JOURNAL OF ECONOMIC THEORY 57, 363–391 (1992)

Evolutionary Stability in Asymmetric Games*

LARRY SAMUELSON

Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706

AND

JIANBO ZHANG

Department of Economics, University of Kansas, Lawrence, Kansas 66044

Received November 2, 1990; revised October 1, 1991

We examine dynamic models of evolutionary selection processes on asymmetric two-player games. Conditions are established under which dynamic selection processes will yield outcomes that respect iterated strict dominance. The addition of a stability requirement ensures that outcomes will be Nash equilibria. However, we find that stable outcomes need not respect weak dominance, and hence need not yield perfect equilibria. We conclude that evolutionary arguments readily motivate such equilibrium oncepts as rationalizability and Nash equilibrium, but appear to provide little basis for even such simple refinements of Nash equilibrium as the recommendation that dominated strategies not be played. *Journal of Economic Literature* Classification Numbers: C70, C72.

I. INTRODUCTION

This paper investigates the ability of evolutionary arguments to provide foundations for common game theoretic solution concepts. We find that evolutionary arguments readily motivate such equilibrium concepts as rationalizability and Nash equilibrium, and provide some grounds, based on stability considerations, for choosing between them. At the same time, evolutionary arguments appear to provide little basis for even such simple refinements of Nash equilibrium as the recommendation that dominated

* We are grateful to an anonymous referee and associate editor for helpful comments and to the CentER for Economic Research at Tilburg University and the Department of Economics at the University of Bonn for the hospitality and support extended to the first author. Financial support from the NWO (Netherlands; Grant B46-276) and DFG (Germany; Sonderforschungsbereich 303) are gratefully acknowledged.

0022-0531/92 $5.00

strategies not be played. A belief in evolutionary foundations for game theory may thus lead to the avoidance of strong equilibrium refinements and to consideration of outcomes in which weakly dominated strategies are played. It appears as if this conclusion can be avoided only if some additional structure can be placed on the evolutionary process, perhaps with the help of explicit models of how players learn.

We examine two-player games. Section III establishes notation and definitions. Section IV shows that if the evolutionary process satisfies a condition called monotonicity, then any strategy which fails to survive the iterated elimination of pure strategies which are strictly dominated by other pure strategies will be eliminated from the population. This holds regardless of whether the process converges. Section IV then shows that if the evolutionary system satisfies a stronger condition called aggregate monotonicity, then any strategy which fails the iterated elimination of strictly dominated strategies, or is not rationalizable, will be eliminated. Aggregate monotonic systems are shown to include the replicator dynamics and simple transformations of the replicator dynamics.

Section V examines asymptotically stable equilibria. We find that asymptotically stable equilibria are "nearly" strict Nash equilibria. Section V also presents a theorem showing that if an evolutionary path converges "quickly" under the replicator dynamics, then its outcome must be a quasi-strict equilibrium. Each of these is a potentially troubling result; the first because a theory which simply recommends strict Nash equilibria is of relatively little use and the second because it suggests that we must either work with the relatively strong equilibrium concept of quasi-strictness or work with systems in which the limiting outcome of the evolutionary path is a poor approximation of behavior along the path.

Section VI attempts to find a middle ground between the rationalizable outcomes of Section IV and the strict Nash equilibria of Section V by relaxing the requirement of asymptotic stability to simply stability. We find that stable points of a monotonic evolutionary process must be Nash equilibria. However, Section VI also shows by example that stable points need not be perfect equilibria or limit *ESS*s. This is an interesting finding in light of the apparent similarity between the trembles involved in the definition of a perfect equilibrium or limit *ESS* and the (possibly small) proportions of the population in an evolutionary system which play each of the strategies of a game.

It is not clear that the stability notion employed in Section VI is satisfactory. Unlike the case of an asymptotically stable equilibrium, it appears as if the cumulative effect of successive mutations could be to lead the system away from a stable outcome, causing "stable" points to exhibit very little stability. Section VI addresses this concern by examining stable outcomes in a model which explicitly allows for mutation. We again find that stable outcomes must be Nash equilibria but need not be perfect. This prompts our conclusion that evolutionary considerations motivate equilibrium concepts such as rationalizability and Nash equilibrium, with stability considerations invoked to choose between the two, but do not motivate stronger solution concepts.

Section VIII shows that stable outcomes will respect weak dominance and hence be perfect in a model in which trembles appear not as mutations but as players making mistakes when playing strategies. This suggests that the choice trembles which appear in the definition of trembling hand perfection differ in important ways from mutations. Section IX concludes.

II. Related Literature

Three types of previous work are related to our results. First, Brown [3] and Robinson [16] used the concept of fictitious play to construct a dynamic process in which players adjust their strategies by choosing best responses to the mixed strategies implcitly defined by accumulated previous play. If the expected payoffs from these accumulated mixed strategies converge (which occurs on zero-sum games), then the limiting payoffs correspond to a Nash equilibrium. This is perhaps the first of many theorems of the type "convergence or stability implies Nash." Interest in fictitious play waned after Shapley [22] presented an example in which the process does not converge. Our work differs both in considering a different class of dynamic processes and in establishing some results for processes which do not converge.

Second, biologists and game theorists have developed an extensive theory of evolutionary games. The standard solution concept in evolutionary game thoery, introduced by Maynard Smith [11] and Maynard Smith and Price [12], is that of an *ESS*, or evolutionarily stable strategy. Attention has also been focused on dynamic evolutionary models, with particular attention devoted to the replicator dynamics, borrowed from biology. A stable outcome under the replicator dynamics must be a Nash equilibrium, producing a second "stability implies Nash" theorem. If an outcome is an *ESS*, then it is asymptotically stable under the replicator dynamics. The converse holds under some (e.g., two player games with two strategies per person) but not all circumstances ([24, section 9.4]). *ESS*s also exhibit considerable structure. For example, an *ESS* is a symmetrically strictly perfect and proper equilibrium [24, section 9.3]. At the same time, Ref. [24, example 9.4.3] shows that there exist outcomes which are stable under the replicator dynamics but not perfect (and hence not an *ESS*).

Some work has been done on extending these results beyond the replicator dynamics. Nachbar [14] shows that if a dynamic process satisfies the condition we call monotonicity below, then a limiting outcome of a converging process must be a Nash equilibrium. In addition, Nachbar shows that if a game is such that the iterated elimination of pure strategies which are strictly dominated by pure strategies yields a unique outcome, then a monotonic adjustment process will converge.

Third, the results cited above apply only to symmetric games. Our work differs in its primary emphasis on asymmetric games. A body of literature has appeared to address asymmetric games. Selten [18] shows that an *ESS* in an asymmetric game must be a strict Nash equilibrium. While it is well known that a strict Nash equilibrium exhibits virtually all desirable properties, a theory which confines attention to strict Nash equilibria is too restrictive to be useful. More importantly, it is not clear how an *ESS* corresponds to the limiting outcomes of dynamic evolutionary processes in asymmetric games.

Selten [19] and [20] offers the more general limit *ESS* (the limit of *ESS*s in games with perturbed strategy choices) as an alternative to the *ESS*. Any *ESS* is a limit *ESS* and limit *ESS*s exist in some games with no *ESS*. The relationship between the limit *ESS* concept and the limiting outcomes of dynamic evolutionary processes in asymmetric games is again unknown.

Progress in the study of evolutionary arguments in asymmetric games now requires examination of the links between equilibrium concepts and the outcomes of dynamic evolutionary processes. Friedman considers general monotonic adjustment processes on asymmetric games, establishing four basic results [7, Propositions 4–6 and Counterexamples 1–3]: Every Nash equilibrium is a rest point of the dynamic system (i.e., a point with the property that if the dynamic system begins at the point then it will not move away). Asymptotically stable outcomes are Nash equilibria. An *ESS* need not be asymptotically stable for all monotonic adjustment processes (in symmetric or asymmetric games). Finally, a regular *ESS* is asymptotically stable for a class of monotonic adjustment processes.

This paper continues the investigation of dynamic evolutionary processes on asymmetric games. Our finding that monotonic adjustment processes on asymmetric games will respect the iterated elimination of pure strategies which are strictly dominated by other pure strategies generalizes Nachbar's result that monotonic processes will converge in symmetric games in which the iterated elimination of pure strategies strictly dominated by other pure strategies yields a singleton outcome. Our finding that stable points must be Nash equilibria is another of the many "stability implies Nash" theorems. Our finding that asymptotically stable outcomes must be essentially strict Nash equilibria strengthens Friedman's result that asymptotically stable outcomes must be Nash equilibria.

Technically, the specifications of the dynamic processes, the definition of stable outcomes, and some of their properties (such as the findings that stable outcomes must be Nash equilibria but may fail to be perfect) generalize from symmetric to asymmetric games in a straightforward way. However, symmetric and asymmetric games are distinguished in two respects. First, the *ESS* concept does not appear to be useful in the latter (nor do obvious alternatives exist). As mentioned above, the *ESS* concept in asymmetric games is equivalent to strict Nash equilibrium. Much of the literature on equilibrium refinements is motivated by a belief that there are interesting equilibria that are not strict; and that one cannot be content with a restriction to strict equilibria. More importantly, one easily finds games (such as (27) below) that have no strict Nash equilibria and hence have no *ESS* but in which evolutionary arguments still appear to have interesting implications. This prompts us to appeal directly to dynamic arguments as the primary form of analysis, unlike the case of symmetric games where considerable work has been done by simply using the *ESS* concept.

Second, asymptotical stability in asymmetric games is also essentially equivalent to strict Nash equilibrium (this is made precise in Theorem 4 below). For reasons analogous to the case of the *ESS* concept, asymptotic stability is then less useful in asymmetric than in symmetric games, forcing us to look at other stability notions. One can thus easily extend techniques from symmetric to asymmetric games, but finds that the results are less useful in the asymmetric case and is prompted to look for alternative techniques. Sections VI–VIII below are in this spirit.

Finally, it is useful to note that attempts have recently been made to construct decision theoretic foundations for solution concepts. Tan and Werlang [23] show that the common knowledge of rationality implies that players will choose rationalizable strategies and that common knowledge of rationality coupled with a consistency condition on beliefs yields Nash equilibrium. Our results provide an alternative, evolutionary foundation for these concepts, with aggregate monotonicity of the learning process implying rationalizability and monotonicity plus stability implying Nash equilibrium. For those who are troubled by the strength of assumptions such as the common knowledge of rationality or the consistency of beliefs, this may be an appealing alternative.

III. Games and Dynamic Processes

Let $(\{1, 2\}, I, J, \pi_1, \pi_2)$ be a two-player normal-form game. The players are denoted 1 and 2. I and J are finite sets of pure strategies for players 1 and 2 with generic elements i and j. $\pi_1 : I \times J \to \mathbb{R}$ and $\pi_2 : I \times J \to \mathbb{R}$ are payoff functions. Let the cardinalities of I and J be denoted n_1 and n_2. Let

x and y be elements of S^{n_1} and S^{n_2}, where S^{n_1} is the (n_1-1)-dimension simplex. x and y are interpreted as vectors identifying the proportions of populations 1 and 2 playing each of the pure strategies in I and J.

We will abuse notation somewhat by letting the expected payoff

$$\sum_{i\in I}\sum_{j\in J}\pi_1(i,j)x_i y_j$$

be written simply as $\pi_1(x, y)$, with $\pi_2(x, y)$ being analogous and with $\pi_1(i, y)$ (for example) being the special case in which player 1 plays pure strategy i. In addition, let A and B be $n_1 \times n_2$ matrices of player 1 and player 2 payoffs, where a_{ij} is the payoff to player 1 if player 1 plays his ith strategy and 2 plays her jth strategy and where b_{ij} is analogous for player 2. Then

$$\pi_1(i, y) = e_i^T A y$$

$$\pi_1(x, y) = x^T A y$$

$$\pi_2(x, j) = x^T B e_j$$

$$\pi_2(x, y) = x^T B y,$$

where e_i is a vector of zeros except for a 1 in the ith place and T denotes transposition.

We now introduce the concept of a selection process. Intuitively, we think of there being two large (formally, infinite) populations of players 1 and 2 who are repeatedly, randomly matched to play single repetitions of the game G. Each player plays a pure strategy, with the distribution of strategies among players being given by x and y. Over time, the proportions of the populations playing the various pure strategies adjust in response to payoff differences. We assume that these changes in population proportions can be described in the following way:

DEFINITION 1. Let $f: S^{n_1} \times S^{n_2} \to \mathbb{R}^{n_1}$ and $g: S^{n_1} \times S^{n_2} \to \mathbb{R}^{n_2}$. Then the system

$$\dot{x}_i = f_i(x, y) \qquad i = 1, \ldots, n_1$$

$$\dot{y}_j = g_j(x, y) \qquad j = 1, \ldots, n_2$$

is a *selection dynamic* if it satisfies, for all $(x, y) \in S^{n_1} \times S^{n_2}$,

(1.1) f and g are Lipschitz continuous, i.e., $\exists k \in \mathbb{R}_+$ s.t. $\forall x, x' \in S^{n_1}$, $\forall y, y' \in S^{n_2}$,

$$\max\{|f(x, y) - f(x', y')|, |g(x, y) - g(x', y')|\} \leqslant k\, |(x, y) - (x', y')|$$

(1.2) $\sum_{i=1}^{n_1} f_i(x, y) = 0 = \sum_{j=1}^{n_2} g_j(x, y)$

(1.3) $\forall x \in S^{n_1}$, $x_i = 0 \Rightarrow f_i(x, y) \geqslant 0$
$\forall y \in S^{n_2}$, $y_j = 0 \Rightarrow g_j(x, y) \geqslant 0$.

We will write $x(t)$ and $y(t)$ to denote the time-t values of x and y, but will suppress t whenever possible.[1] Note that f_i/x_i is the growth rate of the proportion of population 1 playing strategy i. It is convenient to define $f/x \equiv (f_1/x_1, ..., f_{n_1}/x_{n_1})$. The Lipschitz condition contained in (1.1) ensures that for any initial condition, the selection dynamic has a unique solution.

Ideally, one would like to build the selection dynamic up from a precise theory of how individual players switch strategies. Unfortunately, it appears as if such a theory must include a number of ad hoc elements. We attempt to avoid arbitrary choices in the construction of a theory of learning and individual behavior of placing assumptions directly on the selection dynamic. We hope that these properties are general enough to include the dynamic processes produced by a variety of selection or learning theories.[2]

We will make use of the following two properties:

DEFINITION 2. f and g yield a *regular* selection dynamic if (1.1)–(1.3) hold and the following limits exist and are finite:

$$\frac{f_i}{0} \equiv \lim_{x_i \to 0} \frac{f_i}{x_i}$$

$$\frac{g_j}{0} \equiv \lim_{y_j \to 0} \frac{g_j}{y_j}.$$

DEFINITION 3. f is *monotonic* if, for $i, i' \in I$,

$$\pi_1(i, y) > (=)\, \pi_1(i', y) \Rightarrow \frac{f_i(x, y)}{x_i} > (=) \frac{f_{i'}(x, y)}{x_{i'}}, \tag{1}$$

and f is *aggregate monotonic* if, for all $p, p' \in S^{n_1}$,

$$\pi_1(p, y) > \pi_1(p', y) \Rightarrow \sum_{i=1}^{n_1} (p_i - p_i') \frac{f_i(x, y)}{x_i} > 0. \tag{2}$$

Regularity causes the growth rates f/x and g/y, which are continuous on the interior of $S^{n_1} \times S^{n_2}$ as a result of Definition 1, to be continuous on all

[1] Following standard practice, the aggregate selection dynamic is taken to be deterministic even though the players are randomly matched by a stochastic process. Boylan [2] rigorously investigates the stochastic foundations of the deterministic replicator dynamics.

[2] Constructing models of how individuals in an evolutionary system make choices and learn is an important area for research. Fudenberg and Kreps [8] examine learning in extensive form games of perfect information which have the property that in any course of play, each agent moves at most once. See also [4, 10, 13, 21].

of $S^{n_1} \times S^{n_2}$. This has the important implication of ensuring that if $x(0)$ and $y(0)$ are strictly positive, as we will assume, then $x(t)$ and $y(t)$ are strictly positive for all t. This allows us to avoid problems which arise because of extinction. Monotonicity requires that if pure strategy i receives a higher expected payoff than i' and if x_i and $x_{i'}$ are both positive, then x_i grows faster than $x_{i'}$.[3]

Regularity and monotonicity together have the key effect of ensuring that if one takes a sequence of values of x_i and $x_{i'}$ on which the expected profit of x_i is higher than that of $x_{i'}$, and with the expected-profit difference bounded away from zero, then the difference in growth rates of x_i and $x_{i'}$ does not deteriorate to zero along this sequence even if one of x_i or $x_{i'}$ approaches zero.

The interpretation of monotonicity is that, on average, players are able to switch form worse to better strategies. Aggregate monotonicity requires that if the population 2 vector y is such that a mixed strategy p would receive a higher payoff against y than would p', then the system grows faster toward p than toward p'. One readily verifies that an aggregate monotonic system is monotonic but the converse does not hold. Note that while mixed strategies are involved in the definition of aggregate monotonicity, we maintain the convention that individual agants play pure strategies.

The biology literature makes frequent use of a particular selection dynamic which we adapt to our asymmetric model:

DEFINITION 4. The selection dynamic (f, g) is the *replicator dynamics*, denoted (f^*, g^*), if

$$\frac{f_i(x, y)}{x_i} = \pi_1(i, y) - \sum_{k=1}^{n_1} x_k \pi_1(k, y), \tag{3}$$

with an analogous specification for $g_j(x, y)$.

IV. MONOTONICITY AND STRICT ADMISSIBILITY

We now consider the following question. What are the implications of assuming that a regular, monotonic selection process governs the play of a game? Our first results will examine strictly dominated strategies, and we accordingly begin with some definitions.

[3] Note that monotonicity allows i to receive a higher expected payoff than i' *without* x_i growing faster than $x_{i'}$ if one of x_i or $x_{i'}$ is zero. In the case of the replicator dynamics, for example, $x_i = 0$ implies $f_i(x, y) = 0$, so that (1) can hold with $x_i = 0$ and without x_i growing faster than $x_{i'}$.

DEFINITION 5. Strategy $i \in I$ is *strictly dominated* if there exists $x \in S^{n_1}$ such that

$$\pi_1(x, y) > \pi_1(i, y) \qquad \forall y \in S^{n_2}.$$

Let $D_1(X_1, X_2)$ be the set of pure strategies in $X_1 \subseteq I$ that are not strictly dominated by any pure strategies in I given that player 2 chooses strategies from $X_2 \subseteq J$. Let $\bar{D}_1(M_1, M_2)$ be the set of mixed strategies in $M_1 \subseteq S^{n_1}$ that are not strictly dominated by any strategies in M_1 given that player 2 chooses from $M_2 \subseteq S^{n_2}$. Similar definitions apply to player 2.

DEFINITION 6. The strategy $i \in I$ survives *pure strict iterated admissibility* if there exist sequences of the form $I = X_{10}, X_{11}, ..., X_{1T}$ and $J = X_{20}, X_{21}, ..., X_{2T}$, where $X_{1n+1} = D_1(X_{1n}, X_{2n})$, and $X_{2n+1} = D_2(X_{1n}, X_{2n})$ for $n = 1, ..., T-1$, with $X_{1T} = D_1(X_{1T}, X_{2T})$ and $X_{2T} = D_2(X_{1T}, X_{2T})$ and with $i \in X_{1T}$. The strategy $x \in S^{n_1}$ survives *strict iterated admissibility* if D_i is replaced by $\bar{D}_i$ $(i = 1, 2)$ in this definition.

Pearce [15] shows that the set of strategies which survives strict iterated admissibility is nonempty and coincides with the set of rationalizable strategies in two-player games.

We now show that a monotonic selection process will eliminate strategies which do not survive pure strict admissibility. We establish this result for all evolutionary systems, regardless of whether they converge.

THEOREM 1. *Let (f, g) be a monotonic, regular selection dynamic. Suppose $i \in I$ does not survive pure strict iterated admissibility. Then for any evolutionary path $(x(t), y(t))$ with $(x(0), y(0))$ completely mixed, we have*

$$\lim_{t \to \infty} x_i(t) = 0.$$

Proof. Let $I_0 \subseteq I$ be the set of player-one strategies which do not survive pure strict iterated admissibility and are not eliminated in the limit by the selection process. Let $J_0 \subseteq J$ be similarly defined. Suppose the theorem fails, so that $I_0 \cup J_0 \neq \varnothing$. For all $l \in I_0 \cup J_0$, let $k(l)$ be such that $l \in X_{1k(l)} \backslash X_{1k(l)+1}$ (if $l \in I_0$) or $X_{2k(l)} \backslash X_{2k(l)+1}$ (if $l \in J_0$). Let l_0 be the minimizer of $k(l)$ on $I_0 \cup J_0$ and let $k = k(l_0)$. Without loss of generality, we can assume $l_0 \in I_0$ and can rename l_0 to be i_0. Then there exists $i_1 \in I$ such that $\pi_1(i_0, j) < \pi_1(i_1, j)$ for all $j \in X_{2k}$. Since k minimizes $k(l)$, we have $\lim y_j(t) = 0$ for all $j \notin X_{2k}$. Then

$$\pi_1(i_0, y(t)) - \pi_1(i_1, y(t)) = \sum_{j \in X_{2k}} (\pi_1(i_0, j) - \pi_1(i_1, j))\, y_j(t) + \sum_{j \notin X_{2k}} (\pi_1(i_0, j) - \pi_1(i_1, j))\, y_j(t). \qquad (4)$$

As $t \to \infty$, the second part of (4) goes to zero, while the first approaches a negative number. Therefore, there exist $\varepsilon > 0$ and $T > 0$ such that

$$\pi_1(i_0, y(t)) - \pi_1(i_1, y(t)) < -\varepsilon \qquad \forall t > T. \tag{5}$$

By monotonicity and regularity, we then have, for some $\delta > 0$,

$$\frac{\dot{x}_{i_0}(t)}{x_{i_0}(t)} - \frac{\dot{x}_{i_1}(t)}{x_{i_1}(t)} < -\delta \qquad \forall t > T, \tag{6}$$

and hence, $\forall t > T$,

$$\frac{x_{i_0}(t)}{x_{i_1}(t)} \leqslant \frac{x_{i_0}(T)}{x_{i_1}(T)} e^{-\delta(t-T)} \to 0. \tag{7}$$

Therefore, $\lim_{t \to \infty} x_{i_0}(t) = 0$, contradicting the definition of i_0. Thus $I_0 \cup J_0 = \varnothing$ and the theorem holds. ∎

As a special case of this theorem, we obtain Nachbar's [14] result that if pure strict iterated admissibility in a symmetric game removes all but a single strategy for each player, then a monotonic adjustment process will converge to these strategies. In particular, we find that a monotonic selection process always respects the outcome of pure strict iterated admissibility, regardless of the number of strategies which survive such a procedure. If only one strategy remains, then respecting pure strict iterated admissiblity implies convergence.[4]

Pure strict iterated admissibility falls short of rationalizability in two ways. First, strategies may fail to be eliminated which are dominated by mixed strategies. Second, dominated mixed strategies may fail to be eliminated.

We can extend the results of Theorem 1 to mixed strategies, i.e., to strict iterated admissibility, if we strengthen monotonicity to aggregate monotonicity:

THEOREM 2. *Let (f, g) be a regular, aggregate monotonic selection dynamic. Let $x' \in S^{n_1}$ fail strict iterated admissibility. Then for any evolutionary path $(x(t), y(t))$ with $(x(0), y(0))$ completely mixed, there exists a function $\varepsilon(t)$ with $\lim_{t \to 0} \varepsilon(t) = 0$ such that for every t, there exists a pure strategy $i(t)$ in the carrier of x' such that $x_{i(t)}(t) \leqslant \varepsilon(t)$.*

Proof. Let p be a mixed strategy for player 1 (without loss of generality) which fails strict iterated admissibility (SIA), and let $(x(t), y(t))$

[4] Milgrom and Roberts [13], restricting attention to the case of supermodular games, obtain similar results (among other findings).

be an evolutionary path with $(x(0), y(0))$ completely mixed. It is sufficient to show

$$\lim_{t\to\infty} \prod_{i=1}^{n_1} x_i^{p_i}(t) = 0. \tag{8}$$

Suppose (8) does not hold. Define

$$A_1 = \left\{ u \in S^{n_1} \mid u \text{ fails SIA and } \lim_{t\to\infty} \prod_{i=1}^{n_1} (x_i(t))^{u_i} \neq 0 \right\}$$

$$A_2 = \left\{ v \in S^{n_2} \mid v \text{ fails SIA and } \lim_{t\to\infty} \prod_{j=1}^{n_2} (y_j(t))^{v_j} \neq 0 \right\}.$$

Because (8) fails for p, $A_1 \cup A_2$ is nonempty. For $a' \in A_1 \cup A_2$, let $k(a')$ be such that $a' \in M_{1k(a')} \backslash M_{1k(a')+1}$ or $M_{2k(a')} \backslash M_{2k(a')+1}$, depending on whether $a' \in S^{n_1}$ or S^{n_2}. Let a be a minimizer of k on $A_1 \cup A_2$. Without loss of generality, assume $a \in A_1$. Then since $a \in M_{1k} \backslash M_{1k+1}$, there exists $b \in M_{1k}$ such that b strictly dominates a for all y in M_{2k}, i.e.,

$$\pi_1(a, y) - \pi_1(b, y) < 0 \qquad \forall y \in M_{2k}. \tag{9}$$

Let Y consist of all those $y \in S^{n_2}$ with the property that $y_j > 0$ only if $j \in M_{2k}$. Then

$$\pi_1(a, y) - \pi_1(b, y) < 0 \qquad \forall y \in Y.$$

By aggregate monotonicity, we have

$$\sum_i (a_i - b_i) \frac{f_i(x, y)}{x_i} < 0 \qquad \forall y \in Y, x \in S^{n_1}. \tag{10}$$

Because Y is a closed subset of S^{n_2}, regularity ensures that there exists $\varepsilon > 0$ such that

$$\sum_i (a_i - b_i) \frac{f_i(x, y)}{x_i} < -\varepsilon < 0 \qquad \forall y \in Y, x \in S^{n_1}. \tag{11}$$

Now given $y(t)$, define $\hat{y}(t)$ by

$$\hat{y}_j(t) = \begin{cases} \iota_t\, y_j(t) & \text{if } j \in M_{2k} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where ι_t is chosen so that $\hat{y}(t) \in S^{n_2}$. Because $y_j(t) \to 0$ for $j \notin M_{2k}$ (by definition of k), we have $\hat{y}(t) - y(t) \to 0$. Let

$$Z(t) = \prod x_i^{a_i}(t) \Big/ \prod x_i^{b_i}(t). \tag{13}$$

Differentiating (13) gives

$$\begin{aligned}\dot{Z}(t)/Z(t) &= \sum_{i=1}^{n_1} (a_i - b_i)\frac{\dot{x}_i(t)}{x_i(t)} \\ &= \sum_{i=1}^{n_1} (a_i - b_i)\frac{f_i(x(t), \hat{y}(t))}{x_i(t)} \\ &\quad + \sum_{i=1}^{n_1} (a_i - b_i)\left[\frac{f_i(x(t), y(t))}{x_i(t)} - \frac{f_i(x(t), \hat{y}(t))}{x_i(t)}\right]. \end{aligned} \tag{14}$$

The first part in (14) is bounded above by $-\varepsilon$ (from (11)), the second part is bounded by $\varepsilon/2$ when t is large. Thus, there exists $T > 0$ such that

$$\frac{\dot{Z}(t)}{Z(t)} < -\frac{\varepsilon}{2}$$

for $t > T$, or, equivalently,

$$Z(t) = \prod_{i=1}^{n_1} x_i^{a_i}(t) \Big/ \prod_{i=1}^{n_1} x_i^{b_i}(t) < Z(T)\, e^{-0.5\varepsilon(t-T)} \to 0.$$

Therefore,

$$\prod_{i=1}^{n_1} x_i^{a_i}(t) \to 0, \tag{15}$$

which contradicts the fact that $a \in A_1$. Therefore $A_1 \cup A_2$ must be empty. ∎

In light of this result, interest naturally turns to the question of which selection dynamics satisfy aggregate monotonicity. The following theorem shows that aggregate monotonic systems consist of the replicator dynamics and multiples of the replicator dynamics.

THEOREM 3. *The replicator dynamic (f^*, g^*) is aggregate monotonic. In addition, if (f, g) is a regular, aggregate monotonic selection dynamic, then there exist functions $\lambda(x, y) > 0$ and $\beta(x, y) > 0$ such that*

$$\begin{aligned} f_i(x, y) &= \lambda(x, y)\, f_i^*(x, y), &\quad i &= 1, ..., n_1 \\ g_j(x, y) &= \beta(x, y)\, g_j^*(x, y), ... &\quad j &= 1, ..., n_2. \end{aligned}$$

Proof. It is straightforward to verify that the replicator dynamic is aggregate monotonic. Next, let (f, g) be regular and aggregate monotonic. Given $(x, y) \in S^{n_1} \times S^{n_2}$, let $\xi = (e_1^T A y, ..., e_{n_1}^T A y)$.

Case a. Suppose $\xi = c(1, 1, ..., 1)$ for some real number c. Then,

$$\pi_1(i, y) = \pi_1(i', y) \qquad \forall i, i' \in I.$$

Thus, by aggregate monotonicity

$$(e_i^T - e_{i'}^T)\left[\frac{f_1(x, y)}{x_1}, ..., \frac{f_{n_1}(x, y)}{x_{n_1}}\right] = 0,$$

which implies

$$\frac{f_i(x, y)}{x_i} = \frac{f_{i'}(x, y)}{x_{i'}} \equiv h(x, y) \qquad \forall i, i' \in I. \tag{16}$$

By condition (1.2) of Definition 1,

$$0 = \sum_i \dot{x}_i = \sum_i f_i(x, y) = \sum_i x_i\, h(x, y) = h(x, y). \tag{17}$$

Therefore, for any $\lambda(x, y) > 0$, we have

$$f_i(x, y) = x_i\, h(x, y) = 0 = \lambda(x, y)\, f_i^*(x, y), \tag{18}$$

where the last equality holds because f^* is the replicator dynamics and $\xi = c(1, ..., 1)$.

Case b. Suppose $\xi \neq c(1, ..., 1)$ for any $c \in \mathbb{R}$. By aggregate monotonicity. $\forall u \in S^{n_1} - S^{n_1}$,

$$\left[\sum_i u_i = 0, \sum_i \xi_i u_i = 0\right] \Rightarrow \sum_i u_i \frac{f_i(x, y)}{x_i} = 0. \tag{19}$$

Let $X = \operatorname{span}\{u \mid u \in S^{n_1} - S^{n_1},\ \sum_i \xi_i u_i = 0\}$. Then the orthogonal complement of X in $\mathbb{R}^{n_1}$ is span $\{(1, ..., 1), \xi\} \equiv X^\perp$ and we have

$$\left[\frac{f_1(x, y)}{x_1}, ..., \frac{f_{n_1}(x, y)}{x_{n_1}}\right] \in X^\perp. \tag{20}$$

Thus, there exist $\lambda(x, y)$ and $a(x, y)$ such that

$$\left[\frac{f_1(x, y)}{x_1}, ..., \frac{f_{n_1}(x, y)}{x_{n_1}}\right] = \lambda(x, y)\xi + a(x, y)(1, ..., 1). \tag{21}$$

There exists $u \in S^{n_1} - S^{n_1}$ such that $u^T\xi > 0$, since $\xi \neq (c, ..., c)$. Then aggregate monotonicity implies

$$0 < u^T\left[\frac{f_1(x, y)}{x_1}, ..., \frac{f_{n_1}(x, y)}{x_{n_1}}\right] = \lambda(x, y)\, u^T\xi. \tag{22}$$

Therefore $\lambda(x, y) > 0$. Now, by condition (1.2) of Definition 1,

$$0 = \sum_i f_i(x, y) = \sum_i x_i(\lambda(x, y)\xi_i + a(x, y))$$
$$= \lambda(x, y)\, x^T A y + a(x, y).$$

Therefore

$$a(x, y) = -\lambda(x, y)\, x^T A y$$

and

$$f_i(x, y) = x_i(\lambda(x, y)\, \xi_i(x, y) + a(x, y))$$
$$= \lambda(x, y)\, x_i[e_i^T A y - x^T A y]$$
$$= \lambda(x, y)\, f_i^*(x, y).$$

Combining cases a and b, we have $\lambda(x, y) > 0$ such that

$$f_i(x, y) = \lambda(x, y)\, f_i^*(x, y) \qquad \forall (x, y) \in S^{n_1} \times S^{n_2}.$$

Similarly, $g_j(x, y) = \beta(x, y)\, g_j^*(x, y)$. ∎

Note that this proof uses a purely local argument, so that the statement of the theorem could be strengthened to the claim that if a selection dynamic is locally aggregate monotonic then it is locally a multiple of the replicator dynamics.

These results show that monotonic evolutionary paths readily yield results respecting strict admissibility. Given a belief in the evolutionary approach to games, one can thus embrace the proposition that strictly dominated strategies and iteratively strictly dominated strategies should not be played. This already presents results as strong as many that have emerged from models of rationality.

V. Asymptotic Stability

We now ask what additional statements can be made about the outcome of the evolutionary process if we restrict attention to asymptotically stable outcomes. Note that this investigation of stable points represents a departure from the analysis Section IV. We have previously worked only with assumptions on the structure of the selection process, such as monotonicity and regularity. We now introduce an assumption on the *outcome* of the process, namely that it is asymptotically stable. On the one hand, the assumption of asymptotic stability appears to be less primitive than assumptions such as regularity or monotonicity. On the other hand, stability is (at least in principle) observable, so that one can first ascertain whether one is investigating stable behavior and (if so)) then apply the following results.

We begin by defining:

DEFINITION 7. (x^*, y^*) is *asymptotically stable* if there exists a neighborhood U of (x^*, y^*) such that

$$(x(0), y(0)) \in U \Rightarrow \lim_{t \to \infty} (x(t), y(t)) = (x^*, y^*).$$

Asymptotic stability thus requires that any path starting sufficiently close to (x^*, y^*) converge to (x^*, y^*).

We now show that asymptotically stable outcomes are "almost" strict Nash equilibria.

THEOREM 4. *Let (f, g) be a regular monotonic selection dynamic. Let (x^*, y^*) be asymptotically stable. Then (x^*, y^*) is a Nash equilibrium and there does not exist $x' \in S^{n_1}$ such that*

$$\pi_1(x', y^*) = \pi_1(x^*, y^*) \tag{23}$$

$$\pi_2(x', j) = \pi_2(x^*, j') \qquad \forall j, j' \in \operatorname{supp} y^*. \tag{24}$$

In particular, there must exist no alternative best reply x' for player 1 such that player 2 is indifferent over the strategies in the support of y^* given x'.[5]

Proof. Fix (x^*, y^*). It is clear that (x^*, y^*) cannot be asymptotically stable if there exists x' such that

$$\pi_1(x', y^*) > \pi_1(x^*, y^*)$$

since there would then be strategies $i, i' \in I$, a time T, and an $\varepsilon > 0$ such that $x_i^* > 0$ and, for all $t > T$, $\pi_1(i', y(t)) - \pi_1(i, y(t)) > \varepsilon$ (given convergence to (x^*, y^*)). This in turn implies that there exists $\delta > 0$ such that $f_{i'}(x(t), y(t))/x_{i'}(t) - f_i(x(t), y(t))/x_i(t) > \delta$, precluding convergence to (x^*, y^*) and hence precluding the asymptotic stability of the latter. Suppose next there exists x' such that (23) and (24) hold. Consider an initial condition given by $((1-\varepsilon)\, x^* + \varepsilon x', y^*) \equiv (x'', y^*)$. Equations (23) and (24) and monotonicity ensure that

$$f(x'', y^*) = g(x'', y^*) = 0,$$

[5] See [9, p. 282, exercise 1] for a similar result for the special case of the replicator dynamics.

so that no subsequent movement in the dynamic process occurs. Then the path originating at (x'', y^*) does not converge to (x^*, y^*), precluding the asymptotic stability of the latter. ▮

To appreciate the strength of this condition, consider:

COROLLARY 1. *If (x^*, y^*) is pure and asymptotically stable, then (x^*, y^*) is a strict Nash equilibrium.*

Note that Theorem 4 strengthens Friedman's result that asymptotically stable equilibria must be Nash equilibria by showing that asymptotically stable outcomes possess additional structure (strictness in the case of pure strategies).

Theorem 4 poses a dilemma for a research program designed to provide evolutionary foundations for refinements of Nash equilibria. On the one hand, we can work without assuming stability, but find that the implied equilibrium concept is rationalizability. On the other hand, we can work with asymptotically stable equilibria and find that the implied equilibrium concept is virtually as strong as strict Nash equilibria. There appears to be little support for the Nash equilibrium concept and the equilibrium refinements literature, which deals with the gap between Nash and strict Nash equilibria.

We can offer another perspective on strict Nash equilibria:

THEOREM 5. *Let $(x(t), y(t))$ be an evolutionary path produced by the replicator dynamics with a strictly interior origin and converging to (x^*, y^*). Suppose that*

$$\int_0^\infty |x_i(t) - x_i^*| \, dt \equiv \mathscr{X} < \infty \tag{25}$$

for all $i = 1, ..., n_1$, and

$$\int_0^\infty |y_j(t) - y_j^*| \, dt \equiv \mathscr{Y} < \infty \tag{26}$$

for all $j = 1, ..., n_2$. Then (x^*, y^*) is quasi-strict.[6]

The interpretation of (25) and (26) is that the evolutionary path converges relatively quickly, so that the cumulative difference between the path and its limit is finite.

[6] An equilibrium (p, q) in a bimatrix game is quasi-strict if $p_i = 0$ implies that i is not a best reply to q and $q_j = 0$ implies that j is not a best reply to p. See [24].

Proof. Suppose (x^*, y^*) is not quasi-strict. Then there exists i' such that $x^*_{i'} = 0$ and i' is also a best reply to y^*. Consider the function defined by

$$U(t) = \prod_{i=1}^{n_1} x_i^{x_i^*}(t)/x_{i'}(t).$$

Then, as $t \to \infty$, $U(t) \to \infty$ (because $x_{i'}(t) \to 0$ and $\lim_{t \to \infty} \prod_{i=1}^{n_1} x_i^{x_i^*}(t) > 0$). Now

$$\begin{aligned}\frac{\dot{U}(t)}{U(t)} &= \sum_{i=1}^{n_1} (x_i^* - \delta_{ii'})(e_i^T A y(t) - x^T A y(t)) \\ &= x^{*T} A y(t) - e_{i'}^T A y(t) \\ &= (x^{*T} - e_{i'}^T)\, A(y(t) - y^*)\end{aligned}$$

where the last equality appears because x^* and i' are both best replies to y^* and where

$$\delta_{ii'} = \begin{cases} 1 & \text{if } \; i = i' \\ 0 & \text{otherwise} \end{cases}$$

is the Kronecker delta. Solving this differential equation,

$$\begin{aligned}\ln U(T) - \ln U(0) &= \int_0^T (x^* - e_{i'})^T A(y(t) - y^*)\, dt \\ &\leqslant \sum_{ij} |a_{ij}|\, |x_i^* - \delta_{ii'}| \int_0^T |y_j(t) - y_j^*|\, dt \\ &\leqslant \sum_{ij} |a_{ij}|\, |x_i^* - \delta_{ii'}|\, \mathscr{X} < \infty\end{aligned}$$

contradicting the fact that $U(t) \to \infty$ as $t \to \infty$. ∎

This theorem indicates that equilibria which are not quasi-strict appear only if *every* converging evolutionary path under the replicator dynamics converges slowly. The conclusion is then that unless we can restrict attention to quasi-strict equilibria, the cumulative error involved in taking the limit of an evolutionary process to be an estimate of the path of the process will be infinitely large.

This result again poses a challenge. To apply evolutionary game theory, we must hope that the limit serves as a reasonable approximation of the outcomes which appear along the evolutionary path. Theorem 5 shows that unless one is willing to confine attention to quasi-strict Nash equilibria, there is good reason to doubt the reasonableness of this approximation.

A possible alternative interpretation of this result is to let $L(x-x^*, y-y^*)$ be a bounded function identifying the cost of an incorrect prediction of the outcome of the game, where the actual outcome is given by the path $(x(t), y(t))$ and the limit (x^*, y^*) is predicted. Then (25) and (26) are sufficient to ensure that the discounted loss

$$\int_0^\infty e^{-\delta t} L(x(t)-x^*, y(t)-y^*)\, dt$$

does not increase without bound as $\delta \to 0$. If (25 and (26) fail, as will be the case with equilibria which are not quasi-strict, then patient investigators will suffer arbitrarily large losses when using evolutionary game theory to make predictions.

VI. Stable Evolutionary Outcomes

The previous section suggests that if we are to insist on asymptotic stability, then evolutionary arguments provide a foundation for the game theoretic solution concepts of rationalizability (if stability fails) and (nearly) strict Nash equilibrium. These are essentially the weakest and strongest of equilibrium concepts, and we would like to investigate intermediate solution concepts such as Nash equilibrium.

To do this, we investigate a weaker stability requirement.

Definition 8. (x^*, y^*) is *stable* if, for any neighborhood V of (x^*, y^*), there exists a neighborhood U with $(x^*, y^*) \in U \subseteq V$ such that $(x(0), y(0)) \in U \Rightarrow (x(t), y(t)) \in V$ for all t.

Stable points thus have the property that evolutionary paths which start nearby, stay nearby. At the end of this section, we discuss the adequacy of this requirement as a stability notion in a evolutionary context.

We first show that there is a relationship between stabiity and Nash equilibrium:

Theorem 6. *Let (f, g) be a regular, monotonic selection dynamic and let (x^*, y^*) be stable. Then (x^*, y^*) is a Nash equilibrium.*

Proof. Suppose (x^*, y^*) is not a Nash equilibrium. Then there exists a player, say 1, and strategies i and k such that

$$x_i^* > 0$$

$$\pi_1(i, y^*) < \pi_1(k, y^*).$$

Because π is continuous, there exists a neighborhood V_2 of y^* such that $\pi_1(i, y') < \pi_1(k, y')$ for all $y' \in V_2$. Then there exists a enighborhood V_1 of x^* such that, letting $V = V_1 \times V_2$, we have

$$\frac{\dot{x}_i}{x_i} < \frac{\dot{x}_k}{x_k} - \delta$$

for some $\delta > 0$ and for all $(x, y) \in V$. No path originating in the interior of V can then remain in V, precluding the stability of (x^*, y^*). ▌

Intuitively, a stable point (x, y) must be a Nash equilibrium because if a reply to y that is superior to x exists, then the dynamics around (x, y) must lead toward the superior reply and away from (x, y), precluding stability of (x, y). We can interpet Theorem 4 as indicating that the conditions yielding the Nash equilibrium concept match those yielding rationalizability plus a stability requirement.

Because Nash equilibria respect strict interated admissibility, we immediately have:

COROLLARY 2. *Let (f, g) be a regular, monotonic selection dynamic and let (x^*, y^*) be stable. Then (x^*, y^*) survives strict iterated admissibility.*

Attention now turns to refinements of Nash equilibria. In particular, we examine normal form perfection. It is useful to recall that all pure strategies are played by a positive proportion of the population along the path induced by a regular selection dynamic. This is reminiscent of the completely mixed strategy perturbations of the perfect equilibrium concept, and suggests that stable points should be perfect equilibria.

In two-player games, an outcome is a perfect equilibrium if and only if it is a Nash equilibrium in undominated strategies. The following result then reinforces the suspicion that stable points must be perfect:

THEOREM 7. *Let (f, g) be a regular, monotonic selection dynamic and let (x^*, y^*) be a limiting outcome of (f, g) given completely mixed $(x(0), y(0))$. Then (x^*, y^*) cannot attach unitary probability to a weakly dominated strategy.*

Proof. Let i be weakly dominated by the (possibly mixed) strategy x' and let $x_i^* = 1$. Then for every t there exists a strategy j with

$$\frac{\dot{x}_j}{x_j} > \frac{\dot{x}_i}{x_i}.$$

This precludes the possibility that (x^*, y^*) is a limiting outcome. ▌

We can show, however, that stable outcomes need not be perfect.[7] Consider the following game:

$$\begin{array}{ccc} & L & R \\ T & 1,1 & 1,0 \\ B & 1,1 & 0,0 \end{array} \tag{27}$$

It is straightforward to calculate that under the replicator dynamics, any point in which player 2 plays L and player 1 mixes between T and B is a stable outcome and is also the limiting outcome of some evolutionary path (under the replicator dynamics).[8] These are Nash but not perfect equilibria, since player 1 attaches positive probability to the dominated strategy B. The difficulty is that T dominates B, but only weakly, and the payoff difference between T and B disappears as population 2 becomes concentrated on L. It is then possible that population 2 can converge to L sufficiently rapidly that the pressure pushing population 1 toward T dissipates too quickly to drive all of population 1 to T, yielding an outcome in which population 1 is split between T and B.

This argument gives:

THEOREM 8. *Let (x^*, y^*) be stable. Then (x^*, y^*) can attach positive probability to a weakly dominated strategy even if (f, g) is regular and aggregate monotonic.*

Stable points of regular, monotonic selection dynamics will thus be Nash equilibria but need not be perfect equilibria.

Before interpreting these results, we must examine our stability notion. An evolutionary model should capture a learning or selection process that causes agents to adjust their choices in light of their experience and that is buffeted by rare mutations. Our selection dynamic captures the former of these considerations but not the latter. When working with asymptotically stable systems, the failure to capture mutations does not appear to be a difficulty. The asymptotic stability of the system ensures that it will return to its limiting outcome after being disturbed by any (small) mutation.

Stable outcomes are less satisfactory in this regard. A small mutation will not prompt the system to move far away from a stable outcome, but the system need not return to the stable outcome. This is the case with out-

[7] This result does not depend upon asymmetry, as [24, example 9.4.3] illustrates this possibility in a symmetric game.

[8] Under more general evolutionary processes, such outcomes will exist as long as there exist numbers $\alpha, \beta \in (0, 1)$ such that the slope of the direction of movement in the phase diagram is bounded above by some $\varepsilon > 0$ for all $(x(t), y(t))$ satisfying $x_T(t) < \alpha$ and $y_R(t) < \beta$.

comes in (27), where mutations toward R will be followed by convergence to nearby points for which $y_R^* = 0$, but with a slightly larger proportion of population 1 playing T. The effect of a mutation can thus remain permanently. No matter how rare mutations are, successive mutations can then produce a large drift away from the stable point.

These considerations suggest that if we are to work with stable rather than asymptotically stable outcomes, then the model must be altered to explicitly include mutation. Note that mutations appear to allow some hope of finding a foundation for perfect equilibria. In the game given by (27), the effect of successive mutations toward R will apparently be to drive the system to the perfect quilibrium (T, L), suggesting that stable outcomes of evolutionary systems with mutation may be perfect.

VII. Mutations

This section constructs an evolutionary model with mutations. Intuitively, we assume that the agents do not live forever. Instead, agents continually exit the game or "die" and are replaced by new entrants or "births." The strategies played by entrants mirror those of existing agents with probability $\lambda \in (0, 1)$, but are given by an exogeneously specified "mutant" distribution with probability λ.

More formally, we work with the following process:[9]

$$\dot{x}_i = f_i(x, y)(1-\delta_1) + \delta_1(\lambda x_i + (1-\lambda)\xi_i - x_i) \qquad i = 1, 2, ..., n_1 \quad (28)$$

$$\dot{y}_j = g_j(x, y)(1-\delta_2) + \delta_2(\lambda y_j + (1-\lambda)\eta_j - y_j) \qquad j = 1, 2, ..., n_2, \quad (29)$$

where $\xi_i > 0$, $\eta_j > 0$, and

$$\sum_{i=1}^{n_1} \xi_i = \sum_{j=1}^{n_2} \eta_j = 1$$

and where (f, g) is a selection dynamic. The distribution of strategies among surviving members of a population evolves according to the selection dynamic (f, g). However, population k agents die at rate δ_k and are replaced by entrants whose strategies are governed by the existing distribution with probability $(1-\lambda)$ and by the mutant distribution ξ or η with probability λ. While ξ or η could be made to vary over time, it is convenient to take them to be fixed.

[9] See [9, chapter 25] for a similar model of an evolutionary process with transmission errors. Reference [1] examines an analogous process in symmetric games. An alternative model of evolutionary trembles is examined by [6].

Let (x, y) be called a *rest point* of (28) and (29) if (x, y) yields $\dot{x} = 0 = \dot{y}$. Denote such a point $(x(\delta_1, \delta_2), y(\delta_1, \delta_2))$. We are interested in cases in which mutations occur relatively infrequently. The appropriate object of study is then the limit of the rest points of the system as the δ_k approach zero. Let such points be called *limit rest points.* (x^*, y^*) is then a limit rest point if $(x^*, y^*) = \lim_{\delta_1, \delta_2 \to 0}(x(\delta_1, \delta_2), y(\delta_1, \delta_2))$.

Given this evolutionary process with mutation, consider the game given by:

$$\begin{array}{cccc} & & \multicolumn{2}{c}{2} \\ & & L & R \\ \multirow{2}{*}{1} & T & 1, 1 & 1, 1 \\ & B & 1, 1 & 0, 0 \end{array} . \tag{30}$$

This is a variant of game (27) in which either player can play a weakly dominated strategy in a Nash equilibrium. We then have

THEOREM 9.

(9.1) *Let (f, g) be regular and monotonic. Then a limit rest point of* (28) *and* (29) *is a Nash equilibrium.*

(9.2) *A limit rest point of* (28) *and* (29) *need not be a perfect equilibrium. In particular, let $\{\delta_{1n}\}_{n=1}^{\infty}$ and $\{\delta_{2n}\}_{n=1}^{\infty}$ be sequences such that*

$$\lim_{n \to \infty} \frac{\delta_{1n}}{\delta_{2n}} \equiv \Lambda \tag{31}$$

exists. Let (f, g) be the replicator dynamics. Then the system of rest points of (28) *and* (29) *converges to (T, L) in* (30) *as n approaches ∞ only if $\Lambda = (1 - \eta_L)/(1 - \xi_T)$.*

Proof. 9.1. Let (x, y) be a limit rest point of (28) and (29) but not be a Nash equilibrium. Without loss of generality, assume that x attaches probability to a pure strategy which is not a best reply to y. Let $i \in I$ be the strategy in the carrier of I that earns the lowest expected payoff against y. Then there must exist an alternative strategy $i' \in I$ and an $\varepsilon > 0$ such that for all δ_{1n} and δ_{2n} sufficiently small, $\pi_1(i', y(\delta_{1n}, \delta_{2n})) - \pi_1(i, y(\delta_{1n}, \delta_{2n})) > \varepsilon$. Monotonicity and regularity in turn imply that there exists a $\delta > 0$ such that, for all δ_{1n} and δ_{2n} sufficiently small, $f_i(x(\delta_{1n}, \delta_{2n}), y(\delta_{1n}, \delta_{2n}))/x_i(\delta_{1n}, \delta_{2n}) < \delta$. Condition (28) can then hold only if $x_i \to 0$ as δ_{1n} and δ_{2n} approach zero, contradicting the definition of i.

9.2. To conserve on notation, let x (y) denote the proportion of population 1 (2) playing T (L) and let ξ_T and η_L be denoted simply ξ and η. Then the replicator dynamics with transition errors for game (30) is given by:

$$\begin{aligned} \dot{x} &= x(1-x)(1-y)(1-\delta_1) + \delta_1(1-\lambda)(\xi - x) \\ \dot{y} &= y(1-y)(1-x)(1-\delta_2) + \delta_2(1-\lambda)(\eta - y). \end{aligned} \tag{32}$$

Letting $\dot{x} = \dot{y} = 0$, a rest point must satisfy

$$\begin{aligned} x(1-x)(1-y) + \alpha(\xi - x) &= 0 \\ y(1-x)(1-y) + \beta(\eta - y) &= 0, \end{aligned} \tag{33}$$

where

$$\alpha = \frac{(1-\lambda)\,\delta_1}{1-\delta_1} \qquad \beta = \frac{(1-\lambda)\,\delta_2}{1-\delta_2}. \tag{34}$$

From (33), we have

$$\alpha(\xi - x)\,y - \beta(\eta - y)\,x = 0. \tag{35}$$

Let $\delta_{1n} \to 0$, $\delta_{2n} \to 0$, and $\delta_{1n}/\delta_{2n} \to \Lambda$; and let (x^*, y^*) be a limit rest point. From (34), we have $\alpha/\beta \to \Lambda$. From (35), the limiting outcome (x^*, y^*) will then equal (1, 1), the perfect equilibrium, only if $\Lambda = (1-\eta)/(1-\xi)$. ▮

The model with mutations thus provides a motivation for Nash equilibria but does not provide a basis for a theory in which players shun dominated strategies, or play perfect equilibria. A limit *ESS* must be perfect, so our results also show that stable points of evolutionary processes with mutations need not be a limit *ESS*.

VIII. Evolutionary Choice Trembles

The previous section suggests that a model of mutations does not provide an effective theory of perfect equilibria or limit *ESS*. We are then still left with the puzzle that the outcome (T, L) appears to be "more stable" than other outcomes in (27) or (30). In this section we examine an alternative temble-based theory in which the trembles appear not as mutations but mistakes when agents play their strategies.

To construct this theory we now let $x(t)$ and $y(t)$ denote the "intended" strategies of the agents, which are to be interpreted as the population proportions which would arise if no mistakes in play were made. We let $\tilde{x}(t)$ and $\tilde{y}(t)$ be the realized strategy proportions, with

$$\tilde{x}_i(t) = (1-\varepsilon)\, x_i(t) + \varepsilon \sum_{k=1}^{n_1} x_k \xi_{ki} \tag{36}$$

$$\tilde{y}_j(t) = (1-\varepsilon)\, y_j(t) + \varepsilon \sum_{k=1}^{n_2} y_k \eta_{kj}, \tag{37}$$

where $(\xi_{k1}, ..., \xi_{kn_1}) \equiv \xi_k \in S^{n_1}$ identifies how the mistakes made by players intending to play strategy k are distributed among strategies $1, ..., n_1$. $(\eta_{k1}, ..., \eta_{kn_2})$ is similar. The interpretation of (36) and (37) is then that agents play their intended strategies with probability $1-\varepsilon$. With probability ε, however, a mistake is made and the actual play is governed by the functions ξ and η. We assume ξ_k and η_k are strictly positive. We also find it convenient to let $\sum_{k=1}^{n_1} x_k \xi_k \equiv \xi(x)$ and $\sum_{k=1}^{n_2} y_k \eta_k \equiv \eta(y)$.

To specify the evolutionary process, we allow a selection dynamic to adjust players' intended strategies in response to payoffs determined by realized strategies:

$$\dot{x}_i = x_i(\tilde{e}_i^T A \tilde{y} - \tilde{x}^T A \tilde{y}) \tag{38}$$

$$\dot{y}_j = y_j(\tilde{x}^T B \tilde{e}_j - \tilde{x}^T B \tilde{y}), \tag{39}$$

where

$$\tilde{e}_i = (1-\varepsilon)\, e_i + \varepsilon \xi_i$$

$$\tilde{e}_j = (1-\varepsilon)\, e_j + \varepsilon \eta_j,$$

and where A and B are matrices of payoffs and $\tilde{x}(t)$ and $\tilde{y}(t)$ are given by (36) and (37). It is readily verified that (38) and (39) specify a mapping from $S^{n_1} \times S^{n_2} \to S^{n_1} \times S^{n_2}$.[10]

We now examine the limiting outcomes of this selection dynamic with strategy errors.

DEFINITION 9. *Let* $(x(0), y(0))$ *be completely mixed and let* $(x(t), y(t))$ *converge to* (x^*, y^*). *Then* (x^*, y^*) *is an* ε-evolutionary outcome *of* (38) *and* (39).

It is easy to show that the limiting outcome of this model with errors in strategy choices respects weak dominance:

[10] Note that in specifying (38)–(39), we have now restricted attention to the replicator dynamics, unlike the previous section. We do so because the replicator dynamics are used in verifying that (38)–(39) yield a mapping that stays within the appropriate simplices. It is not obvious how to ensure this with general monotonic processes.

THEOREM 10. *Let* (x^*, y^*) *be an* ε*-evolutionary outcome of* (38) *and* (39). *Then*

(10.1) x^* *and* y^* *are not weakly dominated*

(10.2) $(\tilde{x}^*, \tilde{y}^*)$ *is an* ε*-perfect equilibrium*[11]

(10.3) (x^*, y^*) *need not be a Nash equilibrium.*

Proof. 10.1. Suppose that x^* is dominated by p and hence $\tilde{x}^*$ is dominated by $\tilde{p}$. Let

$$u(t) \equiv \prod_{i=1}^{n_1} x_i(t)^{p_i - x_i^*}.$$

Then we have

$$\begin{aligned}
\frac{du}{dt} &= u(t) \sum_{i=1}^{n_1} (p_i - x_i^*) \frac{\dot{x}_i(t)}{x_i(t)} \\
&= u(t) \sum_{i=1}^{n_1} (p_i - x_i^*)[\tilde{e}_i^T A\tilde{y}(t) - \tilde{x}(t)^T A\tilde{y}(t)] \\
&= u(t)[\tilde{p}^T A\tilde{y}(t) - \tilde{x}^{*T} A\tilde{y}(t)] \geqslant 0.
\end{aligned}$$

Because $\tilde{y}(t)$ converges to $\tilde{y}^*$ with $\tilde{y}^*$ in the interior of S^{n_2}, there must exist a T such that for $t \in [T, \infty)$,

$$\tilde{p}^T A\tilde{y}(t) - \tilde{x}^{*T} \mathrm{A}\tilde{y}(t) \geqslant \delta > 0$$

and hence

$$\frac{du}{dt} \geqslant u(t)\delta$$

Because $(x(T), y(T))$ is in the interior of $S^{n_1} \times S^{n_2}$, we have $u(T) > 0$. Hence

$$\lim_{t \to \infty} u(t) = \infty.$$

This gives

$$\lim_{t \to \infty} \prod_{i=1}^{n_1} x_1(t)^{x_i^*} = 0,$$

which is a contradiction because $x_i(t)$ converges to x^* and hence $x_i(t)^{x_i^*}$ converges to $x_i^{*x_i^*} > 0$.

[11] An ε-perfect equilibrium is a Nash equilibrium of a perturbed game in which strategies must be completely mixed. See [24] for details.

10.2. We first show that x^* is a best reply to $\tilde{y}^* = (1-\varepsilon)\, y^* + \varepsilon\eta(y^*)$. Suppose there exists i with $x_i^* > 0$ and i not a best reply to $\tilde{y}^*$. Then there exists i' such that

$$\tilde{e}_i^T A\tilde{y}^* < \tilde{e}_{i'}^T A\tilde{y}^*.$$

Because $\tilde{y}(t)$ converges to $\tilde{y}^*$, there exists a $T > 0$ such that for $t > T$, we have

$$\tilde{e}_i^T A\tilde{y}(t) - \tilde{e}_{i'}^T A\tilde{y}(t) \leqslant -\delta > 0$$

and hence

$$\frac{\dot{x}_i(t)}{x_i(t)} - \frac{\dot{x}_{i'}(t)}{x_{i'}(t)} = \tilde{e}_i^T A\tilde{y}(t) - \tilde{e}_{i'}^T A\tilde{y}(t) \leqslant -\delta < 0$$

and

$$\lim_{t\to\infty} \frac{x_i(t)}{x_{i'}(t)} = 0$$

which contradicts $\lim x_i(t) = x_i^* > 0$. Because x^* is a best reply to $\tilde{y}^*$, $\tilde{x}^*$ is a best reply to $\tilde{y}$ in the restricted strategy set $\tilde{S}^{n_1} \equiv \{\sum_i (x_i \tilde{e}_i \mid \sum_i x_i = 1, x_i \geqslant 0\}$. A similar argument for $\tilde{y}^*$ yields an ε-perturbed game with strategy sets $\tilde{S}^{n_1}$ and $\tilde{S}^{n_2}$ in which $(\tilde{x}^*, \tilde{y}^*)$ is a Nash equilibrium, yielding (10.2).

10.3. Consider the following game:

$$\begin{array}{ccc} & \multicolumn{2}{c}{2} \\ & L & R \\ T & 1,0 & 0,1 \\ B & 0,1 & 1,0 \end{array} \tag{40}$$

Let $(\xi_{TT}, \xi_{TB}) = (\xi_{BT}, \xi_{BB}) = (0.25,\ 0.75)$ and $(\eta_{LL}, \eta_{LR}) = (\eta_{RL}, \eta_{RR}) = (0.5, 0.5)$. Then it is easily verified that $(x^*, y^*) = (((2-\varepsilon)/4(1-\varepsilon)^2, (1-(2-\varepsilon)/4(1-\varepsilon)^2)), (0.5, 0.5))$ is an ε-evolutionary outcome but is not a Nash equilibrium. ▮

We have thus finally established a link between evolutionary outcomes and weak dominance. At the same time, (10.3) indicates that Nash equilibria need not appear. However, this is an artifact of presuming that the tremble probability ε is "large." To see this, we next examine the limits of the ε-evolutionary outcomes which appear as the tremble probabilities become small.

THEOREM 11. *Let* $(x^*(\varepsilon_n), y^*(\varepsilon_n))$ *be a sequence of* ε_n*-evolutionary outcomes of* (51) *and* (52) *with* $\lim_{n\to\infty} \varepsilon_n = 0$. *Then if* $\lim_{n\to\infty}(x^*(\varepsilon_n), y^*(\varepsilon_n)) = (x^*, y^*)$, *then* (x^*, y^*) *is a perfect equilibrium. However,* (x^*, y^*) *need not be a limit ESS.*

Proof. From (10.2), we know that $(\tilde{x}^*(\varepsilon_n), \tilde{y}^*(\varepsilon_n))$ is ε-perfect. Because $(x^*, y^*) = \lim_{n\to\infty}(x^*(\varepsilon_n), y^*(\varepsilon_n)) = \lim_{n\to\infty}(\tilde{x}^*(\varepsilon_n), \tilde{y}^*(\varepsilon_n))$, (x^*, y^*) is then perfect. To show (x^*, y^*) need not be a limit *ESS*, we consider again the example given by (40). Let the error generating process be $\xi_{TT} = \xi_{TB} = \xi_{BT} = \xi_{BB} = \eta_{LL} = \eta_{LR} = \eta_{RL} = \eta_{RR} = 0.5$. Then $(x^*(\varepsilon_n), y^*(\varepsilon_n)) = ((0.5, 0.5), (0.5, 0.5))$ is an ε_n-evolutionary outcome. $(x^*, y^*) = \lim_{n\to\infty}(x^*(\varepsilon_n), y^*(\varepsilon_n)) = ((0.5, 0.5), (0.5, 0.5))$ is then a perfect equilibrium but not a limit *ESS* (because it is not pure, see [17]). ▌

These results suggest that an evolutionary model with strategy errors can provide a foundation for perfect equilibrium. These results are no surprise, since this strategy-choice formulation provides a fairly transparent transition of the definition of trembling hand perfection into evolutionary terms. At the same time, it is not clear that the model provides support for the seemingly similar limit *ESS* concept.[12]

It may initially appear surprising that the models of Section VII and VIII give different results. The first model might be described as one in which a small proportion of agents make mistakes with large probability while in the latter model all agents make mistakes with small probability. These may appear equivalent. However, the difference is that with strategy mistakes, the evolutionary process operates in *intended* strategies. In the game given by (27), it is as if evolutionary adjustment proceeds with population one free from trembles to B, with these trembles added *after* evolution has played its course. This yields an outcome that can be viewed as "(T, L) plus mistakes," and in the limit, as these mistakes become arbitrarily small, the result is (T, L). With mutations it is as if the order in which the two limits (evolutionary adjustment and shrinking mistakes) are taken is reversed. The evolutionary process is continually buffeted by trambles to B. This gives evolutionary outcomes which are different from (T, L). In the limit as mutations go to zero, these errors continue to drive a wedge between (T, L) and the evolutionary limit.

The question remains as to which of the evolutionary models with trembles is most appropriate. Each may apply in some circumstances, though mutations appear to be more easily interpreted in evolutionary terms. The

[12] For an alternative perspective on the limit *ESS*, note that, from [17], the essential difference between a limit *ESS* and perfection is that the latter allows mixed strategies. Doubt has been cast on the ability of evolutionary processes to yield mixed strategy outcomes (e.g., [5]). If mixed strategies do not generally appear, then the limit *ESS* concept is more applicable.

choice of which model is appropriate is potentially important. If one desires an evolutionary foundation for game theoretic solution concepts, then the choice has implications for such matters as whether one embraces the perfect equilibrium concept.

IX. Conclusion

Our results suggest that evolutionary arguments readily motivate the solution concepts of iterated strict admissibility and Nash equilibrium, with the difference between the two being that the latter requires stability. However, we have encountered difficulty in establishing conditions under which evolutionary outcomes respect weak admissibility and must conclude that evolutionary models generally do not provide support for such refinements of Nash equilibrium as the perfect equilibrium concept.

One's initial impression might be that tremble-based arguments, and hence the avoidance of weakly dominated strategies, are built into dynamic evolutionary models. We have seen that this is not the case. Trembles must be explicitly built into the model in order to motivate admissibility, and even then these trembles must involve not the seemingly natural convention of mutations, but instead mistakes in playing strategies. If one is uncomfortable with such mistakes, then evolutionary arguments may drive one to a theory in which weakly dominated strategies are played. It appears as if such a conclusion can be avoided only if additional structure is placed on the evolutionary selection or learning process. Theories of learning are thus an important area for further research.

References

1. R. Boylan, Equilibria resistant to mutuation, mimeo, California Institute of Technology, 1990.
2. R. Boylan, A strong law of large numbers for games with infinitely many randomly matched agents, mimeo, California Institute of Technology, 1990.
3. G. W. Brown, Iterative solutions of games by fictitious play, *in* "Activity Analysis of Production and Allocation" (T. C. Koopmans, Ed.), Wiley, New York, 1951.
4. D. Canning, "Convergence to Equilibrium in Sequences of Games with Learning," Working paper 88/331, European University Institute, 1987.
5. V. P. Crawford, Learning and mixed-strategy equilibria in evolutionary games, *J. Theoret. Biol.* **140** (1989), 537–550.
6. D. Foster and P. Young, Stochastic evolutionary game dynamics, *Theoret. Population Biol.* **38** (1990), 219–232.
7. D. Friedman, Evolutionary games in economics, *Econometrica* **59** (1991), 637–666.
8. D. Fudenberg and D. Kreps, A theory of learning, experimentation, and equilibrium in games, mimeo, MIT and Stanford, 1988.

9. J. Hofbauer and K. Sigmund, "The Theory of Evolution and Dynamical Systems," Cambridge Univ. Press, Cambridge, 1988.
10. J. S. Jordan, Bayesian learning in normal form games, *Games Econ. Behav.* **3** (1991), 60–81.
11. J. Maynard Smith, "Evolution and the Theory of Games," Cambridge Univ. Press, Cambridge, 1982.
12. J. Maynard Smith and G. R. Price, The logic of animal conflict, *Nature* **246**, 15–18.
13. P. Milgrom and J. Roberts, Rationalizability, learning, and equilibrium in games with strategic complementarities, *Econometrica* **58** (1990), 1255–1278.
14. J. H. Nachbar, "Evolutionary" selection dynamics in games: Convergence and limit properties, *Int. J. Game Theory* **19** (1990), 59–90.
15. D. G. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52** (1984), 1029–1050.
16. J. Robinson, An iterative method of solving a game, *Ann. Math.* **54** (1951), 296–301.
17. L. Samuelson, Limit evolutionarily stable strategies in asymmetric games, *Games Econ. Behav.* **3** (1990), 110–128.
18. R. Selten, A note on evolutionarily stable stratefies in asymetric animal conflicts, *J. Theoret. Biol.* **84** (1980), 93–101.
19. R. Selten, Evolutionary stability in extensive two-person games *Math. Soc. Sci.* **5** (1983), 269–363.
20. R. Selten, Evolutionary stability in extensive two-person games—Correction and further development, *Math. Soc. Sci.* **16** (1988), 223–266.
21. R. Selten, "Anticipatory Learning in Two-Person Games," ZIF Discussion paper 4, University of Bielefeld, 1988.
22. L. Shapley, Some topics in two-person games, *Ann. Math. Stud.* **5** (1964), 1–28.
23. T. C.-C. Tan and S. R. da C. Werlang, The Bayesian foundations of solution concepts of games, *J. Econ. Theory* **45** (1988), 370–391.
24. E. van Damme, "Stability and Perfection of Nash Equilibria," Springer-Verlag, Berlin, 1987.