Stochastic Evolutionary Stability in Generic Extensive Form Games of Perfect Information

Christoph Kuzmics*

University of Cambridge
Faculty of Economics and Politics
Sidgwick Avenue
Austin Robinson Building
Cambridge CB3 9DD
UK
email:cak26@econ.cam.ac.uk

December 4, 2002

*I am grateful to Hamid Sabourian and Florian Herold for helpful comments and suggestions.
Abstract

Generic extensive form games of perfect information have a unique sub-game perfect equilibrium. Nöldeke and Samuelson (1993) show that in a stochastic evolutionary model also non-subgame perfect equilibrium-strategies may well survive in the long run. In a different model of evolution in the agent normal form of generic extensive form games of perfect information Hart (2002) shows that under suitable limit-taking, where small mutation rates are accompanied by large population sizes in a particular way, the unique prediction is again the subgame perfect equilibrium. This paper provides a proof of a similar result for the model of Nöldeke and Samuelson (1993).

JEL classification: C62, C72, C73

Keywords: backward induction, learning, experimentation, subgame-perfect equilibrium
Notation

$N$ ... set of nodes
$M(i)$ ... population of individuals at node $i$
$m_i = |M(i)|$
$m = (m_1, m_2, ..., m_N)$
$A(i)$ ... action set at node $i$
$b^i \in A(i)$ ... backward induction action at node $i$
$\lambda_x$ ... conditional probability that if an agent mutates she or he does not mutate to action $x \in A(i)$
$\Omega$ ... state space
$\sigma$ ... probability of a learn draw
$\mu$ ... probability of mutation
$Q^n_\mu$ ... transition probability matrix of Markov chain on $\Omega$
$\pi^n_\mu$ ... invariant distribution
$\Lambda_{i,x}^k$ ... set of states in which the proportion of individuals at node $i$ playing action $x$ is $\frac{k}{m_i}$
$R(i)$ ... set of predecessor nodes of node $i$
$a^j_i \in A(j)$ ... unique action at node $j$ which (eventually) leads to node $i$
$S(i)$ ... set of successor nodes of node $i$
$C^b_x$ ... set of states such that $x \in A(i)$ is not the unique best reply for any agent at node $i$ given their conjectures
$D^b_x$ ... set of states such that $x \in A(i)$ is not a best reply for any agent at node $i$ given their conjectures
$B^i_{\epsilon,m}$ ... set of states, in which more than a fraction of $1 - \epsilon$ individuals at node $i$ play $b^i$
$B^i_{\epsilon,m} \cup \Omega^c$ ... the $\Omega$-complement of $B^i$
$B^i_{\epsilon,m} = \bigcap_{i \in N} B^i_{\epsilon,m}$
t_{kj} ... transition probability of a move from a state in $\Lambda_{i,b^i}^{i,b^i} \setminus C^i_{b^i}$ to any state in the set $\Lambda_{j,b^i}^{i,b^i}$
$j_s = \lceil (1 - \epsilon)m_i \rceil$
for $A \subset \Omega$, $A^c$ denotes its $\Omega$-complement
1 Introduction

Generic extensive form games of perfect information have a unique subgame perfect equilibrium. This equilibrium coincides with the unique trembling hand perfect equilibrium in these games. One might think that random mutations, as in Kandori, Mailath, and Rob (1993) or Young (1993), in evolutionary models should serve the same purpose as perturbations in the rational formulation and lead to virtually all individuals playing their backward induction action. Nöldeke and Samuelson (1993), however, show that this is not generally the case. Indeed they construct an example where the stationary distribution of the Markov chain induced by the mutation-selection dynamics puts positive probability on non-subgame perfect Nash equilibria of the game even when the mutation rate is taken to zero in the limit.

In deterministic models of evolution in extensive form games with perfect information, Demichelis and Ritzberger (2000) show that if any Nash equilibrium component is asymptotically stable then it must be the subgame perfect one (see also Demichelis, Ritzberger and Swinkels (2002)). In general, however, no component of Nash equilibria is asymptotically stable. Cressman and Schlag (1998) find that the subgame perfect component of the set of Nash equilibria is, for any such games, contained in the unique minimal interior asymptotically stable set. In general, however, other non-
subgame perfect components of Nash equilibria are also contained in this minimal interior asymptotically stable set, as Cressman and Schlag (1998) prove by example.

In another stochastic model of mutation and selection for the agent normal form of generic perfect information games Hart (2002) shows that in the limit where the mutation rate tends to zero while population sizes tend to infinity in a way such that the mutation rate times the population sizes are bounded away from zero, indeed only subgame perfect equilibria emerge in the long-run. Hart’s (2002) model differs from Nöldeke’s and Samuelson’s (1993) mainly in two respects. First, in Hart’s (2002) model only one individual per population is called upon to either select a better action or to mutate randomly at any given period in time, while in Nöldeke’s and Samuelson’s (1993) model every individual may change strategy at any given time. Second, Hart’s (2002) individuals, unlike Nöldeke’s and Samuelson’s (1993), do not hold conjectures about what other agents do in the game.

This paper shows that Hart’s (2002) result, with slightly different limit-taking, can be extended to the mutation-selection dynamics of Nöldeke and Samuelson’s (1993).
2 The Model

The selection-mutation mechanism outlined in this section is the one due to Nöldeke and Samuelson (1993). Let \( \Gamma \) be a finite generic extensive form game of perfect information. Generic in the sense that no player has the same payoff at any two different terminal nodes. Under these assumptions there is a unique subgame-perfect equilibrium.

Let \( N \) denote the set of nodes. For all \( i \in N \) let \( M(i) \) denote a population of individuals at node \( i \), i.e. the dynamics are at work on the agent normal form, or as Hart (2002) calls it, the gene normal form. A player who owns \( k \geq 2 \) information sets is supposed to delegate the strategy decision problem to \( k \) independent agents (agent normal form) or to have \( k \) independent genes each of which control one information set. Let \( m_i = |M(i)| \) be the size of the population at node \( i \) and \( m = (m_1, m_2, ..., m_N) \) denote the vector of population sizes. For all \( i \in N \) let \( A(i) \) denote the set of possible actions available to individuals at node \( i \). Let \( b^i \in A(i) \) denote the backward induction action at node \( i \).

The game is supposed to be played recurrently at discrete points in time by every possible combination of agents in each population. Every agent in every population is characterised by a pure action and a conjecture about every other agent’s action in the game. A state \( \omega \) is a specification of a characteristic for every agent in every population. The state space shall be
denoted by $\Omega$.

In every period after the game has been played every agent in every population takes a draw from a Bernoulli distribution with outcomes "learn" and "don’t learn" with probabilities $\sigma$ and $1 - \sigma$, respectively. If the agent receives the learn draw she updates her conjecture such that it coincides with the actions individuals at the various nodes actually take provided they are observable. Conjectures about actions at nodes, which are not reached, will not change. She then chooses an action, which is a best reply to her conjecture. If there is more than one best reply she will choose one according to some fixed probability distribution with full support over all best replies. If her current action is already a best reply she will not change her action. In the other case she neither changes her conjecture or her action.

After the learning phase every individual in every population receives a draw of another Bernoulli variable with outcomes "mutate" and "don’t mutate" with probabilities $\mu$ and $1 - \mu$, respectively. If the agent receives a mutate-draw she will choose an arbitrary characteristic according to a probability distribution with full support over all possible characteristics for this agent (including the one she is holding at the moment). In the other case she does not change her characteristic.

The above mutation-selection mechanism constitutes a Markov chain on the state space $\Omega$ with transition probability matrix denoted by $Q^m\mu$, indicat-
ing that it is different for different population sizes and different mutation
rates. The transition probabilities also vary with different learning proba-
bilities $\sigma$. In the course of this essay, however, $\sigma$ is assumed to be fixed at
a value strictly between 0 and 1.

Clearly, the Markov chain induced by the above selection-mutation dy-
namics is ergodic and irreducible. Hence, it has a unique stationary distri-
bution, which shall be denoted by $\pi^m_\mu$, and satisfies

$$\pi^m_\mu Q^m_\mu = \pi^m_\mu. \quad (1)$$

3 An example

![Figure 1: The extensive game in the example](image)

The three-player extensive form game given in Figure 1 (see also Figure
1.2 in Hart or Figure 1 in Nöldeke and Samuelson) has the unique subgame
perfect equilibrium (R,R,R). There are, however, other Nash equilibria, like
(L,L,L), which are not in the subgame perfect Nash equilibrium component.
Nöldeke and Samuelson show that for any fixed vector of population sizes both equilibria must carry positive probability in the limiting invariant distribution. The argument is the following. Suppose the system is in the state where every individual in every population plays R and conjectures match these actions, i.e. the system is in the subgame-perfect state. Then node 3 is not reached and evolutionary drift can occur. Indeed just by this drift the system will eventually be in the state where everyone at nodes 1 and 2 plays R and everyone conjectures as much and where all individuals at node 3 play L while individuals at nodes 1 and 2 conjecture them to play R. Suppose now that one mutation occurs at node 2, i.e. node 3 will suddenly be reached. Then, Nöldeke and Samuelson argue, with positive probability all agents at node 2 learn before anyone at node 3 learns. In fact, ignoring the small mutation probability for the moment, this will happen with probability \( \sigma^m_2 (1 - \sigma)^m_3 \). In the next period with positive probability all individuals at node 1 learn, update their conjectures and play L. This probability is given by \( \sigma^m_1 \). Hence, it takes only one mutation to get from the subgame perfect component generated by (R,R,R) to another Nash equilibrium component, generated by (L,L,L), by learning only, which happens with positive probability \( \sigma^m_2 (1 - \sigma)^m_3 \sigma^m_1 \), and hence (L,L,L) must be in the domain of the limiting invariant distribution. This probability, however, tends to zero when population sizes go to infinity. This is to say that
when population sizes go to infinity it is not enough to count the number of mutations it takes to get from one absorbing state to another, as these are not the only infinitesimal probability transitions. Any long chain, such as a fraction of a population, of a lot of people learning will also only occur with infinitesimal probability. This argument does, however, not tell us which states will carry positive weight or not in the limiting invariant distribution when in addition to mutation rates going to zero, population sizes tend to infinity. It only illustrates that the analysis requires more than a mutation counting exercise.

The claim I make in this paper is that only subgame-perfect equilibria will be in the domain of the limiting invariant distribution, when the limit is taken with respect to the mutation rate $\mu$ going to zero and population sizes $m_i$ going to infinity, while $m_i \mu^d$ is bounded away from zero for an arbitrary fixed $d > 1$. The precise claim is to be found in section 4. In the following few paragraphs, however, I want to use the example to illustrate why this should be true.

Consider the population of individuals at node 2. Suppose for the moment that (after learning and updating conjectures) $R$ is the unique best-reply there, i.e. at least one individual at node 1 plays $R$ and the population mix at node 3 is such that more than $2/3$ of the population play $R$ there. Under these circumstances individuals at node 2 will play $L$, with some con-
jecture, only by mistake, i.e. by mutation. Now suppose furthermore that all individuals at node 2 play R at the moment and hold true conjectures about, the now unreached, node 3. How many individuals at node 2 do we expect to play L in the next period? Let $X$ denote the number of people mutating from R to L in one period. $X$ is then binomially distributed with parameters $\mu$, the mutation rate, and $m_2$, the population size. In particular, the expected number of mutations to L is given by $E(X) = \mu m_2$. In the limit I consider, this expected number of mutations to L will tend to infinity. This means that even if an action is currently not played at all and is not a best reply, in the next period a very large number (tending to infinity) of individuals is expected to play it. If some more individuals were playing this action already at the moment or the action were a best reply, the expected number of people playing this action in the next period would only be greater. This goes to say that any action anywhere in the tree will essentially always be played by a very large (essentially infinite) number of people from the corresponding population. Given that this is the case, however, all nodes will be reached essentially all the time. Hence, whenever people learn, their updated conjectures will match the truth, i.e. conjectures don’t matter.

Given all this consider individuals at node 3. This node will be reached essentially all the time. Hence, the unique best-reply for individuals at
node 3 is the backward induction action R. Whenever people learn (with probability $\sigma$) they will choose to play action R. Only by mutation will they adopt L. But the expected number of people who receive a learn draw, $\sigma m_2$ is, in the limit, infinitely greater than the expected number of individuals who mutate. Hence, in the long-run, even though there will always be an infinite number of individuals playing L, infinitely more will play R. Therefore, in the long-run, more than any arbitrarily high fraction of the population at node 3 will play their backward induction action R.

Given that almost everyone at node 3 plays R the unique best-reply at node 2 is the backward induction action R as well. By the same argument as before then, in the long-run, more than any arbitrary fraction of individuals at node 2 will play R. Then again, given that, the same must be true for node 1. Hence, even though the system will virtually never be exactly at the subgame perfect equilibrium, it will always be arbitrarily close to it.

The proof of the main result in the paper is very much along the line of thought outlined above. First, I establish a lemma saying that, for any given action at any given node, the probability that not a single individual plays this action, tends to zero in the, above described, limit. Conjectures, therefore, in the limit, must always coincide with the truth as every node is reached. Second, I prove a lemma saying that, if an action is the unique best-reply with probability going to one in the limit, it will be played by
more than any fraction, arbitrarily close to 1, of individuals at that node. Given these two lemmas I prove the main result that the whole system, in the limit, is arbitrarily close to the subgame perfect equilibrium of the game, using a backward induction argument.

4 Results

Nöldeke and Samuelson show that, except for very special classes of games, non-subgame perfect equilibria will be in the support of the limiting distribution, where $m$ is fixed and $\mu \to 0$.

Hart shows for a different selection-mutation dynamics that in the limit where $\mu \to 0$ and $m_i \to \infty$ such that $m_i\mu \geq \delta > 0$ only any $\epsilon$-neighborhood of the backward induction solution is in the support of the limiting distribution (hence has probability 1).

This section shows that Hart’s result can be extended to the mutation-selection dynamics of Nöldeke and Samuelson’s (1993) if the limiting distribution is taken with respect to $\mu \to 0$ while $m_i\mu^d \geq \delta > 0$, for any arbitrary $d$ strictly greater than one.

Let $i \in N$ be an arbitrary node and let $x \in A(i)$ be an arbitrary action available to individuals at node $i$. Let $\Lambda_{k}^{i,x}$ denote the set of states in which the proportion of individuals at node $i$ is playing action $x$ is $\frac{k}{m_i}$. Note that if this set contains a specific state $\omega$ it also contains every state which is
only different from $\omega$ with respect to conjectures. For any $i \in N$ and any $x \in A(i)$ the collection of sets $\{\Lambda_k^{i,x}\}_{k=0}^{m_i}$ is a partition of the state space $\Omega$, i.e. the system at any given time must be in exactly one of these sets. The proof of the following lemma is in the appendix.

**Lemma 1** Let $i \in N$ be an arbitrary node and $x \in A(i)$ an arbitrary action available to individuals at node $i$. Let $\lambda_x$ denote the conditional probability that if an agent mutates she does not mutate to a characteristic that involves playing action $x$. For all $\kappa > 1$ there is a $\bar{\mu}$ such that for all $\mu < \bar{\mu}$:

$$\pi^m_{\mu} \left( \Lambda_0^{i,x} \right) \leq \frac{1}{1 + \frac{1-\mu(1-\lambda_x)^{m_i}}{\kappa \sigma(1-\mu(1-\lambda_x))}}$$

(2)

An immediate corollary is that for any node $i \in N$ and any $x \in A(i)$, $\pi^m_{\mu} \left( \Lambda_0^{i,x} \right)$ converges to zero under suitable limit-taking.

**Corollary 1** Let $i \in N$ be an arbitrary node and $x \in A(i)$ an arbitrary action available to individuals at node $i$.

$$\lim_{\mu \to 0, m_i \mu^d \geq \delta > 0} \pi^m_{\mu} \left( \Lambda_0^{i,x} \right) = 0.$$

(3)

Proof: The expression $\pi^m_{\mu} \left( \Lambda_0^{i,x} \right)$ tends to zero in the case where $\mu$ goes to zero while $m_i \mu^d \geq \delta > 0$, as $(1-\mu(1-\lambda_x))^{m_i}$ goes to zero under these circumstances. The term $(1-\mu(1-\lambda_x))^{m_i}$ does indeed tend to zero due to the fact that, as $\mu \geq \left( \frac{\delta}{m_i} \right)^\frac{1}{d}$,

\[(1-\mu(1-\lambda_x))^{m_i} \leq \left(1 - \left( \frac{\delta}{m_i} \right)^\frac{1}{d} (1-\lambda_x) \right)^{m_i} \]

(4)
\[
\leq \left[ \left( 1 - \left( \frac{\delta}{m_i} \right)^{\frac{1}{3}} (1 - \lambda x) \right)^{m_i \frac{1}{3}} \right]^{m_i - \frac{1}{3}} \tag{5}
\]

and the fact that \( \left( 1 - \left( \frac{\delta}{m_i} \right)^{\frac{1}{3}} (1 - \lambda x) \right)^{m_i \frac{1}{3}} \) tends to \( e^{-\delta \frac{1}{3} (1-\lambda x)} < 1 \) as \( m_i \) tends to infinity. QED

Another corollary follows immediately from the above lemma.

**Corollary 2** Denote by \( \Psi \) the set of states, in which there is a node such that at least one action is not played by any individual at this node, i.e.

\[
\Psi = \bigcup_{i \in N} \bigcup_{x \in A(i)} \Lambda_{i,x}^0. \tag{6}
\]

Then

\[
\lim_{\mu \to 0, m_i, \mu^d \geq \delta \geq 0} \pi_{\mu}^{m_i} (\Psi) = 0. \tag{7}
\]

This is due to the fact that each set in the union has zero probability in the limit and that \( \Psi \) is a finite union of these sets.

The corollary states that for large \( m_i \)'s and hence small \( \mu \) the evolutionary system is almost always in a state where every node in the game is reached. Node \( i \) is reached if there is at least one person at every node between the root and \( i \) who plays the action that leads towards node \( i \).

In this case conjectures after learning almost always coincide with the actual actions.
Evolutionary pressure in form of selection pressure, therefore, is present at all nodes almost all the time. Hence, by backward induction arguments, we expect the system in the limit to be close to all individuals playing their backward induction action.

For any node $i \in N$ denote by $R(i) \subset N$ the set of predecessor nodes of node $i$. For any node $i \in N$ and for every node $j \in R(i)$ let $a_j^i \in A(j)$ denote the unique action at node $j$ which (eventually) leads to node $i$. For any node $i$, let $S(i)$ denote the set of successor nodes of node $i$.

Let $C_{b^i}^i$ be the set of states such that $b^i \in A(i)$ is not the unique best reply for any agent at node $i$ given their conjectures after a potential learn draw. Let $B_{\epsilon,m}^i = \bigcup_{k \geq (1-\epsilon) m} \Lambda_{k}^{i,b^i}$. Let $B_{\epsilon,m}^{i,c}$ denote its complement in $\Omega$. Let generally for a set $A \subset \Omega$, $A^c$ denote its $\Omega$-complement. Then $B_{\epsilon,m} = \bigcap_{i \in N} B_{\epsilon,m}^i$ is the set of states, in which more than a fraction of $(1-\epsilon)$ individuals in every population play their respective backward-induction action.

**Lemma 2** Let $i \in N$ be a final decision node. Then

$$C_{b^i}^i = \bigcup_{j \in R(i)} \Lambda_{0}^{j,a_j^i}$$

This is due to the fact that $b^i$ is the unique best reply for individuals at final node $i$ if and only if node $i$ is reached.
Lemma 3  Let \( i \in N \) be an arbitrary non-final node. Then there is an \( \bar{\epsilon} \) such that for all \( \epsilon < \bar{\epsilon} \):

\[
C_{b^i}^i \subset \left( \bigcup_{j \in R(i)} \Lambda_{0}^{j,a_1^i} \right) \cup \left( \bigcup_{j \in S(i)} B_{\epsilon,m}^{j,c} \right) \tag{9}
\]

This is due to the fact that \( b^i \) is the unique best reply for individuals at intermediate node \( i \) if the node is reached and a sufficient fraction of individuals at successor nodes play their backward induction action. An alternative presentation, in terms of set-complements, of lemma 3 is given by

\[
C_{b^i}^{i,c} \supset \left( \bigcup_{j \in R(i)} \Lambda_{0}^{j,a_1^i} \right)^c \cap \left( \bigcap_{j \in S(i)} B_{\epsilon,m}^{j} \right) \tag{10}
\]

An immediate corollary to lemmas 1 and 2, and corollary 1 is the following.

Corollary 3  Let \( i \in N \) be a final decision node. Then

\[
\lim_{\mu \rightarrow 0, m, \mu^d \geq \delta} \pi_{\mu}^m (C_{b^i}^i) = 0 \tag{11}
\]

The following lemma is proved in the appendix.

Lemma 4  Let \( i \in N \) be a node such that \( \lim_{\mu \rightarrow 0, m, \mu^d \geq \delta} \pi_{\mu}^m (C_{b^i}^i) = 0 \). Then for any \( \epsilon > 0 \):

\[
\lim_{\mu \rightarrow 0, m, \mu^d \geq \delta} \pi_{\mu}^m (B_{\epsilon,m}^i) = 1. \tag{12}
\]

This now enables me to proof the main result of this paper.
Theorem 1 For any $\epsilon > 0$

\[
\lim_{\mu \to 0, \mu^d \geq \delta \forall i \in N} \pi^m_{\mu_i} (B^m_{\epsilon}) = 1. \tag{13}
\]

Proof: To show that $\pi^m_{\mu} (B) \to 1$ I use a backward induction argument. Let $i \in N$ be a final decision node. Then by corollary 3 $\lim_{\mu \to 0, \mu^d \geq \delta} \pi^m_{\mu} (C^i_{\mu}) = 0$ and hence by lemma 4 $\pi^m_{\mu} (B^i_{\epsilon,m}) \to 1$. Then by lemmas 3 and 1 $\pi^m_{\mu} (C^j_{\mu}) \to 0$ for all immediate predecessor nodes $j$ of $i$. Again by lemma 4 it must be true that $\pi^m_{\mu} (B^j_{\epsilon,m}) \to 1$. This in turn, by lemmas 3 and 1, yields that $\pi^m_{\mu} (C^{l}_{\mu}) \to 0$ for all immediate predecessor nodes $l$ of $j$. The result is then proved after a finite repetition of the above argument. QED

The next theorem shows that it is not necessary to have an, in the limit, infinite number of mutants to obtain subgame-perfection as the only stochastically stable outcome, if the learning-rate also, albeit slowly, tends to zero in the limit. In order to distinguish this new scenario from the above I emphasize that the learning-rate is now also changing by denoting the invariant distribution by $\pi^m_{\mu,\sigma_i}$. The proof of the following theorem is in the appendix.

Theorem 2 For any $\epsilon > 0$

\[
\lim_{\xi \to 0, \sigma \to 0, \mu^d \geq \delta \forall i \in N} \pi^m_{\mu,\sigma} (B^m_{\epsilon}) = 1. \tag{14}
\]
5 Discussion

There are three important differences between the stochastic evolutionary learning models of Hart (2002) and of Nöldeke and Samuelson (1993). First, in Nöldeke and Samuelson’s model individuals hold conjectures over other peoples actions in the game, which is not the case in Hart’s model. Second, while in Hart’s model only one person is drawn to potentially change his action, a much larger number of individuals (possibly everyone) is drawn to learn or then to experiment in the model of Nöldeke and Samuelson. Finally, individuals in the model of Nöldeke and Samuelson play best-replies to their conjectures, once they receive a learn draw, while individuals in Hart’s model play better replies in this case.

I prove that, in the model of Nöldeke and Samuelson, only subgame-perfect equilibria are stochastically stable (in generic extensive form games of perfect information), under two limiting scenarios. One is such that the mutation rate tends to zero, while population sizes diverge to infinity in such a way that the expected number of mutants in each period and population, \( \mu m_i \), diverges to infinity as well. The learning-rate \( \sigma \) is fixed. The other limit is such that the learning rate converges to zero, the mutation rate converges to zero faster than the learning rate, and population sizes diverge to infinity such that the expected number of mutants in each period is bounded away from zero. Table i gives an overview of various characteristics of the limiting
conditions of Hart and this paper under the simplifying assumption of equal population sizes $m_i = m$.

<table>
<thead>
<tr>
<th></th>
<th>Hart</th>
<th>m-Hart</th>
<th>NaS 1</th>
<th>NaS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of mutations</td>
<td>$\mu \to 0$</td>
<td>$\mu m \geq \delta$</td>
<td>$\mu m \to \infty$</td>
<td>$\mu m \geq \delta$</td>
</tr>
<tr>
<td>number of learn-draws</td>
<td>$\sigma$ fixed</td>
<td>$\sigma m \to \infty$</td>
<td>$\sigma m \to \infty$</td>
<td>$\sigma m \to \infty$</td>
</tr>
<tr>
<td>proportion of mutations</td>
<td>$\frac{\mu}{m} \to 0$</td>
<td>$\mu \to 0$</td>
<td>$\mu \to 0$</td>
<td>$\mu \to 0$</td>
</tr>
<tr>
<td>proportion of learn-draws</td>
<td>$\frac{\sigma}{m} \to 0$</td>
<td>$\sigma$ fixed</td>
<td>$\sigma$ fixed</td>
<td>$\sigma \to 0$</td>
</tr>
</tbody>
</table>

Table i: A comparison of characteristics of the limiting conditions of Hart (2002), based on his model, and mine, based on the model of Nöldeke and Samuelson. All figures are expectations per period and population. The column m-Hart shows the expected number of mutations, etc., which occur over a span of $m$ periods in Hart’s model.

Clearly Hart’s limiting conditions are more general than both of mine. However, under his conditions lemma 1 does not necessarily hold, i.e. nodes are not necessarily always reached with probability 1 in the limit. But then the presence of conjectures complicates things as they may well be quite different from the truth for many people in many periods (even in the limit). It may still be the case, however, that Hart’s limiting result holds in the model of Nöldeke and Samuelson. This question is open for future research. I believe the limiting conditions I make in this paper are still very
useful exactly because lemma 1 then holds. This lemma will still hold under these conditions in much more general games as long as the total number of pure strategies is finite. This suggests that similarly sharp results can be obtained in more general games as well, such as non-generic games of perfect information or even imperfect information games.

Coming back to the differences between the models of Hart, and Nöldeke and Samuelson, it seems that the difference in the number of people who potentially change their strategy each period between the two models is not crucial. The fact that individuals hold conjectures complicate matters to a degree, hence the slightly weaker limiting conditions than in Hart. The final difference is in the learning dynamics, individuals in Hart’s model play better replies, while in Nöldeke and Samuelson’s model they play best-replies. However, it is easy to see that changing the model from best to better replies will not affect the result. Even under better replies every node is still reached with probability 1 in the limit. Now suppose there are 3 actions, A, B, and C, available to individuals at a final node and that A is strictly preferred to B, which in turn is strictly preferred to C. As C can only be played by mistake, i.e. by mutation, we know that almost everyone at this node will eventually play either A or B. But then, if people who play B learn, they will switch to playing action A, as it is a better reply (to any probable situation) than B is. A-players, however, will play B only by
mutation. Hence, in the long run almost everyone at this node will play A.
This argument generalizes to any finite number of actions available to players
at final nodes. Then, by backward induction, similarly, all individuals at all
nodes will eventually learn to play their backward induction action.

A Proof of Lemma 1

Define $D_i$ as the set of states such that $x \in A(i)$ is not a best reply for
any agent at node $i$ given their conjectures. I.e. these are states such that
for every agent at node $i$ there is another action $y \in A(i)$, different from $x$
(and possibly different for different agents), such that $y$ is a best reply and
is better than $x$.

Given the property of the invariant distribution (1) any probability
\( \pi^m(\omega) \) can be expressed as

\[
\pi^m(\omega) = \sum_{\omega' \in \Omega} \pi^m(\omega') \left( Q^m_{\mu} \right)_{\omega' \omega},
\]

(15)

where \( \left( Q^m_{\mu} \right)_{\omega' \omega} \) is the transition probability that the system moves from $\omega'$
to $\omega$.

Equivalently for any set of states, $\Lambda$,

\[
\pi^m(\Lambda) = \sum_{\omega \in \Lambda} \sum_{\omega' \in \Omega} \pi^m(\omega') \left( Q^m_{\mu} \right)_{\omega' \omega}.
\]

(16)
Changing the order of summation yields

$$
\pi_m^\mu(\Lambda) = \sum_{\omega' \in \Omega} \pi_m^\mu(\omega') \left( Q_m^\mu \right)_{\omega' \Lambda},
$$

(17)

where $$\left( Q_m^\mu \right)_{\omega' \Lambda} = \sum_{\omega \in \Lambda} \left( Q_m^\mu \right)_{\omega' \omega}$$.

We are interested in the set $$\Lambda = \Lambda_{i,x}^{k,x}$$. It is easy to show that for any $$\omega' \in \Lambda_{k,x}^{i,x}$$,

$$
\left( Q_m^\mu \right)_{\omega' \Lambda_{0}^{i,x}} = \begin{cases} p_{k0} & \forall \omega' \in D_x \\ \leq p_{k0} & \text{otherwise} \end{cases},
$$

(18)

where

$$p_{k0} = \sum_{j=0}^{k} \sigma^j (1 - \sigma)^{k-j} \binom{k}{j} (\mu \lambda x)^{k-j} (1 - \mu(1 - \lambda x))^{m_i-k+j}. \quad (19)$$

This is because there are many ways to move from a state where $$k$$ out of $$m_i$$ individuals at node $$i$$ play $$x$$ to a state where none do. Suppose the current state $$\omega$$ is in $$D_x$$. A possible transition is that any $$j \leq k$$ individuals who are currently playing $$x$$ learn and change their action and the remaining $$k - j$$ agents mutate to play anything other than $$x$$, while everyone else does not change their action to $$x$$. $$p_{k0}$$ is then just the sum of all the probabilities of these various possible transitions.

Careful inspection of equation (19) reveals that

$$p_{k0} = (1 - \mu(1 - \lambda x))^{m_i-k} \sum_{j=0}^{k} \binom{k}{j} (\sigma (1 - \mu(1 - \lambda x)))^j ((1 - \sigma)(\mu \lambda x))^{k-j}$$

$$= (1 - \mu(1 - \lambda x))^{m_i-k} (\mu \lambda x + \sigma(1 - \mu))^k. \quad (20)$$
Hence, for all $k$,

$$
p_{k+1,0} = \frac{\mu \lambda_x + \sigma(1 - \mu)}{1 - \mu(1 - \lambda_x)},
$$

which is less than 1 for small $\mu$.

Using equations (17) and (18) yields

$$
\pi^m_m (\Lambda_{i,x}^0) \leq \sum_{k=0}^{m_i} \pi^m_m (\Lambda_{i,x}^k) p_{k0}.
$$

(22)

Rearranging leads to

$$
\pi^m_m (\Lambda_{i,x}^0) \leq \frac{1}{1 - p_{00}} \sum_{k=1}^{m_i} \pi^m_m (\Lambda_{i,x}^k) p_{k0}
$$

(23)

and hence

$$
\pi^m_m (\Lambda_{i,x}^0) \leq \frac{1 - \pi^m_m (\Lambda_{i,x}^0)}{1 - p_{00}} \max_{k \geq 1} \{p_{k0}\}.
$$

(24)

Finally

$$
\pi^m_m (\Lambda_{i,x}^0) \leq \frac{1}{1 + \frac{1 - p_{00}}{\max_{k \geq 1} \{p_{k0}\}}}.
$$

(25)

By equation (21) $\max_{k \geq 1} \{p_{k0}\} = p_{10}$ for $\mu$ small enough. This confirms the intuition that the easiest way to move to $\Lambda_{i,x}^0$ is coming from $\Lambda_{i,x}^1$.

Now, by equations (20) and (21),

$$
p_{10} = (1 - \mu(1 - \lambda_x))^{m_i} \frac{\mu \lambda_x + \sigma(1 - \mu)}{1 - \mu(1 - \lambda_x)}.
$$

(26)

Hence,

$$
\forall \kappa > 1 \exists \bar{\mu} : \forall \mu \leq \bar{\mu}
$$

$$
p_{10} \leq \kappa \sigma (1 - \mu(1 - \lambda_x))^{m_i}.
$$

(27)
Hence, for all $\kappa > 1$ there is a $\bar{\mu}$ such that for all $\mu < \bar{\mu}$,

$$\pi_{\mu}^{m\kappa}(\Lambda_{i,j,x}^{i,x}) \leq \frac{1}{1 + \frac{1}{\kappa \sigma(1-\mu(1-\lambda_{i,x}))^{m\kappa}}}.$$

(28)

QED

B Proof of Lemma 4

Before I state the proof I want to give a brief sketch of what it consists of.

For fixed $m_i$ and fixed $\mu$ let random variable $U$ be distributed according to the ‘projection’ of the invariant distribution $\pi_{\mu}^{m\kappa}$ on the partition of the state space $\Omega$ given by $\Lambda_{i,b}^{i,b'}$ for $k = 0, 1, ..., m_i$. $U$ denotes the proportion of backward induction ($b'$) players at node $i$ under the invariant distribution, i.e. $U$ can assume values in $\{0, \frac{1}{m_i}, \frac{2}{m_i}, ..., 1\}$.

I define random variables $Z_k$, which represent the random one-step net increase (or decrease) in the number of individuals playing $b'$, given $k$ individuals are currently playing $b'$, and given that $b'$ is currently the unique best-reply. Another random variable $W$ is assumed to take values $k + Z_k$ with probability given by $\pi_{\mu}^{m\kappa}(\Lambda_{k}^{i,b'})$ for $k = 0, 1, ..., m_i$. This is to say that $W$ is such as if it were the number of individuals at node $i$ playing $b'$ tomorrow given the number of individuals playing $b'$ today were distributed according to the invariant distribution, and as if $b'$ were today’s unique best-reply in any state.
Another, final, random variable $V$ is introduced, which is equal to $W$ with probability $1 - \pi^m_\mu \left( C_i^b \right)$ and zero otherwise. Under the invariant distribution, $1 - \pi^m_\mu \left( C_i^b \right)$ is the probability that $b^i$ is the unique best reply, in which case $V$ is assumed to follow the Markov chain dynamic, just as $U$ does. With probability $\pi^m_\mu \left( C_i^b \right)$, however, $b^i$ is not the unique best-reply, in which case 0 people playing $b^i$ tomorrow is the worst case scenario, i.e. $U$ will typically not necessarily be zero. Hence $E(U) \geq \left[ 1 - \pi^m_\mu \left( C_i^b \right) \right] E \left( \frac{V}{m} \right)$. This last fact can then be used indirectly to prove the result.

I now state the proof. Let $i \in N$ be a node such that $\lim_{\mu \to 0, m_i \mu \alpha \geq \delta} \pi^m_\mu \left( C_i^b \right) = 0$. Suppose the current state is $\omega \in \Lambda_{k, b^i}^i \setminus C^i_{b^i}$. This is to say that currently $k$ individuals at node $i$ play the backward-induction action $b^i \in A(i)$, which is currently the unique best-reply. Let $X_k$ denote the random number of individuals, who, either by learning or mutation, switch action from $b_i$ to any other action in $A(i)$ given the current state is $\omega \in \Lambda_{k, b^i}^i \setminus C^i_{b^i}$. Let similarly $Y_k$ denote the random number of individuals, who, either by learning or mutation, switch action from any non-$b^i$ action to $b^i$ given the current state is $\omega \in \Lambda_{k, b^i}^i \setminus C^i_{b^i}$.

Both, $X_k$ and $Y_k$, are binomial random variables, i.e. $X_k \sim \text{Bin}(k, \beta)$ and $Y_k \sim \text{Bin}(m_i - k, \gamma)$, where $\beta = \mu \lambda_{b^i}$ and $\gamma = \sigma (1 - \mu) + \mu (1 - \lambda_{b^i})$. The expression $\lambda_{b^i}$ denotes the conditional probability that if an individual mutates she or he does not mutate to a characteristic which involves playing
$b^i$. Let $Z_k = Y_k - X_k$ denote the net ”addition” of $b^i$-players at node $i$. Then $t_{kj} = P(Z_k = j - k)$ is the transition probability of a move from a state in $\Lambda_k^{i,b_i} \setminus C_{b_i}^j$ to any state in the set $\Lambda_j^{i,b_i}$. This is the transition probability from a state with $k$ individuals at node $i$ playing the backward-induction action to any of the states, in which $j$ individuals play the backward-induction action, provided that $b^i$ is currently the unique best reply. This probability is the same for all states in $\Lambda_k^{i,b_i} \setminus C_{b_i}^j$. This is due to the fact that the situation, individuals at node $i$ are facing, is exactly the same for all such states. For all such states it is true that $b^i$ is the unique best reply for all people and that $k$ individuals are currently playing the backward-induction action.

The expected value of $Z_k$ is given by $(m_i - k)\gamma - k\beta$ or, in terms of the parameters of the model,

$$E(Z_k) = m_i [\sigma(1 - \mu) + \mu(1 - \lambda_{b_i})] - k [\sigma(1 - \mu) + \mu], \quad (29)$$

which is positive for all $k$ that satisfy

$$\frac{k}{m_i} < \frac{\sigma(1 - \mu) + \mu(1 - \lambda_{b_i})}{\sigma(1 - \mu) + \mu} < 1. \quad (30)$$

In particular if $k = (1 - \epsilon)m_i$, $E(Z_k)$ is positive if

$$1 - \epsilon < \frac{\sigma(1 - \mu) - \mu(1 - \lambda_{b_i})}{\sigma(1 - \mu) + \mu}. \quad (31)$$

It is easy to see that $E(Z_{m_i}) < 0$. However, for an arbitrary $\epsilon > 0$, $E(Z_k) > 0$ for all $k < (1 - \epsilon)m_i$, provided $\mu$ is small enough.
Let $V$ denote a random variable, which takes values in \{0, 1, ..., $m_i$\}. In particular $V$ shall be zero with probability given by $\pi^m_{\mu_i} (C^{i}_{b_i})$ and shall be another random variable $W$ with the remaining probability. I.e.

$$V = \begin{cases} 
0 & \text{with probability } \pi^m_{\mu_i} (C^{i}_{b_i}) \\
W & \text{otherwise}
\end{cases}$$  \hfill (32)

The random variable $W$ is such that

$$W = k + Z_k \text{ with probability } \pi^m_{\mu_i} \left( \Lambda^{i,b'}_k \right)$$  \hfill (33)

for all $k = 0, 1, ..., m_i$.

Let $U$ denote yet another random variable. $U$ shall assume values in \{0, $\frac{1}{m_i}$, ..., 1\} and shall be distributed according to the invariant distribution $\pi^m_{\mu}$ (on the partition of $\Omega$ given by the sets $\Lambda^{i,b'}_k$ for all $k$). I.e.

$$U = \frac{k}{m_i} \text{ with probability } \pi^m_{\mu_i} \left( \Lambda^{i,b'}_k \right)$$  \hfill (34)

Clearly it must be true that $E(U) \geq E\left(\frac{V}{m_i}\right)$, which is given by

$$E \left( \frac{V}{m_i} \right) = \left[ 1 - \pi^m_{\mu_i} \left( C^{i}_{b_i} \right) \right] E \left( \frac{W}{m_i} \right)$$  \hfill (35)

$$= \left[ 1 - \pi^m_{\mu_i} \left( C^{i}_{b_i} \right) \right] \sum_{j=0}^{m_i} \frac{j + E(Z_j)}{m_i} \pi^m_{\mu_i} \left( \Lambda^{i,b'}_j \right)$$  \hfill (36)

$$= \left[ 1 - \pi^m_{\mu_i} \left( C^{i}_{b_i} \right) \right] \left[ E(U) + \sum_{j=0}^{m_i} \frac{E(Z_j)}{m_i} \pi^m_{\mu_i} \left( \Lambda^{i,b'}_j \right) \right]$$  \hfill (37)

Hence,

$$\sum_{j=0}^{m_i} \frac{E(Z_j)}{m_i} \pi^m_{\mu_i} \left( \Lambda^{i,b'}_j \right) \leq \frac{\pi^m_{\mu_i} \left( C^{i}_{b_i} \right)}{1 - \pi^m_{\mu_i} \left( C^{i}_{b_i} \right)} E(U).$$  \hfill (38)
By assumption the right hand side of inequality 38 converges to zero when \( \mu \) tends to zero while \( \mu^d m_i \geq \delta \).

Let \( j_* = \lfloor (1 - \epsilon) m_i \rfloor \), where \( \lfloor x \rfloor \) denotes the largest integer smaller than \( x \). Let \( \alpha_j = \frac{E(Z_j)}{m_i} \). Then \( \alpha_j = \frac{m_i - j}{m_i} (\sigma (1 - \mu) + \mu (1 - \lambda b_i)) - \frac{j}{m_i} \mu \lambda b_i \).

Then there is a \( \bar{\mu} \) such that for all \( \mu \leq \bar{\mu} \) it is true that there is an \( \bar{\alpha} \) with \( \alpha_j \geq \bar{\alpha} > 0 \) for all \( j \leq j_* \). Suppose, for the sake of simplicity, that \( (1 - \epsilon) m_i \) is an integer. Then \( \alpha_{j_*} = \epsilon (\sigma (1 - \mu) + \mu (1 - \lambda b_i)) - (1 - \epsilon) \mu \lambda b_i \). One might, for instance, set \( \bar{\alpha} = \frac{\epsilon \sigma}{2} \).

Also observe that for \( j > j_* \), \( \alpha_j \geq \alpha_m = -\mu \lambda b_i \).

Hence,

\[
\sum_{j=0}^{m_i} \alpha_j \pi^m_{\mu} (A^i_{j,b}) \geq \sum_{j=0}^{j_*} \bar{\alpha} \pi^m_{\mu} (A^i_{j,b}) + \sum_{j=j_*+1}^{m_i} -\mu \lambda b_i \pi^m_{\mu} (A^i_{j,b}) \geq \bar{\alpha} \pi^m_{\mu} (B^i_{e,m}) - \mu \lambda b_i \pi^m_{\mu} (B^i_{e,m}) \geq \bar{\alpha} \pi^m_{\mu} (B^i_{e,m}) - \mu \lambda b_i \left( 1 - \pi^m_{\mu} (B^i_{e,m}) \right) \geq -\mu \lambda b_i + (\bar{\alpha} + \mu \lambda b_i) \pi^m_{\mu} (B^i_{e,m})
\]

Combining inequality 38 and 42, we obtain

\[
-\mu \lambda b_i + (\bar{\alpha} + \mu \lambda b_i) \pi^m_{\mu} (B^i_{e,m}) \leq \sum_{j=0}^{m_i} \frac{E(Z_j)}{m_i} \pi^m_{\mu} (A^i_{j,b}) \leq \frac{\pi^m_{\mu} (C^i_{b})}{1 - \pi^m_{\mu} (C^i_{b})} E(U).
\]

Taking \( \mu \to 0 \) while \( \mu^d m_i \geq \delta \) in inequality 43, we obtain

\[
\bar{\alpha} \lim_{\mu \to 0, m_i, \mu^d \geq \delta} \pi^m_{\mu} (B^i_{e,m}) \leq 0
\]

Hence, \( \pi^m_{\mu} (B^i_{e,m}) \to 0 \).

QED
Proof of Theorem 2

To prove this theorem I prove the appropriate variants of the lemmas which lead to the proof of theorem 1. First note that lemma 1 still holds. The important thing to see there is that \( \max_{k \geq 1} \{ p_{k0} \} = p_{10} \) even when \( \sigma \) tends to zero, and that, for all \( \kappa > 1, p_{10} \leq \kappa \sigma (1 - \mu(1 - \lambda x))^m \) provided \( \mu \) is much smaller than \( \sigma \).

Then the appropriate variant of corollary 1 holds. This is due to the fact that \( (1 - \mu(1 - \lambda x))^m \) in the limit is bounded from above by \( e^{-\delta(1-\lambda x)} < 1 \) and hence its product with \( \kappa \sigma \) tends to zero as \( \sigma \) goes to zero.

Then the appropriate reformulation of corollary 2 holds without any further work. Lemmas 2 and 3 are about the structure of the game and do not depend on the dynamics and limits employed. Corollary 3, appropriately reformulated, also holds trivially.

The only remaining ingredient to be shown is then the appropriate version of lemma 4. I will not give the details here, but the proof is now based, not on the one-period net increase, but on the multi-period net increase of \( b^i \)-players. This is to do with the fact that the one-period net increase of \( b^i \)-players converges to zero at the same speed as the learning rate \( \sigma \). However, if one looks at the net increase over \( m_i^d \) periods, where \( d < 1 \), then the number of learners (over that period) is of the order of \( \sigma m_i^d \), which for \( d \) close enough to 1, is bounded away from zero, while the number of mutants
over that period is $\mu m_d^t$ which tends to zero under the limiting conditions. The rest follows from that. QED

References


