Individual learning and social rationality

H. Peyton Young*

Department of Economics, Johns Hopkins University, Baltimore, MD 21218, USA

Abstract

Classical game theory assumes that players have a coherent model of the game in which they are engaged, that they make optimum plans assuming that everyone else makes optimum plans, and that all of this is common knowledge. Evolutionary game theory postulates instead that players have limited understanding of their environment, only modest reasoning ability, and no common knowledge. Surprisingly, a fair amount of classical solution theory survives with this change of perspective. We show how high-rationality solutions can emerge in low-rationality environments provided the evolutionary process has sufficient time to unfold. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: C72; C73

Keywords: Learning; Nash equilibrium; Stochastic dynamical systems; Spontaneous order

1. Introduction

Traditional game theory rests on three propositions about individual behavior. First, players are supposed to have a complete model of the interactive situation in which they are engaged. Second, they are assumed to make

* E-mail: pyoung@jhu.edu.

1 This work was supported by the National Science Foundation. For an elaboration of these ideas see my forthcoming book, Individual Strategy and Social Structure, Princeton University Press, Princeton, 1998.
optimum plans of arbitrarily high complexity, taking into account the fact that everyone else is making optimum plans of arbitrarily high complexity. Third, it is assumed that all of this is common knowledge. While this conception of human behavior leads to a mathematically elegant world, it bears only a faint resemblance to the world we know. In recent years an alternative view has emerged, prompted on the one hand by models of animal behavior and on the other by experiments with human subjects. This general approach is usually referred to as 'evolutionary game theory'. Its hallmarks are the dynamics of adaptation in games played by large populations of agents who are boundedly rational. The general program is to look for behavioral patterns that emerge at the social or population level through the process of selection and reinforcement. In this paper I review some of the recent developments in the field and show that a certain amount of high-rationality solution theory reappears along this low-rationality route.

2. Two examples

I begin with two examples that illustrate the kinds of 'social learning' situations we have in mind. Consider the following game of etiquette: two people are approaching a doorway, and they need to coordinate on who goes first. Suppose that they differ in some recognizable way – say one is a man and the other is a woman – and they condition their behavior on who they are. If each defers to the other they stand outside the door wasting time; if both barge ahead they risk a collision. This interaction can be represented as follows:

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
</tr>
<tr>
<td>Yield</td>
<td>0, 0</td>
</tr>
<tr>
<td>Not yield</td>
<td>2, 1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Formally, this is a Battle of the Sexes game, but it differs in important and subtle respects. Battle of the Sexes is usually described as a one-shot decision problem by a couple trying to coordinate on where to go for an evening's entertainment. The etiquette game, by contrast, is played many times by many different people, and its solution is not an individual but a social construct.

As a second example, consider the following technology adoption game. Two types of typewriter keyboards are available on the market – QWERTY (Q) and DVORAK (D). Every secretary must decide which keyboard to learn on, and

\footnotetext[2]{Aumann and Brandenburger (1995) examine various ways of relaxing the common knowledge assumption.}
every employer must decide which type of keyboard to provide in the office. Assume for simplicity that each employer invests in just one kind of typewriter, and that each secretary can be proficient on only one kind of keyboard. Let us also suppose, for the sake of argument, that D is a bit more efficient than Q, so that if employers and secretaries coordinate on D, the payoffs are slightly higher than if they coordinate on Q.³

<table>
<thead>
<tr>
<th>The technology adoption game</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Secretaries</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>Employers</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>Q</td>
</tr>
</tbody>
</table>

On the face of it, the etiquette game and the technology adoption game are very similar: each is a coordination game with two pure-strategy equilibria and one mixed-strategy equilibrium. But there are important differences. First, the technology game has a unique Pareto efficient equilibrium, whereas the etiquette game has two Pareto efficient equilibria. Even more important are differences in the learning environments that are not captured in the payoff matrix. In the etiquette game, the players actually meet at the time of making a decision, so they can communicate their intentions and try to coordinate on the spot. In the technology game, the agents need to coordinate their decisions vis-à-vis the other population, not against any specific individual. Another important difference is the frequency with which the game is played. Most people have thousands or even millions of doorway encounters in their lifetimes, so there is a lot of opportunity to learn from personal experience. For a secretary, however, learning to type might be a once-in-a-lifetime decision; similarly, employers only occasionally invest in new equipment. In this case most of the learning must come from observing (or hearing about) the actions taken by others.

3. Models of learning

This brings us to the question of how people actually do learn and adapt in situations like those described above. Here I shall briefly review some of the learning mechanisms that have been proposed in the literature. They fall into four broad categories: natural selection, imitation, reinforcement, and best response.

³ We shall not enter the thicket of arguments as to whether DVORAK actually is more efficient than QWERTY, though there is no question that it was intended to be more efficient based on motion studies of typists' hand movements. For a historical discussion of typewriter keyboards see David (1985).
3.1. Natural selection

People who use high-payoff strategies are at a reproductive advantage compared to people who use low-payoff strategies, hence the latter decrease in frequency in the population over time. The standard model of this situation is the replicator dynamic, in which the rate of growth of a strategy $x_i$ is a linear function of the difference between the payoff from playing $x_i$ and the average payoff. Note that, in this context, payoffs refer to reproductive success rates, not to individuals' preferences over outcomes.

3.2. Imitation

People copy the behavior of others, especially behavior that is popular or appears to yield high payoffs. Imitation may be driven purely by popularity, e.g., copy the first person you see. Or it may be prompted by a combination of popularity and payoff, e.g., copy the first person you see with a probability that depends negatively on your own payoff, and positively on their payoff. Unlike natural selection, the payoffs in such a model describe how people make choices rather than how rapidly they multiply, so the model is more consistent with learning in social and economic environments.

3.3. Reinforcement

People tend to adopt actions that yielded a high payoff in the past, and to avoid actions that yielded a low payoff. This is the standard learning model in behavioral psychology, and it has increasingly captured the attention of economists. As in imitative models, payoffs describe choice behavior, but it is one's own past payoffs that matter, not the payoffs of others. The basic premise is that the probability of taking a particular action now increases monotonically with the payoff that resulted from taking that action in the past. Models of this genre have been statistically estimated from the behavior of subjects playing games in laboratory settings, but the evidence is still too limited to draw general conclusions about their empirical validity. A drawback of this approach is that people must play the game often enough for reinforcement to be a plausible learning mechanism. For example, it is doubtful that secretaries (or employers) make choices based on their past experience of choosing QWERTY or DVORAK, since they face such a decision only once or twice in a lifetime.

---


3.4. Best response

People adopt actions that optimize their expected payoff given what they expect others to do. In short, they have beliefs or models concerning others' probable behavior, and they optimize given their beliefs. This approach encompasses a variety of learning rules that attribute more or less sophistication to people's attempts to forecast others' behavior. In the simplest models, subjects choose best replies to the observed frequency distribution of actions taken by their opponents in the past (as in fictitious play). More sophisticated variants include weighting the opponents' previous actions by a discount factor, trying to anticipate the opponents' current thinking, and so forth.\(^6\)

Once we have a learning rule that describes the behavior of individuals, we need to embed it in a social environment where large numbers of individuals interact with one another. In particular, we need to specify the ways in which individuals meet and how they acquire information, which determines through the learning rule how they behave. The object is to study the evolution of this process at the population level, that is, to discover what patterns of behavior emerge over time, and whether some patterns are more likely to emerge than others from arbitrary initial conditions. In other words, we need to study the process as a stochastic dynamical system. In the remainder of the paper I shall show how this program can be carried out for a particular family of best-response rules, and how the results connect with solution concepts in classical theory. It will become evident that similar reasoning could be applied to other learning rules; indeed this is a promising area for research.

The adaptive learning model that I propose incorporates four elements of bounded rationality that seem plausible in many learning environments. I shall assume that players form simple predictive models of others' future behavior based on fragmentary information about their past behavior, that they react myopically given their information, and their choices are subject to idiosyncratic shocks.

When we embed this type of learning rule into a population setting, we obtain the following stochastic process, which is a variant of fictitious play. Let \( G \) be an \( n \)-person game with finite action space \( X = \prod X_i \) and von Neumann Morgenstern utility functions \( u_i : X \to R \). The players are divided into \( n \) disjoint populations, \( C_1, C_2, \ldots, C_n \), where the members of \( C_i \) are candidates to play role \( i \) in the game. The process operates as follows. In each time period \( t = 1, 2, 3, \ldots \), the game is played once by a random draw of \( n \) agents from the populations. The probability of a given draw is determined by some probability distribution \( \pi \) on \( C_1 \times C_2 \times \cdots \times C_n \) that reflects the underlying geographic or social distance.

between individuals, and hence the probability that they meet. For simplicity of exposition we shall assume, in what follows, that \( \pi \) is fixed and has full support.\(^7\)

In each period \( t \), the player selected to play role \( i \) chooses an action \( x_i^t \in X_i \). Denote the set of actions in period \( t \) by \( X^t = (x_1^t, x_2^t, \ldots, x_n^t) \in X \), and the complete history of play up through time \( t \) by \( h^t = (x^1, x^2, \ldots, x^t) \). The process has three parameters: \( m \) for memory, \( s \) for sample size, and \( \varepsilon \) for error rate. At the end of period \( t \), the state consists of the history to date truncated to the last \( m \) periods. Denote this truncated history by \( h^t \). At the beginning of period \( t + 1 \), the player selected for role \( i \) draws a sample of size \( s \) from the set of actions taken by the \( j \)-players over the last \( m \) periods, the samples being without replacement and independent over all \( j \neq i \). Let the random variable \( \hat{P}_{ij} \) denote the resulting sample frequency distribution, and let \( \hat{p}_{-i}^t = \prod_{j \neq i} \hat{P}_{ij} \). With probability \( 1 - \varepsilon \) the \( i \)-player chooses a best reply to \( \hat{p}_{-i}^t \), and with probability \( \varepsilon \) he chooses a strategy in \( X_i \) at random. If there are ties in best reply we assume each is chosen with equal probability. The probability of making errors is assumed to be independent across players. These operations define a finite Markov chain \( P^{m,s,\varepsilon} \) on the state space \( H = X^m \). We shall refer to the process \( P^{m,s,\varepsilon} \) as adaptive play with memory \( m \), sample size \( s \), and error rate \( \varepsilon \) (Young, 1993a).

A stationary distribution of \( P^{m,s,\varepsilon} \) is a probability distribution \( \mu \) defined on the space \( H \) of truncated histories that satisfies the stationarity equations

\[
\forall h \in H, \sum_{h' \in H} \mu(h') P_{h'h} = \mu(h).
\]

\( P^{m,s,\varepsilon} \) is irreducible if it transits from any state to any other in a finite number of periods. This is clearly so whenever the error rate \( \varepsilon \) is positive. In this case, the process possesses a unique stationary distribution, which we shall denote by \( \mu^{m,s,\varepsilon} \). The interpretation of \( \mu^{m,s,\varepsilon} \) is the following: for each truncated history \( h \), \( \mu^{m,s,\varepsilon}(h) \) is the relative frequency with which state \( h \) occurs in almost all realizations of the process independently of the initial state. If \( \mu^{m,s,\varepsilon}(h) \) puts high probability on particular histories (which correspond to social patterns of behavior), we could say that the adaptive process tends to favor these patterns over the long run.

Unfortunately, it is extremely cumbersome to compute \( \mu^{m,s,\varepsilon} \) when the state space is large, which is generally the case in practice. It turns out, however, that when the disturbance term \( \varepsilon \) is small, it is not difficult to estimate the mode of \( \mu^{m,s,\varepsilon} \), that is, the state(s) where \( \mu^{m,s,\varepsilon} \) achieves its maximum. In particular, we say that a state \( h \) is stochastically stable if \( \lim_{\varepsilon \to 0} \mu^{m,s,\varepsilon}(h) > 0 \) (Foster and Young, 1998b).

\(^7\)Alternative assumptions are that \( \pi \) evolves over time, that is, the matching probabilities are endogeneous (Mailath et al., 1994; Tesfatsion, 1995) or that some interactions never occur, e.g., people might interact only with their immediate neighbors (Blume, 1993, 1995; Ellison, 1993; Young, 1998b).
1990). Such states exist, and they can be readily computed by finding the minimum of a certain stochastic potential function (Freidlin and Wentzell, 1984; Young, 1993a).

4. Equilibrium selection in $2 \times 2$ coordination games

In some types of games, the adaptive process described above selects patterns of behavior that are closely connected to classical equilibrium concepts. In other games, the process selects behaviors that do not correspond to equilibria, but that are related to somewhat looser solution concepts, including the elimination of iteratively strictly dominated strategies and minimal curb sets. We briefly survey some of these connections here.

Consider first the class of $2 \times 2$ coordination games:

<table>
<thead>
<tr>
<th>Action 1</th>
<th>Action 1</th>
<th>Action 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$, $b_{11}$</td>
<td>$a_{12}$, $b_{12}$</td>
<td></td>
</tr>
<tr>
<td>$a_{21}$, $b_{21}$</td>
<td>$a_{22}$, $b_{22}$</td>
<td></td>
</tr>
</tbody>
</table>

$a_{11} > a_{21}$, $b_{11} > b_{12}$, $a_{22} > a_{12}$, $b_{22} > b_{21}$.

The etiquette and the technology games are particular examples. Every such game has three equilibria – two pure and one mixed. In the etiquette game, both pure equilibria are efficient and the mixed equilibrium is inferior to both, whereas in the technology game only one of the pure equilibria is efficient and the mixed equilibrium is inferior to both. The coordination equilibrium $(1, 1)$ is strictly risk-dominant if

$$(a_{11} - a_{21})(b_{11} - b_{12}) > (a_{22} - a_{12})(b_{22} - b_{21}).$$

Similarly, equilibrium $(2, 2)$ is strictly risk dominant if the reverse inequality holds (Harsanyi and Selten, 1988). The following result shows that only the risk dominant equilibrium is robust against selection pressure when the perturbations are small.

**Theorem 1.** Let $G$ be a $2 \times 2$ coordination game with a strictly risk dominant equilibrium and let $P_{m,a,d}$ be adaptive play. If information is sufficiently incomplete ($s/m \leq \frac{1}{2}$), and if $s$ and $m$ are sufficiently large, the unique stochastically stable state is the history in which the risk-dominant equilibrium is repeated $m$ times in succession.

Kandori et al. (1993) proved the following variant of this result. Consider a symmetric $2 \times 2$ game played by a single population of agents. Assume that, with probability $1 - \epsilon$, each player chooses a best reply to the frequency
distribution of actions taken by all other players in the preceding period, and with probability $\varepsilon$ he chooses an action at random. This is similar to adaptive play with a single population of players, memory equal to one period, and complete sampling ($s = m$). Kandori et al., show that if the game has a strictly risk-dominant equilibrium, then for all sufficiently large population sizes, the stochastically stable state is the one in which everyone plays the risk-dominant equilibrium.

5. Selection with nonvanishing error rates

Before proceeding, it is appropriate to pause and ask whether these results are meaningful, given that stochastic stability is defined in terms of perturbations that are \textit{vanishingly small}. Suppose, for example, that equilibrium (1, 1) is risk dominant but that the process happens to start in the state $h_2$ where everyone has played action 2 for $m$ periods. To tip the process into the state $h_1$ where everyone is playing action 1 requires at least $rs$ errors, where $s$ is the sample size and

\[
\begin{align*}
    r &= (a_{22} - a_{12})/(a_{22} - a_{12} + a_{11} - a_{21}) \\
    &\wedge (b_{22} - b_{21})/(b_{22} - b_{21} + b_{11} - b_{12}).
\end{align*}
\]

(In general, $x \wedge y$ denotes the minimum of $x$ and $y$.) Since the errors are independent, the probability of this event is on the order of $\varepsilon^r$, which could be very small indeed if $\varepsilon$ is small and $s$ is large. Of course, without knowing the details of the situation, we have no way of knowing how large the error rate $\varepsilon$ actually is. However, this does not matter much if the \textit{shape} of the stationary distribution $\mu^{m,s,\varepsilon}$ is close to its limiting value for nonnegligible values of $\varepsilon$. In fact this is often the case. To illustrate, consider the technology game with parameters $m = 4$ and $s = 2$. The probabilities of the various states are shown in Table 1 as a function of $\varepsilon$. Note that even with an error rate $\varepsilon = 0.20$, the probability is over 75% that \textit{almost everyone} is playing D (the risk-dominant equilibrium) at any given time.

This example illustrate that the size of the perturbations does not have to be especially small for the evolutionary bias toward the stochastically stable state to manifest itself.\textsuperscript{8} It should also be noted that the \textit{waiting time} until the process reaches the stochastically stable regime depends crucially on the interaction structure as well as the size of the population and the error rate. The more that people interact in small close-knit groups, the more rapidly the process adjusts (Ellison, 1993; Young, 1998b, Chapter 6).

\textsuperscript{8} For a further discussion of this issue see Binmore and Samuelson (1997) and Young (1998b, Chapter 5).
6. Bargaining games

In the remainder of the paper I shall examine the implications of this approach for several other classes of games. One class of particular interest consists of distributive bargaining games (Nash, 1950, 1953). Consider two individuals who are bargaining over their shares of a fixed pie. A standard model of this situation is the Nash demand game: the row player demands a share \( x \in (0, 1] \); simultaneously, the column player demands a share \( y \in (0, 1] \). They get what they ask for if the demands sum to one or less; otherwise, they get nothing. Let \( u(x) \) be the utility of the row players as a function of the row player's share \( x \), and let \( v(y) \) be the utility of the column players as a function of their share \( y \). Without loss of generality, we may assume that \( u(0) = v(0) = 0 \). For mathematical convenience we shall discretize the state space by only permitting demands \((x, y)\) that are expressible in \( r \) decimal places for some positive integer \( r \). We call \( 10^{-r} \) the precision of the demands.

Theorem 2 (Young, 1993b). Let \( G \) be the discrete Nash demand game with precision \( 10^{-r} \) and let \( P^{m,s,r} \) be adaptive play. When \( s \leq m/2 \), the stochastically stable states of adaptive play are those in which a fixed division \((x, 1 - x)\) is played for \( m \) periods. If, in addition, \( m, s, \) and \( r \) are sufficiently large, every stochastically stable division is arbitrarily close to the Nash bargaining solution \((x^N, 1 - x^N)\), which maximizes \( u(x)v(1 - x) \).

Several variations of this model can be considered. For example, suppose that the players only get what they demand when the demands sum to one exactly; otherwise they get nothing. In other words, they must coordinate on a complete division of the pie with nothing left over. One then obtains a variant of Theorem 2 in which the stochastically stable division is arbitrarily close to the Kalai–Smorodinsky solution, that is, the unique Pareto efficient division \((x^{KS}, 1 - x^{KS})\) such that \( u(x^{KS})/u(1) = v(1 - x^{KS})/v(1) \) (Young, 1998a).

Table 1

<table>
<thead>
<tr>
<th>Proportion of D-players in a population of size 8</th>
<th>( \varepsilon = 0.20 )</th>
<th>( \varepsilon = 0.10 )</th>
<th>( \varepsilon = 0.05 )</th>
<th>( \varepsilon = 0.01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.419</td>
<td>0.659</td>
<td>0.815</td>
<td>0.961</td>
</tr>
<tr>
<td>7/8</td>
<td>0.373</td>
<td>0.277</td>
<td>0.167</td>
<td>0.039</td>
</tr>
<tr>
<td>6/8</td>
<td>0.150</td>
<td>0.053</td>
<td>0.016</td>
<td>0.001</td>
</tr>
<tr>
<td>5/8</td>
<td>0.043</td>
<td>0.009</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4/8</td>
<td>0.011</td>
<td>0.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3/8</td>
<td>0.002</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2/8</td>
<td>0.001</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1/8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
7. Evolutionary selection in finite, normal-form games

What can be said about evolutionary selection in more general games? We cannot say that adaptive learning always selects a Nash equilibrium. Indeed, Shapley's well-known example showing the cyclic behavior of fictitious play also exhibits cycling under adaptive play (Shapley, 1964). Nevertheless, a fair amount can be said about the strategies that are not stochastically stable. Consider an $n$-person game with finite strategy space $X = \prod X_i$. A strategy is strictly dominated for a given player if he has another strategy that yields a strictly higher payoff under any choice of strategies by the other players. Obviously, a rational player will never use a strictly dominated strategy. Moreover, if the players' utility functions are common knowledge, and the fact that they are rational is also common knowledge, then each player can foresee that the other players will never use their strictly dominated strategies. Hence, everyone can eliminate them from their calculations. Once this has been done, there may exist strategies that are strictly dominated with respect to the remaining (smaller) strategy space. By the same reasoning they can be eliminated, and so forth. Ultimately, one arrives at a unique minimal subspace of strategies $X^*$ such that no strategy is strictly dominated for any player.

In the evolutionary framework we can reach a similar conclusion without appealing to rationality or common knowledge. We claim that the stochastically stable states consist only of strategies in $X^*$. The reason is straightforward. Consider first the process with $\varepsilon = 0$, and let $h$ be any state. A strictly dominated strategy (relative to the space $X$) will never be chosen as the best reply to any sample, so within $m$ periods all such strategies will die out. Then any strategy that is strictly dominated with respect to the resulting smaller subspace will die out within the next $m$ periods, and so forth. It follows that the recurrent states in the unperturbed process $P^{m,.0}$ consist only of strategies in $X^*$. Standard arguments show, however, that the recurrent states of the unperturbed process are the only candidates for being stochastically stable in the perturbed process (Young, 1998b, Chapter 3).

For generic $n$-person games, we can actually say a good deal more. Define a product set of strategies to be a set of form $Y = \prod Y_i$, where each $Y_i$ is a nonempty subset of $X_i$. Let $\Delta Y_i$ denote the set of probability distributions over $Y_i$, and let $\prod (\Delta Y_i)$ be the product set of such distributions. Let $B_i(Y_{-i})$ denote the set of $i$'s strategies that are a best reply to some product distribution $P_{-i} \in \prod_{j \neq i} (\Delta Y_j)$, and define the function

$$B(Y) = B_1(Y_{-1}) \times \cdots \times B_n(Y_{-n}).$$

Observe that $B(\cdot)$ maps product sets to product sets.

The product set $Y = \prod Y_j$ is a curb set (closed under rational behavior) if $B(Y) \subseteq Y$ (Basu and Weibull, 1991). It is a minimal curb set if it is curb and
properly contains no curb set. In this case \( Y \) is a fixed point: \( B(Y) = Y \). This
case is a natural generalization of a strict Nash equilibrium, which is a singleton
minimal curb set. Although every minimal curb set is a fixed point of
\( B(\cdot) \), there may exist fixed points of \( B(\cdot) \) that are not minimal. Note also that
the set of rationalizable strategies corresponds to the unique maximal fixed point of
\( B(\cdot) \) (Bernheim, 1984; Pearce, 1984).

To illustrate this concept, consider the following variant of Shapley’s example,
where we have added one more strategy \((D)\) that is a best reply against itself but
not against any combination of the other three strategies.

\[
\begin{array}{cccc}
A & B & C & D \\
\hline
A & 3, 3 & 4, 1 & 1, 4 & -1, -1 \\
B & 1, 4 & 3, 3 & 4, 1 & -1, -1 \\
C & 4, 1 & 1, 4 & 3, 3 & -1, -1 \\
D & -1, -1 & -1, -1 & -1, -1 & 0, 0 \\
\end{array}
\]

The minimal curb sets in this example are \( Y = \{(D, D)\} \) and
\( Z = \{A, B, C\} \times \{A, B, C\} \). When the process starts in an arbitrary initial state
and the error rate is zero, with probability one the process ultimately enters
either regime \( Y \) in which only \( D \) is played, or it enters regime \( Z \) in which \( A, B, \)
and \( C \) are played but \( D \) is never played. Moreover, it can be shown that the
\( Z \)-regime is stochastically stable but the \( Y \)-regime is not.

In general, given a set of histories \( H' \), define the span of \( H' \) to be the product
set of all strategies that appear in at least one of the histories in \( H' \). We say that
\( H' \) is a minimal curb configuration if its span coincides with a minimal curb set.
The following is a variant of a result of Hurkens (1995).\(^9\)

\textit{Theorem 4} (Young, 1998b). Let \( G \) be a generic \( n \)-person finite game and let
\( P^{m,s,0} \) be adaptive play. If \( s/m \) is sufficiently small, the unperturbed process
\( P^{m,s,0} \) converges with probability one to a minimal curb configuration. If \( s \) is
positive and sufficiently small, the perturbed process puts arbitrarily high probabil-
ity on the minimal curb configuration(s) that minimize stochastic potential.

The method for computing stochastic potential is discussed in detail in Young
(1993a). We remark that when the sample size is sufficiently large, there will
typically be a unique minimal curb configuration that minimizes stochastic
potential, in which case the selection criterion is sharp. Note, however, that what
is selected will not necessarily be an equilibrium; rather it may be a repetitive
disequilibrium pattern of play.

---

\(^9\) Sanchirico (1996) considers other learning models that lead to the selection of minimal curb sets.
8. Conclusions

At the beginning of this paper I pointed out that classical solution concepts often rely on high levels of rationality and knowledge on the part of individual players. The adaptive learning model described above assumes rather modest levels of rationality and no common (or even mutual) knowledge on the part of the players. Nevertheless, quite a few concepts from classical solution theory can be recovered via this route. These include the risk-dominant equilibrium in $2 \times 2$ games, the Nash bargaining solution and Kalai–Smorodinsky solution in various types of bargaining games, and the iterated elimination of strictly dominated strategies in finite strategic form games. Similarly, one obtains the subgame perfect equilibrium in certain types of two-person extensive form games (Canning, 1992; Nöldeke and Samuelson, 1993), and efficient equilibria in pure coordination games (Young, 1998a). These examples show that high-rationality solution concepts can emerge in low-rationality environments if we allow the process sufficient time to evolve. In other words, social feedback mechanisms can substitute for high levels of knowledge and deductive powers on the part of individuals. Perhaps similar results can be derived using other types of learning rules, and perhaps they will suggest entirely new solution concepts that never occurred to classical theorists.

References

Van Huyck, J.B., Battalio, R.C., Rankin, F.W., 1995. Evidence on learning in coordination games. Research Report No. 7. TAMU Economics Laboratory, Texas A&N University, College Station, TX.