

EVOLUTIONARY FOUNDATIONS OF SOLUTION  
CONCEPTS FOR FINITE, TWO-PLAYER,  
NORMAL-FORM GAMES

Larry Samuelson  
Department of Economics  
The Pennsylvania State University  
and  
University of Illinois  
1206 South Sixth Street  
Champaign, Illinois 61820

ABSTRACT

This paper develops evolutionary foundations for noncooperative game-theoretic solution concepts. In particular, we envision a game as being repeatedly played by randomly, anonymously matched members of two populations. Agents initially play arbitrarily chosen pure strategies. As play progresses, a learning process or selection mechanism induces agents to switch from less to more profitable strategies. The limiting outcomes of this dynamic process yield equilibria for the game in question, and the plausibility of an equilibrium concept then rests on the characteristics of the selection process from which it arises. The results suggest that if one accepts the evolutionary approach to equilibrium concepts, then one will embrace either rationalizable or perfect equilibria. The choice between the two hinges upon whether the evolutionary process is sufficiently well behaved as to yield convergence. In general, there are robust adjustment processes which converge as well as robust processes which do not converge.

The standard noncooperative game-theoretic equilibrium concept of a Nash equilibrium (Nash (1951)) has been supplemented by a collection of new solution concepts. In the course of evaluating these solution concepts, attention turns to their behavioral foundations. Two approaches to such foundations exist. First, one can view an equilibrium as the result of each agent's calculation of an optimal strategy contingent upon the agent's information. Differing equilibrium concepts are then a reflection of differences in knowledge, primarily knowledge about one's opponent. Research along these lines has revealed that extremely strong assumptions on players' knowledge are required to yield Nash or alternative equilibrium concepts.

These results direct attention to a second foundation for equilibrium concepts. Suppose that a selection mechanism or market forces cause agents who choose relatively profitable strategies to prosper and those who choose relatively unprofitable strategies to falter. As a result, the proportion of agents choosing one of the former strategies increases over time. This provides an evolutionary account of the emergence of an equilibrium which is analogous to the familiar contention that firms can be modeled as profit maximizers because market forces will force non-maximizers out of the market. Differing equilibrium concepts are here a reflection of differences in the selection mechanism that transfers agents from unprofitable to profitable strategies.

This paper investigates the evolutionary foundations of solution concepts. In order to isolate the issues of interest, we confine the analysis to finite, two-player, normal-form games of complete information. In addition, a variety of possibilities arise for modeling the selection or learning process by which strategy adjustments are made. We model the players as naive automata in this process. This allows us to investigate the ability of market forces to cause completely unreasoning agents to act as if they were completely knowledgeable.

We introduce a class of dynamic processes, called ordinal processes, to represent the evolutionary mechanism. We demonstrate that such adjustment processes may not converge. However, there exists a process for any game which does converge and which is robust in the sense that small perturbations in its specification do not vitiate convergence. In the absence of convergence (and with some additional technical conditions), the process will yield rationalizable strategies (cf. Bernheim (1984), Pearce (1984)) with a limiting probability of unity. If the process converges, then the result is not only a Nash but a perfect equilibrium (Selten (1975)).

We take these results to indicate that if one believes in a market-based or evolutionary theory of games, then one is naturally led to embrace the concepts of rationalizability and perfection, with the former (latter) applying to cases in which a single pattern of behavior has not (has) emerged from the evolutionary process. The effect of the evolutionary process is thus profound, as mechanical agents can be induced to act as if they possess and optimize against quite detailed knowledge.

This analysis must be considered preliminary, since many results depend crucially upon the special structure of the simple class of games chosen for study. The analysis must be extended beyond this type of game and to consideration of more sophisticated (perhaps optimal) learning processes before it can be fully assessed.

The following section introduces the model and develops notation. This is followed by the development of the basic evolutionary argument. The penultimate section derives the main results. The final section discusses implications and extensions.

### GAMES AND EQUILIBRIA

A finite, two-player, normal-form game of complete information, hereafter called simply a game, consists of two agents, referred to as agents 1 and 2; two finite sets  $S_1$  and  $S_2$  with elements denoted by  $s_1$  and  $s_2$ , referred to as the agents' pure strategy sets; and two payoff functions  $\pi_1: S_1 \times S_2 \rightarrow \mathbb{R}$  and  $\pi_2: S_1 \times S_2 \rightarrow \mathbb{R}$ , denoted  $\pi_1(s_1, s_2)$  and  $\pi_2(s_1, s_2)$ . We let  $(S_1, S_2, \pi_1, \pi_2)$  denote a game. Players prefer higher to lower payoffs. Let  $n_1$  and  $n_2$  be the cardinalities of  $S_1$  and  $S_2$ . We let  $M_i$  ( $i = 1, 2$ ) denote the set of probability measures on the space  $(S_i, P(S_i))$ , where  $P(S_i)$  is the  $\sigma$ -algebra given by the power set of  $S_i$ . We refer to elements of  $S_i$  ( $M_i$ ) as pure (mixed) strategies. We let  $m_i$  denote an element of  $M_i$  and let  $m_i(s_i)$  denote the probability attached to pure strategy  $s_i \in S_i$  by mixed strategy  $m_i \in M_i$ . We also abuse notation somewhat to let

$$\pi_i(m_1, m_2) = \sum_{S_1} \sum_{S_2} \pi_i(s_1, s_2) m_1(s_1) m_2(s_2).$$

Definition 1. A pair  $(m_1^*, m_2^*) \in M_1 \times M_2$  is a Nash equilibrium if

$$\begin{aligned} \pi_1(m_1^*, m_2^*) &\geq \pi_1(m_1, m_2^*) && \forall m_1 \in M_1 \\ \pi_2(m_1^*, m_2^*) &\geq \pi_2(m_1^*, m_2) && \forall m_2 \in M_2. \end{aligned} \tag{1}$$

A Nash equilibrium  $(m_1^*, m_2^*)$  is strict if the first (second) inequality in (1) is strict for all  $m_1 \neq m_1^*$  ( $m_2 \neq m_2^*$ ). A Nash equilibrium  $(m_1^*, m_2^*)$  is a dominant strategy equilibrium if the first (second) inequality in (1) holds for all  $m_2 \in M_2$  ( $m_1 \in M_1$ ).

The concept of a Nash equilibrium has been criticized for placing restrictions on behavior which are both apparently too weak and too strong. Consider, for example, the following games:

		2	
		$s_2$	$s'_2$
1	$s_1$	2,2	2,2
	$s'_1$	3,1	1,0

		2	
		$s_2$	$s'_2$
1	$s_1$	0,0	1,1
	$s'_1$	1,1	0,0

Two pure strategy Nash equilibria of the first game appear:  $(s_1, s'_2)$  and  $(s'_1, s_2)$ . The former calls for player 2 to play a dominated strategy and hence is regarded by some as unreasonable. Two pure strategy Nash of the second game equilibria again appear:  $(s'_1, s_2)$  and  $(s_1, s'_2)$ , as well as one mixed strategy equilibrium. However, it is not at all obvious that players 1 and 2, each choosing in ignorance of the other's action, will achieve one of these equilibria.

To address these concerns, a host of alternative equilibrium notions have been formulated. We consider two.

Definition 2. Let  $z_i \in R^{n_i}$  be a strictly positive  $n_i$ -tuple whose elements sum to less than unity. An equilibrium  $(m_1^*, m_2^*)$  of game  $(S_1, S_2, \pi_1, \pi_2)$  is perfect if there exists a sequence  $\{(z_1^n, z_2^n)\}_{n=1}^\infty$  such that the "perturbed games"  $(\{m_1 \in M_1: m_1 \geq z_1^n\}, \{m_2 \in M_2: m_2 \geq z_2^n\}, \pi_1, \pi_2)$  yield a sequence of equilibria converging to  $(m_1^*, m_2^*)$  (in the standard topology that  $M_1 \times M_2$  inherits as a subspace of  $R^{n_1} \times R^{n_2}$ ).

Definition 3. A strategy  $m_1' \in X \subset M_1$  is a best response in  $X$  to  $m_2$ , denoted  $m_1' \in BR_1(X, m_2)$  if  $\pi_1(m_1', m_2) \geq \pi_1(m_1, m_2) \forall m_1 \in X$ . Adopting a similar convention for player 2, let  $H_i^0 = M_i$  and

$$H_i^t = \bigcup_{H_j^{t-1}} BR_i(H_i^{t-1}, m_j), \quad \bar{H}_i = \bigcup_{t=1}^\infty H_i^t.$$

Then a pair of strategies  $(m_1^*, m_2^*)$  is a rationalizable equilibrium if

$$m_i^* \in \bar{H}_i, \quad i = 1, 2.$$

A perfect equilibrium is Nash and a Nash equilibrium is rationalizable. Each converse fails. Each game has at least one perfect (and hence Nash and rationalizable) equilibrium. Only  $(s_1', s_2')$  is perfect in the first game above; any outcome in the second is rationalizable.

### EVOLUTIONARY PROCESSES

In order to evaluate these equilibrium concepts, one would like a clear idea of the primitive assumptions about players and their environments which give rise to the various equilibrium concepts. This question has been addressed by Bernheim (1986) and Tan and Werlang (1987). Players are generally considered to be Bayesian rational decision makers in these analyses, by which it is meant that players maximize expected payoffs subject to beliefs about the strategies of their opponents. The assumptions which characterize differing equilibrium concepts then concern the knowledge players have about their opponents and the inferences that this knowledge allows concerning opponents' strategies. Among other results, Tan and Werlang show that if Bayesian rationality is common knowledge, then a rationalizable equilibrium will appear (in a two-player game). Several sets of assumptions are developed under which a Nash equilibrium will appear, each of which is significantly stronger than common knowledge of rationality.

In general, the assumption of common knowledge of rationality is quite strong, not to mention assumptions of additional knowledge. Milgrom and Roberts (1982), for example, suggest that important phenomena can be explained by the failure of rationality to be common knowledge. This directs attention to a second, evolutionary foundation for equilibrium concepts.

Let there exist two infinite sets of agents, one of whose members potentially fill the role of agent 1 in playing a game and one of whose members potentially fill the role of agent 2. The members of these sets are repeatedly and randomly matched to play single iterations of a fixed game. Players either cannot observe or cannot subsequently recall what strategies are played in a given period by opponents with whom they are not matched. This combines with the zero probability of a future rematch with a current opponent to ensure that play is memoryless, so that choices in a given game are affected only by strategic considerations arising within that game.

Agents play pure strategies. Outcomes that appear to involve mixed strategies will arise from having various representatives of the population on one side of the game choose different pure strategies. This provides an alternative to the standard interpretations of mixed strategies as

either explicit randomizations undertaken by agents or as opponents' expectations concerning an agent's pure strategy.

We presume that each population of agents is somehow initially distributed over the pure strategy set. A dynamic adjustment process causes agents to shift from relatively low to relatively high payoff strategies. This presumably reflects the fact that players have (limited) opportunities to observe and communicate with other members of their population, causing players to gradually become aware of strategies' relative payoffs.

This approach borrows ideas from two areas. Biologists have developed the idea of an evolutionarily stable strategy (Maynard Smith (1982)). While this is similar in spirit, the evolutionary or dynamic process employed in this literature is appropriate for a biological model, in which strategy variations are caused by genetic mutations followed by natural selection, but less appropriate for game theoretic applications in the social sciences. The convention of fictitious play (Brown (1951)) also yields a dynamic process which is similar in spirit to our analysis but somewhat specialized.

Research on rational expectations equilibria (REE) has recently considered the question of whether a learning or adjustment process will converge to a REE (Feldman (1987), Jordan (1986)). Our analysis differs in three ways. First, much of the REE analysis has been concerned with cases in which agents must learn some parameter or feature of the economy as well as opponents' behavior, unlike our analysis. Second, these studies often impose smoothness conditions on learning rules that will be violated by our ordinal rule. Most important, since we are interested in characterizing the implications of varying adjustment process formulations for solution concepts, the failure to converge or failure to converge to a particular type of equilibrium is not necessarily a negative result in our case.

One can conceive of learning processes in which agents are uninformed but sophisticated, forming expectations concerning their uncertainty and updating them via Bayes rule in accordance with their observations. Alternatively, one can conceive of processes involving unsophisticated agents who follow arbitrarily specified, mechanistic rules. We pursue the latter approach in an effort to achieve seemingly rational outcomes with a minimum of demands on agents' rationality.

We begin by constructing a continuous-time dynamic process. Let  $p_t \in R_+^{n_1}$  ( $= (p_t(s_1^1), \dots, p_t(s_1^{n_1}))$ ) and  $q_t \in R_+^{n_2}$  be vectors identifying the proportions of populations 1 and 2 playing each of the pure strategies in  $S_1$  and  $S_2$  at time  $t$ . Let  $\theta: R_+^{n_2} \rightarrow R_+^{n_1}$ , denoted  $\theta(q)$  and  $\gamma: R_+^{n_1} \rightarrow R_+^{n_2}$ , denoted  $\gamma(p)$ , be functions identifying the average

payoffs to each of population 1's ( $\theta$ ) or 2's ( $\gamma$ ) pure strategies given that the opposing population is characterized by  $q$  or  $p$ . Let  $\#(\theta(q_t), s_1)$   $\#(\gamma(p_t), s_2)$  identify the number of population 1 or population 2 pure strategies that yield a strictly higher payoff than strategy  $s_1$  or  $s_2$ .

We then define matrices of transition probabilities consisting of bounded functions.

$$\begin{aligned}
 x_t^{ij} &= x^{ij}(\theta(q_t), p_t) \\
 y_t^{ij} &= y^{ij}(\gamma(p_t), q_t),
 \end{aligned}
 \tag{2}$$

where  $x_t^{ij}$  is the time  $t$  instantaneous proportion of population 1 agents playing strategy  $i$  who switch to strategy  $j$ .  $y_t^{ij}$  is similar for population 2 and  $x_t^{ii} = y_t^{ii} = 0$ . Hence,

$$\frac{dp_t(s_1^i)}{dt} = \sum_{j=1}^n (\max[-p_t(s_1^j)x_t^{ij}, 0] - \max[p_t(s_1^i)s_t^{ij}, 0]).
 \tag{3}$$

We assume that initially, some positive proportion of each population plays each of the population's pure strategies, so that  $p_0 > 0$ ,  $q_0 > 0$ .

Coupled with (3) and the boundedness of the  $x_t^{ij}$  and  $y_t^{ij}$ , this ensures that the proportion attached to a particular strategy may approach but will never equal zero. We say that the process converges to an equilibrium  $(\bar{p}, \bar{q})$  if  $p_t$  and  $q_t$  converge to  $\bar{p}$  and  $\bar{q}$  in the standard topologies on  $R^{n_1}$  and  $R^{n_2}$ .

We will be interested in several possible properties of these transition rates. Let  $\theta(q_t)(s_1^i)$  ( $\gamma(p_t)(s_2^i)$ ) be the average payoff to player 1 (2) pure strategy  $s_1^i$  ( $s_2^i$ ) given that the opposing population is characterized by  $q_t$  ( $p_t$ ).

Definition 4. The  $x_t^{ij}$  are said to be (similar definitions apply to the  $y_t^{ij}$ ):

(4.1) Monotonic if  $\theta(q_t)(s_1^j) \leq (>) \theta(q_t)(s_1^i) \Rightarrow x_t^{ij} = (>) 0$ .

(4.2) Strongly monotonic if (4.1) holds and  $\theta(q_t)(s_1^j) < \max_k \theta(q_t)(s_1^k) \Rightarrow x_t^{ij} = 0$ .

- (4.3) Regular if  $x_t^{ij} > 0$  whenever consistent with the maintained monotonicity assumption and  $x_t^{ij} \searrow 0 \Rightarrow \lim_{t \rightarrow \infty} \theta(q_t)(s_1^j) - \theta(q_t)(s_1^i) \leq 0$  or (for strongly monotonic processes)
- $$\lim_{t \rightarrow \infty} \theta(q_t)(s_1^j) - \theta(q_t)(s_1^k) < 0 \text{ for some } k \in S_1.$$
- (4.4) Ordinal if  $x_t^{ij}$  depends only on  $\#(\theta(q_t), s_1^i)$ .
- (4.5) Cardinal if they are not ordinal.

Monotonicity ensures that if agents switch strategies, they will switch only to more profitable strategies. Strong monotonicity adds the presumption that agents who do switch strategies adopt the currently most profitable strategy. Regularity ensures that if a difference in the payoffs of strategies  $i$  and  $j$  potentially causes agents to switch from  $i$  to  $j$ , then some agents will actually make such a switch. Furthermore, the proportion of agents who switch approaches zero only if the payoff differential motivating the potential switch disappears. Some assumptions of this type are necessary if the evolutionary process is to be described as one in which agents learn and exploit more profitable strategies. Finally, only the relative position of a strategy in a payoff ranking affects strategy transitions in an ordinal process. In a cardinal process, the magnitudes of payoff differences may also play a role.

Both cardinal and ordinal processes have appealing features. The former allows transition rates to be continuous in payoffs, and this property may account for the popularity of cardinal processes in studies of rational expectations and other applications. However, as argued by Mertens (1987), noncooperative game theory is traditionally viewed as an ordinal theory, and it is this view that motivates our interest in ordinal evolutionary processes. Friedman and Rosenthal (1986) offer an alternative motivation for ordinality.

#### EVOLUTIONARY FOUNDATIONS

We can now derive some implications of evolutionary processes.

**Proposition 1.** Let the evolutionary process be ordinal, regular, and strongly monotonic. Let  $s_1$  ( $s_2$ ) be a strategy for player 1 (2) which is not rationalizable. Then  $\lim_{t \rightarrow \infty} p_t(s_1) = 0$  ( $\lim_{t \rightarrow \infty} q_t(s_2) = 0$ ).

**Proof.** Let  $P = \{s_1^i : \lim_{t \rightarrow \infty} p_t(s_1^i) \neq 0\}$  and  $Q = \{s_2^j : \lim_{t \rightarrow \infty} q_t(s_2^j) \neq 0\}$ .

Let  $\Delta(P) = \{p \in M_1 : s_1^i \notin P \Rightarrow p(s_1^i) = 0\}$  and  $\Delta(Q) = \{q \in M_2 : s_2^j \notin Q \Rightarrow$

$q(s_2^j) = 0$ . It suffices to show that every  $s_1^i \in P$  ( $s_2^j \in Q$ ) is a best response in  $P$  ( $Q$ ) to some  $q \in \Delta(Q)$  ( $p \in \Delta P$ ). (See Proposition 2 of Pearce (1984).) Consider  $P$  ( $Q$  is analogous). Suppose that  $s_1^i \in P$  is not a best response in  $P$  to any  $q \in \Delta(Q)$ . One then easily verifies that for  $t$  sufficiently large,  $s_1^i$  is not a best response in  $P$  to any  $q_\tau$  such that  $\tau > t$ . Strong monotonicity and regularity then ensure that  $dp_t(s_1^i)/dt < 0$  for all sufficiently large  $t$  while ordinality ensures that  $(1/p_t(s_1^i))(dp_t(s_1^i)/dt)$  is bounded away from zero. Hence,  $\lim_{t \rightarrow \infty} p_t(s_1^i) = 0$ , contradicting  $s_1^i \in P$ .

||

A less stringent condition than strong monotonicity would suffice for this proof, but some assumption is required whose effect is to ensure that enough of the agents who switch strategies switch to strategies significantly better than their current strategy. This necessity arises because a strategy which is not rationalizable can never be the highest-payoff strategy, but it may often not be the lowest. One must ensure that it does not gain enough converts from lower-payoff strategies to replenish those lost to higher-payoff strategies. Strong monotonicity does this while providing the simplest exposition of the issues.

**Proposition 2.** Let the evolutionary process be monotonic and regular. Then if it converges, it converges to a Nash equilibrium.

**Proof.** Let the evolutionary process converge to  $(\bar{p}, \bar{q})$ . Let  $\bar{p}(s_1^i) > 0$  and  $\bar{p}(s_1^j) > 0$ . Then  $\lim_{t \rightarrow \infty} \theta(q_t)(s_1^i) - \theta(q_t)(s_1^j) = \theta(\bar{q})(s_1^i) - \theta(\bar{q})(s_1^j) = 0$ . If not, monotonicity and regularity ensure that one and only one of  $x_t^{ij}$  or  $x_t^{ji}$  is bounded away from zero, precluding convergence to  $\bar{p}$ . Next, let  $\bar{p}(s_1^k) = 0$ . Then  $\lim_{t \rightarrow \infty} \theta(q_t)(s_1^k) - \theta(q_t)(s_1^i) = \theta(\bar{q})(s_1^k) - \theta(\bar{q})(s_1^i) \leq 0$  for  $s_1^i$  such that  $\bar{p}(s_1^i) > 0$ . If not, monotonicity and regularity would give  $x_t^{ik} \geq 0$  and then  $\lim_{t \rightarrow \infty} p_t(s_1^i) > 0 \Rightarrow \lim_{t \rightarrow \infty} p_t(s_1^i)x_t^{ik} > 0$ , again precluding convergence to  $\bar{p}$ .

||

**Proposition 3.** Let the evolutionary process be ordinal, monotonic and regular. Then if it converges, it converges to a perfect equilibrium.

Proof. Let the evolutionary process converge to Nash equilibrium  $(\bar{p}, \bar{q})$  (cf. Proposition 2). If  $\bar{p}$  and  $\bar{q}$  attach positive probability to no dominated strategies, then  $(\bar{p}, \bar{q})$  is a perfect equilibrium (cf. van Damme

(1983)). Let  $s_1^i$  be a dominated strategy for player 1 and  $\bar{p}(s_1^i) > 0$ . Then  $s_1^i$  cannot solve  $\text{argmax } \theta(q_t)(s_1)$ , and hence monotonicity and regularity ensure that  $x_t^{ij} > 0$  for some  $j$  and all  $t$ . Furthermore,  $p_t(s_1^k)x_t^{ki}$  must approach zero for any  $k$  such that  $\lim_{t \rightarrow \infty} \theta(q_t)(s_1^k) < \lim_{t \rightarrow \infty} \theta(q_t)(s_1^i)$  (because  $p_t(s_1^k)$  must approach zero). Hence, we have  $dp_t(s_1^i)/dt < 0$  for all  $t$  sufficiently large. Moreover, ordinality ensures that  $(1/p_t(s_1^i))(dp_t(s_1^i)/dt)$  does not approach zero. Hence,  $\lim_{t \rightarrow \infty} p_t(s_1^i) = 0$ , a contradiction.  $\parallel$

Example 1. Ordinality is required in Proposition 3. Consider:

		2	
		$s_2$	$s_2'$
1	$s_1$	1, 1	1, 1
	$s_1'$	2, 0	-1, -2

The unique perfect equilibrium is  $(s_1', s_2)$ . However, let  $p$  be the proportion of population 1 agents playing strategy  $s_1$  and  $q$  the proportion of population 2 agents playing  $s_2$ . Notice that  $dq_t/dt > 0$ , since strategy  $s_2$  dominates  $s_2'$ , and that  $dp_t/dt > 0$  iff  $q_t < 2/3$ . Let the evolutionary process satisfy  $\gamma(p_t)(s_2') \rightarrow \gamma(p_t)(s_2) \Rightarrow dq_t/dt \rightarrow 0$ , violating ordinality. Then let the initial condition allocate most of population 1's (2's) agents to strategy  $s_1$  ( $s_2$ ), so  $p_0$  is large and  $q_0$  small. The initial dynamics give  $dp_t/dt > 0$  and  $dq_t/dt > 0$ , with  $\gamma(p_t)(s_2)$  approaching  $\gamma(p_t)(s_2')$  and hence  $dq_t/dt$  approaching zero. If this rate approaches zero quickly enough, the system will converge to a Nash but not perfect equilibrium in which  $p = 1$ ,  $0 < q < 2/3$ .

To interpret these results, notice that any monotonic adjustment process must cause some agents to flow away from a dominated strategy. Hence, such a process cannot converge to a Nash equilibrium in which all agents on one side of the market play a dominated strategy. However, if the difference in payoffs between a dominated strategy and other strategies decreases, then a cardinal process may allow the rate at which

players exit the dominated strategy to approach zero so quickly that the process converges to an outcome in which some agents still play this strategy, yielding a Nash but not perfect equilibrium. An ordinal process precludes such an outcome by ensuring that the rate at which players exit a dominated strategy does not approach zero and hence the proportion of players adopting the strategy must approach zero.

The results on convergence to Nash and perfect equilibria can be supplemented by results for some special cases. An equilibrium will be said to be locally stable if there exists some neighborhood of initial conditions around the equilibrium with the property that the process will converge to the equilibrium from any initial condition in this neighborhood. Arguments analogous to those invoked in the proofs of Propositions 2-3 yield:

Proposition 4. Let the evolutionary process be regular, monotonic, and ordinal. If an equilibrium in unique dominant strategies exists, then the process will converge to this equilibrium from any initial position. If a strict Nash equilibrium exists, then this equilibrium will be locally stable.

Example 2. Neither the uniqueness condition in the first statement nor the local restriction in the second can be deleted. Consider the following:

		2	
		$s_2$	$s'_2$
1	$s_1$	1,1	0,1
	$s'_1$	1,0	0,0

		2	
		$s_2$	$s'_2$
1	$s_1$	1,1	0,0
	$s'_1$	0,0	1,1

Any outcome in the first game is a dominant strategy equilibrium and the evolutionary process obviously cannot converge to all of them. The second game has two strict Nash equilibria. Both are locally stable and hence neither is globally stable.

Each of the results given above depends upon monotonicity. This assumption, ensuring that agents never switch to a less profitable strategy, initially appears intuitively compelling. However, one readily conceives of adjustment processes based on limited information which fail this property. Friedman and Rosenthal (1986) deliberately eschew this property. The Nash equilibrium prescriptions for both a finitely-repeated prisoners' dilemma (defect in every period) and the following game  $((s_1, s_2))$ ,

		2	
		$s_2$	$s'_2$
1	$s_1$	10,000, 1	0,0
	$s'_1$	9,900, 1	9,900, 0

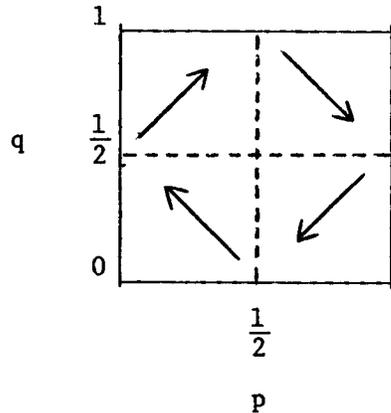
are often considered unconvincing, suggesting that one cannot be completely sanguine concerning monotonic adjustment processes.

The ordinal adjustment process yields a perfect equilibrium if it converges. This allows completely unreasoning agents to achieve outcomes which call for more stringent restrictions on knowledge and rationality than the already restrictive conditions developed for the case of a Nash equilibrium. This raises an obvious question. Does the evolutionary process converge? In order to concisely isolate the salient issues, we now restrict ourselves to cases in which  $n_1 = n_2 = 2$ , so that each side has only two pure strategies.

Example 3. Consider the "matching coins" game:

		2		
		$s_2$	$s_2'$	
1	$s_1$	1, -1	-1, 1	.
	$s_1'$	-1, 1	1, -1	

Letting  $p$  ( $q$ ) be the probability of  $s_1$  ( $s_2$ ), the phase diagram of the evolutionary process is given by

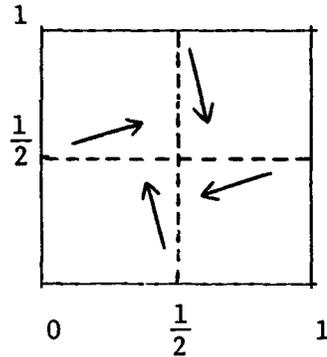


It is then apparent that the evolutionary process may cycle rather than converge.

Convergence is thus not guaranteed. Is it possible? In the above example, we need only specify the transition probabilities given by (2)-(3) so that

$$\frac{1}{p_t} \frac{dp_t}{dt} = \left\{ \begin{array}{lll} -2 & (p_t > \frac{1}{2}, q_t < \frac{1}{2}) & -1 \\ -1 & (p_t < \frac{1}{2}, q_t < \frac{1}{2}) & 2 \\ 2 & (p_t < \frac{1}{2}, q_t > \frac{1}{2}) & 1 \\ 1 & (p_t > \frac{1}{2}, q_t > \frac{1}{2}) & -2 \end{array} \right\} = \frac{1}{q_t} \frac{dq_t}{dt}. \quad (4)$$

Hence, the proportions of agents playing a particular strategy approach 0 or 1 more slowly than they fall away from it. Such a specification is clearly possible. This gives a phase diagram of



which yields convergence.

Notice that convergence does not require the precise numerical values stipulated in (4). Other values that preserve the inequalities concerning  $(1/q_t)(dq_t/dt)$  and  $(1/p_t)(dp_t/dt)$  in each region of the phase diagram will also yield convergence. We can then think of an adjustment process of the type specified in (4)-(5) as an element of  $R^8$  and can say that a process which yields (does not yield) convergence is robust if it is contained in a neighborhood in  $R^8$  which also yields (does not yield) convergence. This example thus yields robust convergence. Robust, nonconverging processes also exist for this game.

Given  $n_1 = n_2 = 2$ , one easily identifies all of the possible phase diagram configurations that various games can produce and observes that in each case a specification of the transition probabilities analogous to (4)-(5) yields a robust converging process. Hence,

**Proposition 5.** Let  $n_1 = n_2 = 2$ . Then for any fixed game, there exists a robust evolutionary process which converges for every initial condition.

Results can be obtained for games with larger strategy sets, though the analysis is considerably more tedious.

It is not immediately obvious how one interprets this result. On the one hand, we can observe that while convergence is not always guaranteed, the existence of robust convergence processes ensures that it is not a knife-edge condition. The evolutionary model thus apparently provides a robust foundation for both rationalizable and perfect equilibria. On the other hand, it is somewhat disconcerting that convergence is obtained only by tailoring the evolutionary process to the game. While a game may have institutional features associated with it that also influence the evolutionary process, a general converging process would be more reassuring. In particular, convergence is obtained by selectively choosing the adjustment rates in various regions of the phase diagram, where the boundaries of these regions correspond to Nash equilibrium strategies. This strikes one as dangerously close to sneaking some knowledge-of-Nash-equilibrium assumption through the back door.

Attention has recently turned to refinements of the perfect equilibrium concept. Two normal-form refinements are the concepts of an iterated dominance equilibrium (Moulin (1979)) and proper equilibrium (Myerson (1978)). We cannot ensure the appearance of such an equilibrium without placing additional restrictions on either the evolutionary process or on initial conditions. The latter restrictions may be interpreted as arising from the type of knowledge of one's opponent offered by Tan and Werlang (1987) as a motivation for Nash equilibria.

#### CONCLUSION

Our results suggest that if one embraces the evolutionary approach to equilibrium concepts, then one will embrace either rationalizable or perfect equilibria. The choice between the two hinges upon whether the evolutionary process is sufficiently well behaved as to yield convergence. In general, there are robust adjustment processes which converge and robust processes which do not converge.

Extensions in two directions are required before these results can be properly assessed. First, the purview must be expanded beyond finite, two-player normal-form games of perfect information. Each of these restrictions plays a role in driving the results of this paper. Second, alternative specifications of the evolutionary process must be considered, with particular emphasis on adjustment processes with more sophisticated agents and possibly optimal learning processes. These extensions may yield differences in specific results but will still allow evolutionary arguments to play a role in evaluating equilibrium concepts.

#### References

- B. D. Bernheim. Rationalizable Strategic Behavior. Econometrica 52 (1984), 1007-1028.

- B. D. Bernheim. Axiomatic Characterizations of Rational Choice in Strategic Environments. Scandinavian Journal of Economics 88 (1986), 473-488.
- G. W. Brown. Iterative Solutions of Games by Fictitious Play. In Activity Analysis of Production and Allocation, ed. T. C. Koopmans. John Wiley (1951).
- M. Feldman. An Example of Convergence to Rational Expectations with Heterogeneous Beliefs. International Economic Review 28 (1987), 635-650.
- J. W. Friedman and R. W. Rosenthal. A Positive Approach to Non-Cooperative Games. Journal of Economic Behavior and Organization 7 (1986), 235-251.
- J. S. Jordan. Convergence to Rational Expectations in a Stationary Linear Game. Mimeo (1986).
- J.-F. Mertens. Ordinality in Non-Cooperative Games. CORE Discussion Paper No. 8728 (1987).
- P. Milgrom and J. Roberts. Predation, Reputation, and Entry Deterrence. Journal of Economic Theory 27 (1982), 280-312.
- H. Moulin. Dominance Soluable Voting Schemes. Econometrica 47 (1979), 1337-1352.
- R. B. Myerson. Refinements of the Nash Equilibrium Concept. International Journal of Game Theory 7 (1978), 73-80.
- J. Nash. Non-cooperative Games. Annals of Mathematics 54 (1951), 286-295.
- D. G. Pearce. Rationalizable Strategic Behavior and the Problem of Perfection. Econometrica 52 (1984), 1029-1050.
- R. Selten. Re-examination of the Perfectness Concept for Equilibria in Extensive Games. International Journal of Game Theory 4 (1975), 25-55.
- J. Maynard Smith. Evolution and the Theory of Games. Cambridge University Press (1982).
- T. C.-C. Tan and S. R. d. C. Werlang. The Bayesian Foundations of Solution Concepts of Games. Journal of Economic Theory, forthcoming (1987).
- E. van Damme. Refinements of the Nash Equilibrium Concept. Springer-Verlag (1983).