# Prediction, Optimization, and Learning in Repeated Games

John H. Nachbar *

February, 1996.

**Abstract**

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 players will learn as the game unfolds to predict the continuation path of play. If this conjecture were true then a convergence result due to Kalai and Lehrer would imply that the continuation path would asymptotically resemble the path of a Nash equilibrium. One would thus have constructed a theory which predicts Nash equilibrium as the necessary long-run consequence of optimization by cautious players. This paper points out that there is an obstacle to such a result in the form of a potential conflict between prediction and optimization.

KEYWORDS: Repeated games, rational learning, Bayesian learning.

*Department of Economics, Box 1208, Washington University, One Brookings Drive, St. Louis, MO 63130; voice: (314) 935-5612; e-mail: nachbar@wuecon.wustl.edu. This work was started while I was a visitor at The Center for Mathematical Studies in Economics and Management Science, Northwestern University. I would also like to acknowledge financial support from the Center for Political Economy at Washington University. The paper has benefited from the comments of many people, including Drew Fudenberg, Ehud Lehrer, David Levine, Bart Lipman, Wilhelm Neuefeind, Yaw Nyarko, Bruce Petersen, Jeroen Swinkels, Suzanne Yee, Bill Zame, a co-editor, and two anonymous referees. The usual caveat applies.

# 1 Introduction.

## 1.1 Overview.

A standard argument for Game Theory's emphasis on Nash equilibrium is 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 which is a best response to his prior belief as to his opponent's strategy. Rather than assume that prior beliefs are in equilibrium, one would like to argue that if beliefs are cautious, in the sense of satisfying some form of full support assumption, then players will learn as the game unfolds to predict the continuation path of play. If this conjecture were true then a convergence result due to Kalai and Lehrer (1993a), hereafter KL, would imply that the continuation path would asymptotically resemble the path of a Nash equilibrium. One would thus have constructed a theory which predicts Nash equilibrium behavior as the necessary long-run consequence of optimization by cautious players.

The message of this paper is that there is an obstacle to such a result in the form of a potential conflict between prediction and optimization. The basic insight is that, in many games, optimality of a repeated game strategy implies restrictions on the set of paths of play which a player can learn to predict. These restrictions do not necessarily rule out prediction of the actual path of play; after all, players optimize and predict the continuation path of play within a Nash equilibrium. However, the restrictions make it difficult, perhaps impossible, to motivate prediction as the consequence of mere caution. Requiring that players both optimize and predict implies that substantial equilibration must occur *before* learning within the repeated game begins.

Subsection 1.2 gives an informal but detailed description of this paper's results and underlying logic. Subsection 1.3 develops a concrete example. Subsection 1.4 comments on some related literature, KL in particular. Readers anxious to begin the formal exposition can turn to Section 2, which contains the basic definitions, and Section 3, which contains this 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 $\sigma_i$ tells Player $i$ to play $\sigma_i(h)$ in period $t+1$, where $\sigma_i(h)$ may be either a pure stage game action or a mixture over actions.[1] 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 which 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 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 arbitrarily 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 non-trivial; 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. However, the set of possible strategies is so large that, provided

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

the opposing player has at least two actions in the underlying stage game, there is *no* belief which would enable a player to learn to predict the continuation path of play for every possible opposing strategy.[2] 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. However, learning in the sense of updating one's prior need not imply that a player is actually 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. However, 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 which players can learn to predict. Moreover, since the ultimate goal is to explain equilibration in terms of repeated interaction, we want to explain prediction without imposing equilibrium-like restrictions on prior beliefs.[3]

[2]Informally, the reason 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 which, 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 which 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 which is asymptotically accurate against all strategies.

[3]This is in contrast to the literature on learning *within* (Bayesian) equilibrium; see Jordan

### 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 which are, for want of a better term, *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 which is a best response (in the standard sense of maximizing the expected present value of the player's stage game payoffs). However, if no conventional strategy is a best response, a player will optimize by choosing a non-conventional 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 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.[4]

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

Then if players have cautious beliefs, Conventional Optimization and our 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 then 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.

Conventional Prediction and Conventional Optimization hold trivially if the product set of conventional strategies consists of a single repeated game Nash equi-

(1991). In that literature, unlike here, players have incomplete information about each other's payoffs, which makes learning non-trivial even when equilibrium is assumed.

[4]I ask that the player learn to predict regardless of which conventional strategy he himself chooses. Weakening this would involve constructing a model in which both a player's strategy choice and the set of paths of play he can predict are determined jointly. The necessary model is likely to be similar to a repeated game version of Aumann (1987). However, the Aumann (1987) construction has proved controversial. KL attempts to finesse such a construction and I am attempting to do so as well.

librium profile. However, such a conventional set assumes away the problem of equilibrium. To satisfy the objective of not imposing equilibrium-like restrictions on prior beliefs, we want beliefs to be cautious not only in the sense that beliefs satisfy a full support condition with respect to the conventional 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 $\Psi$, which assigns product sets of conventional strategies to games satisfies the following properties (the formal definition is in Section 3.1).

1. $\Psi$ *is independent of payoffs.* Payoff information should perhaps be included at least in order to rule out those repeated game strategies which are strictly dominated, perhaps even those which are non-rationalizable. However, this issue is irrelevant to the paper's argument. Indeed, in many of the games considered here, including all of the $2 \times 2$ games, no strategy is even weakly dominated.

2. $\Psi$ *is symmetric.* In particular, $\Psi$ satisfies *Player Symmetry* and *Action Symmetry*. Player Symmetry says that $\Psi$ assigns conventional sets to Player 2 using essentially the same criteria as it uses to assign conventional sets to Player 1. The main implication of Action Symmetry is 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. $\Psi$ *is invariant to trivial changes in strategy.* If a strategy $\sigma_i$ is conventional for Player $i$, then so is any strategy $\sigma_i'$ which is identical to $\sigma_i$ except that, for example, whenever $\sigma_i$ chooses $a_1$, $\sigma_i'$ instead chooses $a_1'$. If strategies can be described by finite computer programs then the program for $\sigma_i'$ can be constructed by taking the program for $\sigma_i'$ and adding a few lines of code to "translate" $\sigma_i$'s action choices. If invariance is violated then a player whose forecasts are persistently wrong may *never* notice that his opponent's behavior is consistent with a computationally trivial variant of behavior which the player *could* learn to predict. This sort of thick-headedness runs counter to what one informally means by a player being cautious. Moreover, because there is a computationally trivial algorithm to expand any conventional set which violates invariance into a conventional set that satisfies it, it is diffi-

cult to justify a violation of invariance on computational grounds, even in the case of players who are only boundedly rational (bounded rationality will be considered briefly in Section 3.4.2).

4. $\Psi$ *is monotone.* Roughly, an increase in the number of stage game actions available to a player should not result in a reduction of his conventional set.

5. $\Psi$ *permits pure strategies.* More accurately, for each non-pure conventional strategy, there should be at least *one* pure strategy in its support which is likewise conventional.[5] 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 $\sigma_i$ for Player $i$ is inherently more complicated than some of the pure strategies in its support. Explicitly, given $\sigma_i$, choose some (arbitrary) ranking for Player $i$'s stage game actions and consider the pure strategy $s_i$ which, after any history $h$, chooses the highest ranked action to which $\sigma_i(h)$ gives positive probability. For example, if Player $i$ has only two actions, Left and Right (ranked in that order), $s_i$ chooses Left after any history such that $\sigma_i$ randomizes. For any standard notion of complexity, $\sigma_i$ is more complicated than $s_i$. Indeed, $\sigma_i$ uses $s_i$ as a kind of pattern and adds to $s_i$ the additional complication of randomization after certain histories. If we view a conventional strategy set as being built up from less to more complicated strategies then, for any conventional randomizing strategy like $\sigma_i$, some pure strategy like $s_i$ should be conventional as well.[6]

Somewhat abusing terminology, I will say that a product set of conventional strategies is *neutral* if there is neutral map $\Psi$ such that the product set is in the image of $\Psi$.

Neutrality is a property of the conventional sets rather than directly of beliefs. For example, the fact that $\Psi$ is independent of payoffs does *not* imply that each

[5] A pure strategy $s_i$ will be said to be in the support of a strategy $\sigma_i$ if, after any history, the action chosen by $s_i$ is also selected with positive probability by $\sigma_i$.

[6] One might object that while players might not deliberately favor randomization, it may be difficult 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 4 and Remark 10, allowing for slight trembling does not materially affect the argument.

player's belief must be independent of payoffs. Similarly, players may have the same conventional set without their beliefs being identical. In fact, we 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 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 which satisfy some standard bound on complexity. Examples of such sets include the strategies which 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 to date"), the strategies which remember only at most the last $\tau$ periods, and the strategies which can be represented as a finite flow chart or program. It bears repeating that taking the conventional set to consist of the strategies which 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 which are simple.

This paper takes the point of view that, while we 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 non-conventional even though those strategies were computationally trivial variants of conventional strategies. To the extent that the theory requires a deviation from neutrality, the theory requires some degree of equilibration before 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 for which neither player has a weakly dominant strategy, if players are sufficiently impatient, then for *any* neutral

conventional set there are *no* beliefs such that Conventional Prediction and Conventional Optimization hold simultaneously for both players. 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 4 in Section 3.3, the Theorem is robust to small deviations from neutrality.

The Theorem implies that if Conventional Optimization holds then Conventional Prediction must fail, hence players may fail to learn to predict the continuation path of play. Alternatively, if Conventional Optimization fails then optimizing players will choose a strategy profile which is not conventional. Hence, even if Conventional Prediction holds, players may again fail to learn to predict the continuation path of play. The next two subsections contain examples in which prediction fails and, as a consequence, the continuation path of play fails to converge to that of a Nash equilibrium. This is not to claim that failure must occur or even that failure is likely, only that failure cannot be ruled out.[7]

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 which arises is that if $s$ is conventional for Player 1 then neutrality implies that at least one of the $s'$ strategies will be conventional for Player 2. Hence, either Conventional Prediction or Conventional Optimization must fail.

It might seem that this argument depends in some essential way on the fact that $s$ was taken to be pure. After all, a player can often avoid doing poorly (in

[7]Computing the likelihood of failure would require imposing objective probability on the learning model. This paper, like KL, attempts to avoid objective probability (the objective notion of conventionality being the one exception) on the grounds that objective probability is a form of *ex ante* equilibration. It may be that this methodological position is too restrictive.

particular, earning less than his minmax payoff) by randomizing. However, not doing poorly is not the same thing as optimizing. In fact, the Theorem extends to include conventional strategy sets which contain randomizing strategies. To see this, note that if a non-pure strategy is a best response to some belief then so is every pure strategy in its support.[8] Suppose that Conventional Prediction holds. Since we have assumed that, for any non-pure conventional strategy, some pure strategy in its support is also conventional, and since we know from the above that no conventional pure strategy is optimal, it follows that no conventional non-pure strategy can be optimal either.[9]

To make the Theorem somewhat more concrete, consider any product conventional set consisting of strategies which satisfy a bound on complexity. Any standard complexity bound yields a neutral conventional set which is at most countable. As noted in the discussion of structural neutrality, it follows that for any such conventional set there are beliefs for which Conventional Prediction holds.[10] 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 violate the complexity bound defining conventionality.[11] 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

[8]This fact, while standard for finite games, is less obvious for discounted repeated games, where the strategy set is infinite. The Appendix provides a proof.

[9]It is natural to ask whether this negative result could be overturned if we allowed players to have a strict preference for randomization in some circumstances. This question will not be pursued here, since it necessarily requires adopting a non-Savage decision theory. Whether there is a reasonable modification of Savage's theory that would allow players to have a taste for randomization is a matter of current debate. See, for example, the discussion in Eichberger and Kelsey (1995) and the references therein.

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

[11]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 best response which is also conventional. There is no contradiction with the Theorem because this sort of closure looks only at beliefs which are degenerate in the sense of assigning all mass to a single strategy. Cautious beliefs, that is, beliefs for which Conventional Prediction holds for neutral conventional sets, are intrinsically non-degenerate.

response that much more complicated. The only way to obtain Conventional Optimization is to add in so many more strategies that Conventional Prediction is lost. In particular, if we take 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 strategy which is conventional. For example, the conventional set might consist of strategies which 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 which 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 $\varepsilon$ optimization*: a strategy is $\varepsilon$ optimal if it is $\varepsilon$ optimal *ex ante* and if, moreover, it induces an $\varepsilon$ optimal continuation strategy in every continuation game (more precisely, in every continuation game which 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 $\varepsilon$ optimization. Therefore, for any neutral conventional set, if Conventional Prediction holds then Conventional Uniform $\varepsilon$ Optimization fails for $\varepsilon$ sufficiently small. This need not prevent a player from choosing a strategy which is only *ex ante* $\varepsilon$ optimal. However, as illustrated in Section 1.3, *ex ante* $\varepsilon$ optimization *per se* may not be enough to guarantee convergence to approximate Nash equilibrium play.

If, on the other hand, the conventional set contains non-pure strategies, then the argument sketched above does not extend. However, 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 strategies which 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 truly boundedly rational, that is, constrained in their basic computational ability, rather than just in the strategies they are permitted to execute. However, a careful study of learning by boundedly rational players lies outside the scope of this paper.

### 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 $\bar{H}$, and "$T$ always," denoted $\bar{T}$. Thus, $\hat{\Sigma} = \{50{:}50, \bar{H}, \bar{T}\}$. Note that $\hat{\Sigma} \times \hat{\Sigma}$ is neutral.

Assume that each player's belief assigns probability one to the set $\hat{\Sigma}$ and assigns positive probability to each of the three elements of $\hat{\Sigma}$. We 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 $\bar{H}$, 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 $\bar{H}$, 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 $\bar{H}$ then Player 1 should start playing $H$ in every period, while if Player 1 learns to predict $\bar{T}$, he should start playing $T$ in every period. However, 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 which gives weight to every strategy in $\hat{\Sigma}$.

If players optimize, players must, therefore, choose non-conventional strategies. 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. However, if the players adopt such pure strategies, then from period 2 onward the path of play will be either $((H,H),(H,H),\dots)$, $((H,T),(H,T),\dots)$, $((T,H),(T,H),\dots)$, or $((T,T),(T,T),\dots)$, 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) equilibrium path of play. (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.)

On the other hand, suppose that, as was discussed in Section 1.2.4, players are *constrained* to choose from among the three strategies in $\hat{\Sigma}$. For $\varepsilon$ low, none of the conventional strategies is uniformly $\varepsilon$ 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 which is merely $\varepsilon$ optimal, rather than uniformly $\varepsilon$ optimal, then each player will strictly prefer either $\bar{H}$ or $\bar{T}$ to 50:50, depending on his prior belief, unless his prior happens to put exactly equal weight on $\bar{H}$ and $\bar{T}$. In the latter case, the player will be indifferent between all three strategies. But, if both players select pure strategies, then we are back to where we were before: 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.

These problems are not special to Matching Pennies and in particular do not depend on the fact that Matching Pennies has no pure strategy equilibrium. Consider, for example, perturbing the stage game to the following.

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

Taking the conventional set to be $\hat{\Sigma} = \{50{:}50, \bar{H}, \bar{T}\}$, as above, we cannot rule out the paths $((H,T),(H,T),\dots)$ and $((T,H),(T,H),\dots)$, which, again, are not equilibrium paths.[12]

The restriction to just $\hat{\Sigma} = \{50{:}50, \bar{H}, \bar{T}\}$ is, of course, extremely limiting. In a satisfactory learning model, players ought to be able to learn to predict paths of play associated with a variety of different strategies, not just 50:50, $\bar{H}$, and $\bar{T}$. The message of this paper is that expanding the conventional set, while desirable on modeling grounds, will not resolve the conflict between prediction and optimization so long as neutrality holds.

## 1.4 Remarks on the Literature.

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

KL, together with its companion paper, Kalai and Lehrer (1993b), does two things. First, KL provides a condition on beliefs which 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.[13] 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.[14]

While the KL sufficient condition for prediction is strong (from the discussion in

[12] 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 our objective of trying to explain equilibration solely through repeated play.

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

[14] 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 $\varepsilon$-Nash equilibrium. See also Sandroni (1995).

Section 1.2.1, we know that any such condition *must* be strong), it has the attractive feature that it imposes essentially no restriction on a 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 Nyarko (1994) and Kalai and Lehrer (1995).

By way of example, consider again the case of Matching Pennies with the conventional set $\hat{\Sigma} = \{50{:}50, \bar{H}, \bar{T}\}$. We 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. However, 50:50 won't be *optimal* for a player unless the player assigns *zero*, not just low, probability to both $\bar{H}$ and $\bar{T}$.[15] 50:50 can thus be optimal for both players only if beliefs are actually in equilibrium at the start of repeated play.

### 1.4.2 Fictitious Play and Related Models of (Semi-) Rational Learning.

Fictitious Play provides an instructive example of what can happen when Conventional Prediction fails. 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 is impatient and optimizes with respect to a prior belief which is a beta distribution over the memoryless behavior strategies, that is, strategies of the form "in any period $t$, go Left with probability $p$, regardless of history." See, for example, Fudenberg and Levine (1995c). The set of memoryless behavior strategies, viewed as the conventional set, is neutral. It is not hard to see that Conventional Prediction holds (although beliefs in this case violate the KL

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

sufficient condition). Hence Conventional Optimization must fail. Indeed, under fictitious play, players choose strategies which are *not* memoryless.

Because each player erroneously believes his opponent is playing a memoryless strategy, players may fail to learn to predict the actual continuation path. In Matching Pennies, for example, each player learns to believe that his opponent is randomizing 50:50 even though, along typical paths of play, neither player actually randomizes. Since, typically, the actual path of play is pure, it does not converge to the stochastic path generated by the unique Nash equilibrium of repeated Matching Pennies. That is, we do not get convergence to Nash equilibrium in the sense used here (and in KL). This is not necessarily a disaster. Since each player comes to believe that the other is randomizing 50:50, beliefs about next period's play converge to the stage game's Nash equilibrium. More importantly, both the marginal and joint frequency distributions of play converge to that of the Nash equilibrium of the stage game. Thus, behavior in fictitious play is consistent with many of the observable consequences of players learning to play a Nash equilibrium.[16]

Unfortunately, fictitious play is not always so well behaved. In $2 \times 2$ stage games, while the marginal frequency distributions of past play, and hence player beliefs about next period's play, always converge to a Nash equilibrium of the stage game, the *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 marginal frequency distributions, and hence player beliefs, fail to converge, a point originally made by Shapley (1962); see also Jordan (1993). What is perhaps more disturbing, in the examples in which convergence fails, the path of play cycles in ways that are obvious to the outside analyst but which the players themselves fail to detect.

These problems with fictitious play stem from the fact that, while players are rational in the sense that they choose best responses to their beliefs, the beliefs themselves are naive: each player believes his opponent adopts a memoryless strategy even though each, in fact, adopts a strategy which is history dependent. It is tempting to dismiss this naiveté as a peculiarity of fictitious play. However, the same reasoning which underlies the Theorem implies that, as an axiom of rational-

[16] An exception is that the path of play under fictitious play will typically exhibit patterns across time which are inconsistent with true randomization.

ity, a strong version of the principle "whatever is optimal for me I should believe possible for my opponent" is paradoxical. Remark 5 in Section 3.3 discusses this more explicitly. Some degree of naiveté is unavoidable.

Despite this, one might still hope to construct a theory of (semi-) rational learning in which players exhibit behavior more sophisticated than that of fictitious play. For recent work along these general lines, see Fudenberg and Levine (1995b), which contains a nice overview in addition to its original contributions, Fudenberg and Levine (1995a), Aoyogi (1994), Sonsino (1995), and Foster and Vohra (1995). In much of this literature, players are modeled as using strategies which are intuitively sensible without necessarily being best responses to well-formed prior beliefs. Justifying these strategies as optimal or near optimal may require adoption of a richer repeated game model or deviation from standard decision theory (see footnote 9 above), or both. See Fudenberg and Levine (1995a) for some additional discussion.

### 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 reminiscent of the unsolvability of the Halting Problem and Gödel's incompleteness theorems; see also Anderlini (1990).

Following Binmore, view a player in a one-shot game as choosing a *decision procedure*, a function which, 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 which, for any game and any opposing decision procedure, chooses an action which is a best response to the action which will be chosen by the opponent's decision procedure.

It is not hard to see that no decision procedure is optimal for Matching Pennies.[17] Perhaps more surprisingly, there may be no optimal decision procedure even in

[17]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 2 establishes that no decision procedure can be optimal even if players can randomize.

games with pure action equilibria. The basic difficulty is that there are so many possible opposing decision procedures that there may be no decision procedure which can optimize with respect to them all. Canning (1992) shows that, for a large set of games with pure action equilibria, one can close the decision problem by limiting players to non-trivial *domains* (subsets) of decision procedures, where "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. However, as Canning (1992) emphasizes, the domains, while non-trivial, necessarily embody rules of equilibrium selection. In games with multiple equilibria, different rules of equilibrium selection give rise to different domains.

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 we 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. However, these differences are somewhat deceptive. While neither player in our model may have accurate knowledge of his opponent *ex ante*, our insistence on prediction means that players will have accurate knowledge *ex post*. If the conventional set is neutral, asking for a conventional strategy which is optimal when Conventional Prediction holds is akin to asking in Binmore's model for a decision procedure which is optimal against all opposing decision procedures. Conversely, the domain restrictions discussed in Canning (1992) are akin to the deviations from neutrality which 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 $\mathcal{H} = \bigcup_{T \geq 0} \mathcal{H}^T$. 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 $\mathcal{Z}$. 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 is denoted $\pi(z,t)$; note that $\pi(z,t) \in \mathcal{H}^t$.

A *(behavior) strategy* for Player $i$ is a function $\sigma_i : \mathcal{H} \to \Delta(A_i)$, where $\Delta(A_i)$ is the set of probability mixtures over $A_i$. Let $\Sigma_i$ be the set of behavior strategies of Player $i$. A *pure strategy* for Player $i$ is simply a behavior strategy which 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 $\sigma_i^*$ will be said to *share the support* of strategy $\sigma_i$ iff, for any history $h$, if $\sigma_i(h)$ assigns positive probability to action $a_i \in A_i$ then so does $\sigma_i(h)$. In the case of a pure strategy, $\sigma_i^* = s_i$, we will say that $s_i$ is *in the support* of $\sigma_i$.

$\Sigma_1 \times \Sigma_2$ denotes the set of *behavior strategy profiles* in the repeated game. For each $t$, a behavior strategy profile $(\sigma_1, \sigma_2)$ 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 $z$ for which the $t$-period initial segment $\pi(z,t)$ 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 $\sigma$-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 $V_i : \Sigma_1 \times \Sigma_2 \to \mathbb{R}$,

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

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 $\Sigma_2$. 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 $(\sigma_2^1, \sigma_1^2)$.

$(\sigma_1, \sigma_2^1)$ is the profile consisting of Player 1's behavior strategy and his belief as to Player 2's behavior strategy. The histories which Player 1 believes are possible are histories $h$ such that $\mu_{(\sigma_1,\sigma_2^1)}(C(h)) > 0$. Similarly 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 $\sigma_2^1$ *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 $\sigma_{20}$ nor the sequence $\alpha_n$ need be unique. Similarly for Player 2. The belief profile $(\sigma_2^1, \sigma_1^2)$ gives weight to all of $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ if $\sigma_2^1$ gives weight to all of $\hat{\Sigma}_2$ and $\sigma_1^2$ 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 0. In the continuation game following $h$, a strategy $\sigma_i$ induces a *continuation strategy* $\sigma_{ih}$ 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 $\sigma_{2h}^1$. 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 many write $\sigma_{i\pi(z,t)}$, $\sigma_{2\pi(z,t)}^1$, 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 $\ell$, no matter how large, and any degree of precision $\eta$, 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 $\ell$ or less with an error of no more than $\eta$. 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* $(\sigma_1, \sigma_2)$ *and* $(\sigma_1^*, \sigma_2^*)$*, a real number* $\eta > 0$*, and an integer* $\ell > 0$*,* $(\sigma_1, \sigma_2)$ plays $(\eta, \ell)$-like $(\sigma_1^*, \sigma_2^*)$ *iff*

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

*for every history* $h$ *of length* $\ell$ *or less.*

**Definition 2.** *Let* $(\sigma_1, \sigma_2)$ *be the strategy profile chosen by the players and let* $\sigma_2^1$ *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* $\ell > 0$*, and* $\mu_{(\sigma_1,\sigma_2)}$ *almost any path of play* $z$*, there is a time* $t(\eta, \ell, z)$ *such that if* $t > t(\eta, \ell, z)$ *then* $(\sigma_{1\pi(z,t)}, \sigma_{2\pi(z,t)})$ *plays* $(\eta, \ell)$*-like* $(\sigma_{1\pi(z,t)}, \sigma^1_{2\pi(z,t)})$*. If* $(\sigma_1, \sigma_2)$ *is pure, then I will write* $t(\eta, \ell)$ *instead of* $t(\eta, \ell, z)$*.*

*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, $(\sigma_1, \sigma_2)$ and $(\sigma_1, \sigma_2^1)$ 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 (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 (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 alluded to 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: $\sigma_2^1$ satisfies *grain of truth* iff $\sigma_2^1 = \alpha\sigma_2' + (1-\alpha)\sigma_2$, where $\sigma_2$ is Player 2's true behavior strategy, $\sigma_2'$ is some other behavior strategy for Player 2 (which, by Kuhn's Theorem, we may reinterpret as a probability distribution over behavior strategies), and $\alpha \in [0,1)$. In the terminology introduced above, $\sigma_2^1$ satisfies grain of truth iff $\sigma_2^1$ 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, we wish $\sigma_1$ 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), we wish the continuation strategy $\sigma_{1h}$ to be a best response to the continuation belief $\sigma_{2h}^1$. If $\sigma_1$ satisfies this dynamic optimization condition then we write $\sigma_1 \in \mathrm{BR}_1(\sigma_1^2)$. For $\delta > 0$, it is readily verified that $\sigma_1 \in \mathrm{BR}_1(\sigma_2^1)$ iff $\sigma_1$ is an *ex ante* best response to $\sigma_2^1$. If $\delta = 0$, $\mathrm{BR}_1(\sigma_2^1)$ will (except in trivial cases) be a proper subset of the set of *ex ante* best responses to $\sigma_2^1$. Henceforth, the term "best response" for Player 1 will be understood to refer to an element $\mathrm{BR}_1(\sigma_1^2)$. It is standard that, for any $\delta \in [0,1)$, $\mathrm{BR}_1(\sigma_2^1) \neq \emptyset$. Similar definitions hold for Player 2.

The following technical lemma extends to discounted repeated games a result which 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 $\sigma_1 \in \mathrm{BR}_1(\sigma_1^2)$ and $\sigma_1^*$ shares the support of $\sigma_1$ then $\sigma_1^* \in \mathrm{BR}_1(\sigma_1^2)$. Similarly for Player 2.*

We will also be interested in approximate best responses. Recall that $\sigma_1$ is an *(ex ante) $\varepsilon$-best response* to $\sigma_2^1$ iff, for any $\sigma_1'$, $V(\sigma_1, \sigma_2^1) + \varepsilon \geq V(\sigma_1', \sigma_2^1)$. Even when $\delta > 0$, *ex ante* $\varepsilon$ optimization is too weak an optimization standard for learning models

of the sort considered here. First, *ex ante* $\varepsilon$ optimization imposes no restriction on behavior far out in the repeated game. Second, *ex ante* $\varepsilon$ 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 occur, even in the near or medium term. We address these problems by strengthening the *ex ante* $\varepsilon$ optimization to what will be called uniform $\varepsilon$ optimization.[18]

**Definition 3.** $\sigma_1 \in \Sigma_1$ *is a* uniform $\varepsilon$-best response *to* $\sigma_2^1 \in \Sigma_2$, *written* $\sigma_1 \in BR_1^{\varepsilon}(\sigma_2^1)$, *iff, for every for every history* $h$ *for which* $\mu_{(\sigma_1,\sigma_2^1)}(C(h)) > 0$, $\sigma_{1h}$ *is an* $\varepsilon$*-best response to* $\sigma_{2h}^1$. *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 which 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 $\hat{\Sigma}_1 \times \hat{\Sigma}_2$. We restrict attention to conventional sets which are not empty: $\hat{\Sigma}_i \neq \emptyset$.

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

Let $\Psi : \overset{\bullet}{G} \times [0,1) \to \overset{\bullet}{\Sigma} \times \overset{\bullet}{\Sigma}$ satisfy $\Psi(G,\delta) \subset \Sigma_1(G) \times \Sigma_2(G)$, where $\Sigma_i(G)$ denotes the set of all of Player $i$'s strategies in the repeated game based on the stage game $G$. Let $\Psi_i : \overset{\bullet}{G} \times [0,1) \to \overset{\bullet}{\Sigma}$ be the coordinate function of $\Psi$ corresponding to Player

[18] Lehrer and Sorin (1994) introduces the concept of $\varepsilon$-*consistent equilibrium*, based on the same idea.

$i$; thus $\Psi(G,\delta) = \Psi_1(G,\delta) \times \Psi_2(G,\delta)$. We interpret $\Psi_i(G,\delta)$ as the conventional set for Player $i$ in the repeated game with stage game $G$ and discount factor $\delta$. We will assume $\Psi_i(G,\delta) \neq \emptyset$ for any $(G,\delta)$.

The definition below makes use of the following construction. For each $i$, let $A_i, A_i' \in \overset{\bullet}{A}$ be action sets with $\#A_i = \#A_i'$. We permit $A_i = A_i'$ as one possibility. Let $\mathcal{H}$ and $\mathcal{H}'$ be the associated sets of histories and let, for each $i$, $\Sigma_i$ and $\Sigma_i'$ be the associated sets of strategies. For each $i$, let $g_i : A_i \to A_i'$ be a bijection. The bijections $g_i$ induce bijections $\mathfrak{g}_i : \Delta(A_i) \to \Delta(A_i)'$, $\mathfrak{h} : \mathcal{H} \to \mathcal{H}'$, and $\gamma_i : \Sigma_i \to \Sigma_i'$, 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 $\alpha_i$ does to $a_i$. $\mathfrak{h}$ is defined by the property that, for any $h \in \mathcal{H}^T$, $\mathfrak{h}(h) \in \mathcal{H}'^T$ and, for any $t \leq T$, the $t$ coordinate of $\mathfrak{h}(h)$ is $(g_1(a_1), g_2(a_2))$ iff the $t$ coordinate of $h$ is $(a_1, a_2)$. (In the special case of the null history, $\mathfrak{h}(h^0) = h^0$.) Finally, $\gamma_i$ is defined by, for any $\sigma_i \in \Sigma_i$, for any $h' \in \mathcal{H}'$,

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

Informally, $\gamma_i(\sigma_i)$ is the strategy which is equivalent to $\sigma_i$ once one makes the appropriate translation between $\Delta(A_i)$ and $\Delta(A_i')$ and between $\mathcal{H}$ and $\mathcal{H}'$.

In the special case in which $A_i = A_i'$ for each $i$, we 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 $\alpha_i$ to all $a_i \in g_i^{-1}(a_i^\star)$. $\gamma_i^\diamond$ is defined by, for any $\sigma_i \in \Sigma_i$, for any $h \in \mathcal{H}$,

$$\gamma_i^\diamond(\sigma_i)(h) = \mathfrak{g}_i\left(\sigma_i(h)\right).$$

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

**Definition 4.** $\Psi : \overset{\bullet}{G} \times [0,1) \to \overset{\bullet}{\Sigma} \times \overset{\bullet}{\Sigma}$ *is* neutral *iff the following properties are satisfied.*

1. $\Psi$ *is independent of payoffs. 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* $\Psi : \overset{\bullet}{A} \times \overset{\bullet}{A} \to$

$\overset{\bullet}{\Sigma}\times\overset{\bullet}{\Sigma}$ *in place of* $\Psi : \overset{\bullet}{G}\times[0,1) \to \overset{\bullet}{\Sigma}\times\overset{\bullet}{\Sigma}$. *Similarly for the coordinate functions* $\Psi_i$.

2. $\Psi$ *is symmetric. Explicitly, the following properties hold.*

   (a) *Player Symmetry. For any* $A, A' \in \overset{\bullet}{A}$, $\Psi_1(A, A') = \Psi_2(A', A)$.

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

3. $\Psi$ *is invariant to trivial changes in strategy. Explicitly, for any* $A_1, A_2 \in \overset{\bullet}{A}$, *for any functions* $g_i^\diamond : A_i \to A_i$, *for any* $\sigma_i \in \Psi_i(A_1, A_2)$, $\gamma_i^\diamond(\sigma_i) \in \Psi_i(A_1, A_2)$.

4. $\Psi$ *is monotone. Explicitly, for any* $A, A' \in \overset{\bullet}{A}$, $A \subset A'$,

$$\Psi_1(A, A) \subset \Psi_1(A, A') \subset \Psi_1(A', A).$$

   *(The inclusions need not be proper.) Similarly for Player 2.*

5. $\Psi$ *permits pure strategies. Explicitly, for any* $A_1, A_2 \in \overset{\bullet}{A}$, *if* $\sigma_i \in \Psi_i(A_1, A_2)$ *then there is a pure strategy* $s_i$ *in the support of* $\sigma_i$ *such that* $s_i \in \Psi_i(A_1, A_2)$.

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

For our purposes, the key part of Monotonicity will be $\Psi_1(A, A') \subset \Psi_1(A', A)$. This says, loosely, that any strategy which is conventional for Player 1 when Player 2 has the larger action set will also be conventional for Player 1 if Player 1 has the larger action set. For the interpretation of, and motivation for, the other properties, see Section 1.2.2.

## 3.2 Conventional Prediction and Conventional Optimization.

**Definition 5.** Conventional Prediction *holds for Player 1 with belief* $\sigma_1^2$ *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* $\sigma_1^2$ *iff*

$$\mathrm{BR}_1(\sigma_2^1) \cap \hat{\Sigma}_1 \neq \emptyset.$$

*Similarly for Player 2.*

**Definition 7.** Conventional Uniform $\varepsilon$ Optimization *holds for Player 1 with belief* $\sigma_1^2$*, iff, for every* $\varepsilon > 0$*,*

$$\mathrm{BR}_1^{\varepsilon}(\sigma_2^1) \cap \hat{\Sigma}_1 \neq \emptyset.$$

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

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

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 $\tilde{a}_2(a_1) = \operatorname{argmin}_{a_2 \in A_2} u_1(a_1, a_2) = \operatorname{argmax}_{a_2 \in A_2} [-u_1(a_1, a_2)]$.

**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 $(\alpha_1, \alpha_2)$. $m_2$ for Player 2 is defined similarly. We 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 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 $\sigma_2^1$ 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 $\varepsilon$-best response to $\sigma_2^1$ 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 $\sigma_2^1$ 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 $\varepsilon$-best response to $\sigma_2^1$ for any $\varepsilon \in [0, \bar{\varepsilon})$ and any $\delta \in [0, 1)$.*

*Similarly 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. If $\#A_1 \leq \#A_2$ then, for any $s_1 \in \hat{\Sigma}_1$, there is an $s_2 \in \hat{\Sigma}_2$ 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, $\hat{\Sigma}_2 \cap \tilde{S}_2(s_1) \neq \emptyset$. Similarly for Player 2 if $\#A_2 \leq \#A_1$.*

*Remark 3.* In many games, the conclusion of Proposition 2 follows from principles more elementary than full neutrality. For example, in the following variant of Chicken, the conclusion of Proposition 2 follows if $\hat{\Sigma}_1 = \hat{\Sigma}_2$, which seems reasonable given the symmetry of the game.

| | $X$ | $Y$ |
|---|---|---|
| $X$ | $0, 0$ | $10, 9$ |
| $Y$ | $9, 10$ | $1, 1$ |

□

We are now in a position to state the paper's main result.

**Theorem.** *Let $G$ be a stage game in which neither player has a weakly dominant strategy. Suppose that $\#A_1 \leq \#A_2$.*

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 $\sigma_2^1$ such that both Conventional Prediction and Conventional Optimization hold simultaneously for Player 1.*

2. *If, moreover, Assumption M holds then for for any $\delta \in (0, 1)$, for any neutral joint conventional set $\hat{\Sigma}_1 \times \hat{\Sigma}_2$, there is no belief $\sigma_2^1$ such that both Conventional Prediction and Conventional Optimization hold simultaneously for Player 1.*

*Similarly for Player 2 if $\#A_2 \leq \#A_1$.*

**Proof.** For the proof of statement 1, choose $\bar{\delta}$ as in Proposition 2. Suppose that Player 1 has beliefs $\sigma_2^1$ and that, for these beliefs, Conventional Prediction holds for Player 1. Consider any $\sigma_1 \in \hat{\Sigma}_1$. By Property 5 of neutrality ($\Psi$ permits pure strategies), there is a pure strategy $s_1 \in \hat{\Sigma}_1$ with $s_1$ in the support of $\sigma_1$. By Proposition 2, there is an $s_2 \in \tilde{S}_2(s_1)$. By Proposition 1, $s_1 \notin \mathrm{BR}_1(\sigma_2^1)$. (Indeed, $s_1 \notin \mathrm{BR}_1^{\varepsilon}(\sigma_2^1)$ for $\varepsilon$ sufficiently small.) By Lemma S, $\sigma_1 \notin \mathrm{BR}_1(\sigma_2^1)$. Since $\sigma_2^1$ and $\sigma_1$ were arbitrary,

it follows by contraposition that Conventional Optimization is violated for Player 1. The proof of statement 2 is almost identical. ■

Since players are assumed to optimize, the Theorem implies that, for at least one of the players, either Conventional Prediction fails or the player, in order to optimize, chooses a strategy which is not conventional. In either event, there is no assurance that both players will learn to predict the continuation path of play. Section 1.3 and Section 1.4.2 give examples where prediction does indeed fail.

*Remark 4.* 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 (non-pure) strategy $\sigma_1$ in a small open neighborhood of some pure strategy $s_1$ and Player 2 chooses a (non-pure) strategy $\sigma_2$ 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 strong (uniform convergence) topology.[19] 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 $\sigma_1$ 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 $q^h$.[20] Let $\overline{q} = \sup_{h \in \mathcal{H}} q^h$. For $\overline{q}$ small, $\sigma_1$ will be close to $s_1$ in the strong 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

[19] Loosely, two strategies are close in the strong topology if, after any history (possibly excepting histories which are impossible under either strategy), the mixture over actions chosen by the first strategy is close (in the standard Euclidean sense) to the mixture chosen by the second.

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

which tremble then one should demand only approximate, in particular uniform $\varepsilon$, optimization rather than full optimization. I will address this point in Remark 10 in Section 3.4.1. □

*Remark 5.* Recalling the discussion of fictitious play in Section 1.4.2, suppose that we attempt to adopt as an axiom of rationality the principle that beliefs be "not naive." "Not naive" presumably means something like: "if, as a result of optimization, Player $i$ chooses strategy $\sigma_i$, then his belief should give some weight to his opponent choosing a strategy analogous (in some appropriate sense) to $\sigma_i$." If "give some weight" is interpreted to mean that the player is able to learn to predict the path of play, then we are flirting with paradox. For example, in the game Chicken given above, the closest analog to any pure strategy $s$ is simply $s$ itself. But $s \in \tilde{S}_2(s)$. Hence, by Proposition 1, $s$ is not a best response to any belief for which the player can learn to predict the path of play when the profile is $(s, s)$. Since $s$ was arbitrary, it follows that, in this game, there is no pure strategy best response, hence no best response whatsoever, to any belief which is "not naive." Since a best response exists for any belief, it follows that there are *no* beliefs which are "not naive." The "not naive" axiom, at least in its strong (predictive) interpretation, is vacuous. □

A consequence of the Theorem is the following.

**Proposition 3.** *Let $G$ be a stage game in which neither player has a weakly dominant strategy. Suppose $\hat{\Sigma}_1 \times \hat{\Sigma}_2$ is both neutral and at most countable and consider any belief profile which gives weight to all of $\hat{\Sigma}_1 \times \hat{\Sigma}_2$. Suppose that $\#A_1 \leq \#A_2$.*

1. *There is a $\bar{\delta} \in (0, 1]$ such that for any $\delta \in (0, \bar{\delta})$, Conventional Optimization fails for Player 1.*

2. *If, moreover, Assumption M holds, then for any $\delta \in (0, 1)$, Conventional Optimization fails for Player 1.*

*Similarly for Player 2 if $\#A_2 \leq \#A_1$.*

**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. ■

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 thought 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 which takes histories (encoded in machine readable form) as input and produces the name of an action as output.[21] The Turing implementable strategies are *precisely* those which can be defined recursively, where I use the term "recursive" in its Recursive Function Theory sense. Equivalently, the Turing implementable strategies are precisely those which 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 which 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 1.2.4.) Let $S_i^T \subset S_i$ be the set of pure strategies for Player $i$ which are Turing implementable. $S_i^T$ is countable.[22] For computability reasons, I will assume that the payoff functions $u_i$ are rational valued and that the discount factor $\delta$ is rational.

**Proposition 4.** *Let $G$ be a stage game in which neither player has a weakly dominant strategy. Suppose $\hat{\Sigma}_1 \times \hat{\Sigma}_2 = S_1^T \times S_2^T$. Suppose $\#A_1 \leq \#A_2$. Consider any belief $\sigma_2^1$ which gives weight to all of $\hat{\Sigma}_2$.*

1. *There is a $\bar{\delta} \in (0,1]$ such that for any rational $\delta \in (0,\bar{\delta})$, Conventional Uniform $\varepsilon$ Optimization fails for Player 1.*

2. *If, moreover, Assumption M holds, then for any rational $\delta \in (0,1)$, Conventional Uniform $\varepsilon$ Optimization fails for Player 1.*

[21] 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 (1990) or Odifreddi (1987).

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

*Similarly for Player 2 if* $\#A_2 \leq \#A_1$.

**Proof.** The result for optimization, rather than uniform $\varepsilon$ optimization, follows from Proposition 3 provide $S_1^T \times S_2^T$ is neutral. Verification of the latter is straightforward and is omitted. The extension to uniform $\varepsilon$ optimization is immediate once one notes that Proposition 1 is stated for uniform $\varepsilon$ optimization and that, therefore, the proof of the Theorem extends to uniform $\varepsilon$ optimization provided $\hat{\Sigma}_i \subset S_i$ (conventional strategies are pure). ■

Proposition 4 states that, when $\hat{\Sigma}_i = S_i^T$, a player would have to choose a non-Turing implementable strategy in order to optimize (indeed, in order even to uniformly $\varepsilon$ optimize). Since we continue to assume that players do in fact optimize, the Church-Turing Thesis notwithstanding, the implication is that at least one of the players chooses a strategy which is not fully consistent with his opponent's belief.

*Remark 6.* Although stated for Turing implementable strategies, Proposition 4 holds for *any* standard bound on complexity: *any* standard complexity bound generates a joint conventional set which 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 which give weight to all of Player 2's conventional strategies, Player 1 has no conventional, or even, for $\varepsilon$ small, uniform $\varepsilon$-best, response. In this sense, Player 1's best response will always be more complicated than the strategies which are conventional for Player 2. This is a variation on the point made in Remark 5 that there is a sense in which beliefs *must* be "naive." □

*Remark 7.* For intuition for Proposition 4, consider the following. Say that a belief $\sigma_1$ which 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 $\varepsilon$-best response. Indeed, one can construct a Turing machine which, after any history, computes a finite approximation to the correct posterior beliefs, then computes a best response with respect to those beliefs 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 which 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 $\varepsilon$-best response does not apply. Proposition 4 verifies that, for $\varepsilon$ sufficiently small, no uniform $\varepsilon$-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 which is not Turing computable is equivalent to facing an opponent playing a strategy which 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, or even, for $\varepsilon$ small, uniformly $\varepsilon$-best, response. □

## 3.4 Extentions.

### 3.4.1 Constrained Rational Players.

The analysis thus far has implicitly maintained the hypothesis that players are free to choose non-conventional 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, at least one of the players will be unable to optimize. However, one might hope that, despite the constraint, players could at least uniformly $\varepsilon$ optimize. If this were true then a small modification of the arguments 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 $\varepsilon$ optimization. Thus, for $\varepsilon$ small, so long as Conventional

Prediction holds, the constraint prevents at least one of the players from choosing a strategy which is uniformly $\varepsilon$ optimal. This does not, of course, prevent a player from choosing a strategy which is *ex ante* $\varepsilon$ optimal. However, as illustrated in Section 1.3, *ex ante* $\varepsilon$ optimization *per se* may not be enough to guarantee convergence to Nash equilibrium play.

If, on the other hand, the conventional set contains non-pure strategies then the proof of the Theorem does not extend. The difficulty is that Lemma S is false for uniform $\varepsilon$ optimization: even if a strategy $\sigma$ is uniformly $\varepsilon$ optimal, some of the pure strategies in its support may not be. Despite this problem, the conclusion of the Theorem does extend, for players who are sufficiently impatient, for conventional sets consisting of the Turing implementable strategies, the benchmark case covered in Proposition 4, even if we modify the definition of Turing machine to permit access to randomization devices (coin tossers).[23]

Let $\Sigma_i^T$ 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, $\Sigma_i^T$ is countable. Hence it makes sense to speak of Player 1's belief giving weight to all of $\Sigma_2^T$. The proof is in the Appendix.

**Proposition 5.** *Let $G$ be a stage game in which neither player has a weakly dominant strategy. Suppose $\#A_1 \leq \#A_2$. There is a $\bar{\delta} \in [0,1)$ such that, for any rational $\delta \in [0,\bar{\delta})$, for any belief $\sigma_2^1$ which gives weight to all of $\Sigma_2^T$, Conventional Uniform $\varepsilon$ Optimization fails for Player 1. Similarly for Player 2 if $\#A_2 \leq \#A_1$.*

*Remark 8.* 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 $s_2' \in S_2^T$ which is close, in the strong topology, to a strategy $s_2 \in S_2$, where $s_2$ is such that, after any history, the mixture of actions chosen by $\sigma_1$ 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 which are neutral and rich in the above sense. For example, it extends to conventional sets formed by the strategies which are implementable by finite automata (roughly, computers with finite memory). □

[23] 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 9.* It is not known to what degree Proposition 5 extends to players who are patient ($\delta$ is high), although it does extend for some non-generic stage games, such as Matching Pennies. □

*Remark 10.* Although, as already noted, the proof used for the Theorem does not generally extend to uniform $\varepsilon$ optimization if conventional sets contain non-pure 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 4. Since strategies in $\hat{\Sigma}_i$ are close, in the strong 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 a $\bar{\varepsilon} > 0$ such that, if Conventional Prediction holds, then Conventional Uniform $\varepsilon$ Optimization fails for any $\varepsilon \in (0, \bar{\varepsilon})$, for any $\overline{q}$ no matter how small (recall that $\overline{q}$ was the maximal tremble). □

### 3.4.2 Boundedly Rational Players.

Bounded rationality refers to fundamental constraints on a player's ability to frame a belief and calculate a best response. There is, unfortunately, no consensus on how bounded rationality should be modeled. To remain close to the basic motivation of this paper, we will imagine a boundedly rational player as a Bayesian uniform $\varepsilon$ optimizer whose conceptual ability is fundamentally limited. For concreteness, we suppose that a boundedly rational player is essentially a Turing machine, and that it is this Turing machine which must frame beliefs and compute a uniform $\varepsilon$ optimum. Similar remarks will apply if a player is instead modeled as, say, a finite automaton.

Since players are Turing machines, each player's belief must be computable. As noted in Remark 7, this implies that each player will be able to uniformly $\varepsilon$ optimize. Remark 7 also indicates that each player's belief, being computable, cannot give weight to all of his opponent's Turing implementable strategies. For example, the belief might assign positive probability only to opposing strategies which 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 we are in the same situation we faced above: a variant of Proposition 4 (or of Proposition 5, if the conventional sets are sufficiently rich in the sense discussed in Remark 8) tells us that at least one of the

players will choose a non-conventional (but still Turing implementable) strategy in order to uniformly $\varepsilon$ optimize. Thus, players may fail to learn to predict the actual path of play and therefore the path of play may fail to converge to that of a Nash equilibrium (although there may be convergence to Nash equilibrium behavior in some weaker sense, as discussed in Section 1.4.2).

If, on the other hand, bounded rationality implies that neutrality fails, then it is possible for Conventional Optimization to hold. (Conventional Prediction is automatic, since, by the above definition of the joint conventional set, beliefs satisfy grain of truth whenever the strategy profile is conventional.) We 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. However, a failure of neutrality, in and of itself, does not assure Conventional Optimization. For Conventional Optimization, neutrality must fail the right way, 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, rather than impedes, Conventional Optimization. Indeed, as suggested by the discussion of invariance in Section 1.2.2, it is not clear why bounded rationality would imply that neutrality would fail at all.

## Appendix

**Proof of Lemma S.** Let $\sigma_1^*$ share the support of $\sigma_1$ and suppose $\sigma_1^* \notin \mathrm{BR}_1(\sigma_2^1)$. Consider any sequence of strategies $\sigma_{1k}$ such that (a) $\sigma_{1k}$ converges to $\sigma_1^*$ in the weak topology, (b) for any $k$, $\sigma_{1k}$ shares the support of $\sigma_1$, and (c) for any $k$, $\sigma_{1k}$ agrees with $\sigma_1$ except for at most a finite number of histories.[24] Because $\sigma_1^*$ is not a best response, and because $V_1$ is continuous when $\Sigma_1 \times \Sigma_2$ is endowed with the weak product topology, there is a $k$ such that $\sigma_{1k} \notin \mathrm{BR}_1(\sigma_2^1)$. Because $\sigma_{1k}$ shares the support of $\sigma_1$, and because $\sigma_{1k}$ agrees with $\sigma_1$ except for at most a finite number of histories, one can show that there is an $\alpha > 0$ and a $\sigma_1^\circ \in \Sigma_1$ such that $\sigma_1 = \alpha\sigma_{1k} + (1-\alpha)\sigma_1^\circ$. Choose any $\sigma_1' \in \mathrm{BR}_1(\sigma_1^2)$. Then, since $\sigma_{1k} \notin \mathrm{BR}_1(\sigma_2^1)$ 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 $\sigma_1 \notin \mathrm{BR}_1(\sigma_2^1)$. The proof then follows by contraposition.[25] ■

**Proof of Proposition 1.** Let

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

Given the definition of $\tilde{a}_2(a_1)$, $w_1(a_1) \geq 0$. 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

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

[24] In particular, one can construct such a sequence by enumerating the finite histories (which form a countable set), 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.

[25] 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.

Let

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

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

To prove the first part of the proposition, choose $\bar{\delta}$ sufficiently small that, under uniform $\varepsilon$ 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 $\delta \in [0, \bar{\delta})$,

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

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, $\ell = 1$), 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*

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

Temporarily fix any $\varepsilon \in [0, \bar{\varepsilon})$ and any $\delta \in [0, \bar{\delta})$. 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 $\varepsilon$ optimization requires

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

or

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

where I have used the fact that, since $s_2 \in \tilde{S}_2(s_1)$, $\max_{a_1 \in A_1} u_1(a_1, a_2^*) - u_1(a_1^*, a_2^*) = w_1(a_1) \geq \underline{w}_1$. However, by the construction of $\bar{\varepsilon}$ and $\bar{\delta}$, there is an $\eta$ 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 $\varepsilon \in [0, \bar{\varepsilon})$,

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

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 $\ell > 0$. Suppose that Player 1 learns to predict the continuation path of play. Then, for any continuation game beginning at time $t+1$, $t > t(\eta, \ell)$, Player 1 assigns some probability $(1-\eta') > (1-\eta)$ to the actual $\ell$-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 there is a probability $\eta'$ that the continuation history might be something else. In an alternate $\ell$-period continuation history, Player 1 could receive at most $\overline{u}_1$ per period. Finally, from date $t+\ell+1$ onwards, Player 1 could receive at most $\overline{u}_1$ per period. Thus beginning at date $t+1$, Player 1 expects to earn *at most*

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

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 $\varepsilon$ optimization requires

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

or

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

By the construction of $\bar{\varepsilon}$, there is an $(\eta, \ell)$ such that this inequality cannot hold for any $\varepsilon \in [0, \bar{\varepsilon})$. This establishes the second part of the proposition. ■

**Proof of Proposition 2.** Consider any neutral map $\Psi : \overset{\bullet}{A} \times \overset{\bullet}{A} \to \overset{\bullet}{\Sigma} \times \overset{\bullet}{\Sigma}$ such that $\Psi(A_1, A_2) = \hat{\Sigma}_1 \times \hat{\Sigma}_2$. Consider any $s_1 \in \hat{\Sigma}_1 = \Psi_1(A_1, A_2)$. Let $s_2 \in S_2$ be defined by, for any history $h$,

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

I will argue that $s_2 \in \hat{\Sigma}_2$.

Suppose that $A_1 = A$, $A_2 = A'$. Consider first the special case in which $A \subset A'$. Then, by assumption, $s_1 \in \Psi_1(A, A')$. By Property 2(a) of neutrality (Player Symmetry), $s_1 \in \Psi_2(A', A)$. By Property 4 of neutrality (monotonicity), $\Psi_2(A', A) \subset \Psi_2(A, A')$. Therefore, $s_1 \in \Psi_2(A, A')$. 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 $g_2^\diamond(a_2) = \tilde{a}_2(a_2)$ if $a_2 \in A$. By Property 3 of neutrality (invariance to trivial changes in strategy), $s_2 = \gamma_2^\diamond(s_1) \in \Psi_2(A, A')$.

If $A \not\subset A'$, we can extend the above argument as follows. By an assumption made when $\overset{\bullet}{A}$ was defined, there exists a set $A^* \in \overset{\bullet}{A}$ 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. Let $G^* = (A^*, A', u_1^*, u_2^*)$ be the game defined by, for any $(a_1, a_2) \in A^* \times A'$, $u_i^*(a_1, a_2) = u_i(g_1^{-1}(a_1), a_2)$. Thus $G^*$ is identical to $G$ up to a renaming of Player 1's actions. Then, by Property 2(b) of neutrality (Action Symmetry), $\gamma_1(s_1) \in \Psi_1(A^*, A')$ and hence, by the same argument as above, $\gamma_1(s_1) \in \Psi_2(A^*, A')$. Also as in the argument above, 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 $g_2^\diamond(a_2) = \tilde{a}_2(a_2)$ if $a_2 \in A^*$. Then, by Property 3 of neutrality, $\gamma_2^\diamond(\gamma_1(s_1)) \in \Psi_2(A^*, A')$ and, by Property 2(b) of neutrality, $\gamma_2^{-1}(\gamma_2^\diamond(\gamma_1(s_1))) \in \Psi_2(A, A')$. One can check that $s_2 = \gamma_2^{-1}(\gamma_2^\diamond(\gamma_1(s_1)))$. ■

**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 (biased) coins. If random state $\xi$ 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 definition 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 $(\alpha_1, a_2)$.[26] Similarly, extend the definition of $w_1$, introduced in the proof of Proposition 1, to

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

As before, $w_1(\alpha_1) \geq 0$. Moreover, $w_1(\alpha_1) = 0$ iff $\alpha_1$ is dominant (weakly or strictly). Since, by assumption, no action (pure or mixed) is weakly dominant, $w_1(\alpha_1) > 0$ for all $\alpha_1$. $\Delta(A_1)$ 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 $\overline{u}_1$ and $\underline{u}_1$ be defined as in Proposition 1. Again, since no action is weakly dominant, $\overline{u}_1 > \underline{u}_1$.

[26] 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 $\varepsilon$ 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 $\delta \in [0, \bar{\delta})$,

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

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 we 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 $\Sigma_2^T$, which is countable, Player 1 learns to predict the path of play generated by $(\sigma_1, s_2^\nu)$. 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, $\ell = 1$), 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 $h = \pi(z, t)$. Discounting payoffs to date $t+1$, Player 1's expected payoff in the continuation game is then *at most*

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

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 $\varepsilon$ optimization requires

$$\begin{aligned} \varepsilon + (1-\eta')\mathbb{E}_{\sigma_1(h)} u_1(a_1, s_2^\nu(h)) + \eta' \overline{u}_1 + \frac{\delta}{1-\delta}\overline{u}_1 & \\ \geq (1-\eta') \max_{a_1 \in A_1} u_1(a_1, s_2^\nu(h)) + \eta' \underline{u}_1 + \frac{\delta}{1-\delta}\underline{u}_1 & \end{aligned}$$

or

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

where I have used Inequality 1 and the fact that $w_1(\alpha_1) \geq \underline{w}_1$. However, by the construction of $\bar{\varepsilon}$ and $\bar{\delta}$, there exist $\nu$ and $\eta$ 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 $s_2^\nu$ satisfying Inequality 1. To avoid bogging down the paper in computability details, I will only sketch the Turing machine construction. Suppose, then, that $\sigma_1$ is implemented by a Turing machine $M$. Let $\nu > 0$ be as given above. From $M$, one can show that one construct a new *deterministic* Turing machine $M^\nu$ which does the following. On input of a history $h$, $M^\nu$ simulates the action of $M$ on $h$. Every time $M$ randomizes, the flow of its program branches in two. $M^\nu$ 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, $M^\nu$ can calculate an approximation, say $\alpha_1'$, to the true mixture over actions, say $\alpha_1 = \sigma_1(h)$. We 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 $\alpha'$ is sufficiently close to $\alpha$. Since $M$ has only a finite number of random states, the accuracy of the estimate $\alpha_1'$ improves geometrically with the depth of the simulation (number of times random states are hit). Moreover, since we can program knowledge of the $p(\xi)$ into $M^\nu$, $M^\nu$ will be able to calculate whether a depth has been reached sufficient to ensure that its estimate $\alpha_1'$ is close enough to $\alpha$ that Inequality 1 holds. Therefore, $M^\nu$ calculates $s_2^\nu(h)$ in finite time.

There are two reasons to have $M^\nu$ approximate $\alpha_1$ rather than to calculate it exactly. First, if $M$ chooses an action only with probability 1, rather than for certain, then $M^\nu$ may be unable to calculate $\alpha_1$ exactly. In particular, if $M^\nu$ attempts to calculate $\alpha_1$ 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 $\alpha$ exactly is desirable because it reduces the complexity of $M^\nu$. In particular, by taking an approximation, the number of computational steps required by $M^\nu$ can be held to a multiple of the number expected

for $M$, and may even be smaller than the worst case for $M$. ■

## References


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

Aoyogi, 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, pp. 627–650. Princeton University Press, Princeton, NJ, Annals of Mathematics Studies, 52.

——— (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

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, pp. 374–376. John Wiley, New York.

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

Cutland, N. J. (1990): *Computability*, vol. 60. Cambridge University Press, Cambridge, UK.

Eichberger, J., and D. Kelsey (1995): "Uncertainty Aversion and Preference for Randomization," University of Birmingham.

Foster, D., and R. Vohra (1995): "Calibrated Learning and Correlated Equilibrium," Wharton School, University of Pennsylvania.

Fudenberg, D., and D. Kreps (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5(3), 320–367.

Fudenberg, D., and D. Levine (1993): "Self-Confirming Equilibrium," *Econometrica*, 61(3), 523–545.

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

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

——— (1995c): "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(3), 368–386.

Kalai, E., and E. Lehrer (1993a): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61(5), 1019–1045.

——— (1993b): "Subjective Equilibrium in Repeated Games," *Econometrica*, 61(5), 1231–1240.

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

Lehrer, E., and R. Smorodinsky (1994): "Repeated Large Games with Incomplete Information," Northewestern University.

Lehrer, E., and S. Sorin (1994): "$\varepsilon$-Consistent Equilibrium," Northewestern University.

Nachbar, J. H., and W. R. Zame (1996): "Non-Computable Strategies and Discounted Repeated Games," *Economic Theory*, forthcoming.

Nyarko, Y. (1994): "Bayesian Learning Leads to Correlated Equilibria in Normal Form Games," *Economic Theory*, 4, 821–841.

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

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," *Procedings of the London Mathematical Society*, 42, 230–165, Corrections, ibid, 1937, *43*, 544-546.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.