Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Games and Economic Behavior 48 (2004) 321–336

GAMES and Economic Behavior

www.elsevier.com/locate/geb

Stochastic evolutionary stability in extensive form games of perfect information

Christoph Kuzmics

J.L. Kellogg School of Management, Northwestern University, 2001 Sheridan Road,
Evanston, IL 60208-2009, USA
University of Cambridge, Faculty of Economics and Politics, Sidgwick Avenue, Austin Robinson Building,
Cambridge CB3 9DD, UK

Received 11 December 2002

Available online 10 December 2003

Abstract

Nöldeke and Samuelson [Games Econ. Behav. 5 (1993) 425] investigate a stochastic evolutionary model for extensive form games and show that even for games of perfect information with a unique subgame perfect equilibrium, non-subgame perfect equilibrium-strategies may well survive in the long run even when mutation rates tend to zero. In a different model of evolution in the agent normal form of these games Hart [Games Econ. Behav. 41 (2002) 227] 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.

JEL classification: C62; C72; C73

Keywords: Backward induction; Learning; Experimentation; Subgame-perfect equilibrium

1. Introduction

The subgame perfect equilibrium in extensive form games of perfect information with a unique subgame perfect equilibrium coincides with the unique trembling hand perfect equilibrium in these games. One might think that random mutations in evolutionary models, as in Kandori et al. (1993) or Young (1993), serve the same purpose as perturbations in the rational formulation, and lead to virtually all individuals playing their

---

E-mail addresses: c-kuzmics@kellogg.northwestern.edu, c.a.kuzmics.99@cantab.net.

0899-8256/$ – see front matter
doi:10.1016/j.geb.2003.10.001

backward induction action. Nöldeke and Samuelson (1993), however, show that this is not generally the case. 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 this must be the subgame perfect one (see also Demichelis et al., 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, for any such games, is contained in the unique minimal interior asymptotically stable set (see also Swinkels, 1993). 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.

Hart (2002) investigates another stochastic model of mutation and selection for the agent normal form of perfect information games with a unique subgame perfect equilibrium. He shows that in the limit where the mutation rate tends to zero while population sizes tend to infinity in such a way that the product of mutation rate and population size is bounded away from zero, the evolutionary process centers on the subgame perfect equilibrium in the long-run. Hart's model differs from Nöldeke's and Samuelson's in four respects. First, in Hart's model, only one individual per population can change action at any given period in time, whereas in Nöldeke's and Samuelson's model every individual may change strategy at any given time. Second, Hart's individuals, in contrast to Nöldeke's and Samuelson's, do not hold conjectures about what other agents do in the game. Third, Hart's individuals, after learning, play a better reply, chosen from the set of currently used actions, whereas individuals in the model of Nöldeke and Samuelson play a best reply to their conjecture. Fourth, Hart analyzes games in the agent normal form, with a different population of individuals for each node of the game, whereas Nöldeke and Samuelson generally investigate games where players can control more than one node, but never more than one along each possible path of play.

This paper shows that Hart's result, with slightly different limit-taking, can be extended to the mutation-selection dynamics of Nöldeke and Samuelson. I investigate the agent normal form version of the model of Nöldeke and Samuelson and the invariant distribution of the induced Markov chain. I show that the invariant distribution puts probability one on any small ϵ-ball around the unique subgame-perfect equilibrium in the limit, where the expected fraction of individuals mutating per period, the mutation rate, tends to zero, while the expected number of mutations per period, which is given by the product of mutation rate and population size, tends to infinity. Of course in this limit population sizes tend to infinity as well, in contrast to the fixed population sizes of Nöldeke and Samuelson.

Hence, while for finite population sizes in Hart's, as well as in Nöldeke's and Samuelson's, model of evolution non-subgame perfect equilibria may well carry positive weight in the long run even when mutation rates tend to zero, for large population sizes one would expect this positive weight on non-subgame perfect equilibria to be small in both models, provided that there is a sufficient number of mistakes per period.

2. The model

The selection-mutation mechanism outlined in this section is the one due to Nöldeke and Samuelson (1993). Let Γ be a finite generic extensive form game of perfect information with a unique subgame-perfect equilibrium as in Hart (2002). Note that this is a stronger requirement than the assumption of a unique subgame perfect equilibrium "outcome." A unique subgame perfect equilibrium implies that applying backward-induction yields a unique best choice at every node in the tree, even in those parts of the tree which are off the equilibrium path. Note also that the assumption of a unique subgame-perfect equilibrium is a little weaker than the usual genericity assumption. In the usual genericity assumption no player obtains the same utility at any two final nodes. However, a game may still have a unique subgame perfect equilibrium in the presence of payoff-ties, provided there are no "crucial" ties. The simplest example of a game which has a unique subgame perfect equilibrium, but payoff-ties, is a decision problem where the single player chooses among three alternatives, A, B, and C, where A yields a payoff of 1, while both B and C yield a payoff of 0. Even though there is a payoff-tie, B and C give the same utility, there is a unique optimal choice, A. The same decision problem with the only difference that payoffs are multiplied by -1, however, does not have a unique subgame perfect equilibrium, and is an example for a "crucial" payoff-tie.

Let N denote the set of nodes. For all $i \in N$ let $M(i)$ denote a finite 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.[1] A player who owns $k \geqslant 2$ 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 $m = (m_1, m_2, \ldots, m_N)$ denote the vector of population sizes. For all $i \in N$ let $A(i)$ denote the finite set of possible actions available to individuals at node i. I assume that $|A(i)| \geqslant 2$ for all $i \in N$. The main results, Theorems 1 and 2, will also hold if at some nodes only one action is available, while some of the lemmas will not. Let $b^i \in A(i)$ denote the backward induction action at node i.

The game is played recurrently at discrete points in time by every possible combination of agents in each population. Every agent in every population is characterized by a pure action and a conjecture attributing a pure action to every other agent in the game. A state ω is a specification of a characteristic for every agent in every population. The state space shall be denoted by Ω.

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 σ and $1 - \sigma$, respectively. If the agent receives the learn draw she updates her conjecture in a way such that it coincides with the actions which 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

---

[1] See Section 5 for a discussion of the case when players control more than one node.

action is already a best reply she will not change her action. In the other case she changes neither her conjecture nor 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 μ 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). Note that this implies that the agent after mutating is not necessarily playing a best reply to her conjecture, as both her action and her conjecture are arbitrary after mutation. In case the agent does not receive a mutate-draw she does not change her characteristic.

The above mutation-selection mechanism constitutes a Markov chain on the state space Ω with transition probability matrix denoted by Q_μ^m, indicating that it is different for different population sizes and different mutation rates. The transition probabilities also vary with different learning probabilities σ. For the main part of this essay, however, σ is assumed to be fixed at a value strictly between 0 and 1.

Clearly, the Markov chain induced by the above selection-mutation dynamics is aperiodic and irreducible. Hence, it has a unique stationary distribution, which shall be denoted by π_μ^m, and satisfies

$$\pi_\mu^m Q_\mu^m = \pi_\mu^m. \tag{1}$$

3. An example

The three-player extensive form game given in Fig. 1 (see also Fig. 1.2 in Hart or Fig. 1 in Nöldeke and Samuelson) has the unique subgame-perfect equilibrium (R, R, R). There are other Nash equilibria, however, 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. 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

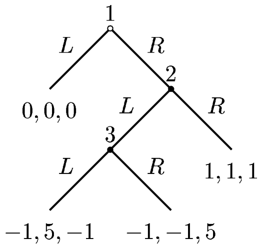

Fig. 1. The extensive game in the example.

one mutation occurs at node 2, i.e. node 3 will suddenly be reached. Then, as 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 σ^{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.[2] This is to say that when population sizes go to infinity, at the same time as mutation rates tend to zero, it is not enough to count the number of mutations it takes to get from one absorbing state to another, since these are not the only infinitesimal probability transitions. Any long chain of a lot of people learning, such as a fraction of a population, will also only occur with infinitesimal probability. Yet this argument does not tell us which states will carry positive weight or not in the limiting invariant distribution when both mutation rates go to zero and 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 μ going to zero and population sizes m_i going to infinity, provided $m_i\mu$ tends to infinity as well. The precise claim is to be found in Section 4. In the following few paragraphs I want to use the above example to illustrate why my claim should be true.

Consider the population of individuals at node 2. Suppose for the moment that R (after learning and updating conjectures) 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 conjecture, only by mistake, i.e. by mutation. 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 μ, the mutation rate, and m_2, the population size. 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 if the action were a best reply, the expected number of people playing this action in the next period would only be greater. This shows 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

---

[2] In fact, the probability of a transition from (R, R, R) to (L, L, L) in a finite period of time is greater than $\sigma^{m_2}(1-\sigma)^{m_3}\sigma^{m_1}$. In any transition, however, a sufficient fraction of individuals at node 2 learns first, and then all individuals at node 1 learn, all before too many individuals at node 3 do. Hence, this probability will tend to zero.

this is the case, all nodes will be reached essentially all the time. Hence, whenever people learn, their updated conjectures will match the truth, i.e. conjectures do not 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 σ) 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, σm_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. 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 in the limit, therefore, 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 non-subgame perfect equilibria, except for very special classes of games, will be in the support of the limiting distribution, where m is fixed and $\mu \to 0$.

Hart demonstrates for a different selection-mutation dynamics that in the limit where $\mu \to 0$ and $m_i \to \infty$ such that $m_i \mu \geqslant \delta > 0$ only any ϵ-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 $m_i \mu \to \infty$.

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 which is playing action x is k/m_i. Note that if this set contains a specific state ω it also contains every state which is only different from ω 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 Ω, i.e. the system at any given time must be in exactly one of these sets. The proof of the following lemma is in Appendix A.

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 λ_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 $\mu < \bar{\mu}$ and for all m_i:*

$$\pi_\mu^m\big(\Lambda_0^{i,x}\big) \leqslant 1 \Big/ \left(1 + \frac{1-(1-\mu(1-\lambda_x))^{m_i}}{\kappa\sigma(1-\mu(1-\lambda_x))^{m_i}}\right). \tag{2}$$

An immediate corollary is that for any node $i \in N$ and any $x \in A(i)$, $\pi_\mu^m(\Lambda_0^{i,x})$ 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.*

$$\lim_{\mu \to 0,\ m_i\mu \to \infty} \pi_\mu^m\big(\Lambda_0^{i,x}\big) = 0. \tag{3}$$

Proof. To show that $\pi_\mu^m(\Lambda_0^{i,x})$ tends to zero in the case where μ goes to zero while $m_i\mu$ tends to infinity, it is enough to prove that $(1-\mu(1-\lambda_x))^{m_i}$ goes to zero under these circumstances. To see this note that $(1-\mu(1-\lambda_x))^{m_i}$ can be written as $[(1-\mu(1-\lambda_x))^{1/\mu}]^{\mu m_i}$, which tends to zero due to the fact that $(1-\mu(1-\lambda_x))^{1/\mu}$ tends to $\mathrm{e}^{-(1-\lambda_x)} < 1$ as μ tends to zero. □

Another corollary follows immediately from the above lemma.

Corollary 2. *Denote by Ψ 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.*

$$\Psi = \bigcup_{i \in N} \bigcup_{x \in A(i)} \Lambda_0^{i,x}. \tag{4}$$

Then

$$\lim_{\mu \to 0,\ m_i\mu \to \infty\ \forall i \in N} \pi_\mu^m(\Psi) = 0. \tag{5}$$

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

The corollary states that for large population sizes, small mutation rates, and a large expected number of mutations per period ($m_i\mu$) 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 the 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 $a_i^j \in A(j)$ 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 $B_{\epsilon,m}^i = \bigcup_{k \geqslant (1-\epsilon)m_i} \Lambda_k^{i,b^i}$. Let $B_{\epsilon,m}^{i,c}$ denote its complement in Ω. Let generally for a set $A \subset \Omega$, A^c denote its Ω-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*

$$C_{b^i}^i = \bigcup_{j \in R(i)} \Lambda_0^{j,a_i^j}. \tag{6}$$

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 $\epsilon \in (0, \bar{\epsilon})$ and for all m:*

$$C_{b^i}^i \subset \left(\bigcup_{j \in R(i)} \Lambda_0^{j,a_i^j} \right) \cup \left(\bigcup_{j \in S(i)} B_{\epsilon,m}^{j,c} \right). \tag{7}$$

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

$$C_{b^i}^{i,c} \supset \left(\bigcup_{j \in R(i)} \Lambda_0^{j,a_i^j} \right)^c \cap \left(\bigcap_{j \in S(i)} B_{\epsilon,m}^j \right). \tag{8}$$

An immediate corollary to Lemma 2 and Corollary 1 is the following.

Corollary 3. *Let $i \in N$ be a final decision node. Then*

$$\lim_{\mu \to 0,\ m_i\mu \to \infty\ \forall i \in N} \pi_\mu^m \left(C_{b^i}^i \right) = 0. \tag{9}$$

The following lemma is proved in Appendix B.

Lemma 4. *Let $i \in N$ be a node such that $\lim_{\mu \to 0,\ m_i\mu \to \infty\ \forall i \in N} \pi_\mu^m (C_{b^i}^i) = 0$. Then for any $\epsilon \in (0,1)$:*

$$\lim_{\mu \to 0,\ m_i\mu \to \infty\ \forall i \in N} \pi_\mu^m \left(B_{\epsilon,m}^i \right) = 1. \tag{10}$$

This now enables me to prove the main result of this paper.

Theorem 1. *For any* $\epsilon \in (0,1)$,

$$\lim_{\mu\to 0,\ m_i\mu\to\infty\ \forall i\in N} \pi_\mu^m(B_{\epsilon,m}) = 1. \tag{11}$$

Proof. Let $F \subset N$ denote the set of all final nodes. Let $P(i) \in N$ be the (unique) immediate predecessor node to node $i \in N$. Let $P(K) = \{P(i) \mid i \in K\}$ be the set of nodes which are an immediate predecessor to a node in $K \subset N$. To show that $\pi_\mu^m(B_{\epsilon,m}) \to 1$; I use a backward induction argument. Let $i \in F$. Then by Corollary 3; $\lim_{\mu\to 0,\ m_i\mu\to\infty\ \forall i\in N} \pi_\mu^m(C_{b^i}^i) = 0$ and hence by Lemma 4; $\pi_\mu^m(B_{\epsilon,m}^i) \to 1$. Then by Lemma 3 and Corollary 1; $\pi_\mu^m(C_{b^j}^j) \to 0$ for all $j \in P(F)$. Again by Lemma 4 it must be true that $\pi_\mu^m(B_{\epsilon,m}^j) \to 1$. This in turn, by Lemma 3 and Corollary 1, yields that $\pi_\mu^m(C_{b^l}^l) \to 0$ for all $l \in P(P(F))$. The result is then proved after a finite repetition of the above argument. □

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 the fact that the learning-rate is now also changing by denoting the transition probability matrix by $Q_{\mu,\sigma}^m$ and the invariant distribution by $\pi_{\mu,\sigma}^m$.

Theorem 2. *For any* $\epsilon > 0$,

$$\lim_{\frac{\mu}{\sigma}\to 0,\ \sigma\to 0,\ m_i\mu\geqslant\delta\ \forall i\in N} \pi_{\mu,\sigma}^m(B_{\epsilon,m}) = 1. \tag{12}$$

The proof of Theorem 2 is available on request from the author. I here give a brief sketch of the proof.

First, the appropriate variant of Lemma 1 holds, i.e. for any node i the probability that any action $x \in A(i)$ is played by nobody in the relevant player population is of the order of σ. Hence the appropriate variant of Corollary 1 holds, i.e. this probability tends to zero when σ does, given the other limiting conditions.

To see why this would be the case consider a single player's decision problem between two alternatives, A and B, where B is the preferred choice. There are two forces at work in the evolutionary dynamics. In every period a fraction σ of A-players will learn to play B, while a fraction of $\lambda_B\mu$ B-players will mutate to play A, both in expectation. Suppose a fraction $(1-\epsilon)$ of individuals plays B at the moment. But then roughly $\lambda_B\mu(1-\epsilon)m_i$ individuals change action from B to A, while approximately $\sigma\epsilon m_i$ individuals change action from A to B. But the smaller ϵ, i.e. the closer the state to everyone playing the backward induction action, the relatively more individuals switch away from B rather than *to* B.

Under the limiting conditions of Lemma 1 and Theorem 1 we had $\mu m_i \to \infty$ and σ fixed. Lemma 1 showed that the two forces were then "balanced" so as to make the event “everyone plays B” probabilistically impossible. It turns out that under the limiting conditions investigated here, when $\mu m_i \geqslant \delta$ and $\sigma \to 0$, we again obtain the same balance.

Given this the appropriate variants of Corollaries 1 and 2 hold 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. 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 σ. However, if one looks at the net increase over $l = \lfloor 1/\sigma \rfloor$ periods, then the number of learners (over that period) is of the order of 1, while the number of mutants over that period is μ/σ which tends to zero under the limiting conditions. The rest follows from that.

5. Discussion

There are four differences between the stochastic evolutionary learning models of Hart (2002) and of Nöldeke and Samuelson (1993). First, whereas in Hart’s model only one person is drawn potentially to change his action, a much larger number of individuals (possibly everyone) is drawn to learn or to experiment in the model of Nöldeke and Samuelson. Second, 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. Third, individuals in the model of Nöldeke and Samuelson play best-replies to their conjectures, once they receive a learn draw, whereas individuals in Hart’s model play better replies. Finally, while in Hart’s model every node is governed by a different population, Nöldeke and Samuelson generally allow populations to control more than one node, but, at most one along each possible path of play.

Theorems 1 and 2 prove that in the agent normal form version of the model of Nöldeke and Samuelson, only subgame-perfect equilibria are stochastically stable (in extensive form games of perfect information with a unique subgame perfect equilibrium), 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, μm_i, diverges to infinity as well. The learning-rate σ 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 in such a way that the expected number of mutants in each period is bounded away from zero. Table 1 offers an overview of various characteristics of the limiting conditions of Hart and this paper under the simplifying assumption of equal population sizes $m_i = m$.

Clearly Hart’s limiting conditions are more general than either 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 precisely because Lemma 1

Table 1
A comparison of characteristics of the limiting conditions of Hart (2002), based on his model, and of this paper, based on the model of Nöldeke and Samuelson (1993)

	Hart	NaS 1	NaS 2
Number of mutations	$\mu m \geqslant \delta$	$\mu m \to \infty$	$\mu m \geqslant \delta$
Number of learn-draws	$\sigma m \to \infty$	$\sigma m \to \infty$	$\sigma m \to \infty$
Proportion of mutations	$\mu \to 0$	$\mu \to 0$	$\mu \to 0$
Proportion of learn-draws	σ fixed	σ fixed	$\sigma \to 0$
Number of mutations relative to learn-draws	$\frac{\mu}{\sigma} \to 0$	$\frac{\mu}{\sigma} \to 0$	$\frac{\mu}{\sigma} \to 0$

Columns NaS 1 and NaS 2 state the various figures for the two limiting conditions employed in this paper for the model of Nöldeke and Samuelson, where NaS 1 and NaS 2 refer to the conditions used in Theorem 1 and Theorem 2, respectively. For these two columns "number of mutations," etc., denotes the corresponding expected number of mutations, etc., per period. To adjust for the fact that in Hart's model only one person per period can change action, whereas potentially all m individuals can change action in every period in Nöldeke's and Samuelson's model, the column for Hart's model gives the expected number of mutations, and so on, that occur over a span of m periods.

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 also be obtained in more general games including games of perfect information with multiple subgame-perfect equilibria 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 complicates matters to a degree, hence the slightly weaker limiting conditions than in Hart. The final difference in the learning dynamics is that individuals in Hart's model play better replies, whereas in Nöldeke and Samuelson's model they play best-replies. It is easy to see, however, that changing the model from best to better replies will not affect the results. 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, since it is a better reply (to any probable situation) than is B. 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.

Throughout the paper I investigate the agent normal form version of the model of Nöldeke and Samuelson. Like Hart, I assume that there is a distinct population for each node of the game tree. Nöldeke and Samuelson in general allow populations to govern more than one node as long as it is at most one node along each possible path of play. However, individuals in their model adapt their strategy locally, i.e. after learning individuals investigate each action separately, and for each node they control choose a best action given their updated conjectures. Also, individuals' conjectures are not about other individuals' strategies, but about their actions at each node. In fact, when individuals update they do so "to match the observed frequency of actions at all information sets that were reached..." (Nöldeke and Samuelson, 1993, p. 430). Hence, Nöldeke and Samuelson do investigate the agent normal form as well.

This paper then demonstrates that in the model of Nöldeke and Samuelson, as in Hart, evolution leads to subgame-perfect equilibrium play if, in contrast to the fixed population assumption in Nöldeke and Samuelson, population sizes are taken to infinity in a suitable way.

Acknowledgments

This paper constitutes part of my PhD thesis written at University of Cambridge under the supervision of Hamid Sabourian. I am grateful to Hamid Sabourian, Florian Herold, Sergiu Hart, an anonymous referee, and an anonymous associate editor for helpful comments and suggestions.

Appendix A. Proof of Lemma 1

Define D_x^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 $\pi_\mu^m(\omega)$ can be expressed as

$$\pi_\mu^m(\omega) = \sum_{\omega' \in \Omega} \pi_\mu^m(\omega')\bigl(Q_\mu^m\bigr)_{\omega'\omega}, \tag{A.1}$$

where $(Q_\mu^m)_{\omega'\omega}$ is the transition probability that the system moves from ω' to ω.

Equivalently for any set of states, Λ,

$$\pi_\mu^m(\Lambda) = \sum_{\omega \in \Lambda} \sum_{\omega' \in \Omega} \pi_\mu^m(\omega')\bigl(Q_\mu^m\bigr)_{\omega'\omega}. \tag{A.2}$$

Changing the order of summation yields

$$\pi_\mu^m(\Lambda) = \sum_{\omega' \in \Omega} \pi_\mu^m(\omega')\bigl(Q_\mu^m\bigr)_{\omega'\Lambda}, \tag{A.3}$$

where $(Q_\mu^m)_{\omega'\Lambda} = \sum_{\omega \in \Lambda} (Q_\mu^m)_{\omega'\omega}$.

We are interested in the set $\Lambda = \Lambda_0^{i,x}$. It is easy to show that for any $\omega' \in \Lambda_k^{i,x}$,

$$\bigl(Q_\mu^m\bigr)_{\omega'\Lambda_0^{i,x}} \begin{cases} = p_{k0} & \forall\, \omega' \in D_x^i, \\ \leqslant p_{k0} & \text{otherwise,} \end{cases} \tag{A.4}$$

where

$$p_{k0} = \sum_{j=0}^{k} \sigma^j (1-\sigma)^{k-j} \binom{k}{j} (\mu\lambda_x)^{k-j} \bigl(1-\mu(1-\lambda_x)\bigr)^{m_i-k+j}. \tag{A.5}$$

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 ω is in D_x^i. A possible transition is that any $j \leqslant k$ 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 Eq. (A.5) reveals that

$$\begin{aligned} p_{k0} &= \bigl(1-\mu(1-\lambda_x)\bigr)^{m_i-k} \sum_{j=0}^{k} \binom{k}{j} \bigl(\sigma\bigl(1-\mu(1-\lambda_x)\bigr)\bigr)^j \bigl((1-\sigma)(\mu\lambda_x)\bigr)^{k-j} \\ &= \bigl(1-\mu(1-\lambda_x)\bigr)^{m_i-k} \bigl(\mu\lambda_x + \sigma(1-\mu)\bigr)^k. \end{aligned} \tag{A.6}$$

Hence, for all $k < m_i$,

$$\frac{p_{k+1,0}}{p_{k0}} = \frac{\mu\lambda_x + \sigma(1-\mu)}{1-\mu(1-\lambda_x)}, \tag{A.7}$$

which is less than 1 for small μ.

Using Eqs. (A.3) and (A.4) yields

$$\pi_\mu^m\bigl(\Lambda_0^{i,x}\bigr) \leqslant \sum_{k=0}^{m_i} \pi_\mu^m\bigl(\Lambda_k^{i,x}\bigr) p_{k0}. \tag{A.8}$$

Rearranging leads to

$$\pi_\mu^m\bigl(\Lambda_0^{i,x}\bigr) \leqslant \frac{1}{1-p_{00}} \sum_{k=1}^{m_i} \pi_\mu^m\bigl(\Lambda_k^{i,x}\bigr) p_{k0} \tag{A.9}$$

and hence

$$\pi_\mu^m\bigl(\Lambda_0^{i,x}\bigr) \leqslant \frac{1-\pi_\mu^m(\Lambda_0^{i,x})}{1-p_{00}} \max_{k\geqslant 1}\{p_{k0}\}. \tag{A.10}$$

Finally,

$$\pi_\mu^m\bigl(\Lambda_0^{i,x}\bigr) \leqslant 1 \Bigg/ \left(1 + \frac{1-p_{00}}{\max_{k\geqslant 1}\{p_{k0}\}}\right). \tag{A.11}$$

By Eq. (A.7), $\max_{k\geqslant 1}\{p_{k0}\} = p_{10}$ for μ small enough. This confirms the intuition that the easiest way to move to $\Lambda_0^{i,x}$ is coming from $\Lambda_1^{i,x}$.

Now, by Eqs. (A.6) and (A.7),

$$p_{10} = \bigl(1-\mu(1-\lambda_x)\bigr)^{m_i} \frac{\mu\lambda_x + \sigma(1-\mu)}{1-\mu(1-\lambda_x)}. \tag{A.12}$$

Hence,

$$\forall m_i \ \forall \kappa > 1 \quad \exists \bar{\mu}\colon \forall \mu \leqslant \bar{\mu}$$
$$p_{10} \leqslant \kappa\sigma\bigl(1-\mu(1-\lambda_x)\bigr)^{m_i}. \tag{A.13}$$

Hence, for all m_i and for all $\kappa > 1$ there is a $\bar{\mu}$ such that for all $\mu < \bar{\mu}$,

$$\pi_\mu^m\bigl(\Lambda_0^{i,x}\bigr) \leqslant 1\bigg/\left(1+\frac{1-(1-\mu(1-\lambda_x))^{m_i}}{\kappa\sigma(1-\mu(1-\lambda_x))^{m_i}}\right). \qquad \square \tag{A.14}$$

Appendix B. Proof of Lemma 4

Let $\{\Omega \times \Omega, P\}$ denote a probability space,[3] where P is such that[4] $P(\omega,\omega') = \pi_\mu^m(\omega)(Q_\mu^m)_{\omega,\omega'}$ for all $(\omega,\omega') \in \Omega \times \Omega$. Define $U : \Omega \times \Omega \to \{0,1,\dots,m_i\}$ such that $U(\omega,\omega')$ is the number of individuals at population $M(i)$ who play b^i in state ω. Similarly, let $V : \Omega \times \Omega \to \{0,1,\dots,m_i\}$ be a random variable such that $V(\omega,\omega')$ is the number of individuals at population $M(i)$ who play b^i in state ω'. Note that

$$\bigl\{\omega \in \Omega \mid U(\omega,\omega') = k\bigr\} = \bigl\{\omega' \in \Omega \mid V(\omega,\omega') = k\bigr\} = \Lambda_k^{i,b^i}.$$

Let $Z : \Omega \times \Omega \to \{-m_i, -m_i+1, \dots, -1, 0, 1, \dots, m_i\}$ denote a third random variable such that $Z(\omega,\omega')$ is the "addition" of b^i-players at population $M(i)$ in the transition from state ω to ω'. Obviously $Z(\omega,\omega') = V(\omega,\omega') - U(\omega,\omega')$.

Note that $P(U = k) = \pi_\mu^m(\Lambda_k^{i,b^i})$ by definition. Also $P(V = k) = \pi_\mu^m(\Lambda_k^{i,b^i})$. To see this let $(\cdot,\omega') = \{(\omega,\omega') \mid \omega \in \Omega\}$. Then $P(\cdot,\omega') = \sum_{\omega\in\Omega} P(\omega,\omega') = \sum_{\omega\in\Omega} \pi_\mu^m(\omega) \times (Q_\mu^m)_{\omega,\omega'} = \pi_\mu^m(\omega')$ by definition of the invariant distribution. But the set of states where $V = k$ is just $\bigcup_{\omega' \in \Lambda_k^{i,b^i}} (\cdot,\omega')$. Hence, $P(V = k) = \pi_\mu^m(\Lambda_k^{i,b^i})$.

Given this we have $E(U) = E(V)$ and, hence, $E(Z) = 0$. The expectation of Z can be written as $E(E(Z \mid U))$ by the law of iterated expectations. Let $C' = \{(\omega,\omega') \mid \omega \in C_{b^i}^i\}$. Obviously $P(C') = \pi_\mu^m(C_{b^i}^i)$. Then

$$E\left(\frac{Z}{m_i}\right) = \pi_\mu^m\bigl(C_{b^i}^i\bigr)E\left(\frac{Z}{m_i}\,\middle|\, C'\right) + \sum_{k=0}^{m_i} \pi_\mu^m\bigl(\Lambda_k^{i,b^i} \setminus C_{b^i}^i\bigr)E\left(\frac{Z}{m_i}\,\middle|\, U = k \wedge (C')^c\right). \tag{B.1}$$

$$\begin{aligned} &= \pi_\mu^m\bigl(C_{b^i}^i\bigr)E\left(\frac{Z}{m_i}\,\middle|\, C'\right) + \sum_{k=0}^{m_i} \pi_\mu^m\bigl(\Lambda_k^{i,b^i}\bigr)E\left(\frac{Z}{m_i}\,\middle|\, U = k \wedge (C')^c\right) \\ &\quad - \sum_{k=0}^{m_i} \pi_\mu^m\bigl(\Lambda_k^{i,b^i} \cap C_{b^i}^i\bigr)E\left(\frac{Z}{m_i}\,\middle|\, U = k \wedge (C')^c\right). \end{aligned} \tag{B.2}$$

---

[3] As the state space is finite, I omit the sigma-algebra, which can be taken as the set of all subsets of $\Omega \times \Omega$, in the description of the probability space.

[4] Given the axioms of a probability measure this is sufficient to uniquely define P.

Now,

$$E\left(\frac{Z}{m_i} \,\middle|\, U = k \wedge C'\right) \geqslant -1$$

as the greatest change in b^i-players can never exceed the total number of individuals at $M(i)$. Similarly

$$E\left(\frac{Z}{m_i} \,\middle|\, U = k \wedge (C')^c\right) \leqslant 1.$$

We then have

$$0 = E\left(\frac{Z}{m_i}\right) \geqslant -2\pi_\mu^m\left(C_{b^i}^i\right) + \sum_{k=0}^{m_i} \pi_\mu^m\left(\Lambda_k^{i,b^i}\right) E\left(\frac{Z}{m_i} \,\middle|\, U = k \wedge (C')^c\right). \tag{B.3}$$

Let

$$\alpha_k = E\left(\frac{Z}{m_i} \,\middle|\, U = k \wedge (C')^c\right).$$

Rearranging then yields

$$2\pi_\mu^m\left(C_{b^i}^i\right) \geqslant \sum_{k=0}^{m_i} \pi_\mu^m\left(\Lambda_k^{i,b^i}\right)\alpha_k. \tag{B.4}$$

Let $k_* = \lfloor (1-\epsilon)m_i \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer smaller than x. By Lemma 5 (see below), there is an $\bar{\alpha} > 0$ such that for all $k < k_*$ $\alpha_k \geqslant \bar{\alpha}$ provided μ is small enough. Also $\alpha_k \geqslant \alpha_{m_i} = -\mu\lambda_{b^i}$ for all k, in particular also for all $k \geqslant k_*$.

Hence,

$$\begin{aligned}
\sum_{k=0}^{m_i} \alpha_k \pi_\mu^m\left(\Lambda_k^{i,b^i}\right) &\geqslant \sum_{k=0}^{k_*} \bar{\alpha}\pi_\mu^m\left(\Lambda_k^{i,b^i}\right) - \sum_{k=k_*+1}^{m_i} \mu\lambda_{b^i}\pi_\mu^m\left(\Lambda_k^{i,b^i}\right) &\text{(B.5)}\\
&\geqslant \bar{\alpha}\pi_\mu^m\left(B_{\epsilon,m}^{i,c}\right) - \mu\lambda_{b^i}\pi_\mu^m\left(B_{\epsilon,m}^i\right) &\text{(B.6)}\\
&\geqslant \bar{\alpha}\pi_\mu^m\left(B_{\epsilon,m}^{i,c}\right) - \mu\lambda_{b^i}\left(1 - \pi_\mu^m\left(B_{\epsilon,m}^{i,c}\right)\right) &\text{(B.7)}\\
&\geqslant (\bar{\alpha} + \mu\lambda_{b^i})\pi_\mu^m\left(B_{\epsilon,m}^{i,c}\right) - \mu\lambda_{b^i}. &\text{(B.8)}
\end{aligned}$$

Combining inequalities (B.4) and (B.8), we obtain

$$(\bar{\alpha} + \mu\lambda_{b^i})\pi_\mu^m\left(B_{\epsilon,m}^{i,c}\right) - \mu\lambda_{b^i} \leqslant \sum_{k=0}^{m_i} \alpha_k \pi_\mu^m\left(\Lambda_k^{i,b^i}\right) \leqslant 2\pi_\mu^m\left(C_{b^i}^i\right). \tag{B.9}$$

Taking $\mu \to 0$ while $m_i\mu$ tends to infinity in inequality (B.9), we obtain

$$\bar{\alpha} \lim_{\mu\to 0,\ m_i\mu\to\infty} \pi_\mu^m\left(B_{\epsilon,m}^{i,c}\right) \leqslant 0. \tag{B.10}$$

Hence, $\pi_\mu^m(B_{\epsilon,m}^{i,c}) \to 0$. □

Lemma 5. *Let $k_* = \lfloor (1-\epsilon)m_i \rfloor$. There is a $\bar{\alpha} > 0$ and a $\bar{\mu} > 0$ such that for all $k < k_*$ $\alpha_k \geqslant \bar{\alpha}$ provided $\mu < \bar{\mu}$. Also $\alpha_k \geqslant \alpha_{m_i} = -\mu\lambda_{b^i}$ for all k.*

Proof. By definition,

$$\alpha_k = \frac{1}{m_i} E\bigl(Z \mid U = k \wedge (C')^c\bigr).$$

To calculate the term $E(Z \mid U = k \wedge (C')^c)$ note that Z can be written as the difference of two random variables X and Y (different from U and V), where $X(\omega, \omega')$ is the number of individuals at $M(i)$ who, in the transition from ω to ω', switch strategy from something other than b^i to b^i, and $Y(\omega, \omega')$ is the number of individuals at $M(i)$ who, in the transition from ω to ω', switch strategy from b^i to any other than b^i. Conditional on $U(\omega, \omega') = k$ and $(\omega, \omega') \in (C')^c$, both X and Y are binomially distributed, i.e. $X \sim \mathrm{Bin}(m_i - k, \sigma(1-\mu) + \mu(1 - \lambda_{b^i}))$ and $Y \sim \mathrm{Bin}(k, \mu\lambda_{b^i})$. Hence, the term $E(Z \mid U = k \wedge (C')^c)$ is the difference between the expectation of these two binomial variables and given by

$$E\bigl(Z \mid U = k \wedge (C')^c\bigr) = \bigl(\sigma(1-\mu) + \mu(1 - \lambda_{b^i})\bigr)(m_i - k) - k\mu\lambda_{b^i}.$$

The term α_k is then negative if and only if

$$\frac{k}{m_i} > \frac{\sigma(1-\mu) + \mu(1-\lambda_{b^i})}{\sigma(1-\mu) + \mu}. \tag{B.11}$$

In particular if $k = \epsilon m_i$, $\alpha_k < 0$ if

$$\epsilon > \frac{\sigma(1-\mu) + \mu(1-\lambda_{b^i})}{\sigma(1-\mu) + \mu}. \tag{B.12}$$

It is easy to see that $\alpha_{m_i} < 0$. However, for an arbitrary $\epsilon > 0$, $\alpha_k > 0$ for all $k < (1-\epsilon)m_i$, provided μ is small enough. Indeed there is a $\bar{\mu} > 0$ and an $\bar{\alpha} > 0$ such that for all $\mu \leqslant \bar{\mu}$ we have that $\alpha_k \geqslant \bar{\alpha}$ for all $k \leqslant k_*$. Suppose, for the sake of simplicity, that $(1-\epsilon)m_i$ is an integer. Then $\alpha_{k_*} = \epsilon(\sigma(1-\mu) + \mu(1-\lambda_{b^i})) - (1-\epsilon)\mu\lambda_{b^i}$. One might, for instance, set $\bar{\alpha} = \epsilon\sigma/2$. Also observe that for all k we have that $\alpha_k \geqslant \alpha_{m_i} = -\mu\lambda_{b^i}$. □

References

Cressman, R., Schlag, K.H., 1998. The dynamic (in)stability of backwards induction. J. Econ. Theory 83, 260–285.

Demichelis, S., Ritzberger, K., 2000. From evolutionary to strategic stability. J. Econ. Theory 113, 5–75.

Demichelis, S., Ritzberger, K., Swinkels, J.M., 2002. The simple geometry of perfect information games. Institute for Advanced Studies, Vienna, Economics Series working paper No. 113, June.

Hart, S., 2002. Evolutionary dynamics and backward induction. Games Econ. Behav. 41, 227–264.

Kandori, M., Mailath, G., Rob, R., 1993. Learning, mutation, and long-run equilibria in games. Econometrica 61, 29–56.

Nöldeke, G., Samuelson, L., 1993. An evolutionary analysis of backward and forward induction. Games Econ. Behav. 5, 425–454.

Swinkels, J.M., 1993. Adjustment dynamics and rational play in games. Games Econ. Behav. 5, 455–484.

Young, H.P., 1993. The evolution of conventions. Econometrica 61, 57–84.