the other or both will think that D is a better option than A.

Note as well what might happen if player 1 experimented with the suboptimal action with probability $1/k^2$ in round k (and 2 did likewise). Then, although there is some experimentation going on, and there is always the chance of another experiment, the number of experiments is almost surely going to be finite. Given the priors we assumed for 1 and 2, there is positive probability in this case that the last experiment will come before enough evidence has accumulated to get one of the first two players to choose across, and we might get "stuck" at the outcome A, A.

This is only an illustrative example, and there is one point in particular where it may mislead the reader. We assumed that players 1 and 2 each played whatever strategy gave them the highest *ex ante* expected payoff (given their personal beliefs) with high probability. We will, in the sequel, not insist on this. Instead, we will insist that a player (eventually) decrease the chances of using a strategy is that strategy seems to be suboptimal by an amount bounded away from zero. But if a player has two options, and if, as time passes, they seem ever closer together in terms of the expected payoff they will engender, then we allow the player to pick either one with high probability, as long as the player experiments with the other sufficiently often. Our reasons for this will become clear as we continue our development. Meanwhile, the astute reader will see why, given this, it makes a difference whether we think of players computing expected payoffs *ex ante* (as we do for the rest of Part I) or *ex post* (in Part II).

I.3. TWO PRINCIPLES OF BEHAVIOR

We will look at behavior that conforms in spirit to two general principles that the example above illustrates: (i) Players experiment infinitely often with every option they have if given the opportunity. (ii) Given more and more data upon which to base their conjectures, players play more and more often those actions which a "naive empiricist" would suggest look (nearly) best in the short run, based on *ex ante* calculations.

I.3.1. Experiment if given the chance

To understand the first principle and one of the issues that it raises, it will help to begin a simple example. Consider the extensive form game depicted in figure I.2.

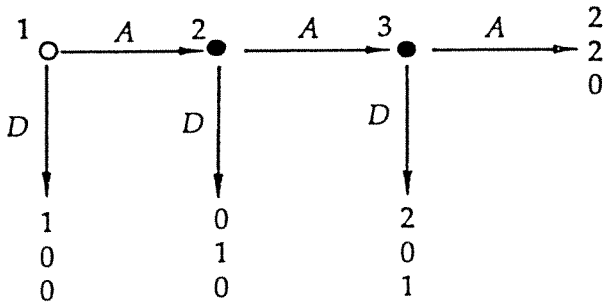

Figure I.2.

Player 1 moves first; if player 1 chooses A, then player 2 moves; if player 2 chooses A, then player 3 moves. Note that the payoffs make (D, D, D) the perfect equilibrium. Imagine that players 1 and 2 begin believing that 2 and 3 are likely to play D, and so 1 and 2 both believe that D is likely to be the better action for themselves. But, in line with discussion of the previous section, both 1 and 2 will sometimes experiment with A, just to see what happens.

The question that we wish to pose here is: Should player 2's rate of experimentation depend on calendar time or on the number of times that he is able to

experiment? The spirit of the first principle, it seems to us, is that player 2's rate of experimentation should be determined by the number of chances he gets to experiment. That is, suppose we imagined that player 1 will experiment with A on round k with probability $1/k$, unless some evidence arrives that disconfirms his initial hypothesis about player 2. Then the number of times that player 2 will be given the chance to experiment will (almost surely) rise to infinity, but the rate of increase will be very slow. Player 2, it seems to us, will "reasonably" wait for those opportunities to build up (and evidence to build up about 3's behavior) before he reduces markedly his rate of experimentation.

With this in mind, we make the following definitions. For every information set h and k length history ζ_k, we let $\kappa(h,\zeta_k)$ be the number of components of ζ_k that are successors of h. In other words, $\kappa(h,\zeta_k)$ records the number of previous instances (according to history ζ_k) that h was hit. And, for fixed behaviors ϕ, define for each information set h and for $\kappa = 1,2,...$

$$\pi(h,\kappa) = \min\{\phi^{n(h)}(k+1,\zeta_k,a) : a \in A(h); \zeta_k \in Z^k \text{ such that } \kappa(h,\zeta_k) \leq \kappa\}.$$

This monster needs some explanation: It is the smallest probability of any action at the information set h if that information set has been hit κ or fewer times previously. Our formalization of the first principle becomes:

Assumption I.1. Behaviors are such that, for each $h \in H$,

$$\sum_{\kappa=1}^{\infty} \pi(h,\kappa) = \infty.$$

From this assumption and the structure of the game, we obtain the following by applying the Borel-Cantelli lemma:

Lemma I.1. For behaviors that conform to assumption I.1, at every information set that occurs infinitely often (with probability one), every action is taken infinitely often (with probability one).

The proof is straightforward. We proceed to a few remarks about the assumption and the lemma.

(1) If we imagined that a player evaluated his overall welfare by a discounted sum of his payoffs, and if he somehow came to regard his decision problem as a

multi-armed bandit problem, then there might well come a time when he gives up on one or more of his options. The classes of behavior we are considering here are definitely not of this sort; our players do not act as if they think they are solving a multi-armed bandit problem with a discounted payoff criterion.

(2) Note that the assumption is a good deal stronger than: "each player, in choosing his normal form strategy, experiments infinitely often". Vicarious experimentation doesn't count. By this we mean: In terms of figure I.2, it isn't enough for player 2 to say: "On round k
probability $1/k$, so I experiment infinitely often." Experiments only count when they really take place. Put differently, if we supposed that player 1 chooses A in round k with probability $1/k$ and if player 2 does likewise, then player 2 will only actually play A a finite number of times (almost surely) and so player 3 will only be called upon to move a finite number of times. In terms of the formal requirement, if h' denotes player 2's information set, then for every κ there are histories ζ_K for arbitrarily large K such that $\kappa(h', \zeta_K) = \kappa$, and hence $\pi(h', \kappa) = 0$. (All this is just saying again what we said to motivate the formalization, although it may be clearer why we were saying all that now.)

(3) In the case of the game in figure I.2, and more generally in any game of complete and perfect information, the lemma can be strengthened to read: Every outcome z occurs infinitely often with probability one. But, in general, this strengthening is unwarranted. To see this, consider the variation on figure I.2 given as figure I.3. Now player 2 moves regardless of whether player 1 chooses A or D, and player 2 moves without knowing what player 1 did. In this case, behavior by player 1 which takes action A with probability $1/k$ in round k and by player 2 which does exactly the same thing *does* conform to assumption I.1. (This is so because now, for h' the information set of player 2, $\kappa(h, \zeta_k) = k$, and hence $\pi(h', \kappa) = 1/\kappa$.) But then in round k there is only probability $1/k^2$ that player 3 is called upon to move; since this series has a finite sum, Borel-Cantelli tells us that player 3 moves only a finite number of times, almost surely.

Note that we have made player 2's move if player 1 chooses D irrelevant. This assignment of payoffs will come into the story more fully later, although it is worth observing now that, as normal form games, figures I.2 and I.3 are strategically equivalent. If they are strategically equivalent, why do we differentiate between them? We do so because they are not strategically equivalent from the spirit of experimentation. Suppose player 2 wishes to experiment infrequently,

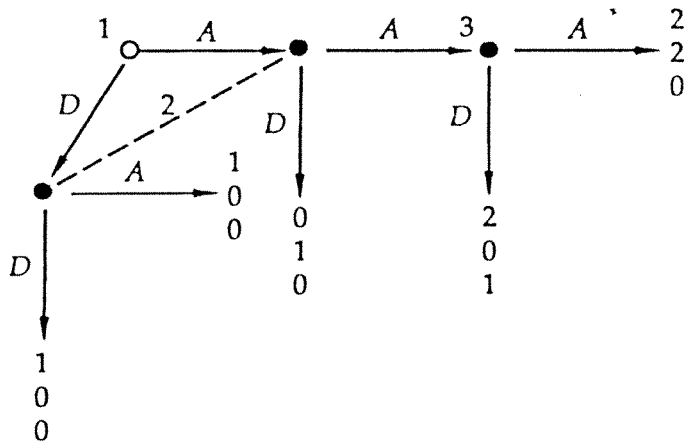

Figure I.3.

but frequently enough so that player 3 is given infinitely many opportunities to move. Moreover, player 2 knows that player 1 will usually choose D, but player 1 will choose A infinitely often. It is harder for player 2 to put player 3 on the move infinitely often in the game in figure I.3 than in the game in figure I.2, since in figure I.2, player 2 can time his experiments to coincide with those of player 1. In the game of figure I.3, player 2 is unable to tell whether player 1 is experimenting until it is "too late" to take advantage of the opportunity this presents. Now it can be objected: Player 2, in the case of the game in figure I.3, ought to be a bit more sophisticated: If he wants player 3 to be given the move infinitely often, and if he believes that player 1 is mostly choosing D, they he might as well *always* try to give 3 the opportunity (unless and until his hypothesis about player 1 is disconfirmed). This is so because, as long as player 1 does play D, it doesn't matter to 2 whether he (2) experiments or not. We will return to this objection later.

(4) In the analysis to follow, life would be much simpler if we could conclude, in place of lemma I.1, that *every information set occurs infinitely often almost surely.* [13] What sort of assumption would guarantee this? Suppose that we defined $\pi(k) = \min\{\phi^n(k+1, \zeta_k, a) : n; \zeta_k; a\}$. That is, $\pi(k)$ is the minimum probability of any action at all in round $k+1$. Suppose as well that L is the maximum

[13] This is one degree weaker than saying that every outcome occurs infinitely often, but, as will become apparent, the weaker statement would be enough for our purposes.

length (in terms of numbers of actions) of any path through the tree. Then if $\sum_{k=1}^{\infty} \pi(k)^{L-1} = \infty$, Borel-Cantelli tells us immediately that every information set occurs infinitely often, almost surely. (Hence for two-player games, assumption I.1 is sufficient for this conclusion.) One could tell a story for this assumption: Players wish to see what each other will do at every information set (which is why they are experimenting in the first place), and they realize the problems inherent in coordinating their experiments, hence they experiment at rates that moot the coordination problem. But this, it seems to us, introduces a little too much cooperation into what is meant to be a noncooperative story. In particular, player 1 could say to player 2 in the context of figure I.3: "You experiment in round k with probability $(1/k)^{1/100}$, and I'll experiment with probability $(1/k)^{99/100}$, and we'll get to see what 3 is going to do." If experimentation is perceived to be costly, then this sort of bargaining is not inconceivable. [14] So we will proceed without assumptions strong enough to get for us the stronger (and very much more convenient) conclusion. The reader will see that we pay quite a price for this.

(5) Suppose that, in the game in figure I.2, player 1 thinks that A is better than D. Since the consequences of D are known, why should player 1 experiment at this point? Similarly, why, in figure I.3, would we ever assume that player 3 would experiment at all? We could imagine a story where players aren't sure about what payoffs they get from their actions, even actions that lead to terminal nodes. Or we could modify assumption I.1, so that it doesn't apply in such

[14] Consider this sort of objection in the context of figure I.1, however. If experimentation is perceived to be costly, and if player 2 is going to experiment with D infinitely often, then mightn't player 1 wish to keep on with A (unless and until the evidence accumulated by virtue of 2's experiments convinces 1 that D is better than A)? Or suppose the payoffs to 1 were different depending on whether 1 or 2 gave the move to 3; suppose in particular that they were a bit bigger (contingent on 3's move) if 2 gives the move to 3 than if 1 does. Then shouldn't 1 foreswear experimentation, and let 2 do all the work, realizing that the evidence will accumulate anyway? We will assume that thoughts this sophisticated do not trouble 1 in his deliberations — we certainly are not constructing an equilibrium in the "experimentation game." Given that this is so, the argument just given against assuming the condition $\sum_{k=1}^{\infty} \pi(k)^{L-1} = \infty$ is a bit less convincing than it may seem at first. Nonetheless, we see a significant difference between assuming that each player, on his own, experiments infinitely often given the opportunity to do so, and an assumption that experimentation rules are chosen individually in a manner that satisfies a socially derived constraint, and we will therefore stick with assumption I.1.

circumstances. The latter would not change our results, but would be costly in terms of notation and exposition, and so we don't do that. The former, on the other hand, is consistent with our basic story, although we would have to increase the notational requirements of assumption I.3 below. We will return to this first story in section III.

I.3.2. Naive empiricism, at least asymptotically

Our first principle of behavior says roughly that everything happens occasionally. Our second principle says that only things that "look" optimal based on history happen often. We present this second principle in two parts; we also give a consistency test that is suggested by the first part of the principle. All of this requires some additional notation:

Recall that $\sigma^{-n}(k+1,\zeta_k)$ is the conjectures of player n about the actions of his opponents at date $k+1$, given history ζ_k. We will write $\sigma^{-n}(k+1,\zeta_k,h)$ for those conjectures at the information set $h \in H^{-n}$.

For any information set h and k-length history ζ_k such that $\kappa(h,\zeta_k) \geq 1$, let $\eta(h,\zeta_k) \in \Delta(A(h))$ be given by $\eta(h,\zeta_k)(a) = \kappa((h,a),\zeta_k)/\kappa(h,\zeta_k)$, where by $\kappa((h,a),\zeta_k)$ we mean the number of components of ζ_k that passed through h and then had action a taken. That is, $\eta(h,\zeta_k)$ is the unweighted empirical distribution at h based on ζ_k.

For any two probability distributions p,q on any finite dimensional probability simplex $\Delta(X)$, let $d(p,q) = \max_{x\in X} |p(x) - q(x)|$.

Let ζ denote a typical sample path or *entire history of play,* an element $(z_1,z_2,...) \in Z^\infty$, with ζ_k standing for the k-long partial history. Given a ζ, let $H^{IO}(\zeta)$ be the set of all information sets that are hit infinitely often along the sample path ζ.

With all this, we can now state the first part of the second principle of behavior:

Assumption I.2. For given ζ and for each player n and information set $h \in H^{IO}(\zeta)$ with $n(h) \neq n$,

$$\lim_{k\to\infty} d(\sigma^{-n}(k+1,\zeta_k,h),\eta(h,\zeta_k)) = 0.$$

Assumption I.2 says that player n's hypothesis is asymptotically the hypothesis that is formed from a naive count of how often various actions were

taken, as (and if) evidence at a particular information set builds up. Note that for any fixed integer K, assumption I.2 can be satisfied in a way that imposes no restriction for n's hypothesis as to what happens at any information set h that is reached K times or fewer.

Most "reasonable" statistical procedures that are based on a hypothesis that the sequence of observations of actions at a particular information set is an exchangeable sequence would have this property. We saw in one example in section I.2, an example which is important for later development, as it will be used constructively. In this statistical procedure, players use Dirichlet priors and posteriors. That is, each player n begins with a prior on how his opponents act, given by an assignment of strictly positive integers to all actions available to the others. We let $j(a,n)$ be the integer assigned by player n to action $a \in A(h)$ for $h \in H^{-n}$. Note that there is no requirement of consistency of these priors. The prior probability hypothesized by player n in the first round that action a will be taken is then simply $j(a,n)/\sum_{a' \in A(h)} j(a',n)$, where h is the information set to which a belongs. These priors are updated very simply: In any round, if information set h is reached and then action a is taken, then $j(a,n)$ is increased by precisely one by all players n. And then hypotheses in the following round are computed just as is the prior, for the new integers. The reader will quickly verify that hypotheses computed in this fashion satisfy assumption I.2.

Now it should be noted that this particular "reasonable" statistical procedure may not be properly specified for the actual situation, since players' behavior in our model is highly nonstationary. It should also be noted that this is just an example. We don't restrict players to a model of exchangeable observations; indeed, they need not use any formal statistical procedure at all. We do insist with assumption I.2, however, that if and when evidence builds up, players' hypotheses are close to what our Dirichlet-using statistician would hypothesize, and their otherwise diverse hypotheses are (therefore) close together.

We do *not* mean to imply that *every* reasonable statistical procedure will have this property. For one thing, we haven't the slightest idea whether a properly specified statistical procedure for some parametric uncertainty in the model, which we haven't introduced, would have this property. More to the point, it seems unlikely that anyone could ever work out the properly specified model for our dynamic system. And there are quite sensible heuristic procedures which our assumption rules out. For example, exponentially weighted moving averages would fail this assumption. Statistical procedures such as this would seem rea-

sonable in environments where the population of players changes slowly through time, as this sort of procedure puts more weight on the recent than on the distant past. Hence we imagine that our players do not consider themselves to be in such an environment. (We will return to this sort of environment and exponentially weighted moving averages later.)

It should also be noted that our players, while perhaps somewhat naive in their statistical procedures, do have some "game theory" behind their hypotheses. More precisely, we assume that players build up their hypotheses about how their opponents play out of *independently implemented* behavior strategies for each opponent. The data could conceivably reveal some correlation in opponents' behavior. (Imagine, for example, that players 1 and 2 start the game with a simultaneous selection of actions, and they alternate between two pairs of strategies in even and odd numbered periods. A third player, according to assumption I.2, seemingly doesn't perceive the periodicity. And, not seeing the periodicity, he also doesn't look at the data and see the correlation that the (unseen) periodicity builds in. Or, rather, if he sees that correlation, he disregards it in forming his hypothesis about the joint actions of players 1 and 2.

This is far from satisfactory. We do not imagine that players would maintain their belief in an asymptotic environment of stationary and independently chosen behavior strategies if the evidence that players accumulate manifestly disconfirms this hypothesis. That is to say, assumption I.2 should be modified to read: "For each history ζ, *unless evidence accumulates along the path* ζ *that contravenes the model of the world that is implicit in players' (asympototic) use of the data,...*"

It is, however, quite difficult to see what players would do if accumulated evidence did contravene this sort of model of the world. So in the development to follow, we take a less than satisfactory middle course between ignoring this problem and dealing fully with it. We will pose a weak consistency test that the accumulated evidence ought to pass, and we restrict attention to cases where this test is in fact passed. [15]

The test we pose is quite weak. Fix a player n, an information set $h \in H^n$, and an action $a \in A(h)$. Let σ^{-n} denote the (strictly positive) hypothesis held by player n about what his opponents will do as a function of date and

[15] The reader is entitled to be somewhat mystified by this plan of action. Things will, we hope, become clear in due course, but it may take until the second example after the proof of theorem I.1 in section 5 before this happens. For now, the reader should bear with us.

partial history. Write $Z(h,a)$ for the subset of Z consisting of all successors of (h,a). For given ζ_k and $z \in Z(h,a)$, use Bayes' rule to compute the conditional probability of reaching z in $Z(h,a)$, given that h is reached, n chooses a at h, and others play according to σ^{-n}, which we will denote by $\sigma^{-n}[z|(h,a),\zeta_k]$. Fix a function $\tau : (0,\infty) \times \{0,1,...\} \to (0,\infty)$ such that $\lim_{k\to\infty} \tau(r;k) = \infty$ for all $r > 0$.

Definition. History ζ passes the consistency test (for the given function τ) if for every player n, information set $h \in H^n$, action $a \in A(h)$, and $z \in Z(h,a)$,

$$\kappa(z,\zeta_k) > \tau\{\sigma^{-n}[z|(h,a),\zeta_k]; \kappa((h,a),\zeta_k)\}.$$

Interpret this test as follows: If player n holds hypothesis $\sigma^{-n}(k+1,\zeta_k)$ and if $\kappa((h,a),\zeta_k)$ times in the past h was reached and a was the action chosen, then player n "expects" that outcome $z \in Z(h,a)$ should have been reached approximately $\kappa((h,a),\zeta_k)\sigma^{-n}[z|(h,a),\zeta_k]$ times. If $\kappa((h,a),\zeta_k)$ goes to infinity, player n would probably become suspicious if the (conditional) fraction of times that z is seen is in the limit anything other than the fraction that is expected. Even if player n is not that suspicious, if $\kappa((h,a),\zeta_k)$ goes to infinity and the expected (conditional) fraction of zs stays bounded away from zero, and yet the observed conditional fraction of zs goes to zero, then player n might begin to suspect that something is awry. The test we have posed is weaker than even this — the (conditional) fraction of times that z is seen may vanish, as long as it doesn't vanish too quickly. It is hard to imagine any "test" of any model that is based on asymptotically independent and stationary strategy choices by one's opponents that would not reject the model if the test just posed is not passed for some function τ.

Remarks. (1) Owing to assumption I.2, we know that $\sigma^{-n}(k+1,\zeta_k)$ will converge to the empirical distribution at information sets that are reached infinitely often. It takes a bit of writing down, but from this one can show that the consistency check could have been formulated with conditional hypotheses that are computed from $\eta(\cdot,\zeta_k)$ instead of from σ^{-n}, without changing the test, as long as assumption I.2 is maintained. We chose to formulate with σ^{-n} because it is notationally a bit easier (we assume σ^{-n} is strictly positive), and because the interpretation of the test is a bit more direct.

(2) This test, being so weak, is certainly only one test of this sort that we might pose. We use this test for necessary conditions, so its weakness is a virtue. (See theorem I.1.) Later, when we get to sufficiency conditions, we would want very much stronger tests than this — we will comment about this when the time comes.

Now we turn to the second half of second principle for behavior. Based on his conjecture $\sigma^{-n}(k+1,\zeta_k)$, player n can evaluate the *ex ante* payoff he will receive for every strategy he might adopt. Since we are working with behavior strategies, it is convenient to decouple the evaluation of different actions at different information sets, and so we proceed as follows: Fix player n, an information set $h \in H^n$, an action $a \in A(h)$, and strictly positive conjectures σ^{-n} for the behavior strategies of others. Since σ^{-n} is strictly positive, it is straightforward to compute the probability that information set h will be reached, denoted $\sigma^{-n}(h)$, the conditional or *ex post* payoff to n if n takes action a at h, conditional on reaching h, denoted $E^n[a|h;\sigma^{-n}]$, and the *ex ante* payoff to n of using action a at h, denoted $E^n[a,h;\sigma^{-n}] = \sigma^{-n}(h) \times E^n[a|h;\sigma^{-n}]$. For fixed h, let $E^n[\star,h;\sigma^{-n}]$ be the maximum of the $E^n[a,h;\sigma^{-n}]$, maximized over $a \in A(h)$, and let

$$L^n[a,h;\sigma^{-n}] = E^n[\star,h;\sigma^{-n}] - E^n[a,h;\sigma^{-n}].$$

It is straightforward to see that L^n measures the *ex ante* loss that n incurs by using a at h in place of the optimal action at h (given his beliefs about what others do).

Assumption I.3. For each player n and information set $h \in H^n$ there is a function $\alpha : (0,\infty) \times \{1,2,...\} \to [0,1]$ such that the behavior of player n at information set h satisfies

$$\phi^n(k+1,\zeta_k,a) \leq \alpha(L^n[a,h;\sigma^{-n}(k+1,\zeta_k)],\kappa(h,\zeta_k)),$$

where α is such that, for every $\lambda \in (0,\infty)$, $lim_{k\to\infty}\alpha(\lambda,k) = 0$.

Remarks. (1) One way of paraphrasing this assumption is that, asymptotically, players are almost certain to play actions that almost maximize their expected payoffs based on their conjectures. This assumption allows a player to continue to select weakly dominated strategies, as long as the probability the player assesses that this will be costly goes to zero. One can explain or excuse this as a manifestation of inattention, in the spirit of Radner's (1980) ϵ-equilibrium. We

are not completely happy with this assumption, but it is important to our development of a theory of Nash equilibrium. Later we will use our unhappiness to justify tightening the assumption in various ways, to restrict the set of behaviors permitted and, accordingly, the behavior that can be observed in the long run.

(2) In particular, because we compute the loss functions using *ex ante* payoffs, the behavior of a player is virtually unconstrained at information sets that are reached with vanishingly small probability. (We will make this exact with some examples in section I.5.) But when a player is called upon to move, even if that move was thought to have small prior probability, it seems as if the player would "reoptimize", using conditional expected payoff calculations. This sort of consideration suggests that assumption I.3 might be changed so that (at least) conditional expected losses are used in place of *ex ante* expected losses, a suggestion that we will follow out in great detail in Part II.

(3) Since there are finitely many players and finitely many information sets, we could take a single function $\alpha(\lambda, \kappa)$ for all n and h. To simplify later notation, we assume that this has been done.

(4) We have made both players' behavior (given by ϕ) and their conjectures (given by the σ^{-n}) part of our basic formulation, with assumptions I.2 and I.3 tying the two pieces together. We could alternatively have had only behavior in the formulation, by computing the function L using directly the naive empirical hypothesis η. Then I.2 can be dropped, and assumption I.3 assumed directly. (Also, it would be necessary to pose the consistency test in terms of η.)

(5) Implicit in this assumption (and in our entire treatment) is the assumption that each player acts as if his opponents' actions in the future are unaffected by his current action, so that the cost of an experiment is properly measured by a short-run calculation. The astute reader might anticipate that a sophisticated player could take advantage of his opponent(s) in such a case: By acting in each period as a "Stackelberg leader", one's opponents will eventually act (mostly) according to their own short-run optimal reaction, hence a sophisticated player, knowing that his opponent(s) conformed to the second principle, would wish to deviate from this principle. Along the same lines, folk-theorem style equilibria, based on trigger strategies or something similar, are precluded by this principle's adherence to short-run optimization. For both these reasons, behavior conforming to the second principle requires either fairly unsophisticated players or random

matching of players from a large population. We will proceed for now with the first interpretation, and return later to the difficulties encountered with the second.

I.4. LOCAL STABILITY

To recapitulate where we are: We imagine players playing a particular game repeatedly, using behavior rules that conform to the two principles given above in assumptions I.1, I.2 and I.3. This generates a stochastic law of motion for play. We want to know what are the "steady states" or "stationary points" of this process, and we will want a definition in the general spirit of local (and also global) stability to such a steady state. In this section we will discuss these notions.

The first choice to make concerns the space in which one looks for stationary points. One way to proceed would be to look in the space of beliefs σ^{-n} for each player n. A second would be to look in the space of behavior strategies σ^n for each player. Yet a third is to look at the space of "outcomes", or probability distributions over the set of terminal nodes Z. We proceed in the third manner for now, and we will return to the second way of proceeding in Part II.

Accordingly, we set the following notation. Let $\Delta(Z)$ be the simplex of probability distributions on Z. We will use δ to denote a typical element of $\Delta(Z)$. For $\phi(k+1,\zeta_k)$ the vector of behavior strategies of the players at date $k+1$ given history ζ_k, we denote by $\delta(\phi(k+1,\zeta_k))$ the resulting probability distribution over outcomes. Also, we denote by $\delta(\zeta_k)$ the empirical distribution of outcomes in ζ_k; that is, $\delta(\zeta_k)(z) = \kappa(z,\zeta_k)/k$, the fraction of times that z is a component of ζ_k. Recall that $d(\cdot,\cdot)$ denotes the sup norm metric on any finite dimensional probability simplex. Finally, for any outcome δ, we use $\text{Supp}(\delta)$ to denote the "extended support" of the outcome. That is, a node $z \in Z$ is in $\text{Supp}(\delta)$ if z is in the support of δ in the usual sense. But also $x \in X$ is in $\text{Supp}(\delta)$ if some terminal successor of x is in the support of δ; $(x,a) \in \text{Supp}(\delta)$ if a is a feasible action from x and some terminal successor of (x,a) is in the support; $h \in H$ is in $\text{Supp}(\delta)$ if some $x \in h$ is in $\text{Supp}(\delta)$, and similarly for (h,a); and even $n \in N$ is said to be in $\text{Supp}(\delta)$ if for some information set $h \in H^n$, $h \in \text{Supp}(\delta)$. (That is to say, $\in \text{Supp}(\delta)$ is rough shorthand for the expression "...is along the path of play, if the outcome is δ.)

Then consider the following notion of local stability.

Definition. *An outcome* $\delta \in \Delta(Z)$ *is said to be locally stable with respect to the behavior rules* ϕ *if, for every* $\epsilon > 0$, *there is some history* ζ_k *such that there is conditional probability greater than* $1 - \epsilon$, *conditional on starting at date* $k+1$ *with history* ζ_k, *that the following two conditions hold jointly:*

(i) $\lim_{k' \to \infty} d(\delta(\phi(k'+1, \zeta_{k'}), \delta) = 0$.

(ii) The consistency test of section I.3.2 is passed for some fixed function τ.

Remarks. (1) We measure the distance between the outcome δ and the distribution of outcomes induced by planned behavior in periods subsequent to period k. Of course, the observed outcomes in any single period won't necessarily be close to δ if players are randomizing.

Since we are dealing with the distance between the outcome δ and the outcome induced by planned behavior, we are dealing to some extent in unobservables. We can deal with the observed outcomes if we average observations over periods. That is, suppose that, at date k', we compared δ with the observed empirical distribution over terminal notes, $\delta(\zeta'_k)$. In the spirit of the strong law of large numbers, the following proposition seems likely:

Proposition I.1. *If* δ *is a locally stable outcome, then for every* $\epsilon > 0$ *we can find a history* ζ_k *such that, conditional on starting from* ζ_k, *there is probability greater than* $1 - \epsilon$ *that* $\lim_{k' \to \infty} d(\delta, \delta(\zeta_{k'})) = 0$.

In fact, this proposition is true, and we provide a proof in appendix 1. Note that this proposition differs from the usual formulation of the strong law because the history-dependent behaviors $\phi(k+1, \zeta_k)$ are not independent or even exchangeable. We will discuss this point further when we get to lemma I.3. Also note that the condition in the proposition is implied by our definition of weak stability. They are certainly not equivalent, a point to which we will return in a bit.

(2) The conclusion is that there is small probability of *ever* seeing behavior that induces a distribution on outcomes that is distant from δ, and the distance from δ vanishes as time passes. Because of the strength of this conclusion, we cannot require that, with probability one, behavior subsequent to ζ_k gives outcomes that approach δ. Given that our players are experimenting as they are, there is always some small chance that (bad) luck will take behavior away from any stable point.

(3) The definition by itself doesn't say too much about the "starting points" ζ_k from which the outcome is locally stable. When we get to the sufficiency results of section 7, we will want to prove results such as: Outcome δ is locally stable starting from all outcomes from a (to be) specified class. Typically, the specification will be that the ζ_k give a lot of evidence about play both on *and off* the path of the outcome.

(4) The second condition, that the consistency test is passed, is not entirely natural. It is put here because we don't imagine that players would stick to the sort of asymptotic naive empirical procedure that we have postulated if this test is failed, and then, since we don't know quite what players would do, we cannot comfortably claim that the outcome is locally stable. Of course, we would, and we will, want to strengthen the test that players are assumed to use, if we do wish to be comfortable in the claim of local stability.

In the next section we will look for necessary conditions for local stability. In order to make clear the nature of the results we will obtain, it is useful to give a variation on the above definition.

Definition. *An outcome $\delta \in \Delta(Z)$ is said to be unstable with respect to the behavior rules ϕ if there exists an $\epsilon > 0$ such that for every history ζ_k and $K > k$, there is zero probability that, starting at date k with history ζ_k, the ensuing behaviors $\phi(k'+1, \zeta_{k'})$ and sample path ζ satisfy:*

(i) $d(\delta(\phi(k'+1, \zeta_{k'}), \delta) < \epsilon$ *for all* $k' > K$.

(ii) The consistency test of section I.3.2 is passed for some fixed function τ.

Remarks. (1) Note that all that is required here is that, with probability one, for each date K there is some subsequent K' where *either* behavior gives rise to an outcome that is uniformly different from δ *or* the consistency test fails. In the first alternative, we do not require that outcomes induced by behavior leave a neighborhood of δ and never return. Indeed, in the spirit of proposition I.1, we might wish to define an unstable outcome as one in which there is zero probability that, from any starting point, $d(\delta, \delta(\zeta'_k))$ goes to zero (or stays within a neighborhood of zero). This, however, is a good deal stronger than the definition given above, and we are unable to give our necessity results (e.g., theorem I.1 following) for this stronger definition.

(2) An outcome δ could fail to be locally stable and, at the same time, not be unstable, according to our two definitions. In the definition of local stability, it was required that, with probability approaching one, behavior would converge in terms of induced outcomes to the outcome δ. Here we insist that for some sufficiently small ϵ, behavior leaves the ϵ neighborhood of δ infinitely often, no matter what is the starting point. So, for example, behavior that "cycled" around δ in cycles which depended on the starting point (the closer the starting point, the tighter the cycle) would fail to be locally stable but would not be unstable.

(3) A second apparent difference between the definition of an unstable outcome and the negation of local stability is that here we require zero probability of staying in a neighborhood of δ. Seemingly, the negation of the previous definition would only require that, for some $\rho < 1$, there is probability less than ρ of staying forever in the neighborhood of δ. But this difference is less real than apparent. If there were ever a starting point ζ_k from which one stayed within ϵ of the outcome with positive probability, then continuity of probability would ensure that starting from some successor to ζ_k there would be probability arbitrarily close to one of staying within ϵ of the outcome. (Lemma I.6 gives insight into the details of the proof.)

(4) Again, condition (ii) is here because we have no clear conception of what players would do if evidence accumulated that clearly disaffirmed the model that we imagine they are using. So one should paraphrase this definition as: An outcome is unstable if, from any starting point, with probability one either the outcome cannot be sustained (in terms of the set plans of players in each period) or some player is led to reject the type of model we have posed here.

I.5. NECESSARY CONDITIONS

We begin with a statement of the the main proposition of this section. The full proof of this result, however, will take some time to come.

Theorem I.1. If an outcome δ is not a Nash equilibrium outcome, then it is unstable with respect to all behavior satisfying Assumptions I.1, I.2, and I.3.

First step of the proof. It is immediate that, for a given δ, there is a unique assignment of probabilities to pairs $(x, a) \in \text{Supp}(\delta)$ which gives rise to the outcome δ. It has nowhere been assumed that this assignment of probabilities respects the informational constraints of the game. For example, imagine a two-by-two simultaneous move game, and an outcome that puts probability one-half on each of the two off-diagonal elements. The necessary assignment of probabilities to moves would require a correlation in the strategies of the two players which is inconsistent with the game form. Accordingly, we say that an outcome δ *respects the game form* if there is some assignment of (independent) behavior strategies to players that gives rise to the outcome δ. And, as a first step, we have to show that if δ doesn't respect the game form it is necessarily unstable.

This is a simple matter of definitions. If δ doesn't respect the game form, then there is some $\epsilon > 0$ such that every outcome δ' such that $d(\delta', \delta) < \epsilon$, δ' doesn't respect the game form. (The set of all outcomes which do respect the game form is compact, since it is the continuous image of a compact set. Minimize the distance in the sup norm between δ and this set, and let ϵ be half that minimum distance.) Then we could never satisfy the definition of local stability.

Accordingly, we hereafter assume that the outcome δ respects the game form.

The rest of the proof of this theorem is delayed because, in order to prove this result, we must state a simple result of independent interest: How much of what one player will do out of equilibrium must be commonly hypothesized by

other players in order to justify restricting to Nash equilibria? We can explain this somewhat cryptic question with two examples. Consider first the game in figure I.1. As we noted earlier, if players 1 and 2 have a commonly held hypothesis of what player 3 will do at his information set, or even if their hypotheses are fairly close together, then the outcome where both 1 and 2 choose A is not a Nash equilibrium. However if player 1 believes that 3 will choose R with high probability, and if player 2 believes that 3 will choose R with low probability, then A, A is an "equilibrium outcome". We see here that an important part of the Nash assumption is that players must hold roughly similar hypotheses about the out of equilibrium actions of others.

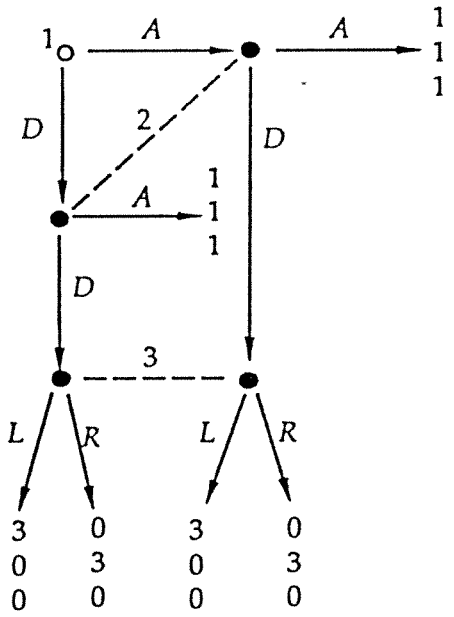

Figure I.4.

But now consider figure I.4. This is very much like the game in figure I.1, except that player 2, when called upon to move, is unaware of what player 1 has done. And, most importantly, player 1 can no longer ensure that player 3 is given the move. Now A, A is a Nash equilibrium outcome. And we can see that at this outcome, players 1 and 2 can have divergent opinions about player 3's actions. In particular, player 1 is "isolated" from player 3 by the strategy of player 2. That is, given that player 2 will choose A, player 1 cannot give the move to player 3. Hence player 1, when evaluating his expected utility from any strategy he might attempt, fixing player 2's strategy at A, is completely indifferent to what player 3 might be doing. Thus his conjectures about what player 3 might be doing are

irrelevant to the optimality of his equilibrium action, and his conjectures about player 3's actions can vary from the actual "equilibrium" strategy of player 3. This example motivates the following concepts and result.

For the given extensive form game, consider an array of conjectures $\{\sigma^{-n} : n = 1,2,...,N\}$ and a strategy vector $\hat{\sigma} = (\hat{\sigma}^1,...,\hat{\sigma}^n)$, where σ^{-n} is the conjecture of player n concerning the strategy choices of his opponents. If for each player n, $\hat{\sigma}^n$ is a best response given the conjecture σ^{-n}, we say that this array of conjectures and strategies forms an *equilibrium in conjectures and actions.*

Consider a strategy vector $\hat{\sigma}$ for the game. For each player n, we say that information set h is *irrelevant* to n under $\hat{\sigma}$ if for every strategy $\tilde{\sigma}^n$ for player n, the strategy vector $(\hat{\sigma}^1,...,\hat{\sigma}^{n-1},\tilde{\sigma}^n,\hat{\sigma}^{n+1},\hat{\sigma}^N)$ does not cause the information set h to be reached with positive probability. If an information set h is not irrelevant, we say that it is *relevant* to player n.

Lemma I.2. The strategy vector $\hat{\sigma}$ is a Nash equilibrium if and only if there is an array of conjectures $\{\sigma^{-n} : n = 1,2,...,N\}$ such that (i) the conjectures and the strategy vector form an equilibrium in conjectures and actions, and (ii) for each player n, the conjectures of player n at relevant information sets agree with the prescriptions given by the strategy $\hat{\sigma}$.

The proof is very nearly immediate. If, for any player n, we change his conjectures at irrelevant information sets, we do not change the optimality of his strategy in response to his conjectures. So we can modify each player's conjectures so that the conjectures all agree with the strategy $\hat{\sigma}$ at all information sets, and we still have an equilibrium in conjectures and actions. But then, by the definition of a Nash equilibrium, we also have a Nash equilibrium.

Next we proceed to a few lemmas needed for the proof of theorem I.1. For the most part, we only sketch the proofs of these lemmas.

The first of these lemmas is a general lemma from probability theory, which extends the strong law of large numbers in a particularly useful way. The motivation for it is as follows. Imagine that a particular outcome is stable with positive probability. That is, behavior doesn't ever leave a given neighborhood of this outcome with positive probability. We would imagine, then, that since observed outcomes are (with positive probability) drawn from distributions in this neighborhood, with this same positive probability the long-run empirical distribution of outcomes should lie in this (or some related) neighborhood.

Lemma I.3. Consider a sequence of random variables $\{z_1, z_2, ...\}$, *each of which has range on some finite set* Z *of size* M. *For each* k, *let* Λ_k *be an event that depends only on* $\{z_1, ..., z_k\}$, *and let* $\Lambda = \cup_{k=1}^{\infty} \Lambda_k$. *Suppose that there is some distribution* δ *over* Z, *an* $\epsilon > 0$, *and a positive integer* K *such that the distribution of* z_{k+1} *conditional on* Λ_k *is within* ϵ *of* δ *(in the sup norm) for all* $k \geq K$. *Then the probability of* Λ *is identical to the probability of* Λ *intersected with the event that the empirical distribution functions of outcomes are, for all large enough* k, *within* $2M\epsilon$ *of* δ.

Roughly put, conditional on the event that each z_k has conditional distribution within ϵ of δ for all large k, there is conditional probability one that the empirical distribution functions of outcomes are, for all large enough k, within $2M\epsilon$ of δ. The proof is left to Appendix 1.

Lemma I.4. Suppose the σ^{-n} *and* $\tilde{\sigma}^{-n}$ *are two conjectures by player* n *about the actions of his opponents such that* $\sigma^{-n} = \tilde{\sigma}^{-n}$ *at any information set* h *that is relevant to player* n *under* σ^{-n}. *Then the set of information sets relevant to player* n *under* σ^{-n} *coincides with the set of information sets relevant to* n *under* $\tilde{\sigma}^{-n}$.

Proof. Call a node x with $n(x) \neq n$ relevant to player n under σ^{-n} if, for some strategy by player n combined with σ^{-n}, there is positive probability of reaching node x. Then if a node is relevant to n under σ^{-n}, so is the information set it belongs to (but not vice versa). Moreover, h is relevant to n under σ^{-n} if and only if some node $x \in h$ is relevant to n under σ^{-n}. Thus it will be sufficient to show that the set of nodes relevant to n under the two conjectures σ^{-n} and $\tilde{\sigma}^{-n}$ coincide.

Take any node x that is relevant to n under σ^{-n} but not under $\tilde{\sigma}^{-n}$. Then every precedessor of x is relevant to n under σ^{-n}. The condition in the lemma implies that $\sigma^{-n} = \tilde{\sigma}^{-n}$ at every node x' that is relevant to n under σ^{-n} and that doesn't belong to player n. Since x is relevant under σ^{-n}, every step along the path to x has positive probability under σ^{-n} (or belongs to n, and so can be taken to have positive probability), hence has positive probability under $\tilde{\sigma}^{-n}$. Thus x is relevant under $\tilde{\sigma}^{-n}$.

For the converse, suppose there is some node x that is relevant to n under $\tilde{\sigma}^{-n}$ but not under σ^{-n}. Then there is some earliest node x (in terms of precedence in the game tree) with this property, and, without loss of generality, we can assume that x is that node.[16] The condition in the lemma implies that every

[16] We either assume that n's conjectures about moves by nature are the same under the two conjectures, or, at least, they have the same support.

node x' that is relevant to n under σ^{-n}, σ^{-n} and $\hat{\sigma}^{-n}$ agree. Thus they agree at all predecessors of x that don't belong to n (since all predecessors of x are relevant under σ^{-n} by assumption, which immediately gives a contradiction.

Now we give a "contrapositive" to lemma I.2. We first require a definition: We say that information set $h \in H^{-n}$ is *ϵ-relevant* to player n under $\hat{\sigma}^{-n}$ if the maximum probability that h is hit under strategy $(\hat{\sigma}^1, ..., \hat{\sigma}^{n-1}, \tilde{\sigma}^n, \hat{\sigma}^{n+1}, \hat{\sigma}^N)$, maximized over $\tilde{\sigma}^n$, is ϵ or more.

Lemma I.5. If δ is not a Nash equilibrium outcome, then there is an $\epsilon > 0$ such that, for any set of behavior strategies $\hat{\sigma} = (\hat{\sigma}^1, ..., \hat{\sigma}^N) \in \Pi_{n=1,...,N} \Pi_{h \in H^n} \Delta(A(h))$ that give outcome within ϵ of δ, there is a player n and an information set $h^* \in H^n \cap Supp(\delta)$ such that, for any conjectures σ^{-n} for n that are within ϵ of $\hat{\sigma}^{-n}$ in the sup norm at every information set h that is ϵ-relevant to player n under σ^{-n}, $\hat{\sigma}^n$ is at least ϵ away (in ex ante payoff) from being a best response to σ^{-n}.

Proof. Suppose that for every integer i there is a strategy vector $\hat{\sigma}_i$ and conjectures for the players σ_i^{-n} such that (i) $\hat{\sigma}_i$ gives an outcome within $1/i$ of δ, (ii) for each $n \in \text{Supp}(\delta)$, $\hat{\sigma}_i$ is within $1/i$ (in the sup norm) of σ_i^{-n} at all information sets that are $1/i$ relevant to n under σ_i^{-n}, and (iii) for each $n \in \text{Supp}(\delta)$, $\hat{\sigma}_i^n$ is within $1/i$ (in *ex ante* payoff) of being a best response by n to σ_i^{-n} at every information set $h \in H^n \cap \text{Supp}(\delta)$. Then, looking along a subsequence, we can assume that $\hat{\sigma}_i$ converges to some $\hat{\sigma}$ and each σ_i^{-n} converges to σ^{-n}. By virtue of (i) and continuity, $\hat{\sigma}$ induces the outcome δ. By virtue of (ii) and continuity, for every $n \in \text{Supp}(\delta)$, behavior at every information set h that is relevant to player n under σ^{-n} is the same under $\hat{\sigma}$ and σ^{-n}. Thus by the previous lemma, behavior at every information set h that is relevant to player n under $\hat{\sigma}$ is the same under $\hat{\sigma}$ and σ^{-n}. From (iii) and continuity, $\hat{\sigma}^n$ is a best response to σ^{-n} for all $n \in \text{Supp}(\delta)$. Since the play of players at information sets $h \notin \text{Supp}(\delta)$ does not affect their *ex ante* payoffs, $\{\hat{\sigma}, \{\sigma^{-n}\}\}$ is an equilibrium in conjectures and actions. Lemma I.2 then implies that δ is a Nash outcome, which is the desired contradiction.

Let us paraphase this: For any outcome that is not a Nash equilibrium outcome, and for any strategy array that gives an outcome sufficiently close to this outcome, there is some player and on-the-path-of-the-outcome information set such that, if the player's conjectures are sufficiently close to the strategy array at all information sets that are sufficiently relevant to this player, this player will

find that his part of the strategy array at this information set is suboptimal by an amount bounded away from zero.

Proof of theorem I.1. Our paraphrase of lemma I.5 presumably suggests to the reader what is the main line of the proof. Suppose an outcome δ that is not Nash is locally stable with nonzero probability. Then with positive probability all subsequent behavior will eventually be close to that outcome, and the empirical evidence will build up (per lemma I.3) so that (per assumption I.2) every player will come to have conjectures "close" to the outcome, at least at information sets in $\text{Supp}(\delta)$. We still must worry about conjectures at information sets not in $\text{Supp}(\delta)$, and we will two sorts of worries: First, the definition of a stable outcome does not require that behavior off of $\text{Supp}(\delta)$ converges. We will finesse this problem by looking along a subsequence along which empirically based conjectures do converge. Denoting by $\hat{\sigma}$ the accumulation point of the convergent subsequence, we can then use lemma I.5 to find a player whose behavior at an information set in $\text{Supp}(\delta)$, once sufficient evidence has been accumulated, will not conform (by assumption I.3) to the outcome δ. Or rather, we can use lemma I.5 to derive this contradiction if we know that every information set that is sufficiently relevant in the sense of lemma I.5 is in fact visited infintely often, so that our player's conjectures about the behavior of others is close to $\hat{\sigma}$ where it matters. Establishing this last part is the key to our argument. It follows from assumption I.1 and the requirement that a stable outcome must pass the consistency test of section I.3.2.

Now for the details. Fix an outcome δ which is not a Nash equilibrium outcome. Recall that we are already assuming that δ respects the game form. Because δ is not a Nash equilibrium outcome we can find an $\epsilon_1 > 0$ such that the statement of lemma I.5 holds for this ϵ_1 and δ. We can moreover take ϵ_1 to be less than or equal to the probability of every outcome $z \in \text{Supp}(\delta)$. Now let ϵ_2 equal one-third of ϵ_1.

Assume that δ is not unstable with respect to behavior ϕ (in order to derive a contradiction). Then for every $\epsilon > 0$ we can find a starting point ζ_k and a K such that, with positive probability, the distribution of outcomes subsequent to ζ_k is, after round K, always within ϵ of δ and the consistency test is passed.

Letting M be the cardinality of Z, let ζ_k^* be a starting point such that there is positive probability, starting at ζ_k^*, that the subsequent distribution of outcomes after K is always within $\epsilon_2/2M$ of δ and the consistency test is passed.

Apply lemma I.3. The event just named has form required in lemma I.3 — the event is the intersection of events $\Lambda_{k'}$ that are $\zeta_{k'}$ measurable.[17] There is, therefore, the same positive probability that, starting at ζ_k^*, the subsequent distribution of outcomes is (after date K) within $\epsilon_2/2M$ of δ, the consistency test is passed, *and* the empirically based conjectures concerning the outcome eventually conforms to a distribution within ϵ_2 of δ. Moreover, we can apply lemma I.1. It is stated there in unconditional form, but it is clear that the conclusion of the lemma is also true conditional on any starting point: There is probability one that, starting from ζ_k, at every information set that occurs infinitely often, every action is taken infinitely often. From now on, restrict attention to sample paths (beginning at ζ_k) along which this statement is correct. Hence we are guaranteed that there is (conditionally) positive probability for the set of sample paths ζ along which the following four things hold:

(1) At every information set that is visited infinitely often, every action is taken infinitely often.

(2) The consistency test is passed.

(3) The empirically based conjectures concerning the outcome are eventually within ϵ_2 of δ.

(4) The distribution of outcomes is closer than $\epsilon_2/2M \leq \epsilon_1/3$ to δ after date K.

We proceed to show that for every sample path such that (1) through (4) hold, a contradiction is derived.

Fix a sample path such that (1) through (4) hold, and let $\eta(h, \zeta_{k'})$ be as before — the empirical distribution over behavior at information set h, based on history up to $\zeta_{k'}$. Since the various probability simplicies are compact, we can look along a subsequence in k' such that $\eta(h, \zeta_{k'})$ converges to some given $\hat{\sigma}$. (For information sets that are never reached at all, let $\eta(h, \zeta_{k'})$ be defined arbitrarily, as long as the arbitrary choice is held constant.) By construction, the continuity of outcomes in strategies, and (4), $\hat{\sigma}$ gives an outcome that is within $\epsilon_1/3$ of δ.

Now according to lemma I.5 there is a player n and information set $h^* \in H^n \cap \text{Supp}(\delta)$ such that: If this player's conjectures σ^{-n} are within ϵ_1 of $\hat{\sigma}^{-n}$ in the sup norm at every information set h that is ϵ_1-relevant to player n under $\hat{\sigma}$, then $\hat{\sigma}^n$ is at least ϵ_1 away from being a best response to those conjectures.

[17] The form of the consistency test was chosen with this in mind.

We claim that every information set h that is ϵ_1-relevant to this player n under σ^{-n} is reached infinitely often. Take any such h. Since it is ϵ_1-relevant, there is some $a \in A(h^*)$ such that if n takes a at h^* and others play according to σ^{-n}, h is reached with probability at least ϵ_1. Since $h^* \in \text{Supp}(\delta)$, by virtue of (4) h^* is reached infinitely often, and by virtue of (1), action a is taken infinitely often. The consistency test then applies immediately.

So for large enough k', n's assessments $\sigma^{-n}(k'+1, \zeta_{k'})$ at all ϵ_1-relevant information sets are within ϵ_1 of $\hat{\sigma}$. Accordingly, the action at h^* prescribed by $\hat{\sigma}^n$ is more than ϵ_1 away from optimal. By going far enough along the subsequence, assumption I.3 ensures that the probability of following the worst of the actions prescribed with positive probability by ϕ^n must be less than ϵ_2. But then the probability of any terminal node which follows this choice must then be less than ϵ_2. Since we chose ϵ_1 so that the least likely terminal node which has positive probability under δ has probability at least $\epsilon_1 = 3\epsilon_2$, the difference between the outcome induced by behavior and δ must be at least ϵ_2, which gives the desired contradiction.

Remark. For purposes of later discussion (in Part II), the reader should note carefully just how much of assumptions I.1 and the consistency test we used in the proof. We only invoked assumption I.1 at the information set h^* and we only invoked the consistency test there. Hence, for purposes of this proof and, indeed, for all of Part I, we can weaken assumption I.1 to read: At information sets that are reached a nonvanishing fraction of the time, every action is taken infinitely often. And we need only insist that the consistency test is passed at such information sets.

We turn now to two examples which illustrate some of the ideas in the results just derived.

The first example is one already given in figure I.1. As we saw in section I.3, while A, A is not a Nash equilibrium outcome, it is possible to get "stuck" there with positive probability for behavior which satisfies assumptions I.2 and I.3, as long as I.1 is violated, even if there is always positive probability that, in an given round, players will experiment. We saw this by supposing that players experimented in round k (or, for player 2, at his k th opportunity) with probability $1/k^2$. Then there will be a finite number of experiments taken almost surely, and a finite number of experiments may be insufficient to overcome 1's and 2's initial divergent hypotheses about how 3 will act. But if experimentation

takes place infinitely often, then at some point, whatever player 3 is doing, there will be enough evidence so that 1 and 2, if they conform to assumption I.2, will come close to agreement about what 3 is doing. And, whatever that is, one of the two of them will play D instead of A.

This is not to say that player 3's actions in this game will converge to anything at all. For the (very nongeneric) payoffs that are given, player 3 is indifferent between L and R no matter what he conjectures, and so he can contemplate, for example, playing L the first one hundred chances he gets, then R the next thousand, then L the next ten thousand, and so on. If he does this, the behavior of players 1 and 2 will shift around quite a bit, and no stable outcome will emerge at all. In other words, we see very quickly from this example that a general theory of global stability is too much to hope for, although we see this based on a highly nongeneric example, and so one still might hope for a theory of global stability for games with generically chosen payoffs.. (We will see later that this is too much to hope for as well.)

Our second example is intended to show why the consistency check is required. We will not attempt to draw the game in question, but simply describe it. Imagine the following modification of the game in figure I.1: If either player 1 chooses D or 1 chooses A and 2 chooses D, three players, called 4, 5 and 6 play a two-by-two-by-two simultaneous move game. They play this game unaware of whether they are given the move because of the action D by 1 or the sequence A, D by 1 and 2. In this game, the moves of each of these three players will be called X and Y. If all three choose X, then player 3 is called upon to choose between L and R. If any one or more of the three chooses Y, then player 3 does not move. The payoffs to 1 and 2 are as follows: As before, each gets 1 if 1 and 2 choose A, A. If either 1 or 2 moves D and any of 4, 5 and 6 choose Y, then 1 and 2 get payoffs drawn from the interval $[1.05, 1.1]$.[18] If either 1 or 2 moves D and 4, 5 and 6 all choose X, then: If 3 chooses L, 1 gets 1003 and 2 gets -1000. And if 3 chooses R, 1 gets -1000 and 2 gets 1003. Players 3, 4, 5 and 6 can be given any payoffs at all.

Consider the following rough description of behavior. Players 1 and 2 play whichever action has the highest expected payoff with probability $(k-1)/k$ on the kth opportunity to move, and experiment with the other with probability $1/k$. Player's 3 play will be irrelevant; have him randomize with probability

[18] We will work at making this a generic example, so the reader will see that nongenericies are not what makes this work.

1/2 on each action, say. Players 4, 5 and 6 participate in the following elaborate dance:

> On the k th opportunity, if k is evenly divisible by 3, players 4 and 5 choose X with probability $1/k$ and Y with probability $(k-1)/k$, and player 6 chooses X with probability $(k-1)/k$ and Y with probability $1/k$. If k is of the form $3i+1$ for some integer i, then interchange the roles of players 4 and 6 in the case just described. And if k is of the form $3i+2$ for some integer i, then interchange the roles of players 5 and 6 in the first case.

The description just given of what 4, 5 and 6 will do can't possibly satisfy assumption I.3 for any assignment of payoffs for these three players and at all partial histories, because it is made completely independent of those payoffs. *Nonetheless, we claim that this behavior will satisfy the assumptions as long as players 1 and 2 find A superior to D and as long as the ratio of the number of experiments by 1 or 2 to the number of rounds goes to zero,* which (if 1 and 2 always find A superior to D) it will do almost surely. This is so because we are using *ex ante* opportunity loss calculations for the players, and so actions taken at information sets that are hit with vanishing frequency are completely unconstrained. This isn't quite precise, but it will be made precise in the next section. For now, accept it as "correct in spirit," subject to later verification (and minor amendment). For the same reason, player 3's actions, whatever they are, satisfy the assumptions, as long as 3 does enough experimenting, and the proportion of rounds with experiments by 1 or by 2 goes to zero.

Setting aside the question of whether 1 and 2 are playing according to the principles of behavior we have set down, what is the affect of this dance by 4, 5 and 6? On the k th time that either 1 or 2 plays D, which will happen with certainty, there is probability $(k-1)/k^3$ of reaching 3. Hence player 3 will be reached only finitely many times almost surely. If players 1 and 2 begin with beliefs about player 3 that are pessimistic (1 believes that 3 will play R with high probability and 2 believes in L with high probability), then there is positive probability that these beliefs will not be disconfirmed. That is, by giving 1 and 2 Dirichlet priors, where 1 initially puts weight M, say, on R and 1 on L, and 2 puts 1 on R and M on L, by taking M sufficiently large, we can have positive probability (actually, probability as close to 1 as we like) that 1 will, in the end, assess probability .9 or more that 3 will choose R given the chance,

and 2 will assess probability .9 or more that 3 will choose L.

Now what will 1 and 2 assess for the actions of 4, 5 and 6? There will almost surely be infinitely many opportunities to see these three players act, and each will pick X around 1/3 of the time and Y 2/3 of the time. Hence, by assumption I.2, the conjectures of players 1 and 2 converge to the mixed strategy: 3, 4 and 5 each play X with probability 1/3 and Y with probability 2/3. Since mixed strategies are *presumed* to be independent, players 1 and 2 will each assess probability 1/27 that player 3 will move. And we have selected the payoffs for 1 and 2 so that, with this assessment that 3 will get the move and with their diverse conjectures on what 3 will do, both 1 and 2 will prefer A to D. So (A, A) would be a locally stable outcome if we ignored the consistency test.

This is so even though (A, A) can never be a Nash equilibrium outcome, for just the reasons that applied to the game in figure 1. What has gone wrong? Simply that 1 and 2 come to believe there is probability 1/27 that, if either of them tries D, player 3 will be called upon to move. That is, 3's information set is 1/27 relevant to them both, asymptotically. But this relevant information set is reached only a finite number of times, *despite* the fact that both 1 and 2 experiment infinitely often. Of course, this means that the consistency test fails.

The way that this came about should be clear. The elaborate dance among players 4, 5 and 6 puts into their actions a lot of correlation that 1 and 2 ignore because, by assumption, their models are based on the premise that players act independently, and 1 and 2 (asymptotically) use statistical procedures that implicitly assume that others' behavior is stationary. Thus 1 and 2 act as if no such correlation exists. When it does, they are badly "fooled". The point of the consistency check is to insure that players are not so badly fooled as to admit as locally stable a non-Nash equilibrium outcome.

Rather than build in the consistency check, we could have modelled things very differently. We could have supposed that the conjectures of player n about the actions of others, given by σ^{-n}, did not necessarily take the form of independent strategy choices by each opponent. Think in terms of the following sort of formalism: For each player n, information set $h \in H$ and action $a \in A(h)$, σ^{-n} might prescribe a subprobability distribution over $Z(h, a)$ — a subprobability distribution because there might be nonzero probability that h is not reached at all — where we impose the constraints that the total mass assigned to $Z(h, a)$ by this distribution doesn't depend on a, and the sum of these total masses, summed over all $h \in H^n$, must be less or equal to one. (If there are moves by

nature in the game, some further restrictions would be built in.) The point is that the distribution over $Z(h,a)$ needn't take the form of a distribution that ensues from independent strategies by opponents. Given this sort of starting formalism, we could:

(1) Continue to assume assumption I.1.

(2) Replace assumption I.2 by an assumption that these subprobability distributions asymptotically agree with the empirically observed distributions. Formally, the distribution over $Z(h,a)$ at stage k with partial history ζ_k should assign probability to $z \in Z(h,a)$ which is close to $[\kappa(z,\zeta_k)\kappa(h,\zeta_k)]/[\kappa((h,a),\zeta_k)k]$. Note the renormalization; reflecting the idea that, while what others do may depend on the choice of a, the choice of action by n at h is a matter of n's free will.

(3) Compute opportunity losses using the subprobability distributions directly, and keep assumption I.3 exactly as before.

The effect of this, together with our definition of local stability, will be to permit as locally stable outcomes that are not Nash equilibrium outcomes in that they allow for correlated behavior out-of-equilibrium. Note that the definition of local stability, together with the strong law (lemma I.3) and assumption I.2 will guarantee that, almost surely, players will assess that their opponents play independently along the path of play at any locally stable outcome. But consider a game with the following form: Player 3 moves first, choosing one of A, U or D. If A is chosen, the game ends. If either U or D are chosen, then players 1 and 2 each choose between two options, doing so independently and without knowledge of whether U or D was chosen. It is easy to assign payoffs to player 3 so that: As long as 3 believes that 1 and 2 are picking their actions independently, either U or D (or both) is the best choice. But if 3 believes that 1's and 2's choices are correlated, then 3 opts for A. (See figure 10 of Fudenberg, Kreps and Levine, 1988. The funny numbering system on players is to facilitate comparison with this figure.) Now if player 3 picks A with frequency approaching one, 1 and 2, using *ex ante* opportunity loss calculations, are free to do whatever they want, no matter what are their payoffs, and so they can use calendar time to correlate their actions. With the formalism above, 3 will perceive the correlation, but not its tie to calendar time, and this will keep 3 playing A. Hence the non-Nash outcome A can be locally stable.

Given our earlier work with Levine, the reader will not find it surprising that we are ambivalent about this. In some situations, there are stories that one

can tell that make plausible correlation in out-of-equilibrium behavior. (See the discussion surrounding figure 10 in Fudenberg, Kreps and Levine, and also Section 5 of Kreps, 1988.) But here these story are at least one degree less plausible than before, since it is player 3 himself who, by the selection of U or D, moves the game to an out-of-equilibrium position. That is, we previously invoked stories that ran: If player n sees an out-of-equilibrium action, n supposes that he has the wrong theory about the game, and if others play according to some different and unknown theory, the marginal distribution on what they do can exhibit correlations. In this setting, the same story would run: Even if the out-of-equilibrium action observed by player n is his own action, then he concludes that he has the wrong theory, and so... For this reason, and because we wished to provide the foundations for Nash equilibrium in our type of story, we have stuck to an assumption that players base their conjectures on models that involve independent strategy choices by their opponents. But it is interesting to contemplate alternatives that look more like the formalism we have just sketched.

I.6. SUFFICIENT CONDITIONS[19]

In this section, we turn to the question: What are sufficient conditions for an outcome to be locally stable?

Theorem I.2. If an outcome is a Nash equilibrium outcome, then it is locally stable for some behavior that satisfies assumptions I.1, I.2 and I.3.

(The proof will come in a bit.) The weakness in this result is that it gives no idea about the range of partial histories ζ_k from which the Nash equilibrium outcome is locally stable. That is, it says little about how special must be the starting point that leads one to converge into the equilibrium outcome (with probability approaching one). After proving theorem I.2, we will say something about the those partial histories from which one gets to the given outcome: Roughly speaking, the partial history must contains lots of evidence consistent with the equilibrium at all information sets that are relevant to players who move along the path of the equilibrium. Moreover, in general the degree to which the evidence must be consistent with the equilibrium increases the longer is the history. This, however, is not true for strict equilibria.[20] If δ is the outcome of a strict equilibrium, then there is an $\epsilon > 0$ such that if $\{\zeta_k; k = 1, 2, ...\}$ is any sequence of initial histories with $d(\delta(\zeta_k), \delta) \leq \epsilon$ for all k, then the probability that play after ζ_k coverges to δ goes to one.

An example will demonstrate the method of proof of theorem I.2. Consider the game depicted in figure 5. This is the standard first example of a game with a Nash equilibrium which is not subgame perfect, namely D, D. If theorem I.2 is correct, then it will have to be possible, in particular, to describe behavior and that will make the outcome D locally stable. In fact, we will do better than this: We

[19] To the reader of version 0.11: From here on through the remainder of Part I, things are increasingly sketchy. The basic ideas of this section will be given by example, although the details of the proof of the main results will have to await another day.

[20] Recall that a *strict equilibrium* is a strategy profile such that each σ^n is a strict best response to σ^{-n} in the reduced normal form of the game.

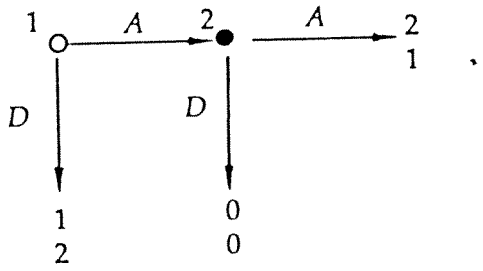

Figure 1.5.

will describe behavior where, for well chosen starting points, there is probability arbitrarily close to one that both players intend, given the opportunity, to choose action D with probability that approaches one as time passes.

This is done as follows. Both players use Dirichlet priors and posteriors in forming their beliefs about the actions of their fellow player. Player 1 begins with a prior that is based on the assignment of 1 to A for 2 and 1000 to D. Player 2 begins with precisely the same prior. Now for each player, the following rule is used:

> On the k th opportunity to move, evaluate (using currently held beliefs) the *a priori* expected value from choosing D and from choosing A. Call these v_D and v_A, respectively. If $v_D + 100/\ln(k+1) > v_A$, then choose D with probability $(k-1)/k$ and choose A with probability $1/k$. Otherwise, choose A with probability $(k-1)/k$ and D with probability $1/k$.

The claim is that these behavior rules, with the standard procedure for computing Dirichlet posteriors, give behavior which satisfies assumptions I.1, I.2 and I.3. Assumption I.1 is automatic, since there is probability $1/k$ of experimenting on the k th opportunity. Assumption I.2, as we've already noted, is true for the Dirichlet inference scheme. And as for assumption I.3, the rule for picking the action will, in the end, pick an action if that action has expected payoff that is clearly better than the other. It will take some time for A to be selected with high probability if A does in fact give clearly better payoff, since $\ln k$ takes a while to get up to 100. But, in the long run, if $v_A - v_D$ is positive and bounded away from zero, A will be picked most of the time.

But it should be fairly clear that things have been arranged so $v_A - v_D$ never gets to a value that is bounded away from zero. For player 1 this is clear — as

time passes, with high probability player 1 sees many D's from player 2, which only confirms the optimality of D. For player 2 things are a bit more delicate. It is always the case that $v_A > v_D$, but if 1 mostly plays D, the difference vanishes, since both v_A and v_D are computed *ex ante*. What we need to show is that, with probability approaching one, as we take sufficiently "good" starting points ζ_k, the decrease in the difference $v_A - v_D$ is faster than the decrease in the factor $100/\ln k$ (where k here refers to the number of opportunities that 2 has had to move). The details of this will be given in a moment in the proof of the theorem, and so we will not repeat them here except to point the reader in the direction that we will use: Let χ_k be 1 if player 1 plays the action assigned probability $1/k$ in round k and zero otherwise. Since the χ_k are independent and the probability of $\chi_k = 1$ is $1/k$. It is a straightforward exercise to show that the series $\{\chi_k/k^{1/2}\}$ is almost surely summable (use the Three Series Theorem), and so $(1/n^{1/2})\sum_{k=1}^{n}\chi_k$ converges to zero almost surely. Hence $(1/n)\sum_{k=1}^{n}\chi_k$, which is the probability that 2's choice is at all relevant, goes to zero almost surely at a rate faster than $1/n^{1/2}$, which is faster than does $1/\ln n$. We can be sure then that there is positive probability that both players always see A as the "better" choice (given the $100/\ln k$ that is added to the expected value of A). And, as we will see in lemma I.8, if there is positive probability of this, then we can make the probability as high as we wish, by taking as the starting point longer and longer initial histories.

Although the details may not be clear until we give the proof of the theorem, the idea should be clear. We might worry that players moving out-of-equilibrium, according to the equilibrium we are trying to sustain, will not conform to the actions prescribed in the equilibrium. After all, in the equilibrium, out-of-equilibrium actions are unconstrained because there is zero probability that the action will matter. But in our story, players always think there is some chance that their actions will be of consequence. The trick, then, is to give a slight advantage (in terms of finding the best action) to the action desired to support the equilibrium. This advantage must vanish with time, so that assumption I.3 holds. But things can be arranged so that, for players who move out-of-equilibrium, the advantage vanishes more slowly that does the degree to which these players think their action is of consequence.

At the risk of saying too much about this, let us sketch another example. Consider a two-by-two simultaneous move game, in which both players get payoff 1 in the upper left-hand cell, and both get 0 in the other three cells. The

bottom right-hand cell is a Nash equilibrium, and the reader may worry that we could never get these players to continue to choose bottom and right. But the same sort of trick will work. For the player selecting the row, have this player evaluate the consequences of choosing the top row and the bottom using his assessment of how likely it is that column player will choose the left or the right column. Call the two values v_t and v_b. In round k, have this player play t with probability $(k-1)/k$ if $v_t > 100/\ln k + v_b$; otherwise he plays b with probability $(k-1)/k$. And do the same thing for the column player. Then if the two start off playing bottom and right, respectively, while top and left always look better in the sense that $v_t > v_b$ (and $v_l > v_r$) the extent to which top beats bottom (and left beats right) vanishes more quickly than $1/\ln k$, with positive probability. That, it will turn out, is sufficient.

We begin the proof of the theorem by formalizing a remark made in section 5: In the definition of a locally stable outcome, the part about "with probability arbitrarily close to one" is something of a red herring.

Lemma I.8. Suppose that for a given outcome δ, we can find a partial history ζ_k and a sequence of positive numbers $\{\epsilon_{k'}\}$ with $\lim_{k'\to\infty} \epsilon'_k = 0$ such that, starting from ζ_k, there is positive probability that the following two properties both hold:

(i) For all $k' \geq k$, $d(\delta, \delta(\phi(k'+1, \zeta'_k)) < \epsilon_{k'}$.

(ii) For a given function τ, the consistency test is passed.

Then the outcome is locally stable.

Proof. The proof relies on the "continuity of probability." Let Λ be the event described by (i) and (ii), and let $\hat{\Lambda}$ be its converse, so that we know that $\hat{\Lambda}$ has probability less than one. We can write $\hat{\Lambda}$ as $\cup_{k'\geq k}\hat{\Lambda}_{k'}$, where $\hat{\Lambda}_{k'}$ is the event that k' is the first time (at k or after) that either (i) is violated or the consistency test fails. Moreover, the $\hat{\Lambda}_{k'}$ are disjoint, so that, for any $\epsilon > 0$, there is some j where $\cup_{k'\geq j}\hat{\Lambda}_{k'}$ has probability less than ϵ. Thus, conditional on the union of $\cup_{k'\geq j}\hat{\Lambda}_{k'}$ with Λ, we can make the probability of Λ as close to one as we wish. Taking any ζ_j in the support of this union (which is ζ_j measurable), this gives us probability as close to one as we wish of convergence of the outcomes induced by behavior to delta and satisfaction of the consistency test, which is the

desired condition. [21]

Proof of theorem I.2. (This will have to wait for a subsequent draft. We hope that the sketch of the examples and the lemma just stated give the general idea. We do note that mixed strategy equilibria makes things slightly complicated, but judicious use of the strong law of large numbers will suffice.)

Now we turn to the question: To get local stability of a given outcome, how special must be the starting point ζ_k? Fix behavior and $\epsilon > 0$. Suppose that we can show that δ is approached and the consistency test passed, with probability $1-\epsilon$, starting from some given ζ_k. Now imagine that we begin from a "m-fold replication" ζ_{mk} of ζ_k — a mk-length history that has the same proportion of outcomes as does ζ_k. Can we conclude that, starting from ζ_{mk}, there is probability at least $1-\epsilon$ of approaching δ and satisfying the consistency test?

The answer is yes for strict equilibria, at least if ϵ is sufficiently small and m is sufficiently large. If σ is strict, then σ^n is a strict best response to $\tilde{\sigma}^{-n}$ for all $\tilde{\sigma}^{-n}$ in some ϵ neighborhood of σ. Thus as $m \to \infty$, the players become more certain that σ^n is the optimal choice, and so they will play σ^n will probability approaching one so long as their conjectures remain in this ϵ-neighborhood.

However in other games the answer can be no. In particular, consider the second example discussed above, a two-by-two simultaneous move game, with payoffs $(1,1)$ in the upper-left corner and $(0,0)$ in all other cells. And take the sort of behavior sketched out above for this game. If we replicate the proportions of a given history ζ_k from which we approach the equilibrium lower-left with positive probability, and if those proportions give any weight to either the upper row or the left-hand column, then for a long enough replication, the other player will be moved to choose left or up. Assumption I.3 guarantees that this is so. The longer is the history, the more, in general, must the starting point "look like" the equilibrium, at least if there are any weakly dominated strategies being used in the equilibrium.

Another question that is suggested by theorem I.2 is: Given a particular Nash equilibrium, what is the range of behavior for which it is locally stable? In the proof of the theorem (and in the two examples), we use behavior that is tailor-made for the particular Nash equilibrium. It would be nice to know that

[21] To be very precise, we would have to worry about versions of conditional probability, but the details are fairly standard.

we can vary the behavior a little and obtain the same conclusion. Once again, strict equilibria are a well-behaved special case:

Proposition I.2. Let δ be the outcome of a strict Nash equilibrium. Then for every behavior rule that conforms to assumptions I.1, 2 and 3, δ is locally stable. Moreover, there is a $\epsilon' > 0$ such that: For every behavior rule that conforms assumptions I.1, 2 and 3, if $\{\zeta_k\}$ is any sequence of increasingly longer initial histories such that $d(\delta, \delta(\zeta_k)) < \epsilon'$ for all k, then as $k \to infty$, the probability, conditional on starting from ζ_k, of converging to δ and satisfying the consistency test goes to one.

(We have not yet written out the details of the proof of this proposition, so it should be considered something of a conjecture. But we are fairly sure it is true. Since a *strict Nash equilibrium* is one in which every choice by every player is strictly best in the reduced normal form of the game, only pure strategy equilibria can be strict, and, in a game in extensive form with the restriction on information sets that we have imposed, only equilibria that hit every information set in the tree (with positive probability) can be strict. So if δ is the outcome of a strict equilibrium, every $h \in \text{Supp}(\delta)$, and so δ completely determines behavior strategies at all information sets.)

To the reader of version 0.11:

We hope, in this section, to provide some further results along these lines. In particular, we wish to investigate equilibrium outcomes for equilibria that are strict along the equilibrium path. The first example given above would be one such. We cannot hope that such equilibria will be locally stable for all behavior rules — it is easy to see that if player 2 in the example always selects the better action, which is A, with high probability, then eventually player 1 will choose A. But we may be able to sharpen (broaden?) the set of starting points ζ_k from which such an outcome is locally stable. In particular, in the example, if we replicated history enough, things might be okay, because, while out-of-equilibrium players become ready to do the wrong thing for a large enough replication, they don't get a chance to do so for a while, and this delay in the opportunity to "correct" the perceptions of in-equilibrium players may be delayed so long that out-of-equilibrium players go back to being content with the action that supports the equilibrium.

Following this section, the rest of Part I will consist of:

(1) Some minor results on global convergence. We saw already that there is no hope of a general result on global convergence, but our previous example was based on nongenericies. We hope to show that global convergence will not hold even for some generically chosen games. In particular, we believe (at some level of belief less than complete) that in games such as matching pennies, we can produce behavior rules that give "cycles", where we put the term in quotes because our dynamics are not stationary — successive trips around the cycle will necessarily take longer and longer. In a more positive vein, we can give some global convergence results for some very simple classes of games — e.g., two-by-two pure coordination games. (Our investigations into this are still very preliminary.)

(2) The myopia of our players is quite hard to swallow. Things would be easier to take if we imagined that these encounters resulted from random and anonymous matchings in a large population. This sort of model seems relatively easy to handle if one takes a large but finite population of players from which random matchings are drawn; in this setting, while myopia is not quite fully rational, it does make sense as a nearly-optimal heuristic. With a continuum of players, some mathematical complications arise, however. We will certainly undertake the extension to a finite population of players, showing that our results go through without too much difficulty. We currently doubt that we will attempt to deal with the mathematical issues raised by a continuum of players.

(3) In a concluding section, we will remark on how our restriction on the game form has helped us, and we will give some conjectures/results on what happens when that assumption goes away. As of the time of writing this, we have little idea what will be said here.

PART II — CONDITIONAL PAYOFFS AND SEQUENTIAL EQUILIBRIUM

II.1. CONDITIONAL PAYOFFS

In Part I, we assumed that players evaluate their loss from experimentation on an *ex ante* basis, so that players are not very concerned about their choice of action at information sets that they think are unlikely to be reached. This corresponds to a situation where players must choose their strategies at the beginning of the game and are unwilling to give much thought to optimizing over unlikely events. In this case we saw that our theory led to a justification of Nash equilibrium outcomes if the rate of experimentation is as high as required by Assumption I.1, and if players' hypotheses about their opponents' strategies converged to the prediction given by the empirical distribution function.

Now we turn to the case where players evaluate their losses on a sequential or a conditional basis. Recall that $E^n[a|h;\sigma^{-n}]$ is the conditional expected payoff to n of taking action a at information set h, conditional on reaching h, and based on the conjecture that n's opponents play according to the strategies σ^{-n}. Let $E^n[\star|h;\sigma^{-n}]$ be the maximum of the $E^n[a|h;\sigma^{-n}]$, maximized over $a \in A(h)$, and let $L^n[a|h;\sigma^{-n}] = E^n[\star|h;\sigma^{-n}] - E^n[a|h;\sigma^{-n}]$; that is, L^n gives the conditional loss from using action a at information set h, conditional on reaching h. If players measure losses on a conditional basis, then Assumption I.3 should be replaced by:

Assumption II.1. For each player n and information set $h \in H^n$ there is a function $\alpha : (0,\infty) \times \{1,2,...\} \to [0,1]$ such that the behavior of player n at information set h satisfies

$$\phi^n(k+1,\zeta_k,a) \leq \alpha(L^n[a|h;\sigma^{-n}(k+1,\zeta_k)],\kappa(h,\zeta_k)),$$

where α is nonincreasing in its first argument and, for every $\lambda > 0$, $\lim_{k\to\infty} \alpha(\lambda,k) = 0$.

Remarks. (1) Because α here is assumed to be nonincreasing and because conditional losses always exceed *ex ante* losses, it is clear that this assumption implies assumption I.3. The motivation behind the monotonicity assumption should be clear: A larger loss should result in a lower probability of taking the action. The assumption, of course, doesn't say this. It says that the upper bound on the probability of taking the action is lower. The reader who finds the first statement compelling should consider the counterarguments we will present in Part III.

(2) As with assumption I.3, there is no loss in generality in assuming that a single function α works uniformly at all information sets, and we will proceed on this basis.

(3) Assumption I.1 required that at every information set, the player moving experiments infinitely often. No distinction was made there between information sets that are reached a nonvanishing fraction of the time and those that are reached with vanishing frequency. So assumption I.1 has a flavor of conditional evaluation. If, in the spirit of assumption I.3, players that move at "unlikely" information sets are unconcerned about their actions, they would presumably also be indifferent to how often they experiment. It is therefore worth remaking the point made at the end of the proof of theorem I.1: We did not use the full strength of assumption I.1 in Part I, but could have made do with an assumption that every action is tried infinitely often at information sets that are hit a nonvanishing fraction of the time. A similar remark applies to the consistency test of section I.3. The test as posed is meant to apply to all information sets, including those that might be reached only a vanishing fraction of the time. But we only used the consistency test for information sets along the path of the given outcome in Part I, so for purposes of Part I we could have posed a less stringent test. In the development of this part of the monograph, however, more of the strength of the consistency test and more of the strength of assumption I.1 will be used.

(4) As with assumption I.1, an unpalatable feature of assumption II.1 is that it permits a player to continue to choose an action that is weakly dominated, as long as the player assesses vanishing probability that doing so will harm him. This is necessary here to get a theory of sequential equilibrium, just as it was needed before for a theory of Nash equilibrium — from our perspective, this is a weakness in those two concepts, at least in the sort of framework that we are exploring here.

The strengthening of Assumption I.3 to assumption II.1 does not on its own lead to a theory of sequential equilibrium. We will soon investigate the additional conditions required for sequential equilibrium to obtain. But first it is interesting to note that, for some games at least, assumption II.1, in conjunction with assumptions I.1 and I.2, is quite powerful.

Proposition II.1. In games of perfect information, the only locally stable outcome(s) for behavior that satisfies assumptions I.1, I.2, and II.1 are those outcomes given by backwards induction.

Proof. As noted in section I.2, in games of perfect information assumption I.1 implies that all information sets are reached infinitely often. Thus all players' conjectures about any opponent will converge if that opponent's behavior strategy converges. The strategies of the last players to move along any branch of the tree must converge to actions that are optimal for this player; eventually the next-to-last players will learn this, and they will then play actions given by backwards induction; and eventually the players just before them will learn this, and so on.

II.2. SEQUENTIAL EQUILIBRIUM — PRELIMINARIES

In general games, however, assumption II.1 is not enough to ensure that only sequential equilibrium outcomes are stable. To make this point, we briefly recapitulate some basic notions from Kreps and Wilson (1982):

Beliefs μ are given as a map from information sets to probability distributions over the nodes of each information set. An assessment is a pair (μ, σ) of beliefs and a strategy. An assessment is consistent if there is a sequence of totally mixed behavior strategies $\sigma_k \to \sigma$ such that the assessments μ_k obtained from σ_k by Bayes' rule converge to μ. An assessment (μ, σ) is sequentially rational if, for each player n, σ^n is a best response to σ^{-n} at every information set $h \in H^n$, where the payoff conditional on reaching h is computed from the assessment μ over the nodes of h and the strategies σ^{-n}. Finally, an assessment (μ, σ) is a *sequential equilibrium* if it is consistent and sequentially rational.

Because we have required that player n's conjectures $\sigma^{-n}(k+1, \zeta_k)$ about the joint distribution of his opponents' strategies are based on the assumption that his opponents randomize independently, we have already gone a long way towards assuming that assessments will be consistent in the sense above. The requirement that each player treats his opponents' play as independent is stronger here than in Part I, because now, in considering the conditional maximization of payoffs, a player's behavior would be different if the player thought that there was only a small probability that his opponents could correlate their play. (If players entertained the possibility of such small probabilities of correlation, we would be led to the concept of c-perfect equilibrium of Fudenberg, Kreps and Levine, 1988). Given the built-in independence assumption, we need only ensure that the players' assessments all converge to the same limit and that this limit is consistent with their conjectures about each others' strategies.

The assumptions of Part I are not strong enough to guarantee that this is the case, however. One problem is illustrated by the game in figure II.1. Here player 1 plays L, which ends the game, with probability $1 - 2/k$, and he plays M and R each with probability $1/k$. Assumption I.2 ensures that the beliefs of players 2 and 3 about player 1's strategy are both close to $(1, 0, 0)$ for large

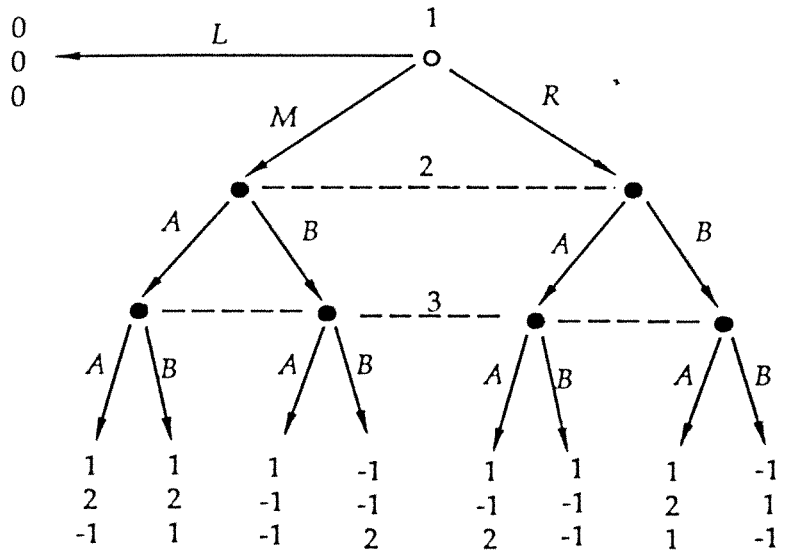

Figure II.1.

k. However the conditionally optimal actions for players 2 and 3 depend on the relative probabilities that they assess for actions M and R, and assumption I.2 does not imply that the relative probabilities assessed by the two players are similar. In the game depicted, any common assessment will lead at least one of the two of them to play A, which will in turn induce player 1 to deviate from L. However if player 2 believes that M is more than twice as likely as R, and player 3 believes that R is more than twice as likely as M, then both 2 and 3 choose B, and player 1 is correct to choose L. Hence L is a stable outcome, even with assumption II.1 in place of I.3.

(It might be of interest to characterize the set of stable outcomes corresponding to assumptions I.1, I.2 and II.1, but we will leave this question for a later draft. Instead, we will go on to pursue more restrictive formulations of our model that will lead us towards sequential equilibrium.)

The natural response to the above example is to strengthen Assumption I.2 to require that the players' beliefs about the relative weights that their opponents give to different actions converge to the empirically observed ratios. (Because the conjectures σ^{-n} are totally mixed, these ratios are all well-defined.)

To do this, we need to take a few preliminary steps. For $r, r' \in [0, \infty)$, let $d'(r, r') = |r/(1+r) - r'/(1+r')|$. That is, d' is a metric on $[0, \infty)$ which compactifies $[0, \infty)$ at ∞. In other words, if $\{r_n\}$ and $\{r'_n\}$ are two sequences of nonnegative numbers, and if $d'(r, r') \to 0$, then, roughly put, either $r_n - r'_n$

goes to zero, or both r_n and r'_n tend to infinity.

Next, recall that $\eta(h, \zeta_k)(a)$ for $a \in A(h)$ is the empirically based estimate of $n(h)$'s strategy at information set h given partial history ζ_k; that is, the ratio of the number of times that $n(h)$ took action a to the number of the times that h was visited. We will write $\sigma^{-n}(k+1, \zeta_k, h)(a)$ for player n's subjective assessment of this probability, for $n \neq n(h)$.

Finally, recall that $H^{IO}(\zeta)$ is the set of information sets that are visited infinitely often along the history ζ.

Assumption II.2. For all histories ζ, players n, information sets h' and h'' both from $H^{-n} \cap H^{IO}(\zeta)$, and actions $a' \in A(h')$ and $a'' \in A(h'')$,

$$\lim_{k \to \infty} d' \left(\frac{\sigma^{-n}(k+1, \zeta_k, h')(a')}{\sigma^{-n}(k+1, \zeta_k, h'')(a'')}, \frac{\eta(h', \zeta_k)(a')}{\eta(h'', \zeta_k)(a'')} \right) = 0.$$

This, of course, is much stronger than assumption I.2, as can be seen immediately in the example of figure II.1. As we noted before, assumption I.2 only requires that 2 and 3 agree asymptotically that player 1 is playing strategy $(1, 0, 0)$. There is no requirement that they agree on the relative frequency with which player 1 plays M vs. R. In assumption II.2, precisely this is required. Players 2 and 3 must asymptotically agree on these relative frequencies, and they must agree that the relative frequencies are whatever is given by history to date.

The reader should note, moreover, that assumption II.2 does even more than this. In the assumption, the information sets h' and h'' can differ, and they can belong to different players. Assumption II.2 requires that players agree with history on the relative rates of one player's propensity to experiment with one action over a second player's propensity to experiment with another, at least at pairs of information sets where a lot of data is accumulated. Of course, precisely this sort of agreement is required in a sequential equilibrium, so this sort of assumption will be needed to provide foundations for sequential equilibrium. Still, this assumption seems to us to be very strong, and we wonder what would happen if it were replaced by something weaker.

At this point, we would like to be able to say whether, for behavior that satisfies assumptions I.1, II.1 and II.2, the only locally stable outcomes are the outcomes of sequential equilibria. We suspect that the answer is no, but we can provide neither counterexample nor proof. The difficulty in providing a proof

(and, although we have not found it, the likely route to a counterexample) is that, at a locally stable outcome, behavior at out-of-equilibrium information sets needn't converge. In Part I, an outcome that is not a Nash equilibrium outcome fails the "optimality test" along the path of the outcome, and local stability of the outcome implies that behavior converges at information sets along the outcome's path. But the optimality test for a sequential equilibrium can be failed either on or off the path of the outcome, so only knowing that behavior along the path converges doesn't give us much to go on for off-the-path behavior.

Accordingly, in order to get a theory of sequential equilibrium, we have to speak not of stable outcomes but of stable strategy profiles. The following pair of definitions replace the definitions of section 4:

Definition. *A profile* $\sigma = (\sigma^1, ..., \sigma^N)$ *of behavior strategies is locally stable with respect to behavior rule* ϕ *if, for every* $\epsilon > 0$, *there is a history* ζ_k *such that conditional on starting at date* k *with history* ζ_k, *there is probability greater than* $1 - \epsilon$ *for the set of histories* ζ *such that:*

(i) $\lim_{k' \to -\infty} d(\phi^{n(h)}(k'+1, \zeta_{k'}, h), \sigma^{n(h)}(h)) = 0$ *for all* $h \in H^{IO}(\zeta)$.

(ii) The consistency test of section 4.2 is passed at ζ *for some fixed test function* τ.

Definition. *A profile* σ *of behavior strategies is unstable with respect to behavior rule* ϕ *if there exists an* $\epsilon > 0$ *such that for every history* ζ_k *and integer* $K \geq k$, *there is zero probability that, starting at date* k *with history* ζ_k, *the ensuing behaviors* $\phi(k'+1, \zeta_{k'})$ *and sample path* ζ *satisfy:*

(i) $d(\phi^{n(h)}(k'+1, \zeta_{k'}, h), \sigma^{n(h)}(h)) \leq \epsilon$ *for all* h *and* k' *such that* $\kappa(h, \zeta_{k'}) \geq K$.

(ii) The consistency test of section 4.2 is passed at ζ *for some fixed test function* τ.

Remarks. (1) As noted previously, we use the same consistency test as in Part I, although we will now use it at information sets that are off the path of the outcome.

(2) Note that if σ is unstable, then it is not locally stable. But, for the same sort of reasons as in Part I, the converse is not true.

Roughly put, our next objective is to show that if a strategy profile σ is not a sequential equilibrium profile, then it is unstable for any behavior that satisfies assumptions I.1, II.1 and II.2. But this is still not quite correct. The problem is that σ may prescribe behavior that is far from sequentially rational at information

sets that are so far off the path of conceivable play that they are irrelevant, even given the experimentation probabilities we are imposing. Imagine, for example, a three player game of the following form. Players 1 and 2 move simultaneously. Each can either have 100 (by moving, say, Left) or 0 (by moving Right). If both move Right, then player 3 is allowed to move, and player 3 can either take 100 or 0. Moves by any player do not affect the payoffs of another player, except insofar as player 3 may not get to move at all. (Suppose he gets 0 if either 1 or 2 select Left.) Now the strategy profile Left for 1, Left for 2, and Right for 3 is not a sequential equilibrium profile. But it is locally stable, even with conditional payoff calculations: If players 1 and 2 choose Right with probability $1/k$ in round k, then player 3 will move only finitely many times. We can select the bound function α in assumption II.1 so that player 3 will pick Right the first, say, thirty times he has the chance. We can then make the probability that player 3 gets fewer than thirty chances as close to one as we like, by starting at histories ζ_k for large k where 3's first chance has yet to come.

The problem of irrational play at "irrelevant" information sets also arose in Part I. Lemma I.2 showed that this did not create any difficulties, as play at irrelevant information sets did not influence the set of best responses along the equilibrium path. When players maximize their conditional payoffs, the notion of relevance must be extended to players who are not on the equilibrium path but who would have a chance to move if there were a unilateral deviation from a "relevant" information set. This consideration leads us to introduce the notion of *sequential relevance*, which we will use to prove a result analogous to Lemma I.2.

As a first step, we define what is meant for an information set $h' \in H^{-n}$ to be *conditionally relevant to player* n *at information set* $h \in H^n$ *under assessment* (μ, π). For every action $a \in A(h)$ we can compute the conditional probability that information set h' is reached if action a is taken at h, conditional on reaching h, if the conditional probability of nodes in h is given by μ and if all players subsequently play in accordance with π. If the maximum conditional probability of reaching h', maximized over $a \in A(h)$, exceeds ϵ, we say that h' is ϵ-conditionally relevant at h under (μ, π). If h' is ϵ-conditionally relevant at h for some $\epsilon > 0$, we say that h' is conditionally relevant at h for given (μ, π).

Next, at an assessment (μ, π), we say that an information set h is ϵ-*sequentially relevant* if (i) the probability of reaching h if players play according

to π exceeds ϵ, or (ii) h is ϵ-conditionally relevant at some h' under (μ, π), for an h' which itself is ϵ-sequentially relevant. If an information set h is ϵ-sequentially relevant for some $\epsilon > 0$, then it is said to be sequentially relevant.

Definition. An sequential r-equilibrium is a consistent assessment (μ, σ) where the actions of each player at every sequentially relevant information set are sequentially rational.

The "r" in the definition is meant to be a mnemonic for "relevant." This new definition squares with the old one according to the following lemma.

Lemma II.1. If (μ, σ) is an r-sequential equilibrium, then there is a sequential equilibrium (μ', σ') such that $\mu = \mu'$ and $\sigma = \sigma'$ at all sequentially relevant information sets (under either assessment). In particular, (μ, σ) and (μ', σ') give the same distribution over terminal nodes.

Proof. We sketch the proof: Let $\{(\mu_n, \sigma_n)\}$ be the sequence of strictly positive strategies and accompanying beliefs which have as limit (μ, σ). We claim that if we change σ_n at information sets that are sequentially irrelevant, compute corresponding beliefs, and then pass to the limit, the limit beliefs at sequentially relevant information sets will continue to be given by μ. Also, changing actions at sequentially irrelevant information sets does not change the sequential rationality of π given μ at sequentially relevant information sets. Thus actions at sequentionally irrelevant information sets are irrelevant, from the point of view of both consistency and sequential rationality, to what goes on at sequentially relevant information sets. The converse to this isn't necessarily true: actions at sequentially relevant information sets can affect beliefs at sequentially irrelevant information sets. But consider: for $n = 1, 2, \ldots$, imagine that players at sequentially irrelevant information sets play the "game", taking as given that actions at sequentially relevant information sets are given by σ_n. For each n, find a sequential equilibrium for the sequentially irrelevant information sets, and then extract a convergent subsequence. This will give the required (μ', σ').

II.3. NECESSARY CONDITIONS

Now we can state the necessity result.

Theorem II.1. If a strategy profile σ is not an sequential r-equilibrium profile, then it is is unstable with respect to all behavior satisfying assumptions I.1, II.1, and II.2.

Sketch of proof. The proof is something of a replica of the proof of theorem I.1, and so we first provide the reader with a sketch. If this sketch makes fairly good sense, there will be little loss in skipping ahead to the Remarks following the proof itself (page xxx).

In the proof of theorem I.1, we supposed that, on a set of positive probability, the consistency test was passed and outcomes eventually were within a close neighborhood of the supposedly not unstable outcome. We were able to use lemma I.1 and lemma I.3 to ensure that, with the same positive probability, those things happened, every experiment was tried infinitely often at information sets that were reached infinitely often, and the empirical distribution of outcomes was close to that of the given outcome. Then we looked along a subsequence where empirically based behavior strategies converged, and supposing that the outcome was not an equilibrium outcome, we derived a contradiction: Some player along the path of the outcome would have conjectures about the strategies of his opponents which, at "relevant" information sets, are approximately the same as those derived empirically. And then this player would wish to deviate.

Here we start in much the same way. Fixing a supposedly not unstable strategy profile, we have a set of positive probability where the consistency test is passed and behavior is eventually close to the fixed profile. Using lemmas I.1 and I.3, we can moreover assume that, on this set, empirically based assessments of behavior strategies at information sets that are hit infinitely often are close to those in the fixed profile, and at every information set that is hit infinitely often, every experiment is tried infinitely often. Taking any realization from the positive probability event, we look along a subsequence where empirically based beliefs converge. If the originally fixed strategy profile is not a sequential r-equilibrium profile, then at some sequentially relevant information set, the player

who is moving is uniformly far from being sequentially rational, if the player's assessments are close to the limit beliefs at all information sets that are sufficiently *conditionally relevant* to the player. We will show that all information sets that are conditionally relevant to this player are hit infinitely often, and assumption II.1 will ensure that this player has an assessment that is close to empirically based beliefs. So we conclude that the player will deviate eventually, which means that the fixed profile wasn't stable after all. Roughly put (and there are a few differences), we follow the proof of theorem I.1, using *strategy profile* in place of *outcome, beliefs* in place of the *strategy profile*, sequentially relevant information sets in place of information sets that lie along the path of play, and *information sets conditionally relevant at a sequentially relevant information set* in place of *information sets that are relevant at information sets along the path of play.*

PART III — EQUILIBRIUM REFINEMENTS AND SOPHISTICATED EXPERIMENTS

III.1. INTRODUCTION: COMPLETE THEORIES OF PLAY

In Part I, we focused on the link between the frequency with which the players experimented and the information they obtained about each other's strategies. We required in assumption I.1 that, at every information set that occurs infinitely often, every action is taken infinitely often, but we did not restrict the way that players experimented, that is, the relative frequency of the various experiments. Now we will study some more "sophisticated" experimentation rules, which assign greater relative probability to experiments that are (heuristically) "more likely" to provide valuable information. Given the results of Part II, most of our conclusions here will be straightforward consequences of the restrictions we place on the experiments, and both the restrictions and their implications in terms of the corresponding sets of locally stable equilibria are closely related to ideas that have already been developed in the literature. Our purpose here is not to propose radically new equilibrium refinements, but rather to use the learning-and-experimentation model as a way to motivate the restrictions we impose and to examine their effects in the context of a "complete theory." This terminology is our shorthand for a methodological point which motivates our work but is also of more general applicability.

In any game, when a player observes another play differently than he had anticipated, his own response will depend on how his explanation for the deviation. (This is an extensive-form description of the idea, but normal-form advocates will note that the same argument can be made *ex ante*.) Thus, one way of thinking about equilibrium refinements is as a combination of sequential rationality and restrictions on player's beliefs at out-of-equilibrium information sets, that is, on

the kinds of stories the player tells himself to explain deviations.

A number of authors (for example, Basu, 1985, Binmore, 1985, Reny, 1987, and Rosenthal, 1981) have noted a fundamental logical puzzle with refinements such as subgame perfection. In these refinements, we maintain the hypothesis that players are "rational" and that certain actions will never be taken by them. And then, when those actions are taken, the theory typically proceeds to assume that players are rational and will not take similar actions. Having been presented with a conterfactual to the theory, the theorist plows ahead blithely with the theory. We interpret these authors as complaining about the lack of a formal theoretical explanation for the observed deviations, without which it is difficult to make sense of the theory. What is needed, it seems, is a formal theoretical explanation, within the model, for any possible observation. This is what we will call a "complete theory" of play in the game.

There do exist complete theories in the literature. Foremost among them is Selten's (1975) notion of a perfect equilibrium, where deviations result from accidental "trembles," with the restriction that the probabilities of trembles at different information sets are independent and commonly known. Perfection in the normal form weakens this complete theory by requiring only independence of trembles between players, but allows each player's trembles at different information sets to be correlated. The notion of c-perfection (Fudenberg, Kreps and Levine, 1988) allows even more correlation in the trembles and also allows different players to have different beliefs about the trembles of a common opponent. Myerson's (1978) notion of properness requires that certain types of mistakes are more likely than others. These refinements are all "complete theories" in our sense because they not only place restrictions on the allowed beliefs at off-the-path information sets, they do so by providing an explanation of why deviations occur. Thus beliefs that are allowed by the theory are those that can be justified by an appropriate choice of tremble probabilities, and disallowed beliefs are those that cannot be so justified.

Other refinements such as the "intuitive criterion" (Cho and Kreps, 1987) and divinity (Banks and Sobel, 1987) have been motivated by the idea that players will interpret deviations as conscious signals which are intended to change the opponent's (receiver's) beliefs about the deviator's type. The style of these refinements is to reject equilibria which would be vulnerable to certain types of conscious signals. While this latter group of refinements often gives answers that we find appealing, they are incomplete in that certain ways of interpreting devi-

ations are disallowed, but we are not told what story the players use to generate the interpretations that are allowed. This makes it difficult to know whether or not the implied model of deviations in internally consistent: Is there a convincing way of justifying the allowed interpretations without simultaneously justifying the interpretations that have been ruled out?

We propose that the best methodology for studying equilibrium refinements is to develop complete theories, ones that provide an explanation for all deviations, and then allow and disallow beliefs accordingly. This is not to say that all complete theories are useful or incomplete theories useless, simply that incomplete theories are difficult to evaluate.

As we have already observed, one class of complete theories is based on deviations being accidental mistakes. A second class, developed in Fudenberg, Kreps, and Levine (1988) is based on the assumption that deviations are the results of the deviator's payoffs being different than had originally been supposed. Just which explanations for deviations are reasonable then depends on what the players' *ex ante* doubts are about each others' payoffs. Although the connection is not made formally explicitly, McLennan's (1985) justifiable equilibria is suggested by a semi-complete theory of the form: Deviations are the manifestation of a player who is confused about the equilibrium that is in force. It is as if there might be two or more populations, playing different equilibria, and with some small probability, individuals from one population are mixed in with the other. (A problem with this story is that there are some deviations which it cannot explain at all, so it is, in a sense, only partially complete.)

This part of the monograph develops a new class of complete theories, one in which deviations are the result of experimentation by the players. In Parts I and II, we placed no restrictions on the relative likelihood of the different experiments, and thus (in Part II) we obtained sequential equlibria as the "stable set" when coupled with the assumptions of conditional payoff maximization and asymptotically correct assessments. Now we wish to assume that the players use somewhat sophisticated experimentation rules. As before, the players follow *ad hoc* rules of behavior and do not do full blown dynamic maximizations, so the increased sophistication of the experimentation rules we develop is not meant to take us anywhere close to full rationality. But we do think that there are "reasonable" heuristics one can impose on experiments, which lead to interesting refinements.

The restrictions that we impose on experimentation probabilities are based

on two general ideas. The first idea is that players should evaluate the relative attractiveness of different experiments by some measure of their option value. By this we mean that, as the reason that players experiment is to guard against the possibility that the experimental action might yield a higher expected utility than the action which currently looks to be the best, the attractiveness of an experiment should be related to an *ad hoc* estimate of the probability that the experimental action will turn out to be better than the current optimum. (As before, we will assume that the players' beliefs are derived from naive empiricism, as opposed to a "correct" statistical procedure.) One implication of this is that dominated strategies are unattractive experiments.

The second idea is that, in assessing the likelihood that an experiment is preferred to the myopically best choice, players asymptotically have infinitely less uncertainty about behavior at information sets that have occurred infinitely more often. Thus, if σ is locally stable, a player on the equilibrium path (one whose information set is in $\text{Supp}(\delta(\sigma))$ should give infinitely less weight to the possibility that play will be different than that suggested by his conjectures σ^{-n} if he takes an action in $\text{Supp}(\delta(\sigma))$ than to the possibility that play will be different than σ^{-n} if the player chooses an action outside of $\text{Supp}(\delta(\sigma))$. We therefore require that the players are aymptotically infinitely more likely to experiment with actions that would be optimal if off-path play differed from expectations than to actions which are only optimal if an opponent deviates along the path. These latter actions are *equilibrium dominated* in the sense of Cho and Kreps (1988), and thus our learning and experimentation model provides a way to justify equilibrium domination in the context of a complete theory.

We begin our development of these ideas in the next section with a simple and well-known arena; that of signaling games. As just noted, we come rather quickly to equilibrium domination. But our theory suggests that one should go beyond equilibrium domination and impose additional constraints in the spirit of Banks and Sobel's (1987) notion of universal divinity. We are lead by the form of our theory to a concept that is slightly different from university divinity, which we call co-divinity. The difference is the following: Universal divinity suggests that deviations are caused by a sender's misperception that the receiver might play a particular mixed-strategy that is a best response for some possible beliefs about the sender's type. In contrast to these point misperceptions, our model implies that it is equally natural for the receiver to anticipate a probability mixture over the receiver's pure strategy best responses, even if that mixture is

not itself ever a best response. (Most readers will find this point a bit obscure; we will explain it in detail later on.) The set of co-divine equilibria is always included in the set of equilibria that satisfy the so-called intuitive criterion; it is neither contained in nor contained by the divine equilibria. To illustrate the difference between universal divinity and co-divinity, we show that, while divinity suffices to select a unique equilibrium in the Spence signalling model (cf. Cho and Kreps, 1988) and in general monotonic signalling games with the Spence single crossing property (Cho and Sobel, 1987), co-divinity will not do so without additional assumptions on the risk preferences of the senders. The reason is that in these games it is never optimal for the receiver to play a mixed strategy, so that if all deviations are caused by point misperceptions, risk preferences are irrelevant. However, if as in co-divinity, a player is uncertain which of two undominated responses the receiver might take, then risk preferences do play a role.

We then turn in section III.3 to a development of these ideas in the context of general games (subject to the restrictions on information sets that we have imposed throughout). Our model suggests that if play converges both on and off the path, then the ranking of the frequency of experiments that we see in section III.2 should be imposed at all of the information sets that are reached infinitely often. This leads to an extension of equilibrium domination that we call *conditional domination.*

In summary, then, if observed deviations from equilibrium are explained as the result of uncertainty about the strategies being played, then plausible restrictions on the nature of that uncertainty lead to fairly strong equilibrium refinements. In our model, strategic uncertainty and experimentation explains all of the observed deviations. This assumption is meant to model situations in which the strategic uncertainty is large relative to the competing explanations for deviations. Thus we should conclude with an emphatic caveat: We think that there are many situations where strategic uncertainty predominates, and where, in consequence, the refinements that we develop are natural and relevant. But there are other situations where "deviations" are at least partially attributable to other factors and, in those situations, other refinements will be more relevant.