ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Prediction, Optimization, and Learning in Repeated Games
Author(s): John H. Nachbar
Source: *Econometrica*, Vol. 65, No. 2 (Mar., 1997), pp. 275-309
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/2171894
Accessed: 09/12/2010 03:01

Econometrica, Vol. 65, No. 2 (March, 1997), 275–309

PREDICTION, OPTIMIZATION, AND LEARNING IN REPEATED GAMES

BY JOHN H. NACHBAR[1]

Consider a two-player discounted repeated game in which each player optimizes with respect to a prior belief about his opponent's repeated game strategy. One would like to argue that if beliefs are cautious, then each player's best response will be in the support, loosely speaking, of his opponent's belief and that, therefore, players will learn as the game unfolds to predict the continuation path of play. If this conjecture were true, a convergence result due to Kalai and Lehrer would imply that the continuation path of the repeated game would asymptotically resemble that of a Nash equilibrium. One would thus have constructed a theory in which Nash equilibrium behavior is a necessary long-run consequence of optimization by cautious players. This paper points out an obstacle to such a theory. Loosely put, in many repeated games, if players optimize with respect to beliefs that satisfy a diversity condition termed *neutrality*, then each player will choose a strategy that his opponent was certain would not be played.

KEYWORDS: Repeated games, rational learning, Bayesian learning.

1. INTRODUCTION

1.1. *Overview*

A STANDARD MOTIVATION FOR GAME THEORY'S emphasis on Nash equilibrium is the conjecture that players will learn to play an equilibrium if they interact repeatedly. This paper focuses on a particular model of learning by optimizing players. In the model considered, two players engage in an infinitely repeated discounted game of complete information. Each chooses a repeated game strategy that is a best response to his prior belief as to his opponent's repeated game strategy. Rather than assume that prior beliefs are in equilibrium, one would like to argue that if beliefs are cautious then each player will choose a strategy that is in the support, loosely speaking, of his opponent's belief and that, therefore, players will learn as the game unfolds to predict the continuation path of play. If this conjecture were true, a convergence result due to Kalai and Lehrer (1993a), hereafter KL, would then imply that the continuation path of the repeated game would asymptotically resemble that of a Nash equilibrium. One would thus have constructed a theory in which Nash equilibrium behavior is a necessary long-run consequence of optimization by cautious players.

[1] This work originated in a conversation with Jeroen Swinkels while I was a visitor at The Center for Mathematical Studies in Economics and Management Science, Northwestern University. The paper has benefited from the comments and suggestions of a number of others, including Richard Boylan, Drew Fudenberg, Ehud Lehrer, David Levine, Bart Lipman, Wilhelm Neuefeind, Yaw Nyarko, Bruce Petersen, Suzanne Yee, Bill Zame, a co-editor, and two anonymous referees. The usual caveat applies. I would like to acknowledge financial support from the Center for Political Economy at Washington University.

This paper points out an obstacle to such a theory. The source of difficulty is that, in many repeated games, for any given strategy to be optimal, the player must believe that certain opposing strategies are so unlikely that the player could not learn to predict the path of play should one of those strategies, for whatever reason, actually be selected. This poses no problem for the existence of Nash equilibrium but it makes it difficult, in the context of learning models, to reconcile optimization with the intuitive notion of cautious belief. Loosely put, the paper's central result is that, in many repeated games, if players optimize with respect to beliefs that satisfy a diversity condition termed *neutrality*, then each player will choose a strategy that his opponent was certain would not be played.

Subsection 1.2 offers a detailed, although still informal, discussion of the paper's motivation, results, and underlying logic. Subsection 1.3 develops a concrete example. Subsection 1.4 comments on some related literature, KL in particular. While the results of this paper do not contradict KL, the results do suggest that the interpretation of KL and related papers requires care. The formal exposition begins with Section 2, which covers basic definitions, and concludes with Section 3, which contains the paper's results.

1.2. *An Informal Exposition*

1.2.1. *Prediction*

Recall that in a repeated game, a (behavior) strategy is a function from the set of finite histories of the repeated game to the set of probability distributions over actions in the stage game (the game being repeated). Thus, given a t-period history h, a strategy σ tells player i to play $\sigma(h)$ in period $t+1$, where $\sigma(h)$ may be either a pure stage game action or a mixture over actions.[2] A player's prior belief is a probability distribution over his opponent's strategies.

A strategy implicitly encodes how a player will behave as he learns from his opponent's past actions. Likewise, a belief records how the player thinks his opponent will behave as he (the opponent) learns. This paper will focus on players who learn via Bayesian updating of their prior beliefs. The assumption of Bayesian learning is satisfied automatically if players optimize. More precisely, if a player adheres to a strategy that is a best response to his belief then, after any t-period history (other than one ruled out by the player's belief or by his own strategy), the player's strategy in the continuation repeated game starting in period $t+1$ will be a best response to his date $t+1$ posterior belief, derived via Bayes's rule, over opposing continuation strategies.

A player's belief as to his opponent's strategy, together with knowledge of his own strategy, induces a probability distribution over paths of play. A player will

[2] In this paper, the term "action" will always refer to the stage game while the term "strategy" will always refer to the repeated game.

be said to *learn to predict the continuation path of play* if, as the game proceeds, the distribution over continuation paths induced by the player's posterior belief grows close to the distribution induced by the actual strategy profile. Here, as elsewhere, the reader is referred to the formal sections of this paper for a precise definition. Note that if players randomize in the continuation game, the actual distribution over continuation paths will be nondegenerate; prediction means that players learn this distribution, not which deterministic path will ultimately be realized.

One might think that players will learn to predict the continuation path of play if each player's prior belief is cautious in the sense of satisfying some form of full support assumption. But the set of possible strategies is so large that, provided the opposing player has at least two actions in the underlying stage game, there is *no* belief that would enable a player to learn to predict the continuation path of play for every possible opposing strategy.[3] This observation may seem counterintuitive since, first, a best response always exists in a discounted repeated game and, second, a best response has the property, noted above, that it is consistent with Bayesian learning. The explanation is that learning in the sense of updating one's prior need not imply that a player is acquiring the ability to make accurate forecasts. Explicit examples where players learn but fail to predict can be found in Blume and Easley (1995).

One response to this difficulty would be to abandon prediction as too burdensome a requirement for learning models. I will have somewhat more to say about this in Subsection 1.4, in the context of the learning model known as fictitious play, but this paper primarily considers an alternate point of view, one implicit in KL, that prediction cannot be lightly abandoned, that prediction may even be part of what one means by rational learning. If one subscribes to this viewpoint, then one must explain why the actual path of play happens to be included in the proper subset of paths that players can learn to predict. Moreover, since the ultimate goal is to explain equilibration in terms of

[3] Informally, the intuition is that, whereas there are only countably many finite histories to serve as data for a player's learning, there are uncountably many continuation strategies. More formally, note that if a player can learn to predict the continuation path of play then, in particular, the player can learn to predict (the distribution over) play in the next period. Let a *one-period-ahead prediction rule* be a function that, for each history, chooses a probability distribution over the opponent's stage game actions. The probability distribution is the rule's prediction for the opponent's action in the next period. For *any* one-period-ahead prediction rule, whether or not derived via Bayesian updating, there exists an opposing strategy that does "the opposite." For example, suppose that in the stage game the opponent has two actions, Left and Right. For those repeated game histories in which the prediction rule forecasts "Left with probability $p \leq 1/2$" in the next period, let the strategy choose "Left with probability 1." Conversely, for those histories in which the prediction rule forecasts "Left with probability $p > 1/2$" in the next period, let the strategy choose "Right with probability 1." This strategy is well-formed (in particular, it is a function from the set of finite histories of the repeated game to the set of probability distributions over stage game actions) and against this strategy the prediction rule always gets the probability wrong by at least $1/2$. Since the prediction rule was arbitrary, it follows that there is no one-period-ahead prediction rule that is asymptotically accurate against all strategies.

repeated interaction, one wants to explain prediction without imposing equilibrium-like restrictions on prior beliefs.[4]

1.2.2. *Conventional Strategies*

The approach proposed here is to suppose that, associated with each player, there is a subset of repeated game strategies. For want of a better term, I will refer to such strategies as *conventional.* I will offer some possible examples below. Players are assumed to have a slight (e.g., lexicographic) preference for conventional strategies. Thus, a player will choose a conventional strategy if there is one that is a best response (in the standard sense of maximizing the expected present value of the player's stage game payoffs). If no conventional strategy is a best response, a player will optimize by choosing a nonconventional best response. For the moment, I put aside the possibility that players might be *constrained* to play conventional strategies.

Suppose that the following properties hold whenever each player's belief is, in some appropriate sense, cautious.

1. *Conventional Prediction.* For any profile of conventional strategies, each player, via Bayesian updating of his prior belief, learns to predict the continuation path of play.[5]

2. *Conventional Optimization.* For each player there is a conventional strategy that is a best response.

Then, if beliefs are cautious, Conventional Optimization and the maintained interpretation of conventionality imply that each player, in choosing a best response, will choose a conventional strategy. Since both players play a conventional strategy, Conventional Prediction implies that each player will learn to predict the continuation path of play. Thus players both optimize and learn to predict the path of play and hence the KL convergence theorem implies that the path of play will asymptotically resemble that of a Nash equilibrium.

While Conventional Prediction and Conventional Optimization hold trivially if the product set of conventional strategies consists of a single repeated game Nash equilibrium profile, such a conventional set assumes away the problem of equilibration. To satisfy the objective of not imposing equilibrium-like restrictions on prior beliefs, one wants beliefs to be cautious not only in the sense that beliefs satisfy some form of full support condition with respect to the conven-

[4] This is in contrast to the literature on learning *within* (Bayesian) equilibrium; see Jordan (1991). In that literature, unlike here, players have incomplete information about each other's payoffs, which makes learning nontrivial even when equilibrium is assumed.

[5] Thus, each player learns to predict the path of play regardless of which strategy he selects. Weakening the definition of Conventional Prediction would require constructing a model in which both a player's strategy choice and the set of paths of play that he can predict are determined jointly. There is a danger in such a model of inadvertently assuming away the problem of equilibration. In any event, KL attempts to finesse constructing such a model and I will attempt to do so as well.

tional strategies but also in the sense that the conventional strategy sets are themselves *neutral* with respect to equilibration.

In this paper, neutrality will mean that the map, call it Ψ, that assigns product sets of conventional strategies to games satisfies the following properties (the formal definition is in Section 3.1).

1. *Ψ depends only on the stage game's strategic form.* In particular, Ψ ignores both stage game payoffs and the discount factor. As I will discuss below, this does *not* imply that player *beliefs* ignore payoff information. One might argue that Ψ should take into account payoff information at least in order to rule out nonrationalizable strategies. Doing so would somewhat restrict the scope of the paper's main Theorem without fundamentally changing the analysis. In many of the repeated games considered below, including all of the games based on 2×2 stage games, *all* strategies are rationalizable.

2. *Ψ is symmetric.* In particular, Ψ satisfies *player symmetry* and *action symmetry*.

(a) Player symmetry specifies that if both players have the same action set in the stage game then if some strategy σ is conventional for player 1, the strategy σ' that is equivalent to σ from player 2's perspective must be conventional for player 2. In conjunction with property 3(b) of neutrality (see below), player symmetry implies that if players have the same action set, their conventional sets will, in fact, be identical; see the Claim established in the Proof of Proposition 2.

(b) Action symmetry implies that if two possible action sets for player i have the same cardinality then, holding the opponent's action set fixed, the associated conventional sets for player i are identical up to a renaming of his stage game actions.

3. *Ψ is invariant to simple changes in strategy.* If a strategy σ is conventional for player i, then so is any strategy σ' that is identical to σ except that:

(a) σ' in effect translates σ's action choices according to some function on the set of player i's actions, or

(b) σ' in effect translates input histories according to some bijection on the set of action profiles.

Such strategy changes are simple in the sense that if σ can be represented as a finite computer program, then a program for σ' can be constructed merely by adding a few lines of additional code to translate action choices, input histories, or both.[6] If invariance is violated, then a player whose forecasts are persistently wrong may *never* notice that his opponent's behavior is consistent with a simple variation on one of the strategies that the player *could* learn to predict. This sort of thick-headedness runs counter to what one informally means by a player being cautious.

[6] Similarly, if σ can be represented by a finite automaton, as in Kalai and Stanford (1988), then an automaton for σ' can be constructed by straightforward changes to the output function and the transition rules, leaving the set of automaton states unchanged.

4. Ψ *is consistent.* Consider any pair of stage game action sets, (A_1, A_2) and (A'_1, A'_2), $A_i \subset A'_i$, where A_i is the action set of player i. Consistency requires the following.

(a) Suppose that strategy σ is conventional for player i when the action sets are (A_1, A_2). Then σ extends to a strategy σ' that is conventional when the action sets are (A'_1, A'_2).

(b) Conversely, suppose that strategy σ' is conventional for player i when the action sets are (A'_1, A'_2) and suppose further that σ' restricts to a well-formed strategy σ when the action sets are (A_1, A_2). Then σ is conventional for (A_1, A_2).

5. Ψ *permits pure strategies.* More accurately, for each conventional nonpure strategy, there should be at least *one* pure strategy in its support that is likewise conventional.[7] If a conventional strategy is fully random (that is, after any history, it assigns positive probability to each of the available stage game actions), this property means only that *some* pure strategy is conventional. One motivation for this is that a randomizing strategy σ for player i is inherently more complicated than some of the pure strategies in its support. Explicitly, given σ, choose some (arbitrary) ranking for player i's stage game actions and consider the pure strategy s that, after any history h, chooses the highest ranked action to which $\sigma(h)$ gives positive probability. For example, if player i has only two actions, Left and Right (ranked in that order), s chooses Left after any history such that σ randomizes. For any standard notion of complexity, σ is more complicated than s. Indeed, σ uses s as a kind of pattern and adds to s the additional complication of randomization after certain histories. If one views a conventional strategy set as being built up from less to more complicated strategies then, for any conventional randomizing strategy like σ, some pure strategy like s should be conventional as well.[8]

A product set of conventional strategies is *neutral* if there is neutral map Ψ such that the product set is in the image of Ψ.

Neutrality is a property of the conventional sets rather than directly of beliefs. For example, as already noted, the fact that Ψ ignores payoffs does not imply that each player's belief ignores payoffs. Similarly, players may have the same conventional set without their beliefs being identical. In fact, I require nothing of beliefs other than that players be able to learn to predict the path of play when the strategy profile is conventional. This property can be satisfied even if beliefs are in many respects quite arbitrary. For example, if the set of conventional strategies is at most countable, then it follows from results in KL that Conventional Prediction will hold provided only that each player's belief assigns

[7] A pure strategy s will be said to be in the support of a strategy σ if, after any history, the action chosen by s is also chosen with positive probability by σ.

[8] One might object that, while players might not deliberately favor randomization, it may be impossible to execute pure strategies because of "trembling." Thus, all conventional strategies should be random. As will be discussed in Section 3, see in particular Remark 3 and Remark 8, allowing for slight trembling does not materially affect the argument.

positive probability to each of his opponent's conventional strategies, regardless of exactly how probability is assigned.

The prototypical examples of neutral, conventional sets are those consisting of strategies that satisfy some standard bound on complexity. Examples of such sets include the strategies that are memoryless (for example, strategies of the form, "in each period, play Left with probability p, Right with probability $1-p$, regardless of the history of date"), the strategies that remember only at most the last τ periods, and the strategies that can be represented as a finite flow chart or program. It bears repeating that taking the conventional set to consist of the strategies that satisfy some complexity bound does not imply that players are constrained to such strategies or that players are in any customary sense boundedly rational. Rather, the implication is merely that players have a slight preference for strategies that are simple.

This paper takes the point of view that, while one might ask a learning theory based on optimization and caution to be *robust* to deviation from neutrality, the theory should not *require* such deviation. For example, it would be disturbing if the theory required either player to view particular opposing strategies as nonconventional even though those strategies were computationally simple variants of conventional strategies. To the extent that the theory requires a deviation from neutrality, the theory requires some degree of equilibration prior to the start of repeated play.

1.2.3. *The Main Result*

The central result of this paper is the following Theorem, stated informally here.

In discounted repeated games based on stage games in which neither player has a weakly dominant action, if players are sufficiently impatient then for *any* neutral conventional set there is *no* belief for either player such that Conventional Prediction and Conventional Optimization both hold. Moreover, for many of these games, including repeated Matching Pennies, Rock/Scissors/Paper, and Battle of the Sexes, the same conclusion holds for *any* level of player patience.

As will be discussed in Remark 3 in Section 3.3, the Theorem is robust to small deviations from neutrality.

The Theorem states that, unless one is willing to violate neutrality, it is impossible in many games to formulate a model of learning that is closed in the sense that Conventional Optimization and Conventional Prediction both hold simultaneously. In particular, if the conventional set is neutral and if Conventional Prediction holds, then each player, in the course of optimizing, will choose a strategy that is *not* conventional. Player beliefs in such a model are naive: each player believes that the other plays a conventional strategy even though, in fact, neither plays a conventional strategy. Section 1.3 develops a simple learning model that exhibits this sort of naiveté in a stark fashion. A somewhat more

sophisticated example is provided by the learning model known as fictitious play, discussed in Section 1.4.2. In both examples, naiveté can lead to a failure of convergence, in any reasonable sense, to even approximate Nash equilibrium behavior. What this naiveté means in general for convergence to Nash equilibrium behavior is not known.

The argument underlying the Theorem runs as follows. For games of the sort described, for any pure strategy s for player 1, there are strategies s' for player 2 such that, under any such profile (s, s'), player 1 gets a low payoff in every period. For example, in repeated Matching Pennies, if s' is a best response to s, then under the profile (s, s'), player 1 gets a payoff of -1 in each period, whereas his minmax payoff is 0 per period. It follows that if s is a best response to player 1's belief, then it must be that player 1 is convinced that player 2 will not choose s', so convinced that, if player 1 chooses s, he cannot, via Bayesian updating of his prior, learn to predict the continuation path of play should player 2, for whatever reason, choose s'. The problem that arises is that if s is conventional for player 1, then neutrality implies that at least one of the s'-type strategies will be conventional for player 2. Hence, either Conventional Prediction or Conventional Optimization must fail.

It might seem that this argument depends in an essential way on the fact that s was taken to be pure. After all, a player can often avoid doing poorly (in particular, earning less than his minmax payoff) by randomizing. But not doing poorly is not the same thing as optimizing. In fact, the Theorem extends to include conventional strategy sets that contain randomizing strategies. To see this, note that if a nonpure strategy is a best response to some belief, then so is every pure strategy in its support.[9] Suppose that Conventional Prediction holds. Since I have assumed that, for any conventional nonpure strategy, some pure strategy in its support is also conventional, and since, by the above argument, no conventional pure strategy is optimal, it follows that no conventional nonpure strategy can be optimal either.[10]

To make the Theorem somewhat more concrete, consider any product conventional set consisting of strategies that satisfy a bound on complexity. Standard complexity bounds yield neutral conventional sets that are at most countable. As noted in the discussion of neutrality, it follows that for any such conventional set there are beliefs for which Conventional Prediction holds.[11] To be optimal with respect to such beliefs, a strategy must be flexible enough to make use of the player's predictive ability. Such a strategy will necessarily be complicated. In fact, the Theorem implies that a player's best response will

[9] This fact, while standard for finite games, is less obvious for discounted infinitely repeated games. The Appendix provides a proof.

[10] It is natural to ask whether this negative result could be overturned if one allowed players to have a strict preference for randomization in some circumstances. This question will not be pursued here since it necessarily requires departing from standard subjective expected utility theory.

[11] It is important to understand that prediction, not countability, is the central issue. The same argument would carry over to a complexity bound that yields an uncountable set *provided* Conventional Prediction continued to hold.

violate the complexity bound defining conventionality.[12] Any attempt to obtain Conventional Optimization by adding more complicated strategies into the conventional set is fruitless as long as neutrality is preserved: adding more complicated strategies just makes the best response that much more complicated. The only way to obtain Conventional Optimization is to add in so many strategies that Conventional Prediction is lost. In particular, if one takes the conventional set to be the set of all strategies (which is uncountable), Conventional Optimization holds, but, as argued above, Conventional Prediction fails.

1.2.4. *Extensions: Constrained and Boundedly Rational Players*

Although the primary focus of this paper is on players who are rational, in particular, on players who have unlimited ability to optimize, it is natural to ask whether the analysis would change fundamentally if players were constrained in some way.

Suppose first that each player's computational ability is unrestricted but that the rules of the repeated game are modified to require each player to choose a conventional strategy. For example, the conventional set might consist of the strategies that can be encoded as a finite list of instructions (a program) and the rules of the game might require players to submit their strategies in this form to a referee, who then executes the strategies on behalf of the players.

Given that players are constrained, the Theorem implies that players will be unable to optimize (assuming that the conventional set is neutral and that Conventional Prediction holds). This is not necessarily a disaster, since one might still hope to find conventional strategies that are approximate best responses. In order to apply convergence results along the lines of those in KL, the appropriate version of approximate optimization is what will be called *uniform ε optimization*: a strategy is ε optimal if it is ε optimal *ex ante* and if, moreover, it induces an ε optimal continuation strategy in every continuation game (more precisely, in every continuation game that the player believes can be reached with positive probability).

If the conventional set consists only of pure strategies, then the argument sketched above extends immediately to uniform ε optimization. Therefore, for any neutral conventional set, if Conventional Prediction holds, then Conventional Uniform ε Optimization fails for ε sufficiently small. This need not prevent a player from choosing a strategy that is only *ex ante* ε optimal. But, as illustrated in Section 1.3, *ex ante* ε optimization *per se* may not be enough to guarantee convergence to approximate Nash equilibrium play.

[12] A potential source of confusion is that it is well known that many of the possible bounds on complexity generate conventional sets with the property that, for any conventional strategy, there is a conventional best response. There is no contradiction with the Theorem because this sort of closure looks only at beliefs that are degenerate in the sense of assigning all mass to a single strategy. A belief for which Conventional Prediction holds for a neutral conventional set is intrinsically nondegenerate.

If, on the other hand, the conventional set contains nonpure strategies, then the argument sketched above does not extend. Section 3.4.1 will show that, nevertheless, the first part of the Theorem, in which players are impatient, does extend for the benchmark case in which the conventional set consists of the strategies that can be represented as a finite program, even if the program has access to randomizers (coin tossers).

Finally, Section 3.4.2 contains some remarks about players who are boundedly rational, that is, players for whom deliberation is costly.

1.3. *An Example*

Consider the game Matching Pennies, given by:

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

For any discount factor, the unique Nash equilibrium strategy profile for repeated Matching Pennies calls for both players to randomize 50:50 in every period, following any history.

Suppose that the conventional set, $\hat{\Sigma}$ for either player, consists of three strategies: randomize 50:50, "H always," denoted $\overline{H}$, and "T always," denoted $\overline{T}$. Thus, $\hat{\Sigma} = \{50:50, \overline{H}, \overline{T}\}$. Note that $\hat{\Sigma} \times \hat{\Sigma}$ is neutral.

Assume that each player's belief assigns probability one to the set $\hat{\Sigma}$ and positive probability to each of the three elements of $\hat{\Sigma}$. I do not require that player beliefs be equal. It follows from results in KL that, for any such beliefs, Conventional Prediction holds. Thus, for example, if Player 2 plays $\overline{H}$, Player 1 will observe a long initial string of H's, hence Player 1's posterior will gradually favor the possibility that Player 2 is playing $\overline{H}$, and so Player 1 will come to predict that Player 2 will continue to play H in subsequent periods.

Now consider Conventional Optimization. Behavior under a best response must respond to the information learned over the course of the repeated game. In particular, if Player 1 learns to predict $\overline{H}$, then Player 1 should start playing H in every period, while if Player 1 learns to predict $\overline{T}$, he should start playing T in every period. None of the three strategies in $\hat{\Sigma}$ have this sort of flexibility. As a consequence, Conventional Optimization fails: none of the conventional strategies is a best response to any belief that gives weight to every strategy in $\hat{\Sigma}$. If players optimize, players must, therefore, choose nonconventional strategies. This model thus exhibits the sort of naiveté discussed in Section 1.2.3.

In this example, the players' naiveté can lead to asymptotic behavior that is far from that of a Nash equilibrium. In particular, note that one optimal strategy for player 1, arguably the most obvious one, is to play H or T in the first period (the choice will depend on player 1's prior belief) and then to switch permanently to H always if player 2 played H in the first period, or to T always if

player 2 played T in the first period. A similar (but mirror image) strategy is optimal for player 2. If the players adopt such pure strategies, then from period 2 onward the continuation path will be either $((H,H),(H,H),\ldots)$, $((H,T),(H,T),\ldots)$, $((T,H),(T,H),\ldots)$, or $((T,T),(T,T),\ldots)$, depending on what happens in the first period (which in turn depends on player beliefs). None of these paths resembles a likely realization of the (random) Nash equilibrium path of play.[13]

Suppose instead that, as was discussed in Section 1.2.4, players are *constrained* to choose from among the three strategies in $\hat{\Sigma}$. For ε low, none of the conventional strategies is uniformly ε optimal, again because none of the conventional strategies exploits the fact that the player learns to predict the path of play. If each player chooses a strategy that is merely *ex ante* ε optimal, rather than uniformly ε optimal, then each player will strictly prefer either $\overline{H}$ or $\overline{T}$ to 50:50, depending on his prior belief, unless his prior happens to put exactly equal weight on $\overline{H}$ or $\overline{T}$. In the latter case, the player will be indifferent between all three strategies. But, if both players select pure strategies, then the path of play will be one of the four discussed in the previous paragraph, none of which resembles a likely realization of the Nash equilibrium path of play.

As this paper's Theorem indicates, the naiveté illustrated above is not limited to Matching Pennies and in particular does not depend on the fact that Matching Pennies has no pure strategy equilibrium. Consider, for example, perturbing the stage game to the following:

	H	T
H	1,1	−1,−1
T	−1,−1	1,1

Once again, assume that each player's belief assigns probability one to the set $\hat{\Sigma} = \{50:50, \overline{H}, \overline{T}\}$ and positive probability to each of the three elements of $\hat{\Sigma}$. Then no element of $\hat{\Sigma}$ is a best response. If each player does indeed optimize, either fully or with respect to the constraint that his strategy be in $\hat{\Sigma}$, then the possible continuation paths from period 2 onward include $((H,T),(H,T),\ldots)$ and $((T,H),(T,H),\ldots)$, neither of which is an equilibrium path.[14]

The conventional set $\hat{\Sigma} = \{50:50, \overline{H}, \overline{T}\}$ used in the above examples is, of course, extremely limited. Section 1.4.2 briefly discusses the behavior of fictitious play, a more satisfactory learning model in which $\hat{\Sigma}$ is taken to be the set of all memoryless strategies.

[13] With probability 1, a realization of the equilibrium path of play will have the property that each of the four possible action profiles (H,H), (H,T), (T,H) and (T,T), appears with a population frequency of 1/4.

[14] In a coordination game such as this, one might expect the players to break out of repeated miscoordination by finding some direct means of communication. While direct communication might be descriptively realistic, appealing to such communication would violate the objective of trying to explain equilibration solely through repeated play.

1.4. *Remarks on the Literature*

1.4.1. *On Kalai and Lehrer* (1993*a*)

KL, together with its companion paper, Kalai and Lehrer (1993b), does two things. First, KL provides a condition on beliefs that is sufficient to ensure that a player learns to predict the path of play. The KL condition is in the spirit of (but is weaker than) assuming that each player puts positive prior probability on the actual strategy chosen by his opponent.[15] Second, KL establishes that if players optimize and learn to predict the path of play, then the path of play asymptotically resembles that of a Nash equilibrium.[16]

While the KL sufficient condition for prediction is strong (from the discussion in Section 1.2.1, any such condition *must* be strong), it has the attractive feature that it imposes essentially no restriction on the player's belief over strategies other than his opponent's actual strategy. It would thus seem that a construction along the lines proposed above, in which the KL sufficient condition is satisfied by means of a full support assumption with respect to some set of conventional strategies, ought to work. That this construction fails stems from the fact that the joint requirement of prediction *and* optimization is far more burdensome than the requirement of prediction alone. This complicates the interpretation of KL and also of related papers such as Kalai and Lehrer (1995).

By way of example, consider again the case of Matching Pennies with the conventional set $\hat{\Sigma} = \{50:50, \overline{H}, \overline{T}\}$. One would like to argue that the path of play will converge to that of the unique Nash equilibrium. The only conventional strategy profile for which this occurs is the one in which both players choose 50:50. Suppose then that both choose 50:50. The KL sufficient condition is satisfied provided only that each player assigns positive probability to the other choosing 50:50. But 50:50 will not be *optimal* for a player unless the player assigns *zero*, not just low, probability to both $\overline{H}$ and $\overline{T}$.[17] In this example, 50:50

[15] The KL prediction result generalizes an earlier theorem of Blackwell and Dubins (1962). For sufficient conditions that are weaker than the KL condition, see Lehrer and Smorodinsky (1994) and Sandroni (1995).

[16] The KL convergence result is intuitive but, for discount factors sufficiently close to 1, it is not immediate. Even if players accurately predict the continuation path of play, they can hold erroneous beliefs about what would happen at information sets off the path of play. KL, see also Kalai and Lehrer (1993b), verifies that an equilibrium with approximately the same path of play can be constructed by altering strategies so as to conform with beliefs at unreached information sets. When there are more than two players, there are additional complications. See also Fudenberg and Levine (1993). In the weak (pointwise convergence) topology, convergence is to the path of a true Nash equilibrium. In the strong (uniform convergence) topology, KL shows convergence only to the path of an ε-Nash equilibrium. See also Sandroni (1995).

[17] As discussed in Section 1.3, if player 1 assigns positive probability to every strategy in $\hat{\Sigma} = \{50:50, \overline{H}, \overline{T}\}$ then no conventional strategy is optimal. If player 1 assigns probability $p \in (0,1)$ to 50:50 and probability $1-p$ to $\overline{H}$, then player 1's best response is $\overline{H}$, not 50:50. Similarly if player 1 assigns probability p to 50:50 and probability $1-p$ to $\overline{T}$, then player 1's best response is $\overline{T}$, not 50:50.

can be optimal for both players only if beliefs are actually in equilibrium at the start of repeated play.

1.4.2. *Fictitious Play and (Semi-) Rational Learning*

For simplicity, I focus initially on stage games with two actions for each player.

The classical fictitious play model of Brown (1951) can be shown to be equivalent to a model in which each player optimizes with respect to the belief that this opponent is playing a memoryless strategy, that is, a strategy of the form "in any period, go Left with probability p, Right with probability $1-p$, regardless of history," with p, which is constant across all periods, drawn from a beta distribution. See, for example, Fudenberg and Levine (1996). The set of memoryless behavior strategies, viewed as the conventional set, is neutral. One can show that Conventional Prediction holds (even though beliefs in this case violate the KL sufficient condition). Hence Conventional Optimization must fail. Thus, while players under fictitious play are rational in the sense that each chooses a best response to his belief, the beliefs themselves are naive: each player believes that his opponent adopts a memoryless strategy even though each, in fact, adopts a strategy that is history dependent.

Despite this naiveté, there are many examples in which players under fictitious play do learn to predict the continuation path of play and hence play does converge to that of a Nash equilibrium. Moreover, even when prediction fails, play may still exhibit Nash equilibrium-like behavior. Consider, for example, Matching Pennies. Under fictitious play, each player in Matching Pennies learns to believe that his opponent is randomizing 50:50 even though the actual path of play is typically nonstochastic. Thus players do not learn to predict the actual path of play and the actual path does not converge, in the sense used here and in KL, to the stochastic path generated by the unique Nash equilibrium of repeated Matching Pennies. Nevertheless, both the empirical marginal and the empirical joint frequency distributions of play converge to that of the Nash equilibrium of Matching Pennies. Thus, behavior under fictitious play is consistent with many (although not all) of the observable consequences of players learning to play the Nash equilibrium of Matching Pennies.

Unfortunately, fictitious play is not always so well behaved. In 2×2 stage games, while empirical marginal frequency distributions of play always converge to a Nash equilibrium of the stage game, the empirical *joint* frequency distribution may be inconsistent with Nash equilibrium. This point has been emphasized by Fudenberg and Kreps (1993), Jordan (1993), and Young (1993). Moreover, there are robust examples of stage games with more than two actions, or more than two players, in which even the empirical marginal frequency distributions fail to converge, a point originally made by Shapley (1962); see also Jordan (1993). What is perhaps more disturbing is that, in the examples in which convergence fails, the path of play cycles in ways that are obvious to the outside analyst but that the players themselves fail to detect.

These problems with asymptotic behavior under fictitious play stem from the fact that player beliefs are naive. While the message of this paper is that some degree of naiveté may be unavoidable, one might still hope to construct theories of (semi-) rational learning in which players are more sophisticated than in fictitious play. For recent work along these general lines, see Fudenberg and Levine (1995b), Fudenberg and Levine (1995a), Fudenberg and Levine (1996), Aoyagi (1994), and Sonsino (1995). A feature of much of this literature is that players are modeled as using strategies that are intuitively sensible without necessarily being best responses to well-formed prior beliefs. Justifying these strategies as optimal or near optimal may require enriching the repeated game model or deviating from standard decision theory, or both.

1.4.3. *Problems with Rationality*

Binmore, in Binmore (1987) and elsewhere, has warned that the concept of rationality in game theory may be vulnerable to problems akin to the unsolvability of the Halting Problem; see also Anderlini (1990).

Following Binmore, view a player in a one-shot game as choosing a *decision procedure*, a function that, taking as input a description of the opponent's decision procedure, chooses as output an action of the game. This formalism is an attempt to capture the idea that a player, in choosing his action, predicts his opponent's action by thinking through the game from his opponent's perspective. Since a player is assumed to know his opponent's decision procedure, the player can predict his opponent's action. The goal is to construct a decision procedure that, for any opposing decision procedure, chooses an action that is a best response to the action chosen by the opponent's decision procedure.

It is not hard to see that no decision procedure is optimal for Matching Pennies.[18] Perhaps more surprisingly, there may be no optimal decision procedure even in games with equilibria in pure actions. The basic difficulty is that there are so many possible opposing decision procedures that there may be no decision procedure that optimizes with respect to them all. Canning (1992) shows that, for a large set of games with equilibria in pure actions, one can close the decision problem by limiting players to *domains* (subsets) of decision procedures. Here "close the decision problem" means that a player finds it optimal to choose a decision procedure within the domain whenever his opponent's decision procedure is likewise within the domain. As Canning (1992) emphasizes, the domains, while nontrivial, necessarily embody rules of equilibrium selection. In games with multiple equilibria, different rules of equilibrium selection give rise to different domains.

[18] If players are constrained to play pure actions, the case originally considered in the literature, then the existence of an optimal decision procedure would imply the existence of a pure action Nash equilibrium, which is false. An argument similar to the one given in footnote 3 establishes that no decision procedure can be optimal even if players can randomize.

The overlap between this paper and the literature just sketched would appear to be small. In this paper, neither player knows the other's decision procedure (indeed, a player's decision procedure for choosing a strategy is not even explicitly modeled), and neither player knows the other's repeated game strategy. Each player merely has a belief as to his opponent's strategy and one would like to permit each player's belief to be inaccurate in the sense of assigning considerable probability mass to strategies other than the one his opponent is actually playing. But while neither player in the present model may have accurate knowledge of his opponent *ex ante*, the insistence on prediction means that players will have increasingly accurate knowledge as the game proceeds. If the conventional set is neutral, asking for a conventional strategy that is optimal when Conventional Prediction holds is akin to asking in Binmore's model for a decision procedure that is optimal against all (or at least a large set of) opposing decision procedures. Conversely, the domain restrictions discussed in Canning (1992) are suggestive of the deviations from neutrality that would have to obtain if Conventional Prediction and Conventional Optimization were to hold simultaneously.

2. SOME BACKGROUND ON REPEATED GAMES

2.1. *Basic Definitions*

Consider a 2-player game $G = (A_1, A_2, u_1, u_2)$, the *stage game*, consisting of, for each player i, a finite *action set* A_i and a *payoff function* u_i: $A_1 \times A_2 \to \mathbb{R}$.

The stage game is repeated infinitely often. After each period, each player is informed of the *action profile* $(a_1, a_2) \in A_1 \times A_2$ realized in that period. The set of *histories* of length T, $\mathcal{H}^T$, is the T-fold Cartesian product of $A_1 \times A_2$. $\mathcal{H}^0$ contains the single abstract element h^0, the null history. The set of all finite histories is $\mathcal{H} = \bigcup_{T \geq 0} \mathcal{H}^T$. I will sometimes write $\mathcal{H}(A_1, A_2)$ to emphasize the dependence of $\mathcal{H}$ on (A_1, A_2). An infinite history, that is, an infinite sequence of action profiles, is called a *path of play*. The set of paths of play is denoted by $\mathcal{Z}$. The projection of a path of play $z \in \mathcal{Z}$ onto its period t coordinate is denoted z_t. The projection of z onto its first t coordinates, that is, the t-period initial segment of z, is denoted $\pi(z, t)$; note that $\pi(z, t) \in \mathcal{H}^t$.

A (*behavior*) *strategy* for player i is a function $\sigma: \mathcal{H} \to \Delta(A_i)$, where $\Delta(A_i)$ is the set of probability mixtures over A_i. I will sometimes write σ_i to emphasize that the strategy is associated with player i. Let Σ_i be the set of behavior strategies of player i. I will sometimes write $\Sigma_i(A_1, A_2)$ to emphasize the dependence of Σ_i on (A_1, A_2). A *pure strategy* for player i is simply a behavior strategy that takes values only on the vertices of $\Delta(A_i)$. Let $S_i \subset \Sigma_i$ be the set of pure strategies for Player i.

Strategy $\sigma^* \in \Sigma_i$ will be said to *share the support* of strategy $\sigma \in \Sigma$ iff, for any history, h, if $\sigma^*(h)$ assigns positive probability to action $a \in A_i$, then so does $\sigma(h)$. In the case of a pure strategy, $\sigma^* = s$, I will say that s is *in the support of* σ.

$\Sigma_1 \times \Sigma_2$ denotes the set of *behavior strategy profiles* in the repeated game. For each t, a behavior strategy profile (σ_1, σ_2) induces a probability distribution over cylinders $C(h)$, where h is a t-period history and $C(h)$ is the set of paths of play for which the t-period initial segment equals h. These distributions can in turn be extended in a natural way to a distribution $\mu_{(\sigma_1, \sigma_2)}$ over $(\mathcal{Z}, \mathcal{F})$, where $\mathcal{F}$ is the smallest σ-algebra containing all the subsets formed by the cylinders; Kalai and Lehrer (1993a) discuss this construction in somewhat more detail.

Fix a discount factor $\delta \in [0,1)$. The payoff to player i in the repeated game is then given by V_i: $\Sigma_1 \times \Sigma_2 \to \mathbb{R}$,

$$V_i(\sigma_1, \sigma_2) = \mathbb{E}_{\mu_{(\sigma_1, \sigma_2)}}\left(\sum_{t=1}^{\infty} \delta^{t-1} u_i(z_t) \right)$$

where $\mathbb{E}_{\mu_{(\sigma_1, \sigma_2)}}$ denotes expectation with respect to the induced probability $\mu_{(\sigma_1, \sigma_2)}$.

2.2. *Beliefs*

Player 1's *ex ante* subjective belief over player 2's behavior strategies is a probability distribution over Σ_2. By Kuhn's Theorem (for the repeated game version, see Aumann (1964)), any such distribution is equivalent (in terms of the induced distribution over paths of play) to a behavior strategy, and vice versa. Thus, following a notational trick introduced in KL, player 1's belief about player 2 can be represented as a behavior strategy $\sigma_2^1 \in \Sigma_2$; similarly for player 2's belief about Player 1. The profile of beliefs for both players is then (σ_2^1, σ_1^2).

(σ_1, σ_2^1) is the profile consisting of player 1's behavior strategy and his belief as to player 2's behavior strategy. The histories that player 1 believes are possible are histories h such that $\mu_{(\sigma_1, \sigma_2^1)}(C(h)) > 0$. Similar definitions hold for player 2.

Suppose that $\hat{\Sigma}_2 \subset \Sigma_2$ is at most countable (finite or infinite, although my notation will be for the infinite case). Let $\sigma_{21}, \sigma_{22}, \sigma_{23}, \ldots, \sigma_{2n}, \ldots$ be an enumeration of $\hat{\Sigma}_2$. Say that belief σ_2^1 *gives weight to all of* $\hat{\Sigma}_2$ if there is a strategy $\sigma_{20} \in \Sigma_2$ and a sequence $\alpha_0, \alpha_1, \ldots, \alpha_n, \ldots$ of real numbers, with $\alpha_0 \geq 0$, $\alpha_n > 0$ for all $n \geq 1$, and $\sum_{n=0}^{\infty} \alpha_n = 1$, such that

$$\sigma_2^1 = \alpha_0 \sigma_{20} + \sum_{n=1}^{\infty} \alpha_n \sigma_{2n}.$$

Neither σ_{20} nor the sequence α_n need be unique. A similar definition holds for player 2. The belief profile (σ_2^1, σ_1^2) gives weight to all of $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ if σ_2^1 gives weight to all of $\hat{\Sigma}_2$ and σ_1^2 gives weight to all of $\hat{\Sigma}_1$.

2.3. *Continuation Games*

A t-period history h defines a *continuation game*, the subgame beginning at period $t+1$. Payoffs for the continuation game starting at date $t+1$ are taken

to be discounted to date $t+1$, rather than back to date 1. In the continuation game following h, a strategy σ_i induces a *continuation strategy* σ_{ih} via

$$\sigma_{ih}(h') = \sigma_i(h \cdot h')$$

for any history h', where $h \cdot h'$ denotes the concatenation of h and h'.

With this notation, a player's posterior belief about his opponent's continuation strategy has a simple representation. If $\mu_{(\sigma_1, \sigma_2^1)}(C(h)) > 0$ then, in the continuation game following h, player 1's *posterior belief*, calculated in standard Bayesian fashion, is σ_{2h}^1; similarly for player 2.

Recalling that $\pi(z,t)$ is the history giving the actions chosen in the first t periods of the path of play z, we may write $\sigma_{i\pi(z,t)}$, $\sigma_{2\pi(z,t)}^2$, and $\sigma_{1\pi(z,t)}^2$.

2.4. *Prediction*

Informally, if the chosen strategy profile is pure, a player will be said to learn to predict the continuation path of play if, for any number of periods l, no matter how large, and any degree of precision η, no matter how small, there is a time t far enough in the future such that, at any time after t, a player predicts every continuation history of length l or less with an error of no more than η. The definition below, in addition to providing a formal statement, extends this idea to cases where one or both players randomize.

The following definition, taken from KL, provides a measure of closeness between two strategy profiles (and hence between the probability distributions over paths of play induced by those profiles).

DEFINITION 1: Given strategy profiles (σ_1, σ_2) and (σ_1^*, σ_2^*), a real number $\eta > 0$, and an integer $l > 0$, (σ_1, σ_2) plays (η, l)-like (σ_1^*, σ_2^*) iff

$$\left| \mu_{(\sigma_1, \sigma_2)}(C(h)) - \mu_{(\sigma_1^*, \sigma_2^*)}(C(h)) \right| < \eta$$

for every history h of length l or less.

DEFINITION 2: Let (σ_1, σ_2) be the strategy profile chosen by the players and let σ_2^1 be player 1's belief. Player 1 *learns to predict the continuation path of play* iff the following conditions hold:

1. $\mu_{(\sigma_1, \sigma_2)}(C(h)) > 0 \Rightarrow \mu_{(\sigma_1, \sigma_2^1)}(C(h)) > 0$ for any finite history h.
2. For any real number $\eta > 0$, any integer $l > 0$, and $\mu_{(\sigma_1, \sigma_2)}$ almost any path of play z, there is a time $t(\eta, l, z)$ such that if $t > t(\eta, l, z)$, then $(\sigma_{1\pi(z,t)}, \sigma_{2\pi(z,t)})$ plays (η, l)-like $(\sigma_{1\pi(z,t)}, \sigma_{2\pi(z,t)}^1)$. If (σ_1, σ_2) is pure, then I will write $t(\eta, l)$ instead of $t(\eta, l, z)$.

The definition for player 2 is similar.

REMARK 1: This is weak learning, weak in the sense that the player is required to make an accurate prediction only about finite continuation histories, not about the infinite tail of the game.

REMARK 2: KL shows that if, instead of (1) in Definition 2, (σ_1, σ_2) and (σ_1, σ_2^1) satisfy the stronger requirement $\mu_{(\sigma_1, \sigma_2)}(E) > 0 \Rightarrow \mu_{(\sigma_1, \sigma_2^1)}(E) > 0$ for all measurable sets of paths E, then part 2 in Definition 2 will be satisfied automatically, and indeed player 1 will be able to make accurate predictions even about the tail of the game. If this strengthened version of part 1 holds, then $\mu_{(\sigma_1, \sigma_2)}$ is said to be absolutely continuous with respect to $\mu_{(\sigma_1, \sigma_2^1)}$; this is the KL sufficient condition to which I alluded in Section 1.4.1.

An observation exploited below is that a sufficient (but not necessary) condition for absolute continuity is that player 1's belief satisfies what KL calls grain of truth: σ_2^1 satisfies *grain of truth* iff $\sigma_2^1 = \alpha\sigma_2' + (1-\alpha)\sigma_2$, where σ_2 is player 2's true behavior strategy, σ_2' is some other behavior strategy for player 2 (which, by Kuhn's Theorem, one may reinterpret as a probability distribution over behavior strategies), and $\alpha \in [0,1)$. In the terminology introduced above, σ_2^1 satisfies grain of truth iff σ_2^1 gives weight to $\{\sigma_2\}$.

2.5. *Optimization*

As usual, $\sigma_1 \in \Sigma_1$ is an (*ex ante*) *best response* to belief $\sigma_2^1 \in \Sigma_2$ iff for any $\sigma_1' \in \Sigma_1$, $V(\sigma_1, \sigma_2^1) \geq V(\sigma_1', \sigma_2^1)$. For learning models along the lines considered here, one wishes σ_1 to be not only *ex ante* optimal but also dynamically optimal in the following sense: for any h such that $\mu_{(\sigma_1, \sigma_2^1)}(C(h)) > 0$ (any h that the player believes will occur with positive probability), one wishes the continuation strategy σ_{1h} to be a best response to the continuation belief σ_{2h}^1. If σ_1 satisfies this dynamic optimization condition, then write $\sigma_1 \in BR_1(\sigma_2^1)$. For $\delta > 0$, $\sigma_1 \in BR_1(\sigma_2^1)$ if σ_1 is an *ex ante* best response to σ_2^1. If $\delta = 0$, $BR_1(\sigma_2^1)$ will (except in trivial cases) be a proper subset of the set of *ex ante* best responses to σ_2^1. Henceforth, the term "best response" for player 1 will be understood to refer to an element of $BR_1(\sigma_2^1)$. It is standard that, for any $\delta \in [0,1)$, $BR_1(\sigma_2^1) \neq \varnothing$. Similar definitions hold for player 2.

The following technical lemma extends to discounted repeated games a result that is well known for finite games. As there does not appear to be a proof readily available in the literature, one is provided in the Appendix.

LEMMA S: *If* $\sigma_1 \in BR_1(\sigma_2^1)$ *and* $\sigma_1^* \in \Sigma_1$ *shares the support of* σ_1, *then* $\sigma_1^* \in BR_1(\sigma_2^1)$; *similarly for Player* 2.

I will also be interested in approximate best responses. Recall that σ_1 is an (*ex ante*) *ε-best response* to σ_2^1 iff, for any σ_1', $V(\sigma_1, \sigma_2^1) + \varepsilon \geq V(\sigma_1', \sigma_2^1)$. Even when $\delta > 0$, *ex ante* optimization is too weak an optimization standard for learning models of the sort considered here. First, *ex ante* ε optimization imposes no restriction on behavior far out in the repeated game. Second, *ex ante* ε optimization may impose little or no restriction on behavior along the actual path of play, as opposed to the paths the player believed most likely to occur,

even in the near or medium term. I address these problems by strengthening the *ex ante* ϵ optimization to what will be called uniform ε optimization.[19]

DEFINITION 3: $\sigma_1 \in \Sigma_1$ is a *uniform ε-best response* to $\sigma_2^1 \in \Sigma_2$, written $\sigma_1 \in BR_1^{\varepsilon}(\sigma_2^1)$, iff, for every history h for which $\mu_{(\sigma_1, \sigma_2^1)}(C(h)) > 0$, σ_{1h} is an ε-best response to σ_{2h}^1. Similarly for $BR_2^{\varepsilon}(\sigma_1^2)$.

3. THE CONFLICT BETWEEN PREDICTION AND OPTIMIZATION

3.1. *Conventionality and Neutrality*

Let $\hat{\Sigma}_1 \subset \Sigma_1$ denote the set of Player 1's strategies that are, for want of a better term, *conventional.* For motivation, see Section 1.2.2. Similarly, the conventional strategies for Player 2 are $\hat{\Sigma}_2 \subset \Sigma_2$. The joint conventional set is $\hat{\Sigma}_1 \times \hat{\Sigma}_2$. Restrict attention to conventional sets that are not empty: $\hat{\Sigma}_i \neq \varnothing$.

As discussed in Section 1.2.2, I wish to confine attention to joint conventional sets that are neutral. The definition of neutrality, given below, will be in terms of a function Ψ that assigns joint conventional sets to repeated games. To formalize the domain of Ψ, begin by fixing a set $\dot{A}$ of finite action sets. I interpret $\dot{A}$ as the universe of possible action sets. For any set K, let $\#K$ denote the cardinality of K. Assume that $\varnothing \notin \dot{A}$ (a player always has at least one action in any game) and that, for any action sets $A, A' \in \dot{A}$, if $\#A \leq \#A'$ then there is an $A^* \in \dot{A}$ such that $\#A^* = \#A$ and $A^* \subset A'$. Take $\dot{A}$ to be the same for both players. Let $\dot{G}$ be the set of possible finite games using action sets drawn from $\dot{A}$ and let $\dot{\Sigma}$ be the associated power set of the set of possible repeated game strategies.

Let Ψ_i: $\dot{G} \times [0,1) \to \dot{\Sigma}$ satisfy $\Psi_i(G, \delta) \subset \Sigma_i(A_1, A_2)$, where (A_1, A_2) are the action sets of G. I interpret $\Psi_i(G, \delta)$ as the conventional set for player i in the repeated game with stage game G and discount factor δ. Assume $\Psi_i(G, \delta) \neq \varnothing$. Let Ψ: $\dot{G} \times [0,1) \to \dot{\Sigma} \times \dot{\Sigma}$ be defined by $\Psi(G, \delta) = \Psi_1(G, \delta) \times \Psi_2(G, \delta)$.

The following constructions will be used in the formal definition of neutrality.

First, for each i, let $A_i, A_i' \in \dot{A}$ be action sets with $\#A_i = \#A_i'$. I permit $A_i = A_i'$ as one possibility. Let $\mathscr{H} = \mathscr{H}(A_1, A_2)$, $\mathscr{H}' = \mathscr{H}(A_1', A_2')$, $\Sigma_i = \Sigma_i(A_1, A_2)$, and $\Sigma_i' = \Sigma_i(A_2', A_2')$. For each i, let g_i: $A_i \to A_i'$ be any bijection. The bijections g_i induce bijections, $\mathfrak{g}_i$: $\Delta(A_i) \to \Delta(A_i)'$, $\mathfrak{h}$: $\mathscr{H} \to \mathscr{H}'$, and γ_i: $\Sigma_i \to \Sigma_i'$, defined as follows. $\mathfrak{g}_i$ is defined by the property that, for any $\alpha_i \in \Delta(A_i)$, for any $a_i \in A_i$, $\mathfrak{g}_i(\alpha_i)$ assigns the same probability to $g_i(a_i)$ that α_i does to a_i. $\mathfrak{h}$ is defined by the property that, for any T, for any $h \in \mathscr{H}^T$, $\mathfrak{h}(h) \in \mathscr{H}'^T$ and, for any $t \leq T$, if the t coordinate of h is (a_1, a_2), then the t coordinate of $\mathfrak{h}(h)$ is $(g_1(a_1), g_2(a_2))$.

[19] Lehrer and Sorin (1994) introduces the concept of *ε-consistent equilibrium*, based on the same idea.

(In the special case of the null history, $\mathfrak{h}(h^0)=h^0$.) γ_i is defined by, for any $\sigma\in\Sigma_i$, for any $h'\in\mathscr{H}'$,

$$\gamma_i(\sigma)(h')=\mathfrak{g}_i(\sigma(\mathfrak{h}^{-1}(h'))).$$

Informally, $\gamma_i(\sigma)$ is the strategy in Σ_i' that is equivalent to $\sigma\in\Sigma_i$ once one translates $\Delta(A_i)$ into $\Delta(A_i')$ and $\mathscr{H}$ into $\mathscr{H}'$.

Next, if $A_i=A_i'$ for each i, then I will also consider, in addition to bijections g_i, functions $g_i^{\diamond}$: $A_i\to A_i$, possibly not 1-1, and associated functions $\mathfrak{g}_i^{\diamond}$: $\Delta(A_i)\to\Delta(A_i)$ and $\gamma_i^{\diamond}$: $\Sigma_i\to\Sigma_i$. $\mathfrak{g}_i^{\diamond}$ is defined by the property that, for any $\alpha_i\in\Delta(A_i)$, for any $a_i^{\star}\in A_i$, the probability assigned by $\mathfrak{g}_i^{\diamond}(\alpha_i)$ to $a_i^{\star}$ equals the sum of the probabilities assigned by α_i to all $a_i\in g_i^{\diamond -1}(a_i^{\star})$. $\gamma_i^{\diamond}$ is defined by, for any $\sigma\in\Sigma_i$, for any $h\in\mathscr{H}$,

$$\gamma_i^{\diamond}(\sigma)(h)=\mathfrak{g}_i^{\diamond}(\sigma(h)).$$

Informally, $\gamma_i^{\diamond}(\sigma)$ is identical to σ except that, whenever σ chooses α_i, $\gamma_i^{\diamond}(\sigma)$ chooses $\mathfrak{g}_i^{\diamond}(\alpha_i)$.

Finally, consider any $A\in\dot{A}$. Let ζ: $A\times A\to A\times A$ be a bijection on the set of action profiles. Let $\mathscr{H}=\mathscr{H}(A,A)$ and let $\Sigma=\Sigma_1(A,A)=\Sigma_2(A,A)$. Then ζ induces bijections $\mathfrak{h}^{\zeta}$: $\mathscr{H}\to\mathscr{H}$ and γ^{ζ}: $\Sigma\to\Sigma$, defined as follows. $\mathfrak{h}^{\zeta}$ is defined by the property that, for any T, for any $h\in\mathscr{H}^T$, $\mathfrak{h}^{\zeta}(h)\in\mathscr{H}^T$ and, for any $t\le T$, if the t coordinate of h is (a,a'), then the t coordinate of $\mathfrak{h}^{\zeta}(h)$ is $\zeta(a,a')$. (In the special case of the null history, $\mathfrak{h}^{\zeta}(h^0)=h^0$.) γ^{ζ} is defined by the property that, for each $\sigma\in\Sigma$, for each $h'\in\mathscr{H}$,

$$\gamma^{\zeta}(\sigma)(h')=\sigma(\mathfrak{h}^{\zeta^{-1}}(h')).$$

Informally, $\gamma^{\zeta}(\sigma)$ is identical to σ except that, upon receiving the history h' as input, $\gamma^{\zeta}(\sigma)$ first translates h' into $\mathfrak{h}^{\zeta^{-1}}(h')$.

DEFINITION 4: Ψ: $\dot{G}\times[0,1)\to\dot{\Sigma}\times\dot{\Sigma}$ is *neutral* iff the following properties are satisfied.

1. *Ψ depends on (G,δ) only through the strategic form of G.* Explicitly, consider any two stage games, $G=(A_1,A_2,u_1,u_2)$ and $G'=(A_1,A_2,u_1',u_2')$, with the same action sets. Then $\Psi(G,\delta)=\Psi(G',\delta')$ for any $\delta,\delta'\in[0,1)$. Abusing notation, write Ψ: $\dot{A}\times\dot{A}\to\dot{\Sigma}\times\dot{\Sigma}$ in place of Ψ: $\dot{G}\times[0,1)\to\dot{\Sigma}\times\dot{\Sigma}$. Similarly for the coordinate functions Ψ_i.

2. *Ψ is symmetric.* Explicitly, the following properties hold.

(a) *Player symmetry.* Consider any $A\in\dot{A}$ and define ζ: $A\times A\to A\times A$ by $\zeta(a,a')=(a',a)$. Then, for any $\sigma\in\Psi_1(A,A)$, $\gamma^{\zeta}(\sigma)\in\Psi_2(A,A)$. Similarly, for any $\sigma\in\Psi_2(A,A)$, $\gamma^{\zeta}(\sigma)\in\Psi_1(A,A)$.

(b) *Action symmetry.* For any $A_1,A_1',A_2,A_2'\in\dot{A}$ with $\#A_i=\#A_i'$ for each i, for any bijections g_i: $A_i\to A_i'$, for any $\sigma\in\Psi_i(A_1,A_2)$, $\gamma_i(\sigma)\in\Psi_i(A_1',A_2')$.

3. *Ψ is invariant to simple changes in strategy.* Explicitly, the following properties hold.

(a) For any $A_1,A_2\in\dot{A}$, for any functions $g_i^{\diamond}$: $A_i\to A_i$, for any $\sigma\in\Psi_i(A_1,A_2)$, $\gamma_i^{\diamond}(\sigma)\in\Psi_i(A_1,A_2)$.

(b) For any $A \in \dot{A}$, for any bijection ζ: $A \times A \to A \times A$, for any $\sigma \in \Psi_i(A, A)$, $\gamma^{\zeta}(\sigma) \in \Psi_i(A, A)$.

4. Ψ *is consistent.* Explicitly, for any $A_1, A_2, A_1', A_2' \in \dot{A}$, $A_i \subset A_i'$ for each i, the following properties hold.

(a) For any $\sigma \in \Psi_i(A_1, A_2)$, there is a $\sigma' \in \Psi_i(A_1', A_2')$ such that $\sigma(h) = \sigma'(h)$ for every $h \in \mathscr{H}(A_1, A_2)$.

(b) For any $\sigma' \in \Psi_i(A_1', A_2')$, if $\sigma'(h) \in \Delta(A_i)$ for every $h \in \mathscr{H}(A_1, A_2)$, then there is a $\sigma \in \Psi_i(A_1, A_2)$ such that $\sigma(h) = \sigma'(h)$ for every $h \in \mathscr{H}(A_1, A_2)$.

5. Ψ *permits pure strategies.* Explicitly, for any $A_1, A_2 \in \dot{A}$, if $\sigma \in \Psi_i(A_1, A_2)$, then there is a pure strategy s in the support of σ such that $s \in \Psi_i(A_1, A_2)$.

A joint conventional set $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ will be called *neutral* if there is a neutral map Ψ such that $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ is in the image of Ψ.

For the interpretation of, and motivation for, these properties, see Section 1.2.2.

3.2. *Conventional Prediction and Conventional Optimization*

DEFINITION 5: *Conventional Prediction* holds for player 1 with belief σ_2^1 iff, for any $(\sigma_1, \sigma_2) \in \hat{\Sigma}_1 \times \hat{\Sigma}_2$, player 1 learns to predict the continuation path of play; similarly for player 2.

DEFINITION 6: *Conventional Optimization* holds for player 1 with belief σ_2^1 iff

$$BR_1(\sigma_2^1) \cap \hat{\Sigma}_1 \neq \varnothing;$$

similarly for player 2.

DEFINITION 7: Given $\varepsilon > 0$, *Conventional Uniform* ε *Optimization* holds for player 1 with belief σ_2^1, iff

$$BR_1^{\varepsilon}(\sigma_2^1) \cap \hat{\Sigma}_1 \neq \varnothing;$$

similarly for player 2.

These properties were discussed in Section 1.2.2 and Section 1.2.4.

3.3. *Main Results*

Consider any action $a_1 \in A_1$ and define

$$\tilde{a}_2(a_1) = \operatorname*{argmax}_{a_2 \in A_2} \left[\max_{a_1' \in A_1} u_1(a_1', a_2) - u_1(a_1, a_2) \right].$$

If the right-hand side is not single-valued, arbitrarily pick one of the values to be $\tilde{a}_2(a_1)$. $\tilde{a}_1(a_2)$ is defined similarly. Loosely, when player 2 chooses action $\tilde{a}_2(a_1)$,

player 1 has maximal incentives *not* to play a_1. $\tilde{a}_2(a_1)$ does not necessarily minimize player 1's payoff from a_1. That is, it is not necessarily true that $\tilde{a}_2(a_1) = \text{argmin}_{a_2 \in A_2} u_1(a_1, a_2) = \text{argmax}_{a_2 \in A_2}[-u_1(a_1, a_2)]$.

DEFINITION 8: Given any pure strategy $s_1 \in S_1$, let $\tilde{S}_2(s_1)$ denote the set of pure strategies for player 2 such that, for any $s_2 \in \tilde{S}_2(s_1)$, if history h is along the path of play generated by (s_1, s_2) (i.e. if $\mu_{(s_1, s_2)}(C(h)) = 1$), then

$$s_2(h) = \tilde{a}_2(s_1(h)).$$

The definition of $\tilde{S}_1(s_2)$ is similar.

Thus, viewed myopically (in terms of period-by-period optimization), s_1 chooses the wrong action in each period against any pure strategy $s_2 \in \tilde{S}_2(s_1)$.

Let m_1 be player 1's minmax value in the stage game:

$$m_1 = \min_{\alpha_2 \in \Delta(A_2)} \max_{\alpha_1 \in \Delta(A_1)} \mathbb{E}_{(\alpha_1, \alpha_2)} u_1(a_1, a_2),$$

where $\mathbb{E}_{(\alpha_1, \alpha_2)} u_1(a_1, a_2)$ is player 1's expected payoff from the mixed action profile (α_1, α_2). m_2 for player 2 is defined similarly. I will sometimes make the following assumption.

ASSUMPTION M: For player 1,

$$\max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1)) < m_1;$$

similarly for player 2.

This assumption is satisfied in Matching Pennies, Rock/Scissors/Paper, Battle of the Sexes, and various coordination games.

A strategy cannot be optimal if a player can learn to predict that its continuation will be suboptimal in some continuation game. As an application of this principle, the next proposition records that, provided there are no weakly dominant actions in the stage game, a pure strategy s_1 cannot be optimal if player 1 can learn to predict the path of play generated by (s_1, s_2) for any $s_2 \in \tilde{S}_2$. The hurdle to a result of this sort is that, even if the player learns to predict the path of play, it can be difficult for a player to learn that he is suboptimizing with respect to his opponent's strategy. For example, a player might think that the low payoffs he (correctly) projects for the near future are simply the price to be paid for the high payoffs he (erroneously) projects for the more distant future. The first part of the proposition assumes away this sort of problem by taking players to be effectively myopic. The second part of the proposition allows players to have any level of patience, but imposes Assumption M. The proof is in the Appendix.

PROPOSITION 1: *Suppose that no action for player* 1 *is weakly dominant in the stage game G.*

1. *There is an* $\bar{\varepsilon} > 0$ *and a* $\bar{\delta} \in (0,1]$ *such that, for any pure strategy* $s_1 \in S_1$ *and any* $s_2 \in \tilde{S}_2(s_1)$, *if player* 1*'s belief* σ_2^1 *allows player* 1 *to learn to predict the continuation path of play generated by* (s_1, s_2), *then* s_1 *is not a uniform* ε*-best response to* σ_2^1 *for any* $\varepsilon \in [0, \bar{\varepsilon})$ *and any* $\delta \in [0, \bar{\delta})$.

2. *If, moreover, Assumption M holds, then there is an* $\bar{\varepsilon} > 0$ *such that, for any pure strategy* $s_1 \in S_1$ *and any* $s_2 \in \tilde{S}_2(s_1)$, *if player* 1*'s belief* σ_2^1 *allows player* 1 *to learn to predict the continuation path of play generated by* (s_1, s_2), *then* s_1 *is not a uniform* ε*-best response to* σ_2^1 *for any* $\varepsilon \in [0, \bar{\varepsilon})$ *and any* $\delta \in [0, 1)$.

Similar results hold for player 2.

The next step in the argument is to make the following observation, the proof of which is in the Appendix.

PROPOSITION 2: *Suppose that* $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ *is neutral. For any pure strategy* $s_1 \in \hat{\Sigma}_1$, *there is a pure strategy* $s_2 \in \hat{\Sigma}_2$ *such that, for any history h (not just histories along the path of play),*

$$s_2(h) = \tilde{a}_2(s_1(h)).$$

In particular, $\hat{\Sigma}_2 \cap \tilde{S}_2(s_1) \neq \emptyset$. *A similar result holds for player* 2.

I am now in a position to state and prove the paper's main result.

THEOREM: *Let G be a stage game in which neither player has a weakly dominant action.*

1. *There is a* $\bar{\delta} \in (0,1]$ *such that, for any* $\delta \in [0, \bar{\delta})$, *for any neutral joint conventional set* $\hat{\Sigma}_1 \times \hat{\Sigma}_2$, *there is no belief* σ_2^1 *such that both Conventional Prediction and Conventional Optimization hold simultaneously for player* 1.

2. *If, moreover, Assumption M holds then, for any* $\delta \in [0,1)$, *for any neutral joint conventional set* $\hat{\Sigma}_1 \times \hat{\Sigma}_2$, *there is no belief* σ_2^1 *such that both Conventional Prediction and Conventional Optimization hold simultaneously for player* 1.

Similar results hold for player 2.

PROOF: For the proof of statement 1, choose $\bar{\delta}$ as in Proposition 1. Suppose that player 1 has beliefs σ_2^1 and that, for these beliefs, Conventional Prediction holds for player 1. Consider any $\sigma_1 \in \hat{\Sigma}_1$. By Property 5 of neutrality (Ψ permits pure strategies), there is a pure strategy $s_1 \in \hat{\Sigma}_1$ with s_1 in the support of σ_1. By Proposition 2, there is an $s_2 \in \hat{\Sigma}_2 \cap \tilde{S}_2(s_1)$. By Proposition 1, $s_1 \notin BR_1(\sigma_2^1)$. (Indeed, $s_1 \notin BR_1^{\varepsilon}(\sigma_2^1)$ for ε sufficiently small.) By Lemma S, $\sigma_1 \notin BR_1(\sigma_2^1)$. Since σ_2^1 and σ_1 were arbitrary, it follows by contraposition that Conventional Optimization is violated for player 1. The proof of statement 2 is almost identical. *Q.E.D.*

For the interpretation of this result, see the Introduction, Section 1.2.3 in particular.

REMARK 3: The Theorem is robust to small deviations from neutrality. More explicitly, because the proof of Proposition 1 relies on strict inequalities, one can show that Proposition 1 extends to situations in which player 1 chooses a (nonpure) strategy σ_1 in a small open neighborhood of some pure strategy s_1 and player 2 chooses a (nonpure) strategy σ_2 in a small open neighborhood of the strategy $s_2 \in \tilde{\Sigma}_2(s_1)$, where s_2 is defined by $s_2(h) = \tilde{a}_2(s_1(h))$ for any h. Here, "open" means in the sup norm (uniform convergence) topology.[20] Using the extended version of Proposition 1, one can then establish that the conclusion of the Theorem continues to hold even if the conclusion of Proposition 2 holds only approximately.

In particular, the Theorem is robust to relaxing Property 5 of neutrality to allow for the possibility that conventional strategies necessarily *tremble*. A trembled version of a pure strategy s_1 is a strategy σ_1 such that, after any history h, $\sigma_1(h)$ chooses $s_1(h)$ with probability $(1 - q^h)$ and chooses some mixture over actions, where the mixture might depend on h, with probability q^h.[21] Let $\bar{q} = \sup_{h \in \mathcal{H}} q^h$. For $\bar{q}$ small, σ_1 will be close to s_1 in the sup norm topology. It is straightforward to show that if Property 5 of neutrality is relaxed to allow small trembles, then versions of Proposition 1 and Proposition 2 continue to hold and therefore the conclusion of the Theorem continues to hold. Of course, if players are *constrained* to play strategies that tremble, then one should demand only approximate, in particular uniform ε, optimization rather than full optimization. I will address this point in Remark 8 in Section 3.4.1.

A consequence of the Theorem is the following.

PROPOSITION 3: *Let G be a stage game in which neither player has a weakly dominant action. Suppose $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ is both neutral and at most countable.*

1. *There is a $\bar{\delta} \in (0,1]$ such that, for any $\delta \in [0, \bar{\delta})$ and any belief σ_2^1 that gives weight to all of $\hat{\Sigma}_2$, Conventional Optimization fails for player* 1.

2. *If, moreover, Assumption M holds, then, for any $\delta \in [0,1)$ and any belief σ_2^1 that gives weight to all of $\hat{\Sigma}_2$, Conventional Optimization fails for player* 1.

Similar results hold for player 2.

PROOF: If players choose a strategy profile in $\hat{\Sigma}_1 \times \hat{\Sigma}_2$, then the belief of either player satisfies grain of truth. It follows that Conventional Prediction holds for both players (see KL, Theorem 3). The result then follows from the Theorem. *Q.E.D.*

[20] The metric for this topology is $d(\sigma_i, \sigma_i') = \sup_{h \in \mathcal{H}} \|\sigma_i(h) - \sigma_i'(h)\|$, where $\|\cdot\|$ is the standard Euclidean norm (view $\sigma_i(h) \in \Delta(A_i)$ as an element of $\mathbb{R}^{\#A_i}$).

[21] This definition of tremble is fairly general; in particular, it allows for trembles that are not i.i.d.

As an application of Proposition 3, suppose that the $\hat{\Sigma}_i$ are defined by a bound on strategic complexity. I will focus on bounds defined in terms of Turing machines, which can be though of as computers with unbounded memory. I will remark briefly below on other possible complexity bounds.

Say that a strategy is *Turing implementable* if there is a Turing machine that takes histories (encoded in machine readable form) as input and produces the name of an action as output.[22] The Turing implementable strategies are *precisely* those that can be defined recursively, where I use the term "recursive" in its Recursive Function Theory sense. Equivalently, the Turing implementable strategies are precisely those that can be defined by a finite flow chart or program. The Church-Turing Thesis, which is generally (although not quite universally) accepted within mathematics, asserts that recursivity captures what one means by "computable in principle." The set of Turing implementable strategies is thus the largest set of computable strategies. It is a natural benchmark for a conventional set that is a large subset of the set of all strategies.

Turing machines, as usually defined, are deterministic and so the Turing implementable strategies are pure. (Randomizing Turing machines will be considered in Section 3.4.1.) Let $S_i^T \subset S_i$ be the set of pure strategies for player i that are Turing implementable. S_i^T is countable.[23] Therefore, player 1's belief can give weight to all of S_2^T. For computability reasons, I will assume that the payoff functions u_i are rational valued and that the discount factor δ is rational.

PROPOSITION 4: *Let G be a stage game in which neither player has a weakly dominant action. Suppose $\hat{\Sigma}_1 \times \hat{\Sigma}_2 = S_1^T \times S_2^T$.*

1. *There is a $\bar{\delta} \in (0,1]$ and an $\bar{\varepsilon} > 0$ such that, for any rational $\delta \in [0, \bar{\delta})$, any $\varepsilon \in [0, \bar{\varepsilon})$, and any belief σ_2^1 that gives weight to all of S_2^T, Conventional Uniform ε Optimization fails for player* 1.

2. *If, moreover, Assumption M holds, then there is an $\bar{\varepsilon} > 0$ such that, for any rational $\delta \in [0,1)$, any $\varepsilon \in [0, \bar{\varepsilon})$, and any belief σ_2^1 that gives weight to all of S_2^T, Conventional Uniform ε Optimization fails for player* 1.

Similar results hold for player 2.

PROOF: The result for optimization, rather than uniform ε optimization, follows from Proposition 3 provided $S_1^T \times S_2^T$ is neutral. Verification of the latter is straightforward and is omitted. The extension to uniform ε optimization is immediate once one notes that Proposition 1 is stated for uniform ε optimiza-

[22] A more formal treatment of Turing implementability for repeated game strategies can be found in, for example, Nachbar and Zame (1996). For general reference on Turing machines and other topics in computability, see Cutland (1980) or Odifreddi (1987).

[23] Any Turing machine has a finite description, hence there are only a countable number of Turing machines, hence only a countable number of strategies are Turing implementable.

tion and that, therefore, the proof of the Theorem extends to uniform ε optimization provided $\hat{\Sigma}_i \subset S_i$ (conventional strategies are pure). *Q.E.D.*

REMARK 4: Although stated for Turing implementable strategies, Proposition 4 holds for *any* standard bound on complexity: *any* standard complexity bound generates a joint conventional set that is (1) neutral and (2) at most countable. Hence Proposition 3 implies that, for conventional sets defined by any standard complexity bound, if player 1 has beliefs that give weight to all of player 2's conventional strategies, player 1 has no conventional best response or even, for ε small, uniform ε-best response. In this sense, player 1's best response will always be more complicated than the strategies that are conventional for player 2.

REMARK 5: For intuition for Proposition 4, consider the following. Say that a belief σ_2^1 that gives weight to all of $\hat{\Sigma}_2 \subset S_2^T$ is *Turing computable* if there is a Turing machine that generates the belief in the form of an enumeration of pairs of probabilities and Turing machine descriptions, which I will refer to as *programs*, with each strategy in $\hat{\Sigma}_2$ implemented by at least one program in the enumeration. If beliefs are Turing computable then, for any $\varepsilon > 0$, there exists a Turing machine implementing a uniform ε-best response. Indeed, one can construct a Turing machine that, after any history, computes a finite approximation to the correct posterior belief, then computes a best response with respect to that approximate posterior for some large truncation of the continuation game. Because of discounting, this best response in the truncation will be an approximate best response in the full continuation. One can show, although I will not do so here, that all the calculations required are well within the scope of a Turing machine.

The problem that arises in Proposition 4 is that a belief that gives weight to all of $\hat{\Sigma}_2 = S_2^T$ is *not* Turing computable because there is no Turing machine that will enumerate a list of strategy programs such that every Turing implementable strategy is implemented by at least one program on the list. This is so even though the set of Turing implementable strategies is countable. The proof, which I omit, is a variation on the diagonalization argument used in Turing (1936) to show that the set of recursive functions is not recursively enumerable.

Since beliefs that give weight to all of S_2^T are not Turing computable, a Turing machine has no way to update beliefs properly, even approximately, after some histories. As a result, the method given above for constructing a uniform ε-best response does not apply. Proposition 4 verifies that, for ε sufficiently small, no uniform ε-best response can be implemented by a Turing machine. Another way to view the same point is to recognize that, by Kuhn's Theorem, having a belief that is not Turing computable is equivalent to facing an opponent playing a strategy that is not Turing implementable. It should not be surprising that, if the opposing strategy is not Turing implementable, one may not have a Turing implementable best response or even, for ε small, uniform ε-best response.

3.4. *Extensions*

3.4.1. *Constrained Rational Players*

The analysis thus far has implicitly maintained the hypothesis that players are free to choose nonconventional strategies in order to optimize. If instead players are constrained to play conventional strategies (see Section 1.2.4 for motivation), then the Theorem implies that, so long as Conventional Prediction holds, neither player will be able to optimize. One might hope that, despite the constraint, players could at least uniformly ε optimize. If this were true then a small modification of the argument in KL would imply asymptotic convergence to approximate Nash equilibrium play.

Proposition 4 has already exploited the fact that if all conventional strategies are pure, then the Theorem's proof, and hence the Theorem itself, extends immediately to cover uniform ε optimization. Thus, for ε small, so long as Conventional Prediction holds, the constraint prevents the players from choosing strategies that are uniformly ε optimal. This does not, of course, prevent a player from choosing a strategy that is *ex ante* ε optimal. But, as illustrated in Section 1.3, *ex ante* ε optimization *per se* may not be enough to guarantee convergence to Nash equilibrium play.

If, on the other hand, the conventional set contains nonpure strategies, then the proof of the Theorem does not extend. The difficulty is that Lemma S is false for uniform ε optimization: even if a strategy σ is uniformly ε optimal, some of the pure strategies in its support may not be. Despite this problem, if players are sufficiently myopic, then the conclusion of the Theorem does extend for conventional sets consisting of the Turing implementable strategies, the benchmark case covered in Proposition 4, even if one modifies the definition of Turing machine to permit access to randomization devices (coin tossers).[24]

Let Σ_i^T denote the set of strategies for player i that can be implemented by a randomizing Turing machine. The proof of the following proposition contains a brief description of how randomizing Turing machines are constructed. Under that construction, Σ_i^T is countable. Hence player 1's belief can give weight to all of Σ_2^T.

The proof is in the Appendix.

PROPOSITION 5: *Let G be a stage game in which neither player has a weakly dominant strategy. There is a $\bar{\delta} \in [0,1)$ and an $\bar{\varepsilon} > 0$ such that, for any rational $\delta \in [0, \bar{\delta})$, any $\varepsilon \in [0, \bar{\varepsilon})$, and any belief σ_2^1 that gives weight to all of Σ_2^T, Conventional Uniform ε Optimization fails for player* 1. *A similar result holds for player* 2.

[24] A randomization device is distinct from the software used by actual computers to generate pseudo random numbers. Since sufficiently complicated Turing machines are capable of pseudo randomization, Proposition 4 already encompasses pseudo randomizing strategies.

REMARK 6: The proof relies on the fact that S_2^T is sufficiently rich in strategies that, for any $\sigma_1 \in \Sigma_1^T$, there is a strategy $s_2' \in S_2^T$ that is close, in the sup norm topology, to a strategy $s_2 \in S_2$, where s_2 is such that, after any history, the mixture of actions chosen by σ_1 is maximally suboptimal (s_2 is thus an element of $\tilde{S}_2(\sigma_1)$, where the latter is defined in the obvious way). The proof of Proposition 5 extends to any subsets of Turing implementable strategies that are neutral and rich in the above sense. For example, it extends to conventional sets formed by the strategies that are implementable by finite automata (roughly, computers with finite memory).

REMARK 7: It is not known to what degree Proposition 5 extends to players who are patient (δ is high), although it does extend for some nongeneric stage games, such as Matching Pennies.

REMARK 8: Although, as already noted, the proof used for the Theorem does not generally extend to uniform ε optimization if conventional sets contain nonpure strategies, the proof does extend in special cases. In particular, suppose that the joint conventional set $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ is a trembled version of a pure neutral joint set $\hat{S}_1 \times \hat{S}_2$; see Remark 3. Since strategies in $\hat{\Sigma}_i$ are close, in the sup norm topology, to strategies in $\hat{S}_i$, and since the Theorem does extend for $\hat{S}_1 \times \hat{S}_2$, a version of the Theorem extends for $\hat{\Sigma}_1 \times \hat{\Sigma}_2$. Somewhat more precisely, one can show that there is an $\bar{\varepsilon} > 0$ such that, if Conventional Prediction holds, then Conventional Uniform ε Optimization fails for any $\varepsilon \in [0, \bar{\varepsilon})$, provided $\bar{q}$ is sufficiently small (recall from Remark 3 that $\bar{q}$ is the maximal tremble).

3.4.2. *Boundedly Rational Players*

A boundedly rational player is one for whom deliberation is costly. There is, unfortunately, no consensus as to how bounded rationality should be modeled. I will assume that a bounded rational player is essentially a Turing machine, and that it is this Turing machine that formulates a belief that fashions a response.

If a player is a Turing machine, then his belief will (almost by definition) be computable. As noted in Remark 5, this implies that each player will be able to uniformly ε optimize. As also noted in Remark 5, since a player's belief is computable, the belief cannot give weight to all of his opponent's Turing implementable strategies. For example, the belief might assign positive probability only to opposing strategies that are implementable by a finite automaton. Define the conventional set for player i to be the strategies to which his opponent assigns positive probability.

If the joint conventional set is neutral, then a variant of Proposition 4 (or, if randomization is permitted, of Proposition 5, provided conventional sets are sufficiently rich in the sense discussed in Remark 6) tells us that, for ε small, players will choose nonconventional (but still Turing implementable) strategies in order to uniformly ε optimize.

If, on the other hand, bounded rationality implies that neutrality fails, then it is possible for Conventional Uniform ε Optimization to hold even for ε arbitrarily small. One might thus be in the ironic position of being able to construct a theory of rational learning along the lines proposed when, but only when, players are only boundedly rational. But a failure of neutrality, in and of itself, does not assure Conventional Uniform ε Optimization. For Conventional Uniform ε Optimization, neutrality must fail the right away, excluding certain strategies but not others. Exactly which strategies will depend on the game. It is not clear why bounded rationality would imply that neutrality would fail in a way that facilitates Conventional Uniform ε Optimization rather than impedes it.

Dept. of Economics, Campus Box 1208, Washington University, One Brookings Drive, St. Louis, MO 63130-4899, U.S.A.

Manuscript received May, 1996.

APPENDIX

PROOF OF LEMMA S: Let σ_1^* share the support of σ_1 and suppose $\sigma_1^* \notin BR_1(\sigma_2^1)$. Endow Σ_1 with the product (pointwise convergence) topology.[25] Consider any sequence of strategies $\sigma_{1k} \in \Sigma_1$ such that (a) σ_{1k} converges to σ_1^* in the product topology, (b) for any k, σ_{1k} shares the support of σ_1, and (c) for any k, σ_{1k} agrees with σ_1 except for at most a finite number of histories.[26] Because σ_1^* is not a best response, and because V_1 is continuous in the product topology, there is a k such that $\sigma_{1k} \notin BR_1(\sigma_2^1)$. Because σ_{1k} shares the support of σ_1, and because σ_{1k} agrees with σ_1 except for at most a finite number of histories, one can show that there is an $\alpha \in (0, 1]$ and a $\sigma_1^\circ \in \Sigma_1$ such that $\sigma_1 = \alpha\sigma_{1k} + (1 - \alpha)\sigma_1^\circ$. Choose any $\sigma_1' \in BR_1(\sigma_2^1)$. Then, since $\sigma_{1k} \notin BR_1(\sigma_2^1)$ and since $\alpha > 0$,

$$V_1(\sigma_1, \sigma_2^1) = V_1(\alpha\sigma_{1k} + (1 - \alpha)\sigma_1^\circ, \sigma_2^1) < V_1(\alpha\sigma_1' + (1 - \alpha)\sigma_1^\circ, \sigma_2^1).$$

It follows that $\sigma_1 \notin BR_1(\sigma_2^1)$. The proof then follows by contraposition.[27]

[25] Since the set of finite histories is countably infinite, Σ_1 can be viewed as the product set $\Delta(A_1)^\infty$, where $\Delta(A_1)$ is viewed as the unit simplex in $\mathbb{R}^{\#A_1}$. Endow $\Delta(A_1)$ with the standard Euclidean topology and endow $\Delta(A_1)^\infty$ with the product topology.

[26] In particular, one can construct such a sequence by enumerating the finite histories and, for each k, defining $\sigma_{1k}(h)$ to equal $\sigma_1^*(h)$ for each of the first k histories, and to equal $\sigma_1(h)$ otherwise.

[27] The proof exploits the continuity of V_1, which follows from the fact that repeated game payoffs are evaluated as a present value. If payoffs were instead evaluated by limit of means, continuity would fail and the Lemma would be false. For example, consider the two-player stage game in which player 2 has only one action and player 1 has two actions, Left, yielding 0, and Right, yielding 1. Under limit of means, it is a best response (to his only possible belief) for player 1 to play the strategy "following any history of length t, play Left with probability 2^{-t}, Right with probability $1 - 2^{-t}$." Under this strategy, player 1 plays Left with positive probability in every period. Thus the pure strategy "play Left always" is in the support of this behavior strategy even though this pure strategy is not a best response.

PROOF OF PROPOSITION 1: Let

$$w_1(a_1) = \max_{a_1' \in A_1} u_1(a_1', \tilde{a}_2(a_1)) - u_1(a_1, \tilde{a}_2(a_1)).$$

$w_1(a_1) \geq 0$. Moreover, $w_1(a_1) = 0$ iff a_1 is dominant (weakly or strictly). Since, by assumption, no action is weakly dominant, $w_1(a_1) > 0$ for all a_1. Let

$$\underline{w}_1 = \min_{a_1 \in A_1} w_1(a_1) > 0.$$

Let

$$\bar{u}_1 = \max_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2),$$

$$\underline{u}_1 = \min_{a_1 \in A_1} \min_{a_2 \in A_2} u_1(a_1, a_2).$$

Since no strategy is weakly dominant, $\bar{u}_1 > \underline{u}_1$.

To prove the first part of the proposition, choose $\bar{\delta}$ sufficiently small that, under uniform ε optimization, player 1 acts to maximize his current period payoff (i.e. he is effectively myopic). In particular, it will turn out that the argument below goes through for $\varepsilon > 0$ and $\bar{\delta} \in (0, 1]$ such that, for any $\varepsilon \in [0, \bar{\varepsilon})$ and any $\delta \in [0, \bar{\delta})$,

$$\varepsilon < \underline{w}_1 - \frac{\delta}{1-\delta}[\bar{u}_1 - \underline{u}_1].$$

Note that such $\bar{\varepsilon}$ and $\bar{\delta}$ do exist.

Consider any pure strategy $s_1 \in S_1$ and any $s_2 \in \tilde{S}_2(s_1)$. Temporarily fix $\eta \in (0, 1)$. Suppose that player 1 learns to predict the continuation path of play. Then, for any continuation game beginning at time $t+1$, $t > t(\eta, 1)$ (that is, $l = 1$), player 1 assigns some probability $(1 - \eta') > (1 - \eta)$ to the actual action chosen by player 2 at date $t+1$. For specificity, suppose that at date $t+1$, player 1 chooses action a_1^* while player 2 chooses action a_2^*. Discounting payoffs to date $t+1$, player 1's expected payoff in the continuation game is then *at most*

$$(1 - \eta')u_1(a_1^*, a_2^*) + \eta'\bar{u}_1 + \frac{\delta}{1-\delta}\bar{u}_1.$$

If player 1 were instead to choose an action a_1 in period $t+1$ to maximize $u_1(a_1, a_2^*)$, his expected payoff in the continuation game would be *at least*

$$(1 - \eta') \max_{a_1 \in A_1} u_1(a_1, a_2^*) + \eta'\underline{u}_1 + \frac{\delta}{1-\delta}\underline{u}_1.$$

Thus, uniform ε optimization requires

$$\varepsilon + (1 - \eta')u_1(a_1^*, a_2^*) + \eta'\bar{u}_1 + \frac{\delta}{1-\delta}\bar{u}_1 \geq (1 - \eta') \max_{a_1 \in A_1} u_1(a_1, a_2^*) + \eta'\underline{u}_1 + \frac{\delta}{1-\delta}\underline{u}_1$$

or

$$\varepsilon + \eta'(\bar{u}_1 - \underline{u}_1) \geq \underline{w}_1 - \frac{\delta}{1-\delta}[\bar{u}_1 - \underline{u}_1],$$

where I have used the fact that, since $s_2 \in \tilde{S}_2(s_1)$, $\max_{a_1 \in A_1} u_1(a_1, a_2^*) - u_1(a_1^*, a_2^*) = w_1(a_1^*) \geq \underline{w}_1$. By the construction of $\bar{\varepsilon}$ and $\bar{\delta}$, there is an η sufficiently small such that this inequality cannot hold for any $\varepsilon \in [0, \bar{\varepsilon})$ and $\delta \in [0, \bar{\delta})$. This establishes the first part of the proposition.

As for the second part of the proposition, suppose that Assumption M holds. Fix any $\delta \in [0,1)$ and choose $\bar{\varepsilon} > 0$ such that, for any $\varepsilon \in [0, \bar{\varepsilon})$,

$$\varepsilon < \frac{1}{1-\delta}\left[m_1 - \max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1))\right].$$

By Assumption M, $m_1 > \max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1))$, hence such $\bar{\varepsilon}$ exist.

Once again, consider any pure strategy $s_1 \in S_1$ and any $s_2 \in \tilde{S}_2(s_1)$. Temporarily fix $\eta > 0$ and an integer $l > 0$. Suppose that player 1 learns to predict the continuation path of play. Then, for any continuation game beginning at time $t+1, t > t(\eta, l)$, player 1 assigns some probability $(1-\eta') > (1-\eta)$ to the actual l-period continuation history beginning at date $t+1$. In that finite continuation history, player 1 receives at most $\max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1))$ per period. On the other hand, player 1 believes that there is a probability η' that the continuation history might be something else. In an alternate l-period continuation history, player 1 could receive at most $\bar{u}_1$ per period. Finally, from date $t+l+1$ onwards, player 1 could receive at most $\bar{u}_1$ per period. Thus beginning at date $t+1$, player 1 expects to earn *at most*

$$\frac{1-\delta^l}{1-\delta}\left[(1-\eta')\max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1)) + \eta'\bar{u}_1\right] + \frac{\delta^l}{1-\delta}\bar{u}_1.$$

In contrast, any best response must expect to earn at least m_1, on average, following any history given positive probability by $\mu_{(s_1,s_2)}$. Thus, under a true best response, player 1 expects to earn *at least*

$$\frac{m_1}{1-\delta}.$$

Thus ε optimization requires

$$\varepsilon + \frac{1-\delta^l}{1-\delta}\left[(1-\eta')\max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1)) + \eta'\bar{u}_1\right] + \frac{\delta^l}{1-\delta}\bar{u}_1 \geq \frac{m_1}{1-\delta}$$

or

$$\varepsilon \geq (1-\eta')\frac{1-\delta^l}{1-\delta}\left[m_1 - \max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(a_1))\right] - \frac{\delta^l + \eta'(1-\delta^l)}{1-\delta}[\bar{u}_1 - m_1].$$

By the construction of $\bar{\varepsilon}$, there is an (η, l) such that this inequality cannot hold for any $\varepsilon \in [0, \bar{\varepsilon})$. This establishes the second part of the proposition. *Q.E.D.*

PROOF OF PROPOSITION 2: Consider any neutral map Ψ: $\dot{A} \times \dot{A} \to \dot{\Sigma} \times \dot{\Sigma}$ such that $\Psi(A_1, A_2) = \hat{\Sigma}_1 \times \hat{\Sigma}_2$. Consider any pure strategy $s_1 \in \hat{\Sigma}_1 = \Psi_1(A_1, A_2)$. Let $s_2 \in S_2$ be defined by

$$s_2(h) = \tilde{a}_2(s_1(h))$$

for every history h. I will argue that $s_2 \in \hat{\Sigma}_2$. I begin with the following observation.

CLAIM: *Consider any $A \in \dot{A}$. If $\sigma \in \Psi_1(A, A)$, then $\sigma \in \Psi_2(A, A)$.*

PROOF: *Define ζ: $A \times A \to A \times A$ by $\zeta(a, a') = (a', a)$. By Property 2(a) of neutrality (player symmetry), $\gamma^\zeta(\sigma) \in \Psi_2(A, A)$. By Property 3(b) of neutrality (invariance property (b)), $\gamma^\zeta(\gamma^\zeta(\sigma)) \in \Psi_2(A, A)$. Finally, note that $\gamma^\zeta(\gamma^\zeta(\sigma)) = \sigma$.* *Q.E.D.*

For ease of notation, henceforth let $A_1 = A$, $A_2 = A'$.

Consider first the special case in which $A \subset A'$. By Property 4(a) of neutrality (consistency property (a)), there is an $s^\circ \in \Psi_1(A', A')$ such that $s^\circ(h) = s_1(h)$ for all $h \in \mathcal{H}(A, A')$. By the Claim, $s^\circ \in \Psi_2(A', A')$. Choose $g_1^\diamond$: $A' \to A'$ to be the identity and choose any function $g_2^\diamond$: $A' \to A'$,

possibly not 1-1, such that $g_2^{\diamond}(a) = \tilde{a}_2(a)$ if $a \in A$. By Property 3(a) of neutrality (invariance property (a)), $\gamma_2^{\diamond}(s^\circ) \in \Psi_2(A', A')$. By Property 4(b) of neutrality (consistency property (b)), there is a strategy $s' \in \Psi_2(A, A')$ such that $s'(h) = \gamma_2^{\diamond}(s^\circ)(h)$ for all $h \in \mathscr{H}(A, A')$. One can verify that $s_2 = s' \in \Psi_2(A, A')$. The argument for $A \supset A'$ is similar.[28]

Suppose, on the other hand, that $A \not\subset A'$ but that $\#A \leq \#A'$. Then one can extend the above argument as follows. By an assumption made when $\dot{A}$ was defined, there exists a set $A^* \in \dot{A}$ with $\#A^* = \#A$ and $A^* \subset A'$. Let g_1: $A \to A^*$ be any bijection and let g_2: $A' \to A'$ be the identity. Then, by Property 2(b) of neutrality (action symmetry), $\gamma_1(s_1) \in \Psi_1(A^*, A')$. By Property 4(a) and the above Claim, there is an $s^\circ \in \Psi_2(A', A')$ such that $s^\circ(h) = \gamma_1(s_1)(h)$ for any $h \in \mathscr{H}(A^*, A')$. Choose $g_1^{\diamond}$: $A' \to A'$ to be the identity and choose any function $g_2^{\diamond}$: $A' \to A'$, possibly not 1-1, such that $g_2^{\diamond}(a) = \tilde{a}_2(a)$ if $a \in A^*$. By Property 3(a), $\gamma_2^{\diamond}(s^\circ) \in \Psi_2(A', A')$. By Property 4(b), there is a strategy $s' \in \Psi_2(A^*, A')$ such that $s'(h) = \gamma_2^{\diamond}(s^\circ)(h)$ for all $h \in \mathscr{H}(A^*, A')$. One can verify that $s_2 = \gamma_2^{-1}(s')$. The argument for $\#A > \#A'$ is similar. *Q.E.D.*

PROOF OF PROPOSITION 5: I begin by sketching how a Turing machine can be made to randomize. Recall that a Turing machine operates by executing a sequence of discrete computational steps. In each such step, a standard (i.e. deterministic) Turing machine reads one bit (consisting of either a 0 or a 1) out of memory, consults its current state (a Turing machine has a finite number of abstract attributes called states), and then, according to a preset deterministic rule that takes as input the value of the bit read from memory and the state, the machine may alter the bit in the current memory location, it may change its state, and it may move to a different memory location. The customary way to handle randomization is to add to the description of a Turing machine a finite number of special states corresponding to one or more coins, possibly biased. If random state ξ is entered, the machine leaves its memory alone but switches with probabilities $p(\xi)$: $(1 - p(\xi))$ to one of two ordinary states. For computability reasons, $p(\xi)$ is assumed to be rational. With randomizing Turing machines, there is a subtlety regarding whether the Turing implementable strategy plays an action for certain after any history or just with probability 1. For the sake of generality, I will allow for the latter. Since the number of random states is finite and since the $p(\xi)$ are rational, each randomizing Turing machine has a finite description and so the set of strategies implemented by such machines is countable.

Extend the domain of $\tilde{a}_2$ to include mixtures over actions by player 1: for any $\alpha_1 \in \Delta(A_1)$,

$$\tilde{a}_2(\alpha_1) = \operatorname*{argmax}_{a_2 \in A_2} \left[\max_{a_1 \in A_1} u_1(a_1, a_2) - \mathbb{E}_{\alpha_1} u_1(a_1, a_2) \right]$$

where $\mathbb{E}_{\alpha_1} u_1(a_1, a_2)$ is player 1's expected payoff from the profile (α_1, a_2).[29] Similarly, extend the domain of w_1, introduced in the proof of Proposition 1, so that, for any $\alpha_1 \in \Delta(A_1)$,

$$w_1(\alpha_1) = \max_{a_1 \in A_1} u_1(a_1, \tilde{a}_2(\alpha_1)) - \mathbb{E}_{\alpha_1} u_1(\alpha_1, \tilde{a}_2(\alpha_1)).$$

$w_1(\alpha_1) \geq 0$. Moreover, $w_1(\alpha_1) = 0$ iff α_1 is dominant (weakly or strictly). Since, by assumption, no action (pure or mixed) is weakly dominant, $w_1(\alpha_1) > 0$ for all α_1. $\Delta(A_1)$ is compact and it is straightforward to show that w_1 is continuous. Therefore,

$$\underline{w}_1 = \min_{\alpha_1 \in \Delta(A_1)} w_1(\alpha_1) > 0.$$

Finally, let $\bar{u}_1$ and $\underline{u}_1$ be defined as in Proposition 1. Again, since no action is weakly dominant, $\bar{u}_1 > \underline{u}_1$.

[28] Briefly, if $A \supset A'$ then, by Property 4(a) and the Claim, there is an $s^\circ \in \Psi_2(A, A)$ such that $s^\circ(h) = s_1(h)$ for all $h \in \mathscr{H}(A, A')$. The argument then proceeds largely as before. The one potential obstacle is the application of Property 4(b), which requires, for $h \in \mathscr{H}(A, A')$, that $\gamma_2^{\diamond}(s^\circ)(h) \in A' \subset A$. This condition is satisfied since, by definition, $\tilde{a}_2$ takes values in A'.

[29] As before, if the right-hand side of the defining expression for $\tilde{a}_2(\alpha_1)$ is not single-valued, arbitrarily pick one of the values to be $\tilde{a}_2(\alpha_1)$.

Choose $\bar{\delta}$ sufficiently small that, under uniform ε optimization, player 1 acts to maximize his current period payoff (i.e. he is effectively myopic). In particular, it will turn out that the argument below goes through for $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0, 1]$ such that, for any $\varepsilon \in [0, \bar{\varepsilon})$ and any $\delta \in [0, \bar{\delta})$,

$$\varepsilon < \underline{w}_1 - \frac{\delta}{1-\delta}[\bar{u}_1 - \underline{u}_1].$$

Note that such $\bar{\varepsilon}$ and $\bar{\delta}$ do exist.

Choose any $\sigma_1 \in \Sigma_1^T$ and temporarily fix a rational number $\nu > 0$. I claim that there is a pure strategy $s_2^\nu \in S_2^T$ with the property that, for any history h,

$$\left| \left(\max_{a_1 \in A_1} u_1(a_1, s_2^\nu(h)) - \mathbb{E}_{\sigma_1(h)} u_1(a_1, s_2^\nu(h)) \right) - w_1(\sigma_1(h)) \right| < \nu. \tag{1}$$

The claim would be trivial if one could set $s_2^\nu(h) = \tilde{a}_2(\sigma_1(h))$. I will discuss the reason for not doing so when I show that there is indeed such an $s_2^\nu \in \Sigma_2^T$.

Temporarily fix $\varepsilon \in [0, \bar{\varepsilon})$, $\delta \in [0, \bar{\delta})$, and $\eta \in (0, 1)$. Let $\sigma_1 \in \Sigma_1^T$ and $s_2^\nu \in S_2^T$ be as above. Since player 1's beliefs give weight to all of Σ_2^T, which is countable, player 1 learns to predict the path of play generated by (σ_1, s_2^ν). In particular, for $\mu_{(\sigma_1, s_2^\nu)}$ almost any path of play z, for any continuation game beginning at time $t+1$, $t > t(\eta, 1, z)$ (that is $l = 1$), player 1 assigns some probability $(1-\eta') > (1-\eta)$ to the actual action chosen by player 2 at date $t+1$, namely $s_2^\nu(h)$, where $h = \pi(z, t)$, the t-period initial segment of z. Discounting payoffs to date $t+1$, player 1's expected payoff in the continuation game is then *at most*

$$(1-\eta')\mathbb{E}_{\sigma_1(h)} u_1(a_1, s_2^\nu(h)) + \eta' \bar{u}_1 + \frac{\delta}{1-\delta}\bar{u}_1.$$

If player 1 were instead to choose an action in period $t+1$ to maximize $u_1(a_1, s_2^\nu(h))$, his expected payoff would be *at least*

$$(1-\eta') \max_{a_1 \in A_1} u_1(a_1, s_2^\nu(h)) + \eta' \underline{u}_1 + \frac{\delta}{1-\delta}\underline{u}_1.$$

Thus uniform ε optimization requires

$$\varepsilon + (1-\eta')\mathbb{E}_{\sigma_1(h)} u_1(a_1, s_2^\nu(h)) + \eta' \bar{u}_1 + \frac{\delta}{1-\delta}\bar{u}_1$$

$$\geq (1-\eta') \max_{a_1 \in A_1} u_1(a_1, s_2^\nu(h)) + \eta' \underline{u}_1 + \frac{\delta}{1-\delta}\underline{u}_1$$

or

$$\varepsilon + \eta'(\bar{u}_1 - \underline{u}_1) \geq (1-\eta')\underline{w}_1 - (1-\eta')\nu - \frac{\delta}{1-\delta}(\bar{u}_1 - \underline{u}_1),$$

where I have used inequality (1) and the fact that $w_1(\alpha_1) \geq \underline{w}_1$. By the construction of $\bar{\varepsilon}$ and $\bar{\delta}$, there exist ν and η sufficiently small such that this inequality cannot hold for any $\varepsilon \in [0, \bar{\varepsilon})$ and $\delta \in [0, \bar{\delta})$.

It remains only to show that there is indeed a Turing implementable strategy s_2^ν satisfying inequality (1). To avoid bogging down the paper in computability details, I will only sketch the Turing machine construction. Suppose, then, that σ_1 is implemented by a Turing machine M. Let $\nu > 0$ be as given above. From M, one can show that one can construct a new *deterministic* Turing machine M^ν that does the following. On input of a history h, M^ν simulates the action of M on h. Every time M randomizes, the flow of its program branches in two. M^ν proceeds by simulating M along *each* branch until M either halts, giving the action chosen by the strategy implemented by M,

or M reaches another random state. Proceeding in this way, M^ν can calculate an approximation, say α_1', to the true mixture over actions, say $\alpha_1 = \sigma_1(h)$. Set

$$s_2^\nu(h) = \tilde{a}_2(\alpha_1').$$

By the continuity of expectation, and the definition of w_1, inequality (1) will hold provided α_1' is sufficiently close to α_1. Since M has only a finite number of random states, the accuracy of the estimate α_1' improves geometrically with the depth of the simulation (number of times random states are hit). Moreover, since one can program knowledge of the $p(\xi)$ into M^ν, M^ν will be able to calculate whether a depth has been reached sufficient to ensure that its estimate α_1' is close enough to α that inequality (1) holds. Therefore, M^ν calculates $s_2^\nu(h)$ in finite time.

There are two reasons to have M^ν approximate α_1 rather than to calculate it exactly. First, if M chooses an action only with probability 1, rather than for certain, then M^ν may be unable to calculate α_1 exactly. In particular, if M^ν attempts to calculate α_1 by the above algorithm, it may never arrive at an answer, and so it may fail to choose an action. Second, even if M always chooses an action, taking an approximation rather than computing α_1 exactly is desirable because it reduces the complexity of M^ν. In particular, by taking an approximation, the number of computational steps required by M^ν can be held to a multiple of the number expected for M, and may even be smaller than the worst case for M. *Q.E.D.*

REFERENCES

ANDERLINI, L. (1990): "Some Notes on Church's Thesis and The Theory of Games," *Theory and Decision*, 29, 19–52.

AOYAGI, M. (1994): "Evolution of Beliefs and the Nash Equilibrium of Normal Form Games," University of Pittsburg.

AUMANN, R. (1964): "Mixed and Behaviour Strategies in Infinite Extensive Games," in *Advances in Game Theory*, ed. by M. Dresher, L. S. Shapley and A. W. Tucker, Annals of Mathematics Studies, 52. Princeton, NJ: Princeton University Press, pp. 627–650.

BINMORE, K. (1987): "Modeling Rational Players, Part I," *Economics and Philosophy*, 3, 179–214.

BLACKWELL, D., AND L. DUBINS (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 38, 882–886.

BLUME, L. E., AND D. EASLEY (1995): "Rational Expectations and Rational Learning," *Economic Theory*, forthcoming.

BROWN, G. W. (1951): "Iterative Solutions of Games by Fictitious Play," in *Activity Analysis of Production and Allocation*, ed. by T. J. Koopmans. New York: John Wiley, pp. 374–376.

CANNING, D. (1992): "Rationality, Computability, and Nash Equilibrium," *Econometrica*, 60, 877–888.

CUTLAND, N. J. (1980): *Computability*, Vol. 60. Cambridge, UK: Cambridge University Press.

FUDENBERG, D., AND D. KREPS (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320–367.

FUDENBERG, D., AND D. LEVINE (1993): "Self-Confirming Equilibrium," *Econometrica*, 61, 523–545.

——— (1995a): "Conditional Universal Consistency," Harvard University.

——— (1995b): "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, 19, 1065–1089.

——— (1996): "Theory of Learning in Games," Harvard University.

JORDAN, J. S. (1991): "Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, 3, 60–81.

——— (1993): "Three Problems in Learning Mixed-Strategy Nash Equilibria," *Games and Economic Behavior*, 5, 368–386.

KALAI, E., AND E. LEHRER (1993a): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1945.

——— (1993b): "Subjective Equilibrium in Repeated Games," *Econometrica*, 61, 1231–1240.

——— (1995): "Subjective Games and Equilibria," *Games and Economic Behavior*, 8, 123–163.

KALAI, E., AND W. STANFORD (1988): "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56, 397–410.

LEHRER, E., AND R. SMORODINSKY (1994): "Repeated Large Games with Incomplete Information," Northwestern University.

LEHRER, E., AND S. SORIN (1994): "ε-Consistent Equilibrium," Northwestern University.

NACHBAR, J. H., AND W. R. ZAME (1996): "Non-Computable Strategies and Discounted Repeated Games," *Economic Theory*, 8, 103–122.

ODIFREDDI, P. (1987): *Classical Recursion Theory*. Amsterdam: North Holland.

SANDRONI, A. (1995): "The Almost Absolute Continuity Hypothesis," University of Pennsylvania.

SHAPLEY, L. (1962): "On the Nonconvergence of Fictitious Play," Discussion Paper RM-3026, RAND.

SONSINO, D. (1995): "Learning to Learn, Pattern Recognition, and Nash Equilibrium," Stanford Graduate School of Business.

TURING, A. (1936): "On Computable Numbers With An Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, 42, 230–265; Corrections, ibid. (1937), 43, 544–546.

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