Econometrica, Vol. 64, No. 4 (July, 1996), 943–956

EVOLUTION WITH STATE-DEPENDENT MUTATIONS

BY JAMES BERGIN AND BARTON L. LIPMAN[1]

Recent evolutionary models have introduced "small mutation rates" as a way of refining predictions of long-run behavior. We show that this refinement effect can only be obtained by restrictions on how the magnitude of the effect of mutation on evolution varies across states of the system. In particular, given any model of the effect of mutations, any invariant distribution of the "mutationless" process is close to an invariant distribution of the process with appropriately chosen small mutation rates.

KEYWORDS: Evolution, mutation, limiting distributions.

1. INTRODUCTION

A RECENT REFORMULATION of some simple evolutionary dynamics has led to a surprising result: the addition of small mutation rates leads to precise long-run predictions. We re-examine the robustness of this result with respect to the nature of the mutation process and show that this precision of prediction is possible only with very strong assumptions on the mutation process.

Consider the evolution of strategic behavior in a finite population of individuals who are repeatedly matched to play some finite normal form game. A variety of evolutionary processes have been analyzed, but to fix ideas, suppose these individuals only change their actions occasionally, always changing to a best response to the current distribution of strategies in the population. Clearly, if the initial distribution of strategies in the population is sufficiently close to a strict Nash equilibrium,[2] the distribution of strategies will converge to this equilibrium and stay there forever. Hence any strict Nash equilibrium is a limit point of such a process, so these dynamics have at least as many possible long-run predictions as there are strict Nash equilibria for the game. Many evolutionary processes will have still other long-run possibilities, such as cycling forever.

[1] We thank Nabil Al-Najjar, Larry Blume, Richard Boylan, George Mailath, J.-F. Mertens, Bernard de Meyer, Yaw Nyarko, Phil Reny, Jeroen Swinkels, seminar participants at Arizona, Columbia, Dublin Economics Workshop, Northwestern, Penn, Queen's, Rochester, Summer in Tel Aviv 1994, UCLA, University of Western Ontario, Windsor, the Midwest Mathematical Economics Society Meetings, and the 1994 NBER/CEME Decentralization Conference, and a co-editor and three anonymous referees for helpful comments. Both authors also acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. Bergin thanks UCD and CORE and Lipman thanks Penn for enjoyable visits during which some of this research was done. Any errors or omissions are our responsibility.

[2] That is, a Nash equilibrium where each player's strategy is the unique best reply to his opponents.

Kandori, Mailath, and Rob (1993) (henceforth KMR) and Young (1993a)[3] reanalyzed such processes, showing that the addition of small probabilities of mutations change the picture significantly. Suppose that there is a small probability $\varepsilon > 0$ that a given agent "mutates"—that is, with probability ε, he deviates in some fashion from the dynamic described above. (We comment on the interpretation of this below.) If every action can be "mutated to," we obtain a Markov process which has a strictly positive probability of moving from any state to any other. Since every such process has a unique invariant distribution to which the system converges from *any* starting point, the addition of mutations makes the limit of the process unique (as a function of ε)! KMR and Young consider a sequence of ε's converging to zero and analyze the limit point of the associated sequence of invariant distributions, which they call a *long-run equilibrium*. They and many others have provided characterizations of this unique long-run equilibrium for various classes of games.

Given the dramatic impact of adding mutations, one naturally wonders what mutations are supposed to represent. The most common answer in the literature, experimentation, is difficult to reconcile with the assumption that the mutation rate is constant across states, across agents, and over time. Surely agents experiment less in states with high payoffs or when they have a great deal of experience.

Of course, if this restriction is not important to the results, there is no need to provide a model of what mutations are. We explore the implications of letting mutation rates vary with the state of the system, but do not allow time varying rates. Within this seemingly slight generalization of KMR/Young, we show that given *any* long-run prediction from the system without mutations and *any* model of the way mutations affect evolution, it is always possible to find small mutation rates such that the long-run behavior of the system with mutations is arbitrarily close to the given prediction without mutations. In other words, any refinement effect from adding mutations is solely due to restrictions on how mutation rates vary across states. This result highlights the importance of developing models or other criteria to determine "reasonable" mutation processes.

The intuition for this result is simple. KMR consider games with two pure strategy equilibria and hence two states into which the system could settle in the long-run in the absence of mutations. In KMR/Young, the long-run behavior of the system with mutations is determined by the number of mutations it takes to force the system from one of these states to the other. The one which is more easily disrupted by mutations in this sense is given zero probability as the mutation rate goes to zero. Here, though, the mutation rates could differ in these two states. So even though one state may be more robust than the other in

[3] See also Blume (1993, 1995), Canning (1992), Ellison (1993), Kandori and Rob (1992), Lagunoff and Matsui (1994), Nöldeke and Samuelson (1993), Robson (1994), Robson and Vega-Redondo (1994), Samuelson (1994), and Young (1993b).

terms of the *number* of mutations, it may be less robust in terms of the *probability* of these mutations, which is, of course, the crucial factor.

Because this reasoning quickly becomes complex as the number of long-run (absorbing) states increases, we do not exploit this intuition directly. Our approach enables us to provide a simple yet comprehensive analysis, treating the way mutations affect evolution very generally and allowing for any of the various ways mutations are added in the literature. It also enables us to highlight where further research is needed. The fact that our theorem holds for almost any formulation of how mutations affect evolution means that the qualitative properties of the effect of mutation is irrelevant to the refinement effect of KMR and Young. Instead, it is their implicit requirement that the magnitude of the effect of mutation—that is, the mutation rate—does not vary "too much" across states which is crucial. More precisely, the key is their requirement that all mutation rates converge to zero at the same rate. Hence to generate more precise predictions, we must generate economically interesting conditions on the mutation rates.

There are several reasons why mutation rates may vary with the state of the system. As noted, if mutation is intended to represent experimentation as suggested in much of the literature, it is only natural to allow experimentation rates to vary with the payoffs being earned by and the experience of the players, both of which vary across states. If mutations are interpreted as computational error, it seems reasonable to suppose that players are more likely to make computational errors in more complex situations. Similarly, traditional formulations of trembles (Selten (1975) and Myerson (1978)) allow tremble rates to vary across information sets. Also, if the underlying mutationless dynamics are "stronger" in some states than others (perhaps because movements in the direction of best replies are more obvious or more rewarding in some states), then the effective mutation rates may differ across states.

Because we wish to show that it is only through restrictions on the mutation rates that adding mutations can refine predictions, we impose as few restrictions *a priori* as possible. In particular, we allow mutation rates for different states to converge to zero at different rates. As mentioned above, these differences in convergence rates are important to the result. If our only change to the KMR/Young framework were to write the mutation rate for state i as some fixed constant a_i times a "system wide" mutation rate which is taken to zero, then the analysis in KMR and Young would be essentially unaffected,[4] as we explain in more detail later.

There are many views one can take on this aspect of our modeling. One view is that sequences of mutation rate vectors with equal mutation rates across states are "nongeneric" and hence unlikely to emerge from a robust model. Another is that the argument Myerson gave for properness applies with equal strength here, so that costlier mutations should have probabilities going to zero relative to the probability of less costly mutations. Yet another argument in

[4] More precisely, the support of the limiting distribution would be independent of the a_i's.

favor of state-dependent and time-varying mutation rates is the fact that in standard Bayesian learning models, different errors go to zero at different rates over time.[5] Alternatively, one might maintain that mutation should be "pure noise," suggesting that all mutation rates should be the same.

Our view is that in the absence of a concrete model of what mutations are and why they arise, one cannot reject any of these arguments. Our purpose is to show that the "model-less" approach to mutation cannot be justified and hence to motivate further research on these questions. In other words, determining what mutation rates are reasonable is not necessary for our result to be of interest. To the contrary, it is the result that makes this determination interesting.

In Section 2.1, we give an example to illustrate our results. In Section 2.2, we present the general model. Our main result is stated and proved in Section 3. We conclude in Section 4.

2. THE MODEL AND MOTIVATING EXAMPLES

2.1. *Examples*

The insight of KMR and Young, following on Foster and Young (1990), is that adding small probabilities of mutation to an otherwise standard evolutionary process can yield a unique long-run prediction. To see this more concretely, consider the following game:

$$\begin{array}{c} \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 8,8 & 0,4 \\ 4,0 & 6,6 \end{pmatrix} \end{array}.$$

Suppose N agents play this game. In each period, each player must choose a single action which he uses when playing against each of the other $N-1$ agents. Suppose the action chosen is a best reply given the actions chosen by the other players in the previous period.

It is not hard to see that the number of agents playing each strategy in a given period completely determines the future evolution of the strategies chosen. Hence we can represent this evolutionary dynamic as a Markov process with a state space given by the number of agents playing, say, strategy 2. So let $S = \{0, 1, \ldots, N\}$. It is easy to calculate the transition matrix $\boldsymbol{P}$ describing how the system moves between states. Letting p_{ij} denote the probability of moving from state i one period to state j the next, there is an integer i^* such that $p_{i0} = 1$ if $i < i^*$, $p_{iN} = 1$ if $i > i^*$, and $p_{i^*, N-i^*} = 1$.[6] Intuitively, if few agents play strategy 2, then all switch to strategy 1, while if many play strategy 2, all switch to 2.

An invariant distribution is a probability distribution on S, say q, which satisfies $q = q\boldsymbol{P}$. Such a distribution may be viewed as a "steady state" for the

[5] See Blume (1995) for analysis of time-varying mutation rates.

[6] This calculation assumes that $(2N-2)/5$ is not an integer for simplicity.

population. For this process, the set of invariant distributions is $\{q \mid q = \theta\delta_0 + (1-\theta)\delta_N\}$, where δ_i is a probability distribution S putting probability 1 on state i. Note that the invariant distributions correspond to probability distributions over states 0 and N—that is, over the two strict Nash equilibria.

Now let us introduce mutations. KMR, Young, and others assume that the probability of a mutation is some fixed ε, independent of time, the current state, or the agent. If an agent does not mutate, he changes strategy or not according to the dynamic described above. If he does mutate, he changes strategies with some fixed probability. For simplicity, we assume that a mutating agent chooses the opposite strategy from what the dynamic without mutations would specify. Mutations are independent events across agents and over time. It is not difficult to calculate the implied transition probabilities as a function of ε. One can use these and Lemma 1 of KMR to show that the unique distribution, say $q^* = (q_0^*, \ldots, q_N^*)$, has $\lim_{\varepsilon \downarrow 0} q_N^* = 1$.

Suppose we relax these assumptions by letting the mutation rate vary with the state. If mutation is intended to model experimentation by the players, then it is difficult to see why the mutation rate in state 0 would be as large as the mutation rate in other states. In state 0, all agents play strategy 1 and so all always earn the highest possible payoff in the game. So why would players experiment in this state? By contrast, in state N, players may experiment in hopes of reaching the other (Pareto preferred) Nash equilibrium. While we do not wish to claim that this is a necessary property of mutation rates, a reasonable model of mutations would surely allow this possibility.

In line with this intuition, suppose that the mutation rate in states i^*, $i^*+1, \ldots, N$ is ε, while the mutation rate in states $0, 1, \ldots, i^*-1$ is ε^α for some positive constant α. The interesting case is where $\varepsilon^\alpha < \varepsilon$ or $\alpha > 1$, although we do not impose this yet. Except for this change, we maintain all the assumptions from above. It is not difficult to show that the unique invariant distribution, say q', satisfies

$$\frac{q'_0}{q'_N} = \varepsilon^{N-i^*+1-\alpha i^*} \frac{k_1[1+f_1(\varepsilon)]}{k_2[1+f_2(\varepsilon)]}$$

where f_1 and f_2 go to zero as $\varepsilon \to 0$. Also, $q'_i \to 0$ as $\varepsilon \to 0$ for all $i \neq 0$ or N. Clearly, if $\alpha > (N-i^*+1)/i^* \equiv \alpha^*$, then the exponent is negative, so this ratio goes to ∞. Hence $q'_0 \to 1$. If $\alpha < \alpha^*$, then $q'_N \to 1$. Finally, if $\alpha = \alpha^*$, then $q'_0/q'_N \to k_1/k_2$ as ε and ξ go to zero.[7] One can show that α^* converges to $3/2$ as $N \to \infty$. Hence for N large, setting $\alpha > 3/2$ reverses the KMR/Young prediction.

To understand these results, consider the process without mutations. Regardless of where the system starts, it will ultimately converge to either state 0 forever or state N forever. The key to the KMR/Young analysis is the comparison between the basin of attraction for state 0—that is, the set of states

[7] If we took the mutation rate in the lower states to be $\beta\varepsilon^\alpha$, we could vary β to trace out all possible probabilities over states 0 and N.

from which the system must ultimately converge to state 0—and the basin of attraction for state N. It is easy to see that the basin of attraction for state 0 consists of the states from 0 up to $i^* - 1$, while the basin of attraction for state N contains the remaining states. It is not hard to show that $i^* < N/2$, so the latter basin contains more than half the states.

When mutation rates are small but nonzero, the system will spend most of its time in either state 0 or state N. For concreteness, suppose the system begins in state 0. It will stay there until mutations shift the system into the basin of attraction for state N. Then the system will drift to state N where it will remain until mutations shift it over to the basin of attraction for state 0. Consider the case where the mutation rates are the same in every state. Because the basin of attraction for state N is larger than that for state 0, the system will shift out of state 0 more quickly than it shifts out of state N. Hence the system will spend most of its time in state N. As the mutation rate goes to zero, the fraction of the time spent in state N converges to one.

When the mutation rates vary across states, the size of the relevant basins of attraction is no longer enough to determine where the system is most of the time. In particular, even though the basin of attraction for state 0 is smaller, it may be "deeper" in the sense that mutations out of this basin are less likely. In this case, the system may well spend most of its time in state 0 instead of state N, reversing the KMR/Young result.

Unfortunately, a general characterization along these lines is quite difficult. The KMR/Young approach is combinatoric, involving the calculation of "minimum cost" paths from one absorbing state to another. As the number of absorbing states increases, the approach rapidly becomes very complex. Instead, we take a more general approach which simplifies the analysis.

2.2. *The Model*

Our theorem does not require a detailed specification of the evolutionary process without mutations. We assume only that it is a finite Markov process. As in Section 2.1, any system with finitely many agents each with finite memory and a finite strategy set can be written as a finite state Markov process. The state space is $S = \{1, \ldots, s\}$ and the transition matrix $\boldsymbol{P}$, where p_{ij} is the probability of a transition from state i to state j. We refer to $\boldsymbol{P}$ as the *mutationless process*.

Let Δ denote the set of probability distributions on S. A long-run prediction about the mutationless process is an invariant distribution.

DEFINITION 1: An *invariant distribution of* $\boldsymbol{P}$ is a vector $q \in \Delta$ such that $q = q\boldsymbol{P}$.

Note that q is a *row* vector. Denote by $\mathcal{I}(\boldsymbol{P})$ the set of invariant distributions of $\boldsymbol{P}$.

Intuitively, if q is our belief about the state of the system and we learn that a period of time has passed (without learning anything directly about the state), then our updated beliefs qP will be unchanged if q is invariant.

The following definitions allow us to characterize the invariant distributions.

DEFINITION 2: A subset C of S is *absorbing in* P if for all $i \in C$, $j \notin C$, $p_{ij} = 0$. C is a *minimal absorbing set of* P if it is absorbing in P and no subset of it is.

DEFINITION 3: A state $i \in S$ is *transient in* P if it is not contained in any minimal absorbing set of P.

It is easy to see that a state $i \in S$ is transient if there is a j such that $p_{ij} > 0$ but j is either transient or in an absorbing set which does not include i.

A well-known fact about Markov processes[8] is the following.

FACT: *There are numbers* μ_i *for each nontransient state* i *such that* $q \in \Delta$ *is an invariant distribution of* P *if and only if* (*a*) $q_j = 0$ *for every transient state* j *and* (*b*) *for every minimal absorbing set* C, *either* $q_j = 0$ *for all* $j \in C$ *or* $q_i / \Sigma_{j \in C} q_j = \mu_i$.

In other words, all probability is concentrated on the nontransient states. Furthermore, given any invariant distribution, the restriction of the distribution to a minimal absorbing set, if well-defined, is unique. Hence all differences between invariant distributions correspond to differences in the relative probabilities of different minimal absorbing sets.

Most mutationless evolutionary dynamics in the literature have the property that every strict Nash equilibrium corresponds to a minimal absorbing set, as in the example in Section 2.1. By the Fact, we see that for any game with more than one strict Nash equilibrium, these dynamics have more than one minimal absorbing set and hence have infinitely many invariant distributions.

The Fact also implies that if the only minimal absorbing set is S, then there is a unique invariant distribution. In this case, P is said to be *irreducible*. When P is irreducible, the empirical frequency distribution over states converges to the unique invariant distribution with probability 1 from any initial condition. While the mutationless processes considered in the literature are typically not irreducible, the processes generated by adding mutations are. This effect is quite intuitive: adding mutations makes it possible for the system to eventually move from any one state to any other.

We call a procedure for determining transition probabilities from mutation rates a "model of mutations." We emphasize that a model of mutations simply describes how mutations affect evolution, not why mutations occur. Rather than adopt a specific model of how the addition of mutations affects evolution, we prove our theorem for *any* model of mutations. In other words, the model of

[8] See, e.g., Iosifescu (1980) for proof.

mutations is held fixed, just as the mutationless process is—the only variable we use in approximating a particular long-run prediction is the vector of mutation rates.

Our theorem will cover any model of mutations satisfying a few simple properties. First, the process with mutations must be irreducible. Second, the process with mutations converges smoothly to the mutationless process as mutation rates go to zero. Third, the mutation rates may be state dependent with only the probability of a mutation in state i relevant for determining movements out of state i via mutation.

To understand the motivation for the third condition, consider the process with mutation and suppose the current state is i. How are the probabilities of transition to other states determined? By assumption, the state summarizes all relevant aspects of the system—the information the players have, the payoffs they receive, etc. For this reason, if a player does not mutate, his behavior is completely determined by the state. Similarly, if a player does mutate, his behavior conditional on this event is completely determined by the state. Finally, it is the state i mutation rate which determines the relative probabilities of these two events. Hence no mutation rate other than the state i mutation rate is relevant for transitions from state i. Technically, this is not necessary for the result—transition probabilities out of one state can depend on the mutation rates for other states. However, the interpretation of mutation rates is less clear in this case. See Remark 2.

Let $\mathcal{M}$ denote the set of Markov matrices on S. We will call a vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_s)$ with $\varepsilon_i \in [0,1]$ for all i a vector of *mutation rates*.

DEFINITION 4: A *model of mutations* for $\boldsymbol{P}$ is a continuous function $\boldsymbol{M}$: $[0,1]^S \to \mathcal{M}$ such that (a) $\boldsymbol{M}(0) = \boldsymbol{P}$, (b) $\boldsymbol{M}(\varepsilon)$ is irreducible for all $\varepsilon \gg 0$, and (c) the elements of the ith row of $\boldsymbol{M}(\varepsilon)$ depend only on ε_i.

These three assumptions have KMR and Young as special cases, as well as all the other models we are aware of in this literature.

To determine the importance of restrictions on mutation rates, we characterize the set of predictions one can generate with arbitrary small mutation rates given some fixed model of mutations. This motivates the following definition.

DEFINITION 5: A probability distribution $q \in \Delta$ is *achievable* with mutation model $\boldsymbol{M}$ if there exists a sequence of strictly positive mutation rate vectors $\varepsilon^n \to 0$ such that $q^n \to q$ where $\{q^n\} = \mathcal{S}(\boldsymbol{M}(\varepsilon^n))$. Let $\mathcal{A}(\boldsymbol{M})$ denote the set of achievable distributions with mutation model $\boldsymbol{M}$.

Intuitively, an achievable distribution is a q solving $q = q\boldsymbol{P}$ which is "robust" in the sense that we can find a sequence $\{M^k\}$ of irreducible transition matrices converging to $\boldsymbol{P}$ such that the sequence of distributions $\{q^k\}$ uniquely solving $q^k = q^k M^k$ converges to q. An important point to note is that the mutation model acts as a constraint on our choice of $\{M^k\}$ in that we require each M^k to

equal $M(\varepsilon^k)$ for some $\varepsilon^k \gg 0$, where $\varepsilon^k \to 0$. Proving our theorem without this restriction is very straightforward.

KMR and Young require ε_i to be independent of i. Hence they, in effect, consider only a single sequence of ε vectors. We do not impose this restriction and so have many sequences.

REMARK 1: The reader may find an analogy useful. When defining trembling-hand perfect equilibria, one first calculates ε-perfect equilibria where agents may make mistakes. One then considers the limit as these mistake probabilities go to zero. Any sequence of mistake probabilities converging to zero is allowed. An ε-perfect equilibrium is analogous to an invariant distribution with mutation rate vector ε and an achievable distribution is analogous to a perfect equilibrium. As with perfection, we allow for any sequence of mutation rate vectors going to zero.

3. THE MAIN RESULT

Our main result is that with state-dependent mutation rates, any invariant distribution of the mutationless process is achievable with any mutation model.

THEOREM 1: *If* $\boldsymbol{M}$ *is any mutation model for* $\boldsymbol{P}$, *then* $\mathscr{A}(\boldsymbol{M}) = \mathscr{I}(\boldsymbol{P})$.

To understand the proof intuitively, suppose that $\boldsymbol{P}$ is the identity matrix, so $\mathscr{I}(\boldsymbol{P}) = \Delta$. Note that any mutation model $\boldsymbol{M}$ defines a correspondence from vectors of mutation rates to Δ giving the set of invariant distributions as a function of the mutation rates. Given any $\eta > 0$, we can analyze this function for the set of ε vectors summing to η. Given this restriction, we can "rescale" the mutation rates by η so that they add to 1, changing this to a correspondence from Δ to itself. Call this correspondence G_η and a typical "rescaled" mutation rate vector x. We show that G_η is onto—that is, for every $q \in \Delta$, there is an x such that $q = G_\eta(x)$. Hence we can find a vector of mutation rates, summing to as small a number as desired, which generate q as the unique invariant distribution.

An important point is that we use very little information about how G_η behaves on the interior. This is a necessary aspect of the proof because it is difficult to characterize how the invariant distribution varies with ε for $\varepsilon \gg 0$.

PROOF OF THEOREM 1: Let Δ° denote the interior of Δ (the set of distributions with strictly positive probability on all points) and $\partial\Delta$ the boundary (the set of distributions with at least one state given zero probability). Lemma 1 gives sufficient conditions for a correspondence to be onto.

LEMMA 1: *If* $G\colon \Delta \to \Delta$ *is an upper semicontinuous correspondence which maps* $\partial\Delta$ *to itself, is single-valued on* Δ° *and convex-valued on* $\partial\Delta$, *and has no fixed points on the boundary, then* $G(\Delta) = \Delta$.

PROOF: Suppose $G(\Delta) \neq \Delta$. Clearly, since G is upper semicontinous, $G(\Delta)$ is closed. Hence $\Delta \setminus G(\Delta)$ is open. Choose any element, say b, from the interior of this set.

Define a function $\rho_b: G(\Delta) \to \partial\Delta$ as follows. Given any $x \in G(\Delta)$, draw a straight line from b through x to the boundary and let $\rho_b(x)$ be the point where this line intersects the boundary. Clearly, ρ_b is a continuous function on $G(\Delta)$. Note also that for any $x \in \partial\Delta$, $\rho_b(x) = x$.

Define a correspondence $G_b: \Delta \to \Delta$ by $G_b(x) = \rho_b(G(x))$. Since G is upper semicontinuous and ρ_b is continuous, G_b is upper semicontinuous as well. Furthermore, since G is single-valued on Δ°, G_b is as well. For any $x \in \partial\Delta$, $G(x) \subseteq \partial\Delta$. Hence $G_b(x) = G(x)$ for any $x \in \partial\Delta$. Since $G(x)$ must be convex, we see that G_b is convex-valued.

By the Kakutani fixed point theorem, then, there is some x^* such that $x^* \in G_b(x^*)$. Since ρ_b maps to the boundary of the simplex, we must have $G_b(x^*) \subseteq \partial\Delta$, so $x^* \in \partial\Delta$. But as just argued, for any $x \in \partial\Delta$, $G_b(x) = G(x)$. Hence $x^* \in \partial\Delta$ and $x^* \in G(x^*)$. But by hypothesis, G has no fixed point on the boundary. Hence $G(\Delta) = \Delta$. *Q.E.D.*

To complete the proof, we focus first on the case where all minimal absorbing sets are singletons. At the end of the proof, we explain the extension to the general case. We prove the theorem by constructing a particular correspondence which satisfies the conditions of Lemma 1.

Let $N \subseteq S$ denote the absorbing states of $\boldsymbol{P}$ and let $T = S \setminus N$. For convenience, we often write $\varepsilon = (\varepsilon_N, \varepsilon_T)$. Let Δ_N denote the set of probability distributions on N. Also, let Δ_N° denote the interior of Δ_N and $\partial\Delta_N$ its boundary. Fix any mutation model $\boldsymbol{M}$. Given any $\varepsilon \gg 0$, let $I(\varepsilon) \in \Delta$ denote the unique invariant distribution for the process $\boldsymbol{M}(\varepsilon)$.

Lemma 2 will be used to establish upper semicontinuity of the correspondence we will construct shortly.

LEMMA 2: *If* $\varepsilon^k \to \varepsilon^*$, $\varepsilon^k \gg 0$ *for all* k, *then for every convergent subsequence* $\{I(\varepsilon^{k_j})\}$ *of* $\{I(\varepsilon^k)\}$,

$$\lim_{j \to \infty} I(\varepsilon^{k_j}) \in \mathscr{I}(\boldsymbol{M}(\varepsilon^*)).$$

Therefore, $\mathscr{A}(M) \subseteq \mathscr{I}(P)$.

PROOF: Let $q^k = I(\varepsilon^k)$ and $M^k = \boldsymbol{M}(\varepsilon^k)$. Without loss of generality, suppose $q^k \to q$. Because $\boldsymbol{M}$ is continuous, $M^k \to \boldsymbol{M}(\varepsilon^*)$ (entry by entry). By definition, $q^k M^k = q^k$, so passing to the limit, $q\boldsymbol{M}(\varepsilon^*) = q$ or

$$\lim_{k \to \infty} I(\varepsilon^k) \in \mathscr{I}(\boldsymbol{M}(\varepsilon^*)).$$ *Q.E.D.*

Lemma 3 will be used to show that the correspondence we will construct will satisfy the boundary properties required in Lemma 1.

LEMMA 3: *If* $\varepsilon_i = 0$ *for some* $i \in N$ *and* $\varepsilon_l > 0$ *for all* $l \in T$, *then for all* j *with* $\varepsilon_j > 0$ *and all* $q \in \mathcal{I}(M(\varepsilon))$, *we have* $q_j = 0$.

PROOF: Suppose that $\varepsilon_i = 0$ for some $i \in N$ and $\varepsilon_l > 0$ for all $l \in T$. By assumption, the jth row of $M(\varepsilon)$ depends only on ε_j. Since $M(0) = P$, $\varepsilon_i = 0$ implies that the ith row of $M(\varepsilon)$ is identical to the ith row of P. Because $i \in N$, this row must contain a 1 in the ith place and 0's elsewhere. Therefore, i must be an absorbing state of $M(\varepsilon)$. Consider any j with $\varepsilon_j > 0$. Fix any $\varepsilon^* \gg 0$ such that $\varepsilon_l^* = \varepsilon_l$ for all l with $\varepsilon_l > 0$. By our assumptions on mutation models, $M(\varepsilon^*)$ is irreducible, so the probability of transiting from state j to state i in a finite number of periods is strictly positive in $M(\varepsilon^*)$. If every state in this sequence has a strictly positive mutation rate in ε, then $M(\varepsilon)$ still gives this sequence of states strictly positive probability since the relevant rows must be the same as the corresponding rows of $M(\varepsilon^*)$. In this event, state j must be transient in $M(\varepsilon)$. If some state in this sequence has a zero mutation rate, then, since it must also be in N by hypothesis, it, too, is absorbing, again establishing that state j is transient in $M(\varepsilon)$. Hence, for any $q \in \mathcal{I}(M(\varepsilon))$, Fact 1 implies that $q_j = 0$. *Q.E.D.*

We are now ready to construct the key correspondence. Define $\alpha: \Delta \to \Delta_N$ by

$$\alpha_i(q) = q_i + \frac{1}{\#T} \sum_{j \in T} q_j.$$

Next, for any $\eta > 0$ and $\varepsilon_T \gg 0$, we construct a correspondence $G_{\eta, \varepsilon_T}: \Delta_N \to \Delta_N$ by

$$G_{\eta, \varepsilon_T}(x) = \alpha(\mathcal{I}(M(\eta x, \varepsilon_T))).$$

For notational simplicity, we often omit the η, ε_T subscripts. In other words, fix the mutation rates for the transient states at ε_T and fix the *total* of the mutation rates for the nontransient states at η. Given a vector x in Δ_N, we construct a vector of mutation rates by "rescaling" x by η and then adding to the vector the mutation rates for the transient states. We then use this to calculate the set of invariant distributions associated with these mutation rates. Finally, we use α to convert these distributions to probability distributions on N. For convenience, we occasionally write $\mathcal{I}(x)$ for $\mathcal{I}(M(\eta x, \varepsilon_T))$, so that $G(x) = \alpha(\mathcal{I}(x))$.

It is not hard to show that G satisfies the requirements of Lemma 1. It is single-valued on the interior because $(\eta x, \varepsilon_T) \gg 0$ when $x \gg 0$. Lemma 2 and the continuity of α imply that G is upper semicontinuous.

To characterize G at some $x \in \partial\Delta_N$, note that there must be some $i \in N$ with $x_i = 0$. Hence by Lemma 3, for every $j \in T$ and every j with $x_j > 0$, we must have

$q_j = 0$ for all $q \in \mathcal{S}(x)$. But then for $i \in N$ and $q \in \mathcal{S}(x)$,

$$\alpha_i(q) = q_i + \frac{1}{\#T} \sum_{j \in T} q_j = q_i.$$

That is, for $q \in \mathcal{S}(x)$, $\alpha(q)$ is simply the identity mapping projecting q onto N, since q is zero off N. Because the set of invariant distributions for any given transition matrix is convex, this means that G is convex-valued on the boundary. Furthermore, fix any j such that $x_j > 0$. Lemma 3 implies that $q_j = 0$ and so $\alpha_j(q) = 0$ for all $q \in \mathcal{S}(x)$. Hence G maps the boundary to the boundary and has no fixed points on the boundary. Therefore, by Lemma 1, G is surjective.

To complete the proof of the Theorem, note that the assumption that all minimal absorbing sets are singletons implies that

$$\mathcal{S}(\boldsymbol{P}) = \{q \in \Delta \mid q_i = 0, \forall i \in T\}.$$

Clearly, then, $\mathcal{S}(\boldsymbol{P})$ can be viewed as one-to-one with Δ_N. We first focus on strictly positive distributions, so fix any $q \in \Delta_N^\circ$.

Since G maps the boundary of Δ_N to the boundary, surjectivity implies that $G_{\eta, \varepsilon_T}(\Delta_N^\circ) = \Delta_N^\circ$ for every $\eta > 0$ and $\varepsilon_T \gg 0$. Fix any sequences η^k and ε_T^k going to zero from above. Clearly for each k, there is an $x^k \gg 0$ such that $\{q\} = G_{\eta^k, \varepsilon_T^k}(x^k) = \alpha(I(\eta^k x^k, \varepsilon_T^k))$. For $i \in N$, let $\varepsilon_i^k = \eta^k x_i^k$. Then

$$q_i = \lim_{k \to \infty} \alpha_i(I(\varepsilon^k)) = \lim_{k \to \infty} I_i(\varepsilon^k) + \frac{1}{\#T} \sum_{j \in T} I_j(\varepsilon^k) = \lim_{k \to \infty} I_i(\varepsilon^k)$$

where the last equality follows from $\varepsilon^k \to 0$. Hence any $q \in \Delta_N^\circ$ is achievable.

Therefore, the interior of $\mathcal{S}(\boldsymbol{P})$ is contained in $\mathcal{A}(\boldsymbol{M})$. It is easy to see from the definition of $\mathcal{A}(\boldsymbol{M})$ that it must be closed. Hence this establishes that $\mathcal{S}(\boldsymbol{P}) \subseteq \mathcal{A}(\boldsymbol{M})$. Since the converse was established by Lemma 2, this shows that the two sets are equal.

To see how to handle the case where the minimal absorbing sets are not singletons, recall from Fact 1 that the invariant distributions *within* each minimal absorbing set are uniquely defined. Hence we only need to ensure that we can achieve any distribution *across* minimal absorbing sets. To do so, redefine N above to be the collection of minimal absorbing sets instead of the set of absorbing states and replace S with $S^* = N \cup T$. We can then focus on mutation rates with the property that for any two states i and j in the same minimal absorbing set of $\boldsymbol{P}$, $\varepsilon_i = \varepsilon_j$. Instead of generating invariant distributions on S, we can focus only on the induced distributions on S^*. It is not hard to rewrite Lemma 3 and the construction of G to conclude that any distribution across minimal absorbing sets is achievable, so that any invariant distribution of $\boldsymbol{P}$ is achievable. *Q.E.D.*

In short, with no restrictions on how mutation rates vary across states, mutations have no refinement effect at all. Hence the "model-less" approach to

mutations does not seem justifiable. We have argued that without a model, there is no obvious reason to impose any particular constraints on mutation rates, including the constraint that mutation rates all go to zero at the same rate.

Different rates of convergence are important to the result. To see why, suppose that we follow KMR/Young in assuming that the mutation model $\boldsymbol{M}$ is polynomial in the mutation rates. More precisely, the transition probability from state i to state j given mutation rate vector ε is some polynomial in ε_i. Suppose we impose the restriction that the mutation rates go to zero at the same rate in the simplest possible way by requiring the mutation rates to be constant multiples of one another—that is, that $\varepsilon_i = a_i \xi$ for some constant $a_i > 0$ where ξ converges to zero. It is not hard to show that the "resistances" computed by KMR/Young in their analysis of the limiting behavior of the system are independent of the a_i's. Hence the support of the limiting invariant distribution as ξ goes to zero must be independent of the a_i's. (This is essentially the proof Samuelson (1994) gives for his Theorem 4.)

This fact can be interpreted in a variety of ways. On the one hand, one might argue that the mutation rates should all go to zero at the same rate and that the polynomial restriction is not unreasonable. If so, one would conclude that the KMR/Young technique is quite robust. On the other hand, as noted in the introduction, there are many possible views on the "reasonableness" of various assumptions on mutation rates, none of which seem to have any priority over the others in the absence of a concrete model describing what mutations are and why they occur. For example, it seems quite natural to adapt Myerson's (1978) argument to conclude that more costly "experimentation errors" should have probabilities going to zero more quickly.

Also, Blume (1995) finds a similar multiplicity of long-run outcomes when the mutation rate for a state is quickly decreasing in the number of times the system has visited that state. Intuitively, if the system stays in one state for a long time, we would expect very little further experimentation while there, so that state's mutation rate would be very low. Hence the states that get visited often in the early periods may have much smaller mutation rates from then on. This will have an effect very similar to what we get by computing invariant distributions for a fixed vector of mutation rates and then taking these rates to zero with some mutation rates going to zero more quickly than others.

REMARK 2: Our assumptions on mutation models can be substantially relaxed with minor changes in the proof of Theorem 1. First, the assumption that $\boldsymbol{M}(\varepsilon)$ is irreducible for all $\varepsilon \gg 0$ can be weakened to requiring only that for every $\varepsilon \gg 0$, $\boldsymbol{M}(\varepsilon)$ has a unique minimal absorbing set containing all nontransient states. Second and more importantly, our assumption that the ith row of $\boldsymbol{M}(\varepsilon)$ only depends on ε_i can be relaxed to the following less intuitive requirement: If $A, B \subseteq S$ are disjoint, absorbing sets in $\boldsymbol{P}$, then

(i) $\varepsilon_i = 0$, $\forall i \in A \Rightarrow A$ is absorbing in $\boldsymbol{M}(\varepsilon)$, and

(ii) $\varepsilon_i > 0$, $\forall i \in A \cup T$ and $\varepsilon_i = 0$, $\forall i \in B \Rightarrow i$ is transient in $M(\varepsilon)$, $\forall i \in A$,

where T is the set of transient states of P. This condition essentially assumes the result of Lemma 3.

4. CONCLUSIONS

We have shown that if mutation rates can vary in any fashion with the state, then, regardless of what one assumes about how mutation affects evolution, adding small mutation probabilities does not refine the set of long-run predictions at all. In this sense, it is only restrictions on the mutation rates that can allow mutation to refine predictions. One could interpret this result as saying that it is futile to try to refine the set of long-run predictions by adding mutation. An alternative interpretation is that the nature of the mutation process must be analyzed more carefully to derive some economically justifiable restrictions. Blume (1995) provides some ideas along these lines. It is an open question whether and what kinds of interesting restrictions will emerge.

Dept. of Economics, Queen's University, Kingston, Ontario, Canada, K7L 3N6

Manuscript received June 1994; final revision received August, 1995.

REFERENCES

BLUME, L. (1993): "The Statistical Mechanics of Strategic Interaction," Cornell University Working Paper.

——— (1995): "How Noise Matters," Cornell University Working Paper.

CANNING, D. (1992): "Learning Language Conventions on Common Interest Signaling Games," Columbia University Working Paper.

ELLISON, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047–1071.

FOSTER, D., AND P. YOUNG (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology*, 38, 219–232.

IOSIFESCU, M. (1980): *Finite Markov Processes and Their Applications*. New York: John Wiley and Sons.

KANDORI, M., G. MAILATH, AND R. ROB (1993): "Learning, Mutation, and Long-Run Equilibria in Games," *Econometrica*, 61, 29–56.

KANDORI, M., AND R. ROB (1992): "Evolution of Equilibria in the Long Run: A General Theory and Applications," University of Pennsylvania Working Paper.

LAGUNOFF, R., AND A. MATSUI (1994): "The Evolution of Cooperation in Mechanisms for Public Projects," University of Pennsylvania Working Paper.

MYERSON, R. (1978): "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7, 73–80.

NÖLDEKE, G., AND L. SAMUELSON (1993): "An Evolutionary Analysis of Backward and Forward Induction," *Games and Economic Behavior*, 5, 425–454.

ROBSON, A. (1994): "The Adam and Eve Effect and Fast Evolution of Efficient Equilibria in Coordination Games," University of Western Ontario Working Paper.

ROBSON, A., AND F. VEGA-REDONDO (1994): "Random Pairing and Long-Run Equilibria in Games," University of Western Ontario Working Paper.

SAMUELSON, L. (1994): "Stochastic Stability in Games with Alternative Best Replies," *Journal of Economic Theory*, 1, 35–65.

SELTEN, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

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

——— (1993b): "An Evolutionary Model of Bargaining," *Journal of Economic Theory*, 59, 145–168.