Unilateral Deviations with Perfect Information*

Pradeep Dubey† and Ori Haimanko‡

October 2000

Abstract

For extensive form games with perfect information, consider a learning process in which, at any iteration, each player unilaterally deviates to a best response to his current conjectures of others' strategies; and then updates his conjectures in accordance with the induced play of the game. We show that, for generic payoffs, the outcome of the game becomes stationary in finite time, and is consistent with Nash equilibrium. In general, if payoffs have ties or if players observe more of each others' strategies than is revealed by plays of the game, the same result holds provided a rationality constraint is imposed on unilateral deviations: no player changes his moves in subgames that he deems unreachable, unless he stands to improve his payoff there. Moreover, with this constraint, the sequence of strategies and conjectures also becomes stationary, and yields a self-confirming equilibrium.

Key Words: extensive form games with perfect information, self-confirming and Nash equilibria, unilateral deviations, objective updates, convergence in finite time.

JEL Classification: C 72.

---

*The authors are grateful to the Cowles Foundation for Research in Economics, where part of this work was carried out.

†Center for Game Theory, Department of Economics, SUNY at Stony Brook, NY 11794, USA. E-mail: pradeepkdubey@yahoo.com

‡CORE, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. E-mail: haimanko@core.ucl.ac.be

1 Introduction

Consider a finite extensive form game with perfect information. Let the players begin with initial conjectures of each others' strategies[1] and "true" strategies of their own. This gives rise to a play of the game, i.e., to a sequence of moves from start to finish, dictated by the true strategies. Let each player now update his conjectures in accordance with the revealed play. And then let him, if he can, unilaterally deviate to any strategy that maximizes his payoff in the light of his updated conjectures. A new set of conjectures and strategies is obtained. Iterate the process. In Theorem 2 we show that, for generic payoffs at the terminal nodes, the outcome of the game (i.e., the set of plays which occur with positive probability) becomes stationary in a finite number of steps and is consistent with Nash equilibrium[2] (NE).

Now consider the general scenario. Suppose payoffs involve ties (the non-generic case), in the sense that there exists a player who obtains the same payoff from two different outcomes in the game. Or else suppose that some players are able to observe more of their rivals' true strategies than is revealed by plays of the game. In either case the outcomes may fluctuate forever. To stabilize them, what is needed is a touch of "rationality": precisely, on irrelevant subgames which are not reachable given his conjectures, the player who deviates alters his old strategy only if he stands to gain by it there[3]. With this constraint on unilateral deviations, the sequence of conjectures and strategies becomes stationary, and constitutes a self-confirming equilibrium (SCE) in the sense of Fudenberg and Kreps (1995) (see Theorem 1). The outcome generated by the SCE remains Nash. Indeed, in the presence of perfect information, SCE and NE outcomes always coincide (see Lemma 1).

When players fully observe true strategies of their rivals, the limit is a full-fledged NE. This result was already established in Dubey (1981). We obtain it here as a byproduct. Our main point however is that, for convergence to NE *outcomes*, it suffices that players have the capacity to observe *just* the plays of the game.

[1]Throughout, "strategy" means "pure strategy."

[2]The NE, which sustain the outcome, need *not* be subgame perfect.

[3]He could, of course, behave in an arbitrary and eccentric manner on these subgames, without in the least affecting the improvement in his payoff. Our "rationality" constraint rules out such eccentricity. Notice that for generic payoffs in the game, when there are no ties, the constraint simply says: the player who deviates does not act on irrelevant subgames in such a way as to render himself *worse-off* there.

The learning process of our model is quite robust. Updates (of conjectures) and deviations (in strategy) need not be synchronous across players. It is essential only that there be infinitely many "rounds," and that in every round each player get at least one opportunity to undertake them. Players may have different numbers of opportunities in any round, or even idiosyncratic time lags in observations. Nevertheless convergence is guaranteed.

It may be worth pointing out some differences between our learning process and those of the standard models of learning, most notably fictitious play. Myopic best response is not required at each stage in our set-up. Moreover, it is only his latest observation of the play that needs to be recorded by any player. He does not have to track the history of past plays to form estimates of the empirical frequencies of his rivals' strategies. This is on account of perfect information, which eliminates the need for mixed strategies.

Finally, for those who feel that NE which are not subgame perfect are not meaningless, it may be noted that our approach provides a new proof of the existence of pure strategy NE without recourse to backward induction or subgame perfection (see Remark 3).

2 Nash and Self-Confirming Equilibria

Let us recall the definition of an extensive form game with perfect information. There is a finite tree with a distinguished node α^* called the *root*, which represents the start of the game. The root α^* orients the tree: node β *follows* node α if $\alpha \neq \beta$ and α is on the unique path from α^* to β. Nodes that have no followers are called *terminal nodes* and represent the end-points of the game. Denote the set of players by $N = \{1, ..., n\}$ and chance by 0. Then each non-terminal node α is labelled by some $i \in N \cup \{0\}$, signifying a *position* in the game at which i must make a move; and the *moves* available to i at α are identified with the arcs that "issue-out" of α, i.e., lead to an immediate follower of α. Let P^i denote the set of all positions of i. (For ease of induction arguments later, we will allow for P^i to be the empty set.) A *strategy* of i specifies a move at each position in P^i. Players $i \in N$ have freedom of choice over their strategies. In contrast, chance is a strategic dummy: at each node α in P^0, chance picks all its moves in accordance with a probability distribution that is specified exogenously at α.

Finally, to complete the description of the extensive form game, each player $i \in N$ has a payoff function u^i defined on the terminal nodes.

Let S^i denote the set of all strategies of $i \in N \cup \{0\}$, i.e., the Cartesian product of the sets of moves taken across all positions in P^i. (If P^i is empty, we will view S^i as consisting of the "empty" strategy, the use of which has no influence on outcomes or payoffs in the game.) Since the moves of chance are picked independently at different positions in P^0, each $s^0 \in S^0$ occurs with probability $\sigma(s^0) \equiv$product of the probabilities of the moves in s^0. Without loss of generality we assume that σ has full support on S^0.

A *play* of the game is the set of all nodes on a path from the root α^* to a terminal node. It may be identified with the terminal node of the path, since each uniquely determines the other.

Put $S \equiv S^1 \times ... \times S^n$. Then, if players' choices constitute a strategy-profile $s \in S$ and chance selects $s^0 \in S^0$, a play $\pi(s, s^0)$ of the game is induced. We denote by $\prod(s) \equiv \{\pi(s, s^0) \mid s^0 \in S^0\}$ the set of plays that occur with positive probability under s, and call it the *outcome* of s. Any terminal node α in $\prod(s)$ is reached under s with probability $p(\alpha, s) = \sum \sigma(s^0)$, where the summation is taken over all $s^0 \in S^0$ such that α is the terminal node of $\pi(s, s^0)$. It is evident that the expected payoff to player i from $s \in S$ is

$$u^i(s) \equiv \sum_{\alpha \in \prod(s)} p(\alpha, s)\, u^i(\alpha) = \sum_{s^0 \in S^0} \sigma\left(s^0\right) u^i\left(\pi\left(s, s^0\right)\right) \equiv u^i\left(\prod(s)\right).$$

For $s \equiv (s^1, ..., s^n) \in S$ and $t \in S^i$, let $(s \mid t)$ denote the strategy-profile obtained from s when s^i is replaced with t. Then s is defined to be a *Nash equilibrium* (NE) of the game if, for each $i \in N$,

$$u^i(s) \geq u^i(s \mid t) \tag{1}$$

for all $t \in S^i$.

A weaker notion is that of a *self-confirming equilibrium* (SCE) described by Fudenberg and Kreps (1995). For each $i \in N$, consider $s_i \equiv (s_i^1, ..., s_i^n) \in S$. The interpretation is that s_i^j is player i's conjecture of j's strategy for $j \in N \backslash \{i\}$, whereas s_i^i is his own "true" strategy. Then $(s_1, ..., s_n) \in S^N$ is an SCE if, for each $i \in N$,

(i)

$$u^i(s_i) \geq u^i(s_i \mid t), \tag{2}$$

for all $t \in S^i$; and

(ii)

$$s_i^j(\alpha) = s_j^j(\alpha), \tag{3}$$

for all $j \in N \backslash \{i\}$ and all $\alpha \in P^j \cap \prod (s_1^1, ..., s_n^n)$.

Thus, in an SCE, each player makes a best response to his conjectures of others' strategies; and, moreover, all conjectures coincide with the true strategies on the SCE outcome $\prod (s_1^1, ..., s_n^n)$. If it happens that all conjectures coincide with true strategies everywhere, i.e., $s_i^j = s_j^j$ for all $i \neq j$, then an SCE $s \in S^N$ yields the NE $(s_1^1, ..., s_n^n) \in S$, and we identify the two.

We shall say that an outcome is an SCE (or, NE) outcome if there exists an SCE (or, NE) which gives rise to it. Our first lemma shows that the notions of NE and SCE are equivalent in terms of outcomes, in the context of extensive form games with perfect information.

Lemma 1 *The set of NE outcomes is equal to the set of SCE outcomes.*

Proof. Let $s = (s^1, ..., s^n) \in S$ be an NE. Define $\widetilde{s} \in S^N$ by $\widetilde{s}_i^j = s^j$ for all $i, j \in N$. Clearly $\widetilde{s}$ is an SCE with the same outcome as the NE.

Now suppose $\widetilde{s} = (\widetilde{s}_1, ..., \widetilde{s}_n) \in S^N$ is an SCE with outcome $\prod (\widetilde{s}_1^1, ..., \widetilde{s}_n^n) \equiv \prod^*$. Given any two nodes α and β, let $Path(\alpha, \beta)$ denote the set of nodes on the path from α to β. (Thus, if β does not follow α and $\beta \neq \alpha$, $Path(\alpha, \beta)$ is the empty set; otherwise α and β are both in $Path(\alpha, \beta)$.) Given $i, j \in N$ and $\alpha \in P^i \cap \prod^*$, define

$$X(j, i, \alpha) = \left\{ \beta \in P^j \mid Path(\alpha, \beta) \cap \prod{}^* = \{\alpha\} \right\}.$$

Note that the (disjoint) union of $X(j, i, \alpha)$, taken over all $i \in N$ and $\alpha \in P^i \cap \prod^*$, is precisely P^j. This allows us to define $s = (s^1, ..., s^n) \in S$ as follows: for any $j \in N$, let $s^j(\beta) = \widetilde{s}_i^j(\beta)$ if $\beta \in X(j, i, \alpha)$ for some i and α, such that either $i \neq j$ or $\alpha = \beta$; and let $s^j(\beta)$ be arbitrary if $i = j$ and $\alpha \neq \beta$. It is easy to check that s is an NE with outcome $\prod^*$. Thus every SCE outcome is an NE outcome. ■

3 The Learning Process

Given any $s = (s^1, ..., s^n) \in S$, player i may unilaterally deviate to $t \in S^i$ if it improves his payoff, i.e., $u^i(s \mid t) > u^i(s)$. But such a t may involve eccentric behavior on irrelevant parts of the game tree. To rule this out, let us define the response $t \in S^i$ to satisfy the *rationality constraint* vis-a-vis $s \in S$ if, for every subgame[4] Γ having a node in P^i as its root,

[4] A subgame is defined by any node, considered as the root, and all its followers.

$$t_\Gamma \neq s^i_\Gamma \Rightarrow u^i_\Gamma \left(s_\Gamma \mid t_\Gamma\right) > u^i_\Gamma \left(s_\Gamma\right),$$

where s^i_Γ (or, t_Γ) stands for the restriction of s^i (or, t) to Γ, and u^i_Γ is i's payoff function in Γ. Denote the set of all such responses of i to s by $I^i(s)$.

Notice that $s^i \in I^i\left(s^1, ..., s^i, ..., s^n\right)$, i.e., we do not *yet* insist that i must change his strategy if he can achieve a higher payoff. The essential idea behind $I^i(s)$ is only to exclude changes in strategy made, so to speak, "in vain," i.e., to exclude changes in any subgame that provide no gain there. This is not a severe exclusion. Indeed, let $s^{-i} \equiv \left(s^1, ..., s^{i-1}, s^{i+1} ..., s^n\right) \in \times_{j \in N \setminus \{i\}} S^j$ and $r \in S^i$ be arbitrary. For any $t \in S^i$ such that $u^i\left(s^{-i}, t\right) > u^i\left(s^{-i}, r\right)$ there exists a $\widetilde{t} \in I^i\left(s^{-i}, r\right)$ with $\prod\left(s^{-i}, \widetilde{t}\right) = \prod\left(s^{-i}, t\right)$, and hence, of course, $u^i\left(s^{-i}, \widetilde{t}\right) = u^i\left(s^{-i}, t\right)$. Thus, in his quest for better outcomes via unilateral deviations, player i is not encumbered at all by the restriction to strategies in $I^i\left(s^{-i}, r\right)$. In some sense only inefficient deviations are excluded from $I^i\left(s^{-i}, r\right)$. Think of r as the past historical strategy of i and assume that it *costs* i to make an alteration in his moves at any position in P^i. Then, upon being confronted with s^{-i}, i will wish to change from r to a best response to s^{-i} at the least cost, and any such change will lead to a strategy in $I^i\left(s^{-i}, r\right)$. More generally, i may weigh the cost of the change against the gain in payoff, and choose a strategy which maximizes the net: any optimal choice will also inevitably lie in $I^i\left(s^{-i}, r\right)$.

Our learning process is described by a random infinite sequence $\{s(l)\}_{l=1}^{\infty}$ of strategies and conjectures, where $s(l) \equiv \left(s_1(l), ..., s_n(l)\right) \in S^N$. We call it a *learning sequence*. The play of the game induced at the l^{th} iteration depends on the true strategies $\left(s^1_1(l), ..., s^n_n(l)\right)$ in $s(l)$ as well as the strategy $s^0(l)$, picked by chance. We assume that chance is an automaton which repeatedly plays its mixed strategy σ, i.e., at every iteration, it picks $s^0 \in S^0$ with probability $\sigma\left(s^0\right)$ independently of previous iterations.

We will say that player i is *admissible* in $\{s(l)\}_{l=1}^{\infty}$ if for every l

$$s^i_i(l+1) \in I^i\left(s^{-i}_i(l+1), s^i_i(l)\right). \tag{4}$$

This simply says that whenever player i changes his strategy, the change must satisfy the rationality constraint vis-a-vis his conjectures at that time and his previous strategy.

Next we say that players are *updating objectively* if they revise their conjectures based *only* upon (possibly partial) knowledge of others' *true* strategies. Formally, i is updating objectively in $\{s(l)\}_{l=1}^{\infty}$ if there exists a random sequence[5] $A(1), ..., A(l), ...$ of subsets of $\bigcup_{j\in N\setminus\{i\}} P^j$ such that

$$s_i^j(l+1)(\alpha) = s_j^j(l)(\alpha) \tag{5}$$

for all l and all $\alpha \in A(l) \cap P^j$ and all $j \in N\setminus\{i\}$; and

$$s_i^j(l+1)(\alpha) = s_i^j(l)(\alpha) \tag{6}$$

for all l and all $\alpha \notin A(l)$ and all $j \in N\setminus\{i\}$.

In the next section we shall refine the notion of updating in a more meaningful way. But at the moment we are aiming for the following technical lemma, stated in as much generality as possible. First let us introduce some terminology which we will use repeatedly. Let X be a finite set. Given a random sequence $\{x(l)\}_{l=1}^{\infty}$ with $x(l) \in X$ for every l, we define an associated random variable $L(\{x(l)\}_{l=1}^{\infty})$ by

$$L(\{x(l)\}_{l=1}^{\infty}) = \min\{k \mid x(l) = x(k) \text{ for all } l \geq k\},$$

if the set of such k is not empty, and $L(\{x(l)\}_{l=1}^{\infty}) = \infty$ otherwise. We say that the sequence $\{x(l)\}_{l=1}^{\infty}$ *becomes stationary* (or, *converges in finite time*) if $L(\{x(l)\}_{l=1}^{\infty}) < \infty$ with probability 1. The values which the random variable $x(L(\{x(l)\}_{l=1}^{\infty}))$ takes with positive probability will be called *limit values* of the sequence $\{x(l)\}_{l=1}^{\infty}$.

Lemma 2 *Let $\{s(l)\}_{l=1}^{\infty}$ be a learning sequence, in which each player is admissible and is updating objectively. Then it becomes stationary.*

Proof. This will be by induction on the number of arcs k in the game tree. Note that if $\bigcup_{i\in N} P^i$ is empty, i.e., the game consists of just chance moves or just a terminal node, then the lemma holds vacuously. In particular it holds for $k = 0$. Assume it to be true whenever $k \leq q$ and take the case $k = q+1$.

Suppose that the sequence $\{s(l)\}_{l=1}^{\infty}$ does not become stationary. Note that if $\left\{(s_i^i(l))_{i\in N}\right\}_{l=1}^{\infty}$ were to become stationary, then so would $\{s(l)\}_{l=1}^{\infty}$ since the players are updating objectively; thus $\left\{(s_i^i(l))_{i\in N}\right\}_{l=1}^{\infty}$ also does not become stationary. It follows that there exist $i \in N$ and $\alpha \in P^i$, such

[5] Every random variable is defined on the same domain. Thus $\{A(l)\}_{l=1}^{\infty}$ and $\{s(l)\}_{l=1}^{\infty}$ may vary with $\{s^0(l)\}_{l=1}^{\infty}$, and possibly other extraneous random factors.

that, with positive probability, i picks at least two arcs at α infinitely often in $\left\{(s_i^i(l))_{i\in N}\right\}_{l=1}^{\infty}$. Consider subgames $\Gamma_1, ..., \Gamma_m$ $(m \geq 2)$ which start at immediate followers of α, i.e., the ends of arcs issuing-out of α. In each such Γ_j the sequence $\left\{s_{\Gamma_j}(l)\right\}_{l=1}^{\infty}$ clearly satisfies the hypothesis of the lemma; and hence, by the inductive assumption, it becomes stationary. Since i is updating objectively, his conjectures of others' strategies in the subgames will also become stationary. But then, with probability 1, i would pick only one arc at α infinitely often because he is admissible in $\{s(l)\}_{l=1}^{\infty}$, a contradiction. ■

4 Convergence to Nash Outcomes

We now refine our general updating process. Consider a player i who is updating objectively via the sets $\{A(l)\}_{l=1}^{\infty}$ in the sequence $\{s(l)\}_{l=1}^{\infty}$. We say that i *observes plays* if there is an infinite set $J(i)$ such that

$$A(l) \supset \left(\bigcup_{j\in N\setminus\{i\}} P^j\right) \bigcap \pi\left(s_1^1(l), ..., s_n^n(l), s^0(l)\right) \tag{7}$$

for all $l \in J(i)$. In other words, player i is able to infinitely often observe at least the moves made by others along realized plays. The most natural scenario is that he observes just these moves infinitely often. (Think of board games in which players can only see the changing configurations of pieces on the board.) To this end, we say that i *observes only plays* if the set inclusion " $\supset$ " in (7) is replaced with equality for all $l \in J(i)$, and with the reverse inclusion "$\subset$" for all other l.

We say that i *observes strategies* if he observes plays and, for infinitely many l,

$$A(l) = \bigcup_{j\in N\setminus\{i\}} P^j, \tag{8}$$

i.e., i observes the entire strategy of every other player infinitely often.

We say that player i is *improving admissibly* in a learning sequence $\{s(l)\}_{l=1}^{\infty}$ if he is admissible, and, for infinitely many l,

$$s_i^i(l+1) \in I^i\left(s_i^{-i}(l+1), s_i^i(l)\right) \cap BR\left(s_i^{-i}(l+1)\right), \tag{9}$$

where $BR\left(s_i^{-i}(l+1)\right) = \arg\max_{t\in S^i} u^i\left(s_i^{-i}(l+1), t\right)$. As was said, the intersection is always non-empty. Note that, unlike $J(i)$ in the definition of observation of plays, the (infinite) set of l for which (9) holds is not required to be deterministic.

Theorem 1 Let $\{s(l)\}_{l=1}^{\infty}$ be a learning sequence in which each player is improving admissibly and updating objectively. If each player observes plays then the sequence converges in finite time and all its limit values are SCE (with NE outcomes). If, moreover, each player observes strategies, then the limit values are NE.

Proof. According to Lemma 2, the sequence $\{s(l)\}_{l=1}^{\infty}$ converges in finite time. Let $\overline{s} \in S^N$ be a limit value of the sequence. We have to show that $\overline{s}$ is an SCE when players observe plays.

Let L be such that

$$L = L\left(\{s(l)\}_{l=1}^{\infty}\right) \text{ and } s(L) = \overline{s} \tag{10}$$

with positive probability. From now, our arguments will be conditioned on (10). That $\overline{s}$ satisfies (2) in the definition of an SCE follows from (9) in the definition of admissible improvements for $l \geq L$. Also, from (5) in the definition of objective updates, and from (7), it follows that all conjectures in $\overline{s}$ coincide with the true strategies on plays in the outcome $\prod(\overline{s}_1^1, ..., \overline{s}_n^n)$ that are observed by players after L^{th} iteration of the game. Since chance picks its moves according to a non-vanishing stationary distribution, independently across iterations, every strategy of chance in the game is realized infinitely often with probability 1. Thus players observe every play in $\prod(\overline{s}_1^1, ..., \overline{s}_n^n)$ after the L^{th} iteration, with probability 1. Therefore (3) in the definition of an SCE is also satisfied, and so $\overline{s}$ is indeed an SCE.

If players observe strategies then, conditional on (10), their conjectures coincide with true strategies at some iteration following L, and hence $\overline{s}$ is an NE. ■

Remark 1 (Better versus Best Responses) As is evident from its proof, the conclusion of Theorem 1 would still hold if each player chose only to strictly improve his payoff whenever possible, instead of maximizing it (i.e., going to a best response) for infinitely many l. ■

Remark 2 (Range of the Learning Process) Suppose that the vector of initial strategies and conjectures is already an SCE and players observe plays. Consider an infinite repetition of the SCE. In this sequence each player is improving admissibly and updating objectively, and thus our learning process can be stationary from the start. Therefore any SCE, or NE, can be the outcome of the class of learning processes that we consider. ■

It is easy to check the "tightness" of Theorem 1. If of any of the conditions on the sequence $\{s(l)\}_{l=1}^{\infty}$ in Theorem 1 is dropped, the sequence will fail to converge, or its limit value will not be an SCE. In particular, the following example highlights the need for the rationality constraint on deviations.

Example 1 Consider the game tree of Figure 1 and assume that players *observe exactly plays*, i.e., (7) holds with equality for all l. (The first component of any payoff vector refers to player 1.)

Figure 1

In every iteration of the process, player 1 chooses his best response to his conjecture of player 2's strategy. When the conjecture is $(\widetilde{r}, l)$, however, player 1 is faced with indifference between L and R. We stipulate that he chooses L (or, R) if, at his last conjecture of $(\widetilde{r}, l)$ in the process, he had chosen R (or, L). At the start, player 1 conjectures $(\widetilde{r}, l)$ and chooses L.

Player 2 flagrantly violates the rationality constraint by rendering himself worse-off at irrelevant nodes, though choosing a best response all the time.

Thus he chooses $(\tilde{r}, r)$ when his conjecture is L, and $\left(\tilde{l}, l\right)$ when his conjecture is R. He starts with conjecturing R.

It is easy to see that the following cycle repeats forever in our sequence (since period 5≡period 1):

period	conj. of 1	strategy of 1	conj. of 2	strategy of 2	outcome of the game[7]
1	$(\tilde{r}, l)$	L	R	$\left(\tilde{l}, l\right)$	$(5, -1)$
2	$\left(\tilde{l}, l\right)$	L	L	$(\tilde{r}, r)$	$(3, 1)$
3	$(\tilde{r}, l)$	R	L	$(\tilde{r}, r)$	$(4, -2)$
4	$(\tilde{r}, r)$	R	R	$\left(\tilde{l}, l\right)$	$(3, 2)$
5	$(\tilde{r}, l)$	L	R	$\left(\tilde{l}, l\right)$	$(5, -1)$

Notice that the fluctuating choices of player 1, when he is indifferent, are crucial to this example; without it, the violation of the rationality constraint by player 2 would not even be revealed through plays of the game. Indeed, if there are no chance moves and if payoffs at terminal nodes are all distinct, then - even without the rationality constraint - the outcomes of our sequence converge to an NE outcome, as we show in Theorem 2 below. To allow for chance moves, we extend the condition on payoffs as follows: for every pair of outcomes $\overline{\prod}, \widetilde{\prod}$ in the game, and each $i \in N$,

$$u^i(\overline{\prod}) = u^i(\widetilde{\prod}) \Rightarrow \overline{\prod} = \widetilde{\prod}. \tag{11}$$

The condition (11) is clearly generic, i.e., it holds for an open and full measure set of payoff vectors in $R^{N \times T}$, where T ≡set of terminal nodes. Notice that, if (11) holds, then we also have, for every subgame Γ, every pair of outcomes $\overline{\prod}, \widetilde{\prod}$ in the game, and each $i \in N$,

$$u^i_\Gamma(\overline{\prod}_\Gamma) = u^i_\Gamma(\widetilde{\prod}_\Gamma) \Rightarrow \overline{\prod}_\Gamma = \widetilde{\prod}_\Gamma,$$

where $\overline{\prod}_\Gamma$ and $\widetilde{\prod}_\Gamma$ are the restrictions of $\overline{\prod}$ and $\widetilde{\prod}$ to Γ, and (recall) u^i_Γ is i's payoff function in Γ.

Theorem 2 Assume that the game satisfies condition (11). Let $\{s(l)\}_{l=1}^{\infty}$ be a learning sequence in which players observe only plays and $s^i_i(l) \in$

[9] Since there are no chance moves in the game, any outcome consists of a single play, which is identified with the payoff vector in its terminal node.

$BR\left(s_i^{-i}(l)\right)$ for all $i \in N$ and all l. Then the sequence of outcomes $\{\prod(s_1^1(l), ..., s_n^n(l))\}_{l=1}^{\infty}$ converges in finite time, and its limit values are NE outcomes.

First we will derive the fact that the sequence of outcomes becomes stationary.

Lemma 3 *Assume that the game satisfies condition (11). If* $\{s(l)\}_{l=1}^{\infty}$ *is a learning sequence in which players update objectively and observe at most plays (i.e., (7) holds for every* $i \in N$ *and all* l *with the set inclusion "* $\subset$ *"), and* $s_i^i(l) \in BR\left(s_i^{-i}(l)\right)$ *for all* $i \in N$ *and all* l*, then the sequence of outcomes* $\{\prod(s_1^1(l), ..., s_n^n(l))\}_{l=1}^{\infty}$ *becomes stationary.*

Proof. As in Lemma 2, this will be by induction on the number k of arcs in the game tree. If $k = 0$ the claim is vacuously true. Assume it to be true whenever $k \leq q$ and take the case $k = q + 1$.

Suppose that the root α^* is in P^i for some $i \in N$. We show first that the move chosen by player i at α^* becomes stationary after a certain iteration. Consider subgames $\Gamma_1, ..., \Gamma_m$ $(m \geq 2)$ which start at the ends of arcs issuing-out of α^*, and identify the moves of i at α^* with $1, ..., m$. For every iteration l, define a vector $v(l) = (v_1(l), ..., v_m(l)) \in R^m$, where $v_h(l)$ is the maximal payoff that i can achieve in the subgame Γ_h given his conjectures at iteration l. Consider the iterations $l_1 < l_2 < ...$ at which i changes his move at α^*. Let h_r be the fixed move chosen by i at α^* during all the iterations $l_r \leq l < l_{r+1}$; note that

$$v_{h_r}(l) = \max_{h=1}^{m} v_h(l) \tag{12}$$

for all such l. Since i observes at most plays, all but the h_r^{th} coordinate of $v(l)$ stay fixed during $l_r \leq l \leq l_{r+1}$, i.e.,

$$v_h(l) = v_h(l_{r+1}) \tag{13}$$

for $l_r \leq l \leq l_{r+1}$ and $h \neq h_r$.

Note that $v_{h_r}(l_{r+1}) < v_{h_r}(l_r)$, since otherwise (11), (12), and (13) would imply that $v_{h_r}(l_{r+1}) > \max_{h \neq h_r} v_h(l_{r+1})$, and thus a best response of i at iteration l_{r+1} would entail the choice of h_r at α^*, contrary to the definition of l_{r+1}. Thus, for any r, $v(l_{r+1}) \leq v(l_r)$ and the inequality is strict for one coordinate. This proves that the sequence $(l_1, l_2, ...)$ is of finite length, since the set of all possible payoffs to i (from outcomes in the game) is finite, and thus the sequence $\{s_i^i(l)(\alpha^*)\}_{l=1}^{\infty}$ becomes stationary.

Suppose now that, contrary to the assertion of the lemma, the sequence $\{\prod(s_1^1(l),...,s_n^n(l))\}_{l=1}^{\infty}$ does not become stationary. Since $\{s_i^i(l)(\alpha^*)\}_{l=1}^{\infty}$ becomes stationary, there exist L and j such that

$$L = L\left(\left\{s_i^i(l)(\alpha^*)\right\}_{l=1}^{\infty}\right) \text{ and } s_i^i(L)(\alpha^*) = j \tag{14}$$

with positive probability, and $\{\prod(s_1^1(l),...,s_n^n(l))\}_{l=L}^{\infty}$ does not become stationary conditional on (14). Note that, under (14), $\prod(s_1^1(l),...,s_n^n(l))$ is the union of α^* and $\prod_{\Gamma_j}(s_1^1(l),...,s_n^n(l))$ for $l \geq L$, and thus the sequence $\left\{\prod_{\Gamma_j}(s_1^1(l),...,s_n^n(l))\right\}_{l=L}^{\infty}$ also does not become stationary. However, also conditional on (14), the restriction of $\{s(l)\}_{l=L}^{\infty}$ to Γ_j satisfies the assumptions of the lemma in this subgame, and thus (by the induction hypothesis) the sequence $\left\{\prod_{\Gamma_j}(s_1^1(l),...,s_n^n(l))\right\}_{l=L}^{\infty}$ does becomes stationary, a contradiction. Therefore the sequence $\{\prod(s_1^1(l),...,s_n^n(l))\}_{l=1}^{\infty}$ becomes stationary.

Finally, suppose that $\alpha^* \in P^0$. It is easy to see that restrictions of the sequence $\{s(l)\}_{l=1}^{\infty}$ to subgames $\Gamma_1,...,\Gamma_m$, starting at the ends of arcs issuing-out of α^*, satisfy the induction hypothesis, and thus outcomes of the sequence $\{s(l)\}_{l=1}^{\infty}$, restricted to $\Gamma_1,...,\Gamma_m$, become stationary. Since the union of $\prod_{\Gamma_h}(s_1^1(l),...,s_n^n(l))$ over $h = 1,...,m$ and the root α^* is $\prod(s_1^1(l),...,s_n^n(l))$, it follows that the latter becomes stationary as well. ■

Proof of Theorem 2. The sequence $\{\prod(s_1^1(l),...,s_n^n(l))\}_{l=1}^{\infty}$ of outcomes becomes stationary by Lemma 3. Then so does the sequence $\left\{\left(s_1^{-1}(l),...,s_n^{-n}(l)\right)\right\}_{l=1}^{\infty}$ of conjectures, since players update objectively and observe only plays. Let L and $\overline{s} \in S^N$ be such that, with positive probability, from the L^{th} iteration onwards $\prod(\overline{s}_1^1(l),...,\overline{s}_n^n(l))$ is the constant value of the first sequence, and $\left(\overline{s}_1^{-1}(l),...,\overline{s}_n^{-n}(l)\right)$ is the constant value of the second sequence. As in the proof of Theorem 1 one can now show that $\overline{s}$ is an SCE, and its outcome is an NE outcome. Thus the outcomes of $\{s(l)\}_{l=1}^{\infty}$ converge in finite time to an NE outcome. ■

Remark 3 (Existence of NE without Backward Induction) It is worth noting that the convergence in Theorem 2 is to an NE outcome, though none of the NE that sustain it may be subgame-perfect[10]. Thus we have provided a new proof of the existence of pure strategy NE, which is different from the standard backward induction proof. Indeed, first consider generic

[10] An NE is *subgame-perfect* if its restriction to any subgame constitutes an NE of that subgame.

payoffs with no ties, and suppose players observe only plays and choose best responses at every iteration. Then, by Theorem 2, our process yields an NE outcome *without* the rationality constraint on unilateral deviations. To complete the proof, take limits on payoffs when there are ties.... Note that the dropped constraint had the last remaining vestiges of "subgame perfection".

Of course, by further fine-tuning a best response strategy and requiring it to constitute a best-response in every subgame, the proof of Theorem 1 also shows the existence of a subgame-perfect NE. ■

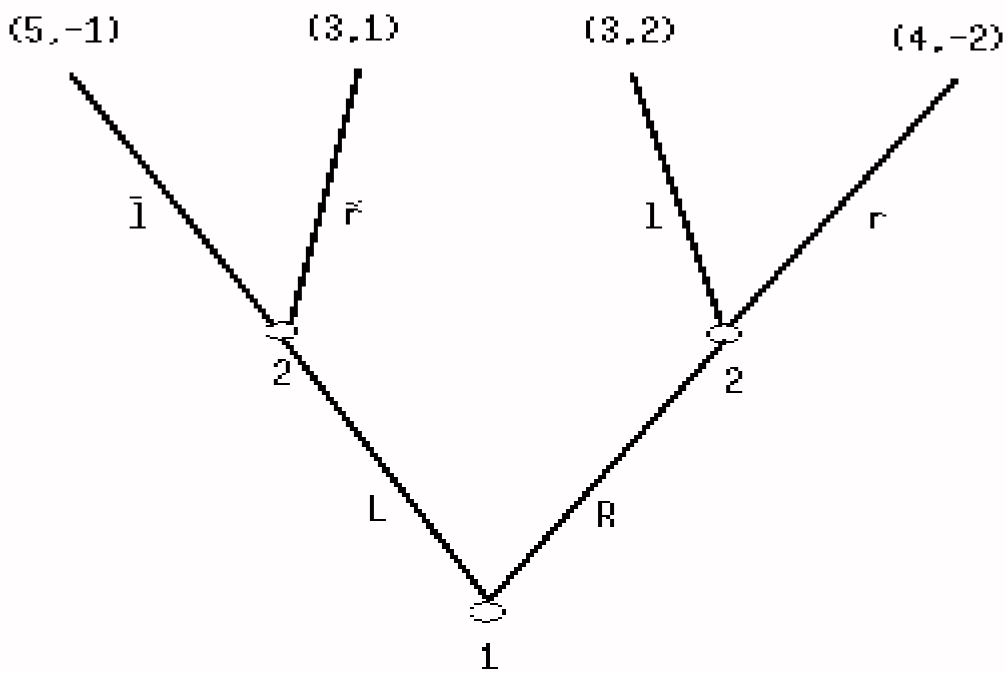
Remark 4 (The Empirical Distribution of Plays) If the outcome of the game becomes stationary, then so do players' strategic choices at all their positions in the outcome. Since chance picks its moves independently across all its positions and also across iterations of the game, it is evident that, in the setting of either Theorem 1 or Theorem 2, the empirical distribution of plays (selected from the outcome) will converge to an NE distribution. ■

Remark 5 (Absence of Chance Moves) Suppose there are no chance moves in the game. Further suppose that players follow a deterministic rule in observing others strategies and deviating to their responses. The canonical case we have in mind is that they observe exactly plays or exactly strategies at every iteration, and deviate to a best reply that entails the minimum number of changes of moves (with a recipe for breaking ties, in the degenerate case when payoffs are not distinct across outcomes, e.g., via an a priori ranking of their strategies). Then, in the setting of either Theorem 1 or Theorem 2, the number of iterations needed to achieve stationarity is completely determined by the initial choice of strategies and conjectures. This is obvious from our analysis, but seems to us worthy of record. ■

References

1. Dubey, P. (1981) "Best Response with Perfect Information," working paper # 81-77, International Institute for Applied Systems Analysis, Austria.

2. Fudenberg, D., and D. Kreps (1995) "Learning in Extensive Form Games: I. Self-Confirming Equilibrium," *Games and Economic Behavior* **8,** pp. 20 - 55.


(5,-1)
(3,1)
(3,2)
(4,-2)
l̄
r̄
l
r
2
2
L
R
1