Notes, Comments, and Letters to the Editor

Clever agents in adaptive learning

Alexander Matros

ELSE, Department of Economics, University College London, Gower Street, London WC1E 6BT, UK

Received 13 July 2000; final version received 22 March 2002

Abstract

Saez-Marti and Weibull (J. Econom. Theory 86 (1999) 268) investigate the consequences of letting some agents play a myopic best reply to the myopic best reply in Young’s (J. Econom. Theory 59 (1993) 145) bargaining model, which is how they introduce “cleverness” of players. I analyze such clever agents in general finite two-player games and show Young’s (Individual Strategy and Social Structure, Princeton University Press, Princeton, NJ, 1998) prediction to be robust: adaptive learning with clever agents does select the same minimal curb set as in the absence of clever agents, if their population share is less than one. However, the long-run strategy distribution in such a curb set may vary with the share of clever agents.

© 2003 Elsevier Science (USA). All rights reserved.

JEL classification: C72; C73

Keywords: Evolution; Game theory; Bounded rationality; Markov chain; Stochastic stability

1. Introduction

While bounded rationality and learning models have been studied extensively in the last few years, game theory has been unsuccessful in explaining where the bounds on rationality should be placed in a model of boundedly rational agents.

Recently, Young [5–7] suggested an evolutionary model which explains how agents can make their choices on basis of their own preferences and a sample of occurrences in the recent past only. A two-player game is played repeatedly by the members of two large populations and agents in the same population have the same preferences. In every round, two agents, one from each population, are randomly
selected to play the game. Each of the two agents simultaneously chooses a strategy in the game, and each agent has access to a random sample, drawn from the recent history of play. They use their sample as a predictor of the behavior of the agent they face, and almost always play a best reply to the empirical strategy distribution of the opponent population in the sample. Occasionally, agents “mutate”, however, and instead choose a strategy that is not a best reply to any possible sample from the recent history of play.

Using Young’s [6] model, Saez-Marti and Weibull [3] consider agents that are “clever” in a certain sense. They study the effect of letting a share of one of the populations know the preferences of the opponent population, denoting these agents as “clever”, in the Nash Demand Game. Saez-Marti and Weibull [3] first assume, that the population without clever agents plays its best reply to a sample of past strategies played by the other population and, second, that clever agents play a best reply to the opponent population’s best reply to the clever agent’s sample. In other words, clever agents try to anticipate their opponent’s choice on the basis of the sample of strategies played by their own population. They show that Young’s prediction is robust to the introduction of any share of clever agents less than one.

The purpose of the present paper is to demonstrate that this robustness holds for generic finite two-player games. Young [7] proves that the adaptive learning process in a generic class of finite games settles down in a minimal curb configuration which minimizes the stochastic potential in such games when the mutation rate goes to zero. My main result is that the adaptive learning with any share of clever agents less than one converges to the same minimal curb configuration as Young’s adaptive learning process, when the mutation rate goes to zero. However, I show that the presence of clever agents may influence the long-run strategy distribution inside the minimal curb configuration.

Hurkens [2] develops a model in which agents receive samples with replacement from the recent past. His main result—the learning process leads the agents to playing strategies from a minimal curb set—is robust to the introduction of “sophisticated” agents. Hurkens’ sophisticated agents are analogous to clever agents. However, the replacement assumption, that is, an agent can keep resampling a single previous strategy by another agent without realizing that she looked at it before, implies that the learning process does not discriminate between minimal curb sets even when the sample size is large. While Young’s model on which my paper builds identifies precisely which of the minimal curb sets is selected in the long run.

I also analyze the question, which was asked in [3], how well clever agents fare among non-clever agents given that there are fixed population shares of clever and non-clever agents. More specifically, the Matching-Pennies Game is considered and found that the gain of clever agents depends on their share in the population. Moreover, clever agents do not only outsmart the agents in the other population but, indirectly, also the non-clever agents in their own population. As a result, non-clever agents in both populations earn expected negative payoffs while the “clever subpopulation” on average earns expected positive payoffs in this zero-sum game. The larger the share of clever agents, the larger the gain for the population. On the
margin, an additional clever agent gains less as the share of clever agents increases, here called “decreasing returns to cleverness”.

Saez-Marti and Weibull [3] show that the “clever” population gets the whole pie when playing with the “non-clever” population in the Nash demand game. However, I demonstrate that “cleverness” of the whole population does not always guarantee an advantage. For example, in the strict demand game, where two players must coordinate to get exactly the size of the pie, otherwise they get nothing, the population without clever agents obtains the whole pie.

The paper is organized as follows. In Section 2, I describe the unperturbed and perturbed versions of the adaptive play with clever agents. In Section 3, I derive general results for the stationary distribution of this process. In Section 4, the specific nature of the limiting distribution inside a minimal curb set is studied in detail. Section 5 concludes. Proofs are given in the appendix.

2. Adaptive play with clever agents

In the evolutionary model described below, I consider clever agents, introduced in [3] for the Nash bargaining game, in two-player games. The basic setting without clever agents is Young’s [7] model.

Let \( G \) be a two-player game with finite strategy space, \( X_1 \times X_2 \), and payoff functions \( u_i : X_1 \times X_2 \rightarrow \mathbb{R}, i = 1, 2 \). I assume that there exist two finite populations of agents. In each discrete time period, \( t = 1, 2, \ldots \), one agent is drawn at random from each of the populations to play the game. Agents in population 1 (2) can only play role 1 (2) in the game. Population 1 consists of clever and non-clever agents, in fixed population shares \( \lambda \) and \( 1 - \lambda \), respectively, while the agents in population 2 are only non-clever. All agents are equally likely to be drawn to play. An agent in role \( i \) chooses a strategy \( x^{t}_{i} \) from the set \( X_i \) at time \( t \) according to a rule defined below. The play at time \( t \) is the vector \( x^{t} = (x^{t}_{1}, x^{t}_{2}) \). The history of play up to time \( t \) is the sequence \( h^{t} = (x^{t-m+1}, \ldots, x^{t}) \) of \( m \) last plays.

Strategies are chosen as follows. Fix integers \( s \) and \( m \), where \( 1 \leq s \leq m \). At time \( t + 1 \), each agent drawn to play the game inspects a sample \( (x^{t}_{1}, \ldots, x^{t}_{s}) \) of size \( s \), taken without replacement from the history of size \( m \) of play up to time \( t \), where \( t_1, \ldots, t_s \in \{t - m + 1, t - m + 2, \ldots, t\} \). The draws of samples are statistically independent across agents and time. A non-clever agent chooses a best reply to the opponent population’s empirical strategy distribution in her sample. Clever agents are assumed to know the preferences of the other population and they use this knowledge to choose a best reply to the anticipated choice by their opponent. More precisely, a clever agent—these always play in role 1—inspects her own population’s play in her sample \( (x^{t}_{1}, \ldots, x^{t}_{s}) \), and calculates player 2’s best reply to this sample. Then, the clever agent chooses a best reply to this predicted strategy. If there is more than one best reply, an agent chooses each of these with fixed positive probabilities. These probabilities will be specified later on.

Consider the sampling process to begin in period \( t = m + 1 \) from some arbitrary initial sequence of \( m \) plays \( h^m \). We then obtain a finite Markov chain on the state
space \((X_1 \times X_2)^m = H\) of sequences of length \(m\) drawn from strategy space \(X_1 \times X_2\), with an arbitrary initial state \(h^m\). As we will see below, the resulting process is \textit{ergodic}; thus, in the long-run, the initial state is irrelevant. Given a history \(h^t = (x^{t-m+1}, \ldots, x^t)\) at time \(t\), the process moves to a state of the form \(h^{t+1} = (x^{t-m+2}, \ldots, x^t, x^{t+1})\), in the next period, a state called a \textit{successor} of \(h^t\).

The process moves from the current state \(h^t \in H\) to a successor state \(h'\) in each period, according to the following transition rule. For each \(x_i \in X_i\), let \(p_i(x_i|h)\) be the conditional probability that agent \(i\) chooses \(x_i\), given that the current state is \(h^t\). We assume that \(p_i(x_i|h)\) is independent of \(t\) and \(p_i(x_i|h) > 0\) if and only if there exists a sample \(s\) such that \(x_i\) is a best reply to this sample for a non-clever agent or \(x_i\) is a best reply to the opponent’s best reply to this sample for a clever agent in population \(1\). If \(x = (x_1, x_2)\) is the rightmost element of \(h^t\), the probability of moving from \(h^t\) to \(h'\) is \(R^{m,s,\lambda,0}_{hh'} = p_1(x_1|h)p_2(x_2|h)\) if \(h'\) is a successor of \(h^t\) and \(R^{m,s,\lambda,0}_{hh'} = 0\) if \(h'\) is not a successor of \(h^t\). Following Saez-Martí and Weibull [3], I call the process \(R^{m,s,\lambda,0}_{hh'}\) unperturbed \textit{adaptive play with clever agents} with memory \(m\), sample size \(s\), and share \(\lambda\) of clever agents in population \(1\).

The perturbed process can be described as follows. In each period, there is a small probability \(\varepsilon > 0\) that any drawn agent in role \(i\) experiments by choosing a random strategy from \(X_i\) instead of applying the best reply rule. The event that \(i\) experiments is assumed to be independent from the event that the other agent playing this game in the opponent role \(j\), experiments. For every \(i\), let \(q_i(x_i|h)\) be the conditional probability that \(i\) chooses \(x_i \in X_i\), given that \(i\) experiments and the perturbed process is in state \(h\). Assume that \(q_i(x_i|h)\) is independent of \(t\) and \(q_i(x_i|h) > 0\) for all \(x_i \in X_i\) and all \(h^t\). Suppose that the perturbed process is in state \(h^t\) at time \(t\). The probability is \(\varepsilon(1-\varepsilon)\) that exactly one of the agents playing the game experiments and that the other does not. Conditional on this event, the transition probability of moving from \(h^t\) to \(h'\) is \(Q^1_{hh'} = q_i(x_i|h)p_j(x_j|h)\), where \(i \neq j\), if \(h'\) is a successor of \(h^t\) and \(Q^1_{hh'} = 0\), if \(h'\) is not a successor of \(h^t\). Similarly, \(\varepsilon^2\) is the probability that both drawn agents experiment. Conditional on this event, the transition probability of moving from \(h^t\) to \(h'\) is \(Q^2_{hh'} = q_1(x_1|h)q_2(x_2|h)\), if \(h'\) is a successor of \(h\) and \(x\) is the rightmost element of \(h'\), and \(Q^2_{hh'} = 0\), if \(h'\) is not a successor of \(h\). This gives the following transition probability of the perturbed Markov process:

\[
R^{m,s,\lambda,\varepsilon}_{hh'} = (1-\varepsilon)^2 R^{m,s,\lambda,0}_{hh'} + 2\varepsilon(1-\varepsilon)Q^1_{hh'} + \varepsilon^2 Q^2_{hh'}.
\]

The process \(R^{m,s,\lambda,\varepsilon}_{hh'}\) is a denoted (perturbed) adaptive play with clever agents with memory \(m\), sample size \(s\), share \(\lambda\) of clever agents in population \(1\) and error rate \(\varepsilon\).

As is usual in evolutionary models, two forces drive the perturbed Markov process. The first—\(R^{m,s,\lambda,0}_{hh'}\)—is the \textit{selection} rule. The second—\(Q^1_{hh'}\) and \(Q^2_{hh'}\)—is the \textit{mutation}. Note that (1) if \(\lambda = 0\), then \(R^{m,s,0,\varepsilon}_{hh'}\) is Young’s [7] adaptive learning, (2) if \(\Gamma^r\) is the Nash bargaining game, we are in the framework of Saez-Martí and Weibull [3].
3. How cleverness does not matter

In this section, I discover when the introduction of a positive share $\lambda$ of clever agents does not change the long-run prediction of the model without clever agents, starting with useful definitions.

A product set of strategies is a set of form $C = C_1 \times C_2$, where each $C_i$ is a non-empty subset of $X_i$, $i = 1, 2$. Let $\Delta C_i$ denote the set of probability distributions over $C_i$, and let $\Delta C_1 \times \Delta C_2$ denote the product set of such distributions. Let $BR_i(C_j)$ denote the set of strategies in $X_i$ that are player $i$’s best replies to some distribution $\rho_j \in \Delta C_j$, $i \neq j$. Denote $BR(C) = BR_1(C_2) \times BR_2(C_1)$.

**Definition 1** (Basu and Weibull [1]). A non-empty Cartesian product set $C = C_1 \times C_2 \subseteq X$ is closed under best replies (or $C$ is a curb set) if $BR(\Delta C_1 \times \Delta C_2) \subseteq C$. Such a set is a minimal curb set if it does not properly contain a curb set.

It is straightforward to show that $BR(\Delta C_1 \times \Delta C_2) = C$ for any minimal curb set $C$. Following Young [7], a span of a subset $H' \subseteq H = (X_1 \times X_2)^m$, denoted by $S(H')$, is the product set of all pure strategies that appear in some history in $H'$. $H'$ is a minimal curb configuration if its span is a minimal curb set.

We say that a recurrent class of the process $R^{m,s,\lambda,0}$ is a set of states such that there is zero probability of moving from any state in the class to any outside state and there is a positive probability of moving from any state in the class to any other state in the class.

I will work with generic games and need to introduce a generic condition, which is a common condition in economics models. For discussion of this issue see, for example, Samuelson [4, p. 30].

Given a two-player game $\Gamma$ on the finite strategy space $X_1 \times X_2$, let $BR^{-1}_i(x_i)$ denote the set of all probability mixtures $\rho_j \in \Delta_j = \Delta X_j$, where $j \neq i$, such that $x_i$ is a best reply to $\rho_j$. I will work with Young’s [7] generic condition.

**Definition 2.** $\Gamma$ is a non-degenerate in best replies if for every player $i$ and every $x_i \in X_i$, either $BR^{-1}_i(x_i)$ is empty or contains a non-empty subset open in the relative topology of $\Delta_j$, where $j \neq i$.

The following result is a variant of a theorem of Hurkens [2] and a generalization of a theorem of Young [7].

**Theorem 1.** Let $\Gamma$ be a non-degenerate in best replies two-player game on the finite strategy space $X_1 \times X_2$. If $s/m$ is sufficiently small and $\lambda \in [0, 1)$, the unperturbed process $R^{m,s,\lambda,0}$ converges to a minimal curb configuration with probability one.

**Proof.** See the appendix.

If $\lambda = 1$, then Theorem 1 can fail. Consider the game in Fig. 1.
In this game, $A$ is a best reply to $a$ and $B$ a best reply to $b$. For any history $h$, an agent in player position 2 has three opportunities. She can have $a$ or $b$ as the only best reply to a sample of player 1 from history or she can be indifferent between $a$ and $b$: Hence, there are only strategies $A$ and $B$ in any sample of player 1 in the long-run if $\lambda = 1$. However, strategy $C$ also belongs to a minimal curb set, for it is a best reply to, for example, mixed strategy $2^{-7}a + 5^{-7}b$.

A process is said to be irreducible if and only if there is a positive probability of moving from any state to any other state in a finite number of periods. We will need the following definitions:

**Definition 3** (Young [7]). $R(\varepsilon)$ is a regular perturbed Markov process if $R(\varepsilon)$ is irreducible for every $\varepsilon \in (0, \varepsilon']$, and for every state $h, h' \in H$, $R_{hh'}(\varepsilon)$ approaches $R_{hh'}(0)$ at an exponential rate, i.e. $\lim_{\varepsilon \to 0} R_{hh'}(\varepsilon) = R_{hh'}(0)$ and if $R_{hh'}(\varepsilon) > 0$ for some $\varepsilon > 0$, then $0 < \lim_{\varepsilon \to 0} \frac{R_{hh'}(\varepsilon)}{R_{hh'}(0)} < \infty$ for some $r_{h \to h'} > 0$. The real number $r_{h \to h'}$ is the resistance of the transition $h \to h'$.

**Lemma 1.** An adaptive play with clever agents is a regular perturbed Markov process.

**Proof.** $R^{m,s,\lambda,\varepsilon}$ is a regular perturbed Markov process for the same reason as shown by Young [7] when he considers adaptive play. $\square$

**Definition 4** (Young [5]). Let $\mu(\varepsilon)$ be the unique stationary distribution of an irreducible process $R(\varepsilon)$. A state $h$ is stochastically stable if $\lim_{\varepsilon \to 0} \mu_h(\varepsilon) > 0$.

Let process $R^{m,s,\lambda,0}$ have recurrent classes $E_1, \ldots, E_K$. For each pair of distinct recurrent classes, a $pq$-path is a sequence of states $\zeta = (h_p, \ldots, h_q)$ beginning in $E_p$ and ending in $E_q$. The resistance of this path is the sum of the resistances on the edges composing it. Let $r_{pq}$ be the least resistance over all $pq$-paths. Construct a complete directed graph with $K$ vertices, one for each recurrent class. The weights on the directed edge $E_p \to E_q$ is $r_{pq}$. A tree rooted at $E_I$ is a set of $K - 1$ directed edges such that, from every vertex different from $E_I$, there is a unique directed path in the tree to $E_I$. The resistance of such a rooted tree is the sum of resistances $r_{pq}$ on its $K - 1$ edges.

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>12.1</td>
<td>0.0</td>
</tr>
<tr>
<td>$B$</td>
<td>4.0</td>
<td>4.1</td>
</tr>
<tr>
<td>$C$</td>
<td>8.1</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Fig. 1.
The stochastic potential of a recurrent class $E_l$ is the minimum resistance over all trees rooted at $E_l$. I will use the following result in the main theorem.

**Theorem 2** (Young [7]). Let $R(\varepsilon)$ be a regular perturbed Markov process and let $\mu(\varepsilon)$ be the unique stationary distribution of $R(\varepsilon)$, for $\varepsilon > 0$. Then, $\lim_{\varepsilon \to 0} \mu(\varepsilon) = \mu(0)$ exists and is a stationary distribution of $R(0)$. The stochastically stable states are precisely the states contained in the recurrent classes of $R(\varepsilon)$, having minimum stochastic potential.

We are now in a position to state the main result.

**Theorem 3.** Let $\Gamma$ be a non-degenerate in best replies two-player game on the finite strategy space $X_1 \times X_2$. If $s/m$ and $\varepsilon$ are sufficiently small, $s$ and $m$ are sufficiently large and $\lambda \in (0, 1)$, the perturbed process $R^{m,s,\lambda,\varepsilon}$ puts arbitrarily high probability on the minimal curb configuration(s) that minimize the stochastic potential of the perturbed process $R^{m,s,0,\varepsilon}$.

**Proof.** See the appendix.

This theorem shows that strategies taken by agents in the two populations are the same for the perturbed process $R^{m,s,\lambda,\varepsilon}$ with $\lambda \in (0, 1)$ and the perturbed process $R^{m,s,0,\varepsilon}$, without clever agents. In other words, the same recurrent classes will be chosen in the long run by the perturbed process $R^{m,s,\lambda,\varepsilon}$ for all $\lambda \in [0, 1)$. However, in the next section, it will be shown by means of an example that the distribution of strategies taken by agents in the two populations is different for different values of $\lambda$.

4. How cleverness matters

4.1. Intra-curb effects with $\lambda \in [0, 1)$

Although the presence of clever agents does not influence the choice of the limiting curb set if $\lambda \in [0, 1)$, as we saw in the previous section, I show here that it can influence the distribution of strategies *inside* the limiting curb configuration. This is clarified by means of the following example. Consider the Matching-Pennies game with the payoff matrix in [Fig. 2](#).

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>1, -1</td>
<td>-1, 1</td>
</tr>
<tr>
<td>$B$</td>
<td>-1, 1</td>
<td>1, -1</td>
</tr>
</tbody>
</table>

Fig. 2.
Choose the parameters in adaptive learning with clever agents as follows: memory size, \( m = 2 \); sample size, \( s = 1 \); proportion \( \lambda \in [0, 1] \) of clever agents in population 1.\(^1\) Each state can be represented by a \( 1 \times 4 \) block of \( A \)'s, \( B \)'s, \( a \)'s, and \( b \)'s, where the first two squares represent the previous two strategies of the agent in population 1 and the last two represent the previous two strategies of the agent in population 2. For example, state \( ABab \) means that the agent in population 1 chose \( A \) two periods ago, and \( B \) one period ago, while the agent in population 2 chose \( a \) two periods ago, and \( b \) one period ago. There are 16 possible states for the process.

The asymptotic properties of the finite Markov process \( R^{m,s,\lambda,\varepsilon} \) can be studied algebraically as follows. Let \( z_1 = AAaa, \ldots, z_{16} = BBbb \) be an enumeration of the states, let \( R \) be a transition matrix of the Markov process \( R^{m,s,\lambda,\varepsilon} \) on the finite state space \( \{(A, B) \times \{a, b\}\}^2 \), and let \( \mu(e, \lambda) = \left( \begin{array}{c} \mu_{z_1}(e, \lambda) \\ \vdots \\ \mu_{z_{16}}(e, \lambda) \end{array} \right) \) be a column vector of probability distribution on the finite state space \( \{(A, B) \times \{a, b\}\}^2 \). Consider a system of linear equations

\[
R \cdot \mu(e, \lambda) = \mu(e, \lambda), \quad \text{where } \mu(e, \lambda) \geq 0 \quad \text{and} \quad \sum_{l=1}^{16} \mu_{z_l}(e, \lambda) = 1.
\]

(2)

It is well known that this system (for the irreducible process \( R^{m,s,\lambda,\varepsilon} \)) always has exactly one solution \( \mu(e, \lambda) \), called a stationary distribution of the process \( R^{m,s,\lambda,\varepsilon} \). From Theorem 2, it follows that

\[
\lim_{\varepsilon \to 0} \mu(e, \lambda) = \mu(0, \lambda),
\]

(3)

where \( \mu(0, \lambda) \) is the stationary distribution of \( R^{m,s,\lambda,0} \). Note that the process \( R^{m,s,\lambda,\varepsilon} \) has only one recurrent class—the whole state space \( \{(A, B) \times \{a, b\}\}^2 \)—and without loss of generality, we can only analyze the unperturbed process \( R^{m,s,\lambda,0} \). Solving the system of linear equations:

\[
R \cdot \mu(0, \lambda) = \mu(0, \lambda),
\]

(4)

gives the stationary distribution \( \mu(0, \lambda) \) for different values of \( \lambda \), where \( R \) is the matrix described in Fig. 3. Empty squares in the matrix correspond to zeros. If an agent is indifferent between two pure strategies, then she is assumed to play either of them with equal probability.

Now, we can calculate the expected payoffs for both populations. It is enough only to consider the expected payoffs for the agent drawn from population 1 to play the game, since this is a zero-sum game.

A clever agent correctly predicts the only best reply of the other agent from population 2 if the agents in the two last periods in population 1 played the same strategy. Hence, a clever agent always receives payoff \( u^i_i(\lambda, z) = 1 \) in the following 8 states: \( AAaa, AAab, AAbba, AAbb, BBaa, BBab, BBba, BBbb \). In the remaining 8 states, the expected payoff to a clever agent is zero. Therefore, the total expected

\(^1\) For our purpose, we only need to consider case \( \lambda \in [0, 1] \).
payoff to a clever agent is

\[ u^c_1(\lambda) = \sum_{j=1}^{16} \mu_{z_j}(0, \lambda)u^c_1(\lambda, z_j) = \sum_{j=1}^{4} \mu_{z_j}(0, \lambda) + \sum_{j=13}^{16} \mu_{z_j}(0, \lambda) \geq 0, \quad (5) \]

where \( z_1 = AAaa, \ldots, z_4 = AAbb, z_{13} = BBaa, \ldots, z_{16} = BBbb \).

A non-clever agent in population 1 plays a best reply to an opponent's probability distribution in the sample. As a result of this behavior, she always receives payoff \(-1\) in the following two states: \( AAaa \) and \( BBbb \), and payoff 1 in the following two states: \( AAbb \) and \( BBaa \). In the remaining 12 states, the expected payoff to a non-clever agent is zero. Therefore, the total expected payoff to a non-clever agent in population 1 is

\[ u^n_1(\lambda) = \sum_{j=1}^{16} \mu_{z_j}(0, \lambda)u^n_1(\lambda, z_j) = \mu_{z_1}(0, \lambda) - \mu_{z_4}(0, \lambda) - \mu_{z_{13}}(0, \lambda) + \mu_{z_{16}}(0, \lambda). \quad (6) \]

The expected payoff to a clever agent is positive if at least one of the states \( z_1, \ldots, z_4, z_{13}, \ldots, z_{16} \) shows up in the stationary distribution \( \mu(0, \lambda) \), which is the case for any value of \( \lambda \in (0, 1) \). Hence, "cleverness" is an advantage in the Matching-Pennies game, because the expected payoff to a clever agent is higher than the expected payoff to a non-clever agent from the same population. What is the expected average payoff of population 1? Fig. 4 shows how the expected payoffs to a clever agent, a non-clever agent and population 1 (on average) depend on the share \( \lambda \) of clever agents in population 1.

There are different stationary distributions for different values of \( \lambda \). The expected payoff to a non-clever agent in population 1 is negative and falling as the share of clever agents increases. The expected average payoff to population 1 is positive for

\[ \text{ARTICLE IN PRESS} \]
\begin{align*}
\lambda > 0 \text{ and depends positively on the share of clever agents. Accordingly, all agents in population 2 earn a negative expected payoff, which is decreasing in } \lambda. \text{ Clever agents outsmart agents from the other population. The larger is the share of clever agents in population 1, the smaller is the expected payoff to each clever agent. In this sense, the marginal return to cleverness is decreasing. Moreover, the presence of clever agents in population 1 imposes a negative externality on the non-clever agents in the same population.}

\textbf{4.2. Clever population against non-clever population, } \lambda = 1
\end{align*}

Consider an extreme case, where all agents in population 1 are clever. In this case, strategies that agents choose in the two populations may differ for the perturbed process $R^{m,s,1,\varepsilon}$ with all clever agents in population 1 and the perturbed process $R^{m,s,0,\varepsilon}$ without clever agents. The question now arises whether the clever population does better off than the non-clever population in the same player position. The answer depends on the game. The following examples illustrate this point. I start from the strict demand game and then compare that game with the Nash demand game, studied in [3].

Consider two finite populations, 1 and 2, who periodically bargain pair-wise over their shares of a common pie. Let $x$ denote the share of player 1, and let $y$ denote the share of player 2. Suppose that all agents in population 1 have the same concave,
increasing, and differentiable utility function, which is a function of the share \( x \)
\[
u : [0, 1] \to \mathbb{R},
\]
and all agents in population 2 have the same concave, increasing, and differentiable utility function as a function of the share \( y \)
\[
v : [0, 1] \to \mathbb{R}.
\]
Without loss of generality, we can normalize \( u \) and \( v \) so that \( u(0) = v(0) = 0 \).

In each period \( t = 1, 2, \ldots \), one agent is drawn at random from each population. They play the strict demand game, later SDG: player 1 demands some number \( x \in (0, 1) \), and simultaneously, player 2 demands some number \( y \in (0, 1) \). The outcomes and payoffs are as in Fig. 5.

To keep the state space finite, we shall discretize demands. Let a finite set \( D(\delta) = \{\delta, 2\delta, \ldots, 1 - \delta\} \) be the space of demands. Furthermore, let \( R^{m,s,1,e} \) be an adaptive play with all clever agents in population 1. Let \( (x', y') \) denote the amounts demanded by the agents in populations 1 and 2, respectively, in period \( t \). At the end of period \( t \), the state is
\[
h_t' = ((x', y', m-1), \ldots, (x', y', 0)).
\]
At the beginning of period \( t + 1 \), the current clever agent, playing the game, draws a sample of size \( s \) from the \( x \)-values in history \( h_t' \). Simultaneously and independently, the agent in population 2 also draws a sample of size \( s \) from the \( x \)-values in history \( h_t' \).

A conventional division is a state of the form
\[
h_x = ((x, 1 - x), \ldots, (x, 1 - x)),
\]
where \( 0 < x < 1 \). We say that a division \( (x, 1 - x) \) is stochastically stable for a given precision \( \delta \), if the corresponding convention \( h_x \) is stochastically stable.

The main result of this part is the following.

**Proposition 1.** Assume all agents in population 1 to be clever. Then, for any precision \( \delta > 0 \), there exists one stable division, which converges to \( (x, y) = (0, 1) \) as \( \delta \to 0 \).

**Proof.** See the appendix.

The proposition shows that to be an agent from the non-clever population is better than to be an agent from the clever population in the strict demand game. It contrasts to Saez-Marti and Weibull [3], who prove that if \( \lambda = 1 \), then as \( \delta \to 0 \), the stable division converges to \( (x, y) = (1, 0) \) in the Nash demand game.

<table>
<thead>
<tr>
<th>Demands</th>
<th>Outcomes</th>
<th>Payoffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x + y = 1 )</td>
<td>( x, y )</td>
<td>( u(x), v(y) )</td>
</tr>
<tr>
<td>( x + y \neq 1 )</td>
<td>( 0, 0 )</td>
<td>( 0, 0 )</td>
</tr>
</tbody>
</table>

Fig. 5.
5. Concluding remarks

In this paper, I answer some questions asked in [3]. They study the consequences of letting some agents play a myopic best reply to the myopic best reply in Young’s [6] bargaining model, which is how they introduce “cleverness” of players. Saez-Marti and Weibull [3] ask whether their results can be generalized. I use the “cleverness” approach from their paper to analyze generic two-player games in Young’s [7] set-up. The resulting Markov process is denoted as an adaptive play with clever agents.

Saez-Marti and Weibull [3] prove that an introduction of any share of clever agents less than one in the special case will not change the long-run behavior for the Nash demand game. I have shown this result to be robust in generic two-player games: adaptive learning with clever agents will settle down in a minimal curb configuration, which minimizes the stochastic potential for adaptive learning without clever agents. However, the share of clever agents does matter inside the minimal curb configuration, as shown in the Matching-Pennies game, where the gain of clever agents depends on the share of these agents in the population.

Furthermore, in the extreme case, if the share of clever agents equals one, then we have a discontinuity in the following sense. In the case of all clever agents in one of the populations, the stochastically stable states might differ from previous ones. Saez-Marti and Weibull [3] find this discontinuity for the Nash demand game.

I also study whether it is advantageous to be a member of the population consisting of clever agents only, and show the answer to be ambiguous even in coordination games. On the one hand, Saez-Marti and Weibull [3] show that the “clever” population gets the whole pie in the Nash demand game. On the other hand, in the strict demand game, where two players must coordinate to get exactly the size of the pie, otherwise they both get nothing, the population without clever agents obtains the whole pie.

Acknowledgments

This paper is a revised version of Chapter 2 of my Ph.D. Thesis at the Stockholm School of Economics. I thank Martin Dufwenberg, Jens Josephson, Martin Meier, Vladimir Rodionov, Maria Saez-Marti, Jana Vyrastekova, Peyton Young, Karl Wärneryd, the associate editor, conference participants at the Stockholm School of Economics and the Tenth International Conference at Stony Brook for helpful comments, and Tatiana Damjanovic for assistance on computer issues. Special thanks to Kerstin Niklasson and Roland Nilsson for their fine editing. I am indebted to Jörgen Weibull, who has provided personal encouragement for this paper. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged. Remaining errors are solely due to the author.
Appendix

Proof of Theorem 1. This proof is similar to the proof in [7, Theorem 7.2]. We need only show that clever agents will change nothing.

Given a positive integer $s$, we say that the probability distribution $\rho_i \in \Delta_i$ has precision $s$ if $s \rho_i$ is an integer for all $x_i \in X_i$. We shall denote the set of all such distributions by $\Delta_i^s$. For each subset $Y_i \subseteq X_i$, let $\Delta_i^s(Y_i)$ denote the set of distributions $\rho_i \in \Delta_i^s$ such that $\rho_i(x_i) > 0$ implies $x_i \in Y_i$. For each positive integer $s$, let $BR_s^i(X_i)$ be the set of pure-strategy best replies by a non-clever agent in population $i$ to some product distribution $\rho_{-i} \in \Delta_{-i}^s(X_{-i}) = \Delta_i^s(Y_i)$, where $j \neq i$. Similarly, $BR_s^i(Y_{-i})$ denotes the set of all best replies by a non-clever agent in population $i$ to some product distribution $\rho_{-i} \in \Delta_{-i}^s(Y_{-i})$. Let $BR_s^i(X_i)$ be the set of pure-strategy best replies by a clever agent in population $1$ to some product distribution $\rho_1 \in \Delta_1^s(X_1)$.

For each product set $Y$ and an agent in population $i$, define the mappings

$$\beta_i(Y) = Y_i \cup BR_s^i(Y_{-i}) \quad \text{and} \quad \beta'_i(Y) = Y_i \cup BR_s^i(Y_1),$$  \hspace{1cm} (A.1)

and let $\beta(Y) = [\beta_1(Y) \cup \beta'_1(Y)] \times \beta_2(Y)$, where $\beta'_1(Y)$ is the mapping for the clever agents from population $1$ and $\beta_i(Y)$ is the mapping for the non-clever agent $i$. Note that $\beta(Y) = [\beta_1(Y) \cup BR_s^1(Y_1)] \times \beta_2(Y)$. Similarly, for each integer $s \geq 1$ let

$$\beta^s_1(Y) = Y_i \cup BR_{s}^i(Y_{-i}), \quad \beta^s_i(Y) = Y_i \cup BR_{s}^i(Y_1)$$  \hspace{1cm} (A.2)

and

$$\beta^s(Y) = [\beta^s_1(Y) \cup \beta^s_1(Y)] \times \beta^s_2(Y).$$  \hspace{1cm} (A.3)

In the same way, as it appears in the proof in [7, Theorem 7.2], we can show that $\beta^s(Y) = \beta(Y)$ for all sufficiently large $s$.

Consider the process $R_{m,s,x}^{i,0}$. We show that if $s$ is large enough and $s/m$ is small enough, the spans of the recurrent classes correspond one to one with the minimal curb sets of game $\Gamma$.

Fix a recurrent class $E_k$ of $R_{m,s,x}^{i,0}$, and choose any $h^0 \in E_k$ as the initial state. We shall show that the span of $E_k$, $S(E_k)$, is a minimal curb set. As shown in the proof in [7, Theorem 7.2], there is a positive probability of reaching a state $h^1$ where the most recent $s$ entries involve a repetition of some fixed $x^* \in X$, because there is a positive probability that a non-clever agent will be chosen from population $1$ in every period.

Note that $h^1 \in E_k$, because $E_k$ is a recurrent class. Let $\beta^{(j)}$ denote the $j$-fold iteration of $\beta$ and consider the nested sequence

$$\{x^*\} \subseteq \beta(\{x^*\}) \subseteq \beta^{(2)}(\{x^*\}) \subseteq \cdots \subseteq \beta^{(j)}(\{x^*\}) \subseteq \cdots.$$  \hspace{1cm} (A.4)

Since $X$ is finite, there exists some point at which this sequence becomes constant, say

$$\beta^{(j)}(\{x^*\}) = \beta^{(j+1)}(\{x^*\}) = Y^*.$$  \hspace{1cm} (A.5)

By construction, $Y^*$ is a curb set.

The proof that $Y^*$ is, in fact, a minimal curb set is the same as in the proof in Young [7, Theorem 7.2]. □
Proof of Theorem 3. It follows immediately from Theorems 1 that minimal curb configurations are recurrent classes of the regular perturbed Markov process $R^{m,s,\lambda,\varepsilon}$. By Theorem 2, one (or some) of this minimal curb configuration(s) is (are) stochastically stable. We must show that this minimal curb configuration is the same as in the absence of the clever agents.

Take any two recurrent classes, two minimal curb configurations, $E_p$ and $E_q$. Note that every mistake made in population 2 can only influence the behavior of the non-clever agents in population 1. It means that all mistakes made in population 2 have the same effect for both processes $R^{m,s,\lambda,\varepsilon}$, $\lambda \in (0,1)$, and $R^{m,s,0,\varepsilon}$.

Suppose that $l$ mistakes in a row in population 1 are necessary to move process $R^{m,s,0,\varepsilon}$ from recurrent class $E_p$ to recurrent class $E_q$. The clever agents in process $R^{m,s,\lambda,\varepsilon}$ anticipate this. Hence, if there were less than $l$ mistakes in population 1, then the clever agent in role 1 expects an agent in role 2 to play as if they are in recurrent class $E_p$. There must be at least $l$ mistakes in population 1 to change these expectations of the clever agent. It means that there must be at least $l$ mistakes in population 1 to move process $R^{m,s,\lambda,\varepsilon}$ from recurrent class $E_p$ to recurrent class $E_q$. At the same time, there is a positive probability that only non-clever agents will be chosen from population 1 in every period. Therefore, it is enough to make exactly $l$ mistakes in population 1 to move process $R^{m,s,\lambda,\varepsilon}$ from recurrent class $E_p$ to recurrent class $E_q$. \[ \square \]

Proof of Proposition 1. We will need the following definitions.

Definition A.1. The basin of attraction of state $h$ is the set of states $h'$ such that there is a positive probability of moving from $h'$ to $h$ in a finite number of periods under the unperturbed process $R^{m,s,1,0}$.

For every real number $r$, let $\lfloor r \rfloor$ denote the least integer greater than or equal to $r$.

Lemma A.1. For every $x \in D(\delta)$, the minimum resistance of moving from convention $h_x$ to a state in some other basin of attraction is $\lfloor sr_0(x) \rfloor$, where

$$ r_0(x) = \frac{1}{1 + \frac{\pi(1-\delta)}{\pi(1-x)}}. \quad (A.6) $$

Proof. Suppose that the process is in the convention $h_x$, where $x \in D(\delta)$. Let $\pi$ be a path of the least resistance from $h_x$ to a state that is in some other basin of attraction. Clearly, $\pi$ must pass through some state $w$ such that some best reply of agent in population 2 to a sample from $w$ is different from $1/x$. Let $w$ be the first such a state.

To compute the least number of mistakes necessary to exit from convention $h_x$, it suffices to consider, for every $x' \neq x$, the least number of initial mistakes $x'$ by the agents in population 1 that will cause an agent in population 2 to reply with $1-x'$. 

\[ \text{ARTICLE IN PRESS} \]
The number of mistakes in population 2 does not matter in this setting, because both agents only look at the $x$-values in $h'$. Choose an arbitrary $x' \neq x$. Suppose that the agents in population 1 make $j$ successive demands of $x'$ that cause some agent’s best reply in population 2 to switch to $1 - x'$ instead of $1 - x$. We can assume that $j \leq s$. When the agent in population 2 samples these $j$ mistaken demands $x'$, together with $s - j$ of the previous “conventional” demands $x$, she switches to $1 - x'$ provided that
\[
\frac{j}{s} v(1 - x') \geq \frac{s - j}{s} v(1 - x),
\]
that is
\[
j \geq \frac{v(1 - x)}{v(1 - x') + v(1 - x)s}.
\]
Over all feasible $x' \neq x$, the minimum value of $j$ occurs when $x' = \delta$ and
\[
j = \frac{1}{1 + \frac{v(1 - \delta)}{v(1 - x)}}.
\]
Hence, the lowest number of mistakes to exit from the $h_x$-basin of attraction is $\lceil sr_\delta(x) \rceil$, where
\[
r_\delta(x) = \frac{1}{1 + \frac{v(1 - \delta)}{v(1 - x)}}.
\]
This completes the proof of Lemma A.1.

**Lemma A.2.** A division $(x, 1 - x)$ is stochastically stable if and only if $x$ maximizes the function $r_\delta(x)$ on $D(\delta)$.

**Proof.** It follows from Theorem 2.

**Corollary A.1.** The division $(\delta, 1 - \delta)$ is stochastically stable.

**Proof.** $x$ maximizes the function $r_\delta(x)$ on $D(\delta)$ at $x = \delta$.

Proposition 1 follows immediately from the corollary.

**References**