

# Clever agents in adaptive learning

ALEXANDER MATROS\*  
DEPARTMENT OF ECONOMICS  
STOCKHOLM SCHOOL OF ECONOMICS

October 3, 2000

ABSTRACT. Saez-Marti and Weibull [4] investigate the consequences of letting some agents play a myopic best reply to the myopic best reply in Young's [8] bargaining model. This is how they introduce "cleverness" of players. We analyze such clever agents in *general* finite two-player games. We show that Young's [9] prediction is 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 strategies distribution in such a curb set may vary with the share of clever agents.

*Key words:* Evolution; game theory; bounded rationality; Markov chain; stochastic stability.

*Journal of Economic Literature Classification:* C72, C73.

## 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 ought to be placed in a model of boundedly rational agents.

Recently, Young [7] – [9] suggested an evolutionary model which explains how agents can make their choices based only on their own preferences and a sample of what happened in the recent past. A two-player game is played repeatedly by the members of two large populations. 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 drawn agents simultaneously chooses a strategy in the game. Every agent has access to a random sample, drawn from the

---

\*Previous version - licentiate paper - February 15, 2000. I would like to thank Martin Dufwenberg, Jens Josephson, Martin Meier, Vladimir Rodionov, Maria Saez-Marti, Jana Vyrastekova, Peyton Young, Karl Wärneryd, conference participants at the Stockholm School of Economics and the Tenth International Conference at Stony Brook for helpful comments, and Tatiana Damjanovic for assistance regarding computer issues. Special thanks go 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 by the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged. Remaining errors are due solely to the author.

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 opponent population's empirical strategy distribution in the sample. However, occasionally agents "mutate", and instead choose a strategy that is not a best reply to any possible sample from the recent history of play.

Saez-Marti and Weibull [4], using Young's [8] model, consider agents that are "clever" in a certain sense. They study the effect of letting a share of one of the populations know the opponent population preferences, denoting these agents as "clever", in the Nash Demand Game. Saez-Marti and Weibull [4] assume, firstly, that the population without clever agents plays its best reply to a sample of past strategies played by the other population and, secondly, clever agents play a best reply to their opponent population's best reply to the clever agent's sample. In other words, the 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 less than one of clever agents.

The purpose of the present paper is to demonstrate that this robustness holds for generic finite two-player games. Young [9] proves that the *adaptive learning* process settles down in a minimal curb configuration which minimizes the stochastic potential in such games when the mutation rate goes to zero. Our main result is that the *adaptive learning with* any share less than one of *clever agents* converges to the same minimal curb configuration as the Young's adaptive learning process when the mutation rate goes to zero. However, we show that the presence of clever agents may influence the long-run strategies distribution inside the minimal curb configuration.

We also analyze the question, which was posted in Saez-Marti and Weibull [4], how well clever agents fare among non-clever given fixed population shares of clever and non-clever agents. More specifically, we consider the Matching-Pennies Game and find that the gain of clever agents depends on the share of these agents in the population. Moreover, clever agents outsmart not only the agents in the other population but also indirectly 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 is the gain to the population. On the margin, an additional clever agent gains less as the share of clever agents increases. We can call it "decreasing returns to cleverness".

Saez-Marti and Weibull [4] show that the "clever" population gets the whole pie playing with the "non-clever" population in the *Nash Demand Game*. However, we demonstrate that "cleverness" does not guarantee an advantage in asymmetric coordination games even if all agents in one population are clever. In the *Strict Demand Game*, where two players must coordinate to get in sum exactly the size

of the pie, otherwise they both get nothing, the population without clever agents obtains the whole pie.

Young [9] shows that for  $2 \times 2$  coordination games, in the limiting case when the mutation rate goes to zero, adaptive play converges to the *risk-dominant* convention - a repetition of the *risk-dominant* equilibrium. He also analyzes the case with two different sample sizes for two populations. We demonstrate that clever agents matter: if the sample sizes are different in the two populations, then the stochastically stable convention may differ from Young's [9] prediction. In the extreme case, when *all agents* in one population are clever, the outcome is *the risk-dominant convention* for the other population.<sup>1</sup> The intuition for this result is as follows. The clever agent in population 1 predicts the strategy of her opponent and chooses the predicted strategy too, because of the specific structure of  $2 \times 2$  coordination games.

The paper is organized as follows. In Section two we describe the unperturbed and perturbed versions of adaptive play with clever agents. In Section three we derive general results for the stationary distribution of this process. In Section four the specific nature of the limiting distribution inside of a minimal curb set is studied in details. In Section five we investigate the properties of  $2 \times 2$  coordination games when agents have asymmetric information. We conclude in Section six. Proofs are given in the Appendix.

## 2. ADAPTIVE PLAY WITH CLEVER AGENTS

In the evolutionary model described below we consider clever agents, introduced in Saez-Marti and Weibull [4] for the Nash bargaining game, in *two-player games*. The basic setting without clever agents is Young's [9] model.

Let  $\Gamma$  be a two-player game with finite strategy space  $X_1 \times X_2$  and payoff functions  $u_i : X_1 \times X_2 \rightarrow \mathbf{R}$ ,  $i = 1, 2$ . We assume that there exist two finite populations of agents. In each discrete time period,  $t = 1, 2, \dots$ , 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. We assume that all agents are equally likely to be drawn to play. An agent in role  $i$  chooses at time  $t$  a strategy  $x_i^t$  from the set  $X_i$  according to a rule that will be defined below. The *play* at time  $t$  is the vector  $x^t = (x_1^t, x_2^t)$ . The *history of play* up to time  $t$  is the sequence  $h^t = (x^{t-m+1}, \dots, x^t)$ .

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 of size  $s$ , taken without replacement from the history of play up to time  $t$ . The draws of samples

---

<sup>1</sup>To find the risk-dominant convention we have to take in consideration only payoffs of the non-clever population. The formal definition is on page 16.

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. We assume that clever agents know the preferences of the other population and that 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, and calculates player 2's best reply to this sample. Then the clever agent chooses a best reply to this predicted strategy. If there are more than one best reply, then an agent chooses each of them with positive probability. We will specify these probabilities later on.

We can think of the sampling process as beginning 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 some an arbitrary initial state  $h^m$ . As we will see below, the resulting process is *ergodic*; thus, in the long-run, the initial state is irrelevant. Given a history  $h^t = (x^{t-m+1}, \dots, x^t)$  at time  $t$ , the process moves in the next period to a state of the form  $h^{t+1} = (x^{t-m+2}, \dots, x^t, x^{t+1})$ . Such a state is called a *successor* of  $h^t$ .

The process moves from the current state  $h$  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  $h$ . We assume that  $p_i(x_i | h)$  is independent of  $t$  and that  $p_i(x_i | h) > 0$  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  $h'$ , the probability of moving from  $h$  to  $h'$  is  $R_{hh'}^{m,s,\lambda,0} = p_1(x_1 | h)p_2(x_2 | h)$  if  $h'$  is a successor of  $h$  and  $R_{hh'}^{m,s,\lambda,0} = 0$  if  $h'$  is not a successor of  $h$ . Following Young [10], we call the process  $R^{m,s,\lambda,0}$  unperturbed *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 strategy at random 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's 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  $h$ . We assume that  $q_i(x_i | h)$  is independent of  $t$  and that  $q_i(x_i | h) > 0$  for all  $x_i \in X_i$  and for all  $h$ . Suppose that the perturbed process is in state  $h$  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  $h$  to  $h'$  is  $Q_{hh'}^1 = q_i(x_i | h)p_j(x_j | h)$ , where

$i \neq j$ , if  $h'$  is a successor of  $h$  and  $x$  is the rightmost element of  $h'$  and  $Q_{hh'}^1 = 0$ , if  $h'$  is not a successor of  $h$ . Similarly,  $\varepsilon^2$  is the probability that both drawn agents experiment. Conditional on this event, the transition probability of moving from  $h$  to  $h'$  is  $Q_{hh'}^2 = q_1(x_1 | h)q_2(x_2 | h)$ , if  $h'$  is a successor of  $h$  and  $x$  is the rightmost element of  $h'$  and  $Q_{hh'}^2 = 0$ , if  $h'$  is not a successor of  $h$ . This gives the following transition probability of the perturbed Markov process:

$$R_{hh'}^{m,s,\lambda,\varepsilon} = (1 - \varepsilon)^2 R_{hh'}^{m,s,\lambda,0} + 2\varepsilon(1 - \varepsilon)Q_{hh'}^1 + \varepsilon^2 Q_{hh'}^2. \quad (1)$$

The process  $R^{m,s,\lambda,\varepsilon}$  is 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 usual in evolutionary models, two forces drive the perturbed Markov process. The first -  $R_{hh'}^{m,s,\lambda,0}$  - is the *selection* rule. The second -  $Q_{hh'}^1$  and  $Q_{hh'}^2$  - is the *mutation*. Note that 1) if  $\lambda = 0$ , then  $R_{hh'}^{m,s,0,\varepsilon}$  is Young's [9] adaptive learning, 2) if  $\Gamma$  is the Nash bargaining game, then we are in the framework of Saez-Marti and Weibull [4].

### 3. HOW CLEVERNESS DOES NOT MATTER

In this section we find out when introduction of positive share  $\lambda$  of clever agents does not change the long-run prediction of the model without clever agents. We start from 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  $p_j \in \Delta C_j$ ,  $i \neq j$ . Denote  $BR(C) = BR_1(C_2) \times BR_2(C_1)$ .

**Definition 1.** (Basu and Weibull [1]) A non-empty Cartesian product set  $C = C_1 \times C_2 \subset X$  is **closed under best replies** (or  $C$  is a **curb set**) if  $BR(\Delta C_1 \times \Delta C_2) \subset C$ . 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 [9], we say that 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 state outside and there is a positive probability of moving from any state in the class to any other state in the class.

We will work with generic games and we need to introduce a generic condition. These kinds of conditions are common in economics models. For discussion of this issue see, for example, Samuelson [5, pp.30].

Given a two-player game  $\Gamma$  on the finite strategy space  $X_1 \times X_2$ , let  $BR_i^{-1}(x_i)$  denote the set of all probability mixtures  $p_j \in \Delta_j = \Delta X_j$ , where  $j \neq i$ , such that  $x_i$  is a best reply to  $p_j$ . We will work with the Young's [9] generic condition.

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

The following result shows that a prediction in generic games for the unperturbed process  $R^{m,s,\lambda,0}$  with  $\lambda \in (0, 1)$  is the same as in Young [9] for the unperturbed process  $R^{m,s,0,0}$ . In other words recurrent classes for process  $R^{m,s,\lambda,0}$  are independent from  $\lambda$ .

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

**Proof:** See the Appendix.

If  $\lambda = 1$ , then Theorem 1 can fail. Consider the game in Figure 1.

	$a$	$b$
$A$	12, 1	0, 0
$B$	4, 0	4, 1
$C$	8, 1	3, 0

FIGURE 1

In this game,  $A$  is a best reply to  $a$  and  $B$  is 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 the history or she can be indifferent between  $a$  and  $b$ . In the last case we assume that an agent in player position 2 will randomize 50 : 50 between two choices. A clever agent in player position 1 will also calculate these three opportunities and play  $A$  against  $a$ ,  $B$  against  $b$ , and  $A$  in the third case. Hence there are only strategies  $A$  and  $B$  in any sample of player 1 in the long-run if  $\lambda = 1$ . However strategy  $C$  belongs to a minimal curb set too, because it is a best reply to, for example, mixed strategy  $\frac{2}{7}a + \frac{5}{7}b$ .

We say that a process is *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 [10])  $R(\varepsilon)$  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.  $\lim_{\varepsilon \rightarrow 0} R_{hh'}(\varepsilon) = R_{hh'}(0)$  and if  $R_{hh'}(\varepsilon) > 0$  for some  $\varepsilon > 0$ , then  $0 < \lim_{\varepsilon \rightarrow 0} \frac{R_{hh'}(\varepsilon)}{\varepsilon^{r_{h \rightarrow h'}}} < \infty$  for some  $r_{h \rightarrow h'} \geq 0$ . The real number  $r_{h \rightarrow h'}$  is the **resistance** of the transition  $h \rightarrow h'$ .

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

**Proof:**  $R^{m,s,s',\lambda,\varepsilon}$  is a regular perturbed Markov process for the same reason as shown by Young [9] when he considers adaptive play. **End of proof.**

**Definition 4.** (Young [7]) Let  $\mu(\varepsilon)$  be the unique stationary distribution of an irreducible process  $R(\varepsilon)$ . A state  $h$  is **stochastically stable** if  $\lim_{\varepsilon \rightarrow 0} \mu_h(\varepsilon) > 0$ .

Let process  $R^{m,s,\lambda,0}$  have recurrent classes  $E_1, \dots, E_K$ . For each pair of distinct recurrent classes, a  $pq$ -path is a sequence of states  $\zeta = (h_p, \dots, h_q)$  that begins in  $E_p$  and ends in  $E_q$ . The *resistance* of this path is the sum of the resistances on the edges that compose 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  $E_p \rightarrow E_q$  is  $r_{pq}$ . A tree rooted at  $E_l$  is a set of  $K - 1$  directed edges such that, from every vertex different from  $E_l$ , there is a unique directed path in the tree to  $E_l$ . The *resistance of such a rooted tree*  $\mathfrak{S}(E_l)$  is the sum of the resistances  $r_{pq}$  on the  $K - 1$  edges that compose it. The *stochastic potential*  $\rho(E_l)$  of a recurrent class  $E_l$  is the minimum resistance over all trees rooted at  $E_l$ . We will use the following result in our main theorem.

**Theorem 2.** (Young [9]) 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  $\lim_{\varepsilon \rightarrow 0} \mu(\varepsilon) = \mu(0)$  exists and is a stationary distribution of  $R(0)$ . The stochastically stable states are precisely the states that are 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 nondegenerate 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 that agents take 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 that agents take in the two populations is different for different values of  $\lambda$ .

#### 4. HOW CLEVERNESS MATTER

**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, here we show that it can influence the distribution of strategies *inside* the limiting curb configuration. We clarify that by means of the following example. Consider the Matching-Pennies game with the following payoff matrix in Figure 2.

	$a$	$b$
$A$	$1, -1$	$-1, 1$
$B$	$-1, 1$	$1, -1$

FIGURE 2

Choose the parameters in adaptive learning with clever agents as follows: memory,  $m = 2$ ; sample size,  $s = 1$ ; proportion  $\lambda \in [0, 1]$  of clever agents in population 1.<sup>2</sup> 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 agent's in population 1 previous two strategies and the last two represent the agent's in population 2 previous two strategies. For example, the 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  $z_1 = AAaa, \dots, z_{16} = BBbb$  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  $\mu(\varepsilon, \lambda) = \begin{pmatrix} \mu_{z_1}(\varepsilon, \lambda) \\ \vdots \\ \mu_{z_{16}}(\varepsilon, \lambda) \end{pmatrix}$  be a column vector

---

<sup>2</sup>For our purpose we need to consider only case  $\lambda \in [0, 1)$ .



of probability distribution on the finite state space  $(\{A, B\} \times \{a, b\})^2$ . Consider a system of linear equations

$$R \cdot \mu(\varepsilon, \lambda) = \mu(\varepsilon, \lambda), \text{ where } \mu(\varepsilon, \lambda) \geq 0 \text{ and } \sum_{l=1}^{16} \mu_{z_l}(\varepsilon, \lambda) = 1. \quad (2)$$

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

$$\lim_{\varepsilon \rightarrow 0} \mu(\varepsilon, \lambda) = \mu(0, \lambda), \quad (3)$$

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 analyze only the unperturbed process  $R^{m,s,\lambda,0}$ . Solving the system of linear equations:

$$R \cdot \mu(0, \lambda) = \mu(0, \lambda), \quad (4)$$

gives the stationary distribution  $\mu(0, \lambda)$  for different values of  $\lambda$ , where  $R$  is the matrix described in Figure 3. Empty squares in the matrix correspond to zeros. We assume here that if an agent is indifferent between two pure strategies, then she plays both of them with the same probability.

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

A clever agent predicts correctly the only best reply of the other agent from population 2 if two last periods agents in population 1 played the same strategy. Hence a clever agent always receives payoff  $u_1^c(\lambda, z_j) = 1$  in the following 8 states:  $AAaa, AAab, AAbA, AAbb, BBaa, BBab, BBba, BBbb$ . In the remaining 8 states the expected payoff to a clever agent is zero. Therefore, the total expected payoff to a clever agent is

$$u_1^c(\lambda) = \sum_{j=1}^{16} \mu_{z_j}(0, \lambda) u_1^c(\lambda, z_j) = \sum_{j=1}^4 \mu_{z_j}(0, \lambda) + \sum_{j=13}^{16} \mu_{z_j}(0, \lambda) > 0, \quad (5)$$

where  $z_1 = AAaa, \dots, z_4 = AAbb, z_{13} = BBaa, \dots, z_{16} = BBbb$ .

FIGURE 3.

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 2 states:  $AAaa$  and  $BBbb$ , and payoff 1 in the following 2 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

$$u_1^n(\lambda) = \sum_{j=1}^{16} \mu_{z_j}(0, \lambda) u_1^n(\lambda, z_j) = \mu_{z_1}(0, \lambda) - \mu_{z_4}(0, \lambda) - \mu_{z_{13}}(0, \lambda) + \mu_{z_{16}}(0, \lambda). \quad (6)$$

The expected payoff to a clever agent is positive if at least one of the states  $z_1, \dots, z_4, z_{13}, \dots, z_{16}$  shows up in the stationary distribution  $\mu(0, \lambda)$ . This is the case for any value of  $\lambda \in (0, 1]$ . Hence a “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.<sup>3</sup> What is the expected average payoff of population 1? Figure 4 shows how the expected payoffs to a clever agent, a non-clever agent and population 1 (in average) depend on the share  $\lambda$  of clever agents in population 1.

---

<sup>3</sup>We think that it should be a general result that the expected payoff of the clever agent is equal to or higher than the expected payoff of the non-clever agent from the same population inside of any minimal curb set.

FIGURE 4.

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  $\lambda > 0$  and depends positively on the share of clever agents. Accordingly, all agents in population 2 earn a negative expected payoff, which is decreasing in  $\lambda$ . 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.

**4.2. Clever population against non-clever population,  $\lambda = 1$ .** Consider an extreme case, with all clever agents in population 1. 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 to be clever, have more information about the opponent, is better off than to be non-clever. The answer depends on a game. The following examples illustrate this point. We start from the *Strict Demand* game and afterwards compare that game with the *Nash Demand* game, which was studied in Saez-Marti and Weibull [4].

**The Strict Demand Game.** Consider two finite populations, 1 and 2, who periodically bargain pair-wise over their shares of a common pie. Let  $x$  denote the

share that player 1 gets, and let  $y$  denote the share that player 2 gets. Suppose that all agents in population 1 have the same concave, increasing, and differentiable utility function, which is a function of the share  $x$

$$u : [0, 1] \rightarrow \mathbb{R}, \quad (7)$$

and all agents in population 2 have the same concave, increasing, and differentiable utility function as a function of the share  $y$

$$v : [0, 1] \rightarrow \mathbb{R}. \quad (8)$$

Without loss of generality we can normalize  $u$  and  $v$  so that  $u(0) = v(0) = 0$ .

In each period  $t = 1, 2, \dots$ , 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 Figure 5.

Demands	Outcomes	Payoffs
$x + y = 1$	$x, y$	$u(x), v(y)$
$x + y \neq 1$	$0, 0$	$0, 0$

FIGURE 5

To keep the state space finite, we shall discretize demands. Let a finite set  $D(\delta) = \{\delta, 2\delta, \dots, 1 - \delta\}$  be the space of demands. Furthermore, let  $R^{m,s,1,\varepsilon}$  be an adaptive play with all clever agents in population 1. Let  $(x^t, y^t)$  denote the amounts demanded by the agents in population 1 and population 2 respectively in period  $t$ . At the end of period  $t$ , the state is

$$h^t = ((x^{t-m+1}, y^{t-m+1}), \dots, (x^t, y^t)). \quad (9)$$

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  $h^t$ . Simultaneously and independently, the agent in population 2 draws a sample of size  $s$  also from the  $x$ -values in  $h^t$ .

A *conventional division* is a state of the form

$$h_x = ((x, 1 - x), \dots, (x, 1 - x)), \quad (10)$$

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.

**Definition 5.** *The basin of attraction of state  $h$  is the set of states  $h'$  such that there is a positive probability of moving in a finite number of periods from  $h'$  to  $h$  under the unperturbed process  $R^{m,s,1,0}$ .*

For every real number  $r$  let  $\lceil r \rceil$  denote the least integer greater than or equal to  $r$ .

**Lemma 2.** *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  $\lceil sr_\delta(x) \rceil$ , where*

$$r_\delta(x) = \frac{1}{1 + \frac{v(1-\delta)}{v(1-x)}}.$$

**Proof:** See the Appendix.

**Lemma 3.** *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 1.** *The division  $(\delta, 1 - \delta)$  is stochastically stable.*

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

The main result of this part follows immediately from the corollary.

**Proposition 1.** *Assume that all agents in population 1 are clever. Then for every  $\delta > 0$  there exists one stable division, and this division converges to  $(x, y) = (0, 1)$  as  $\delta \rightarrow 0$ .*

**The Nash Demand Game.** In this subsection we refresh an example which was analyzed by Saez-Marti and Weibull [4]: the *Nash Demand Game*. The set up is the same as in the case of the Strict Demand Game, but now players will also get a demanding part of the whole pie, if their common demand is less than 1. For the case  $\lambda = 1$ , Saez-Marti and Weibull [4] show that *for any precision  $\delta > 0$  there exist at least one and at most two stable divisions, and as  $\delta \rightarrow 0$  they converge to  $(x, y) = (1, 0)$ .*

5.  $2 \times 2$  COORDINATION GAMES AND ASYMMETRIC SAMPLING

So far, we have assumed that agents have the same amount of information in a sense of having the same sample size. What happens if the two populations differ in sample sizes? In this section we analyze this question for  $2 \times 2$  *Coordination Games* by assuming that the agents in population 1 have sample size  $s$  and that the agents in population 2 have sample size  $s'$ . Let the error rate,  $\varepsilon$ , be the same for all agents, and let the memory,  $m$  be the same for both populations and satisfy  $m \geq \max\{2s, 2s'\}$ . Taken together, these factors define a Markov process  $R^{m,s,s',\lambda,\varepsilon}$  with memory  $m$ , sample sizes  $s$  and  $s'$ , the share  $\lambda$  of clever agents in population 1, and error rate  $\varepsilon$ . We denote any game in this class of  $2 \times 2$  games by  $G$ .

It was a lot of studies recently which analyzed the question: which equilibrium would be observed in  $2 \times 2$  coordination games in the long-run. Two papers, Kandori, Mailath, and Rob [3] and Young [7], pioneer this topic. We will compare our results with predictions from the existing literature. Young [7] shows that the stochastically stable states of the perturbed process  $R^{m,s,s,0,\varepsilon}$  correspond one to one with the risk-dominant conventions.

Consider a two-player coordination game  $G$  with payoff matrix in Figure 6.

	$a$	$b$
$A$	$u_{Aa}, v_{Aa}$	$u_{Ab}, v_{Ab}$
$B$	$u_{Ba}, v_{Ba}$	$u_{Bb}, v_{Bb}$

FIGURE 6

$G$  is a *Coordination Game* with pure strategy Nash equilibria  $(A, a)$  and  $(B, b)$  if the following inequalities hold:

$$u_{Aa} > u_{Ba}, \quad u_{Bb} > u_{Ab}, \quad v_{Aa} > v_{Ab}, \quad v_{Bb} > v_{Ba}. \quad (11)$$

The following concept from Young [7] will play a crucial role in our study of  $2 \times 2$  coordination games.

**Definition 6.** A **convention** is a state of the form  $h_x = (x, x, \dots, x)$ , where  $x$  is a strict Nash equilibrium of  $G$ .

**Definition 7.** (Harsanyi and Selten [2]) Equilibrium  $(A, a)$  is **risk-dominant** if

$$(u_{Aa} - u_{Ba})(v_{Aa} - v_{Ab}) \geq (u_{Bb} - u_{Ab})(v_{Bb} - v_{Ba}).$$

We say that a convention  $h_x$  is *risk-dominant* if the strict Nash equilibrium  $x$  is risk-dominant. It turns out that the risk-dominant convention is stochastically stable in Young's [9] framework if  $s = s'$ . The following proposition is a corollary of Theorems 1 and 3.

**Proposition 2.** *If  $\max\{s, s'\} \leq m/2$ , then from any initial state, the unperturbed process  $R^{m,s,s',\lambda,0}$  converges with probability one to a convention and locks in.*

*If  $\lambda \in (0, 1)$  and  $s = s'$ , then Young's [9] result holds. More precisely, if  $s/m \leq 1/2$ , and  $s$  and  $m$  are sufficiently large, the stochastically stable states of the perturbed process  $R^{m,s,s',\lambda,\varepsilon}$ , as  $\varepsilon \rightarrow 0$ , correspond one to one with risk-dominant conventions.*

Let us examine what happens if the first and the second populations have different sample sizes.

**Proposition 3.** *Suppose that  $\lambda \in (0, 1)$ ,  $s < s'$ ,  $s'/m \leq 1/2$ , and  $s$ ,  $s'$  and  $m$  are sufficiently large. Then the stochastically stable states of the perturbed process  $R^{m,s,s',\lambda,\varepsilon}$ , as  $\varepsilon \rightarrow 0$ , correspond one to one with risk-dominant conventions.*

**Proof:** See the Appendix.

The intuition behind this result is that if the sample size of population 1 is less than the sample size of population 2, then agents in population 1 need less mistakes to switch to another strategy. It means that the role of the first population becomes crucial for finding the minimum stochastic potential. In population 1, all agents have the same sample size and clever agents, as before, play a best reply to a best reply of an agent from population 2. In a  $2 \times 2$  coordination game a clever agent in population 1 chooses the same strategy as an agent in population 2 does with positive probability. So a clever agent needs the same number of mistakes to switch to another strategy as a non-clever agent in population 2 does if they have the same sample sizes. Hence we are back to the Young's [7] framework.

Define

$$\alpha = (u_{Aa} - u_{Ba}) / (u_{Aa} - u_{Ab} - u_{Ba} + u_{Bb}) \quad (12)$$

and

$$\beta = (v_{Aa} - v_{Ab}) / (v_{Aa} - v_{Ab} - v_{Ba} + v_{Bb}). \quad (13)$$

**Proposition 4.** *Suppose that  $\lambda \in (0, 1)$ ,  $s > s'$ ,  $s/m \leq 1/2$ , and  $s$ ,  $s'$  and  $m$  are sufficiently large, and  $\varepsilon$  is sufficiently small. Then the state  $h_{(A,a)}$  of the perturbed process  $R^{m,s,s',\lambda,\varepsilon}$  is stochastically stable if and only if*

$$\min\{\lceil \alpha s \rceil, \lceil \beta s' \rceil\} \geq \min\{\lceil (1 - \alpha)s \rceil, \lceil (1 - \beta)s' \rceil\}. \quad (14)$$

**Proof:** See the Appendix.

What does proposition 3 say if the sample size  $s \gg s'$ ? According to the proposition, the state  $h_{(A,a)}$  is stochastically stable if and only if inequality (14) holds. For sufficiently large  $s'$ , the inequality becomes  $\beta \geq 1 - \beta$ . This means that only the pay-offs of the agents in population 2 matter. So, if the sample size of the first population is “much larger” than the sample size of the second population then risk-dominant convention for population 2 will be stochastically stable.

**Definition 8.** *The convention  $h_{(A,a)}$  is risk-dominant for population 2 if:*

$$v_{Aa} - v_{Ab} \geq v_{Bb} - v_{Ba}. \quad (15)$$

*If inequality (15) is reversed, then convention  $h_{(B,b)}$  is risk-dominant for population 2.*

Observe that (15) is equivalent to

$$\beta \geq 1 - \beta. \quad (16)$$

The following proposition considers the extreme case when population 1 consists entirely of clever agents. It turns out that there is a discontinuity at this end of the spectrum: the whole population will move to risk-dominant equilibrium of the population without clever agents.

**Proposition 5.** *Assume that every agent in population 1 is clever. If  $\max\{s, s'\} \leq m/2$ , and  $s, s'$  and  $m$  are sufficiently large, the stochastically stable states of the perturbed process  $R^{m,s,s',1,\varepsilon}$ , as  $\varepsilon \rightarrow 0$ , correspond one to one with the risk-dominant convention(s) of the population without clever agents.*

**Proof:** See the Appendix.

**Corollary 2.** *Consider a two-player symmetric coordination game,  $G^S$ , with the following payoff matrix in Figure 7*

$$\begin{array}{c} \begin{array}{cc} & a & b \\ A & \boxed{u, u} & \boxed{c, d} \\ B & \boxed{d, c} & \boxed{v, v} \end{array} \end{array} \quad (17)$$

FIGURE 7



where  $u > d$ ,  $v > c$ . Find  $\alpha$  and  $\beta$  from (12) and (13):

$$\alpha = \frac{u - d}{u - d - c + v} = \beta. \quad (18)$$

This means that the stochastically stable state(s) of the perturbed process  $R^{m,s,s',\lambda,\varepsilon}$  correspond one to one with the risk-dominant convention(s), for any share of clever agents in population 1.

**Corollary 3.** If  $d = c$  in (17), then a two-player symmetric coordination game becomes a two-player doubly symmetric coordination game.<sup>4</sup> It follows straightforward that the stochastically stable state(s) of the perturbed process  $R^{m,s,s',\lambda,\varepsilon}$  correspond one to one with the Pareto dominant convention(s), for any share of clever agents in population 1.<sup>5</sup>

The results for two-player  $2 \times 2$  coordination games are summarized in Figure 8, showing the stochastically stable states for different  $\lambda$ ,  $s$ , and  $s'$ .

	$\lambda = 0$	$\lambda \in (0, 1)$	$\lambda = 1$
$s = s'$	<i>risk-dominant</i>	<i>risk-dominant</i>	<i>risk-dominant(2)</i>
$s > s'$	*	*	<i>risk-dominant(2)</i>
$s < s'$	*	<i>risk-dominant</i>	<i>risk-dominant(2)</i>

FIGURE 8

Here

1. *risk-dominant* means that the outcome is *risk-dominant* convention.
2. \* means that the convention  $h_{(A,a)}$  is stochastically stable if inequality (14) holds. If inequality (14) is reversed, then the convention  $h_{(B,b)}$  is stochastically stable.
3. *risk-dominant(2)* means that the outcome is a *risk-dominant* convention for population 2.

<sup>4</sup>A symmetric two-player game is doubly symmetric if  $A^T = A$ , where  $A$  is the symmetric payoff matrix. See also Weibull [6].

<sup>5</sup>We say that convention  $h_{(A,a)}$  is Pareto dominant if  $u \geq v$ , and convention  $h_{(B,b)}$  is Pareto dominant if the reversed inequality holds.

The question now arises whether the population with more information about the opponent, population 1, is better off. The answer depends on the structure of the game.

Consider the following games in Figure 9, and the process  $R^{m,s,s',\lambda,\varepsilon}$  with  $s = s'$ .

	$a$	$b$	
$A$	3, 2	0, 0	
$B$	0, 0	1, 4	
	Game 1		

	$a$	$b$
$A$	3, 4	4, 3
$B$	0, 0	5, 2
	Game 2	

FIGURE 9

Then  $(A, a)$  is a risk-dominant equilibrium and  $h_{(B,b)}$  is a risk-dominant convention for population 2 in both games. Hence if  $\lambda \in [0, 1)$  then convention  $h_{(A,a)}$  is stochastically stable (Proposition 2), otherwise if  $\lambda = 1$  then convention  $h_{(B,b)}$  is stochastically stable (Proposition 5). Cleverness (or  $\lambda = 1$ ) is a disadvantage in Game 1 (the clever population switches from 3 in  $h_{(A,a)}$  to 1 in  $h_{(B,b)}$ ) and an advantage in Game 2 (the clever population switches from 3 in  $h_{(A,a)}$  to 5 in  $h_{(B,b)}$ ).

## 6. CONCLUDING REMARKS

In this paper we answer some questions which were posted in Saez-Marti and Weibull [4]. They study the consequences of letting some agents play a myopic best reply to the myopic best reply in Young's [8] bargaining model. This is how they introduce "cleverness" of players. Saez-Marti and Weibull [4] ask whether their results can be generalized. We use the "cleverness" approach from their paper to analyze *generic* two-player games in Young's [9] set-up. The resulting Markov process is denoted adaptive play with clever agents.

Saez-Marti and Weibull [4] prove that an introduction of any share of clever agents less than one will not change the long-run behavior in the special case, for the Nash Demand Game. We have shown that this result is 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 of 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 to 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 can be different from the previous ones. Saez-Marti and Weibull [4] find this discontinuity for the Nash Demand Game.

We also study whether it is advantageous to be a member of the population consisting of only clever agents, and show that the answer is ambiguous even in coordination games. On the one hand, Saez-Marti and Weibull [4] 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 in sum exactly the size of the pie, otherwise they both get nothing, the population without clever agents obtains the whole pie. Hence "cleverness" does not guarantee an advantage in general coordination games even if all agents in one of the populations are clever.

We analyze properties of  $2 \times 2$  coordination games when agents have asymmetric information. It is a classical result (see Young [7] and Kandori, Mailath, and Rob [3] - the first, among others) for the symmetric setting that in the limiting case, when the mutation rate goes to zero, the *risk-dominant* equilibrium will be observed as the long-run outcome in such games. We prove that if the sample sizes are different in the two populations, then the long-run outcome may differ from this prediction.

### Appendix.

**Proof of Theorem 1:** This proof is similar to the proof in [9, *Theorem 7.2*]. We need only to show that clever agents will not change anything.

Given a positive integer  $s$ , we say that the probability distribution  $p_i \in \Delta_i$  has *precision*  $s$  if  $sp_i$  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  $Y_i \subset X_i$ , let  $\Delta_i^s(Y_i)$  denote the set of distributions  $p_i \in \Delta_i^s$  such that  $p_i(x_i) > 0$  implies  $x_i \in Y_i$ . For each positive integer  $s$ , let  $BR_i^s(X_{-i})$  be the set of pure-strategy best replies by a non-clever agent in population  $i$  to some product distribution  $p_{-i} \in \Delta_{-i}^s(X_{-i}) = \Delta_j^s(X_j)$ , where  $j \neq i$ . Similarly,  $BR_i^s(Y_{-i})$  denotes the set of all best replies by a non-clever agent in population  $i$  to some product distribution  $p_{-i} \in \Delta_{-i}^s(Y_{-i})$ . Let  $BR_1^s(X_1)$  be the set of pure-strategy best replies by a clever agent in population 1 to some product distribution  $p_1 \in \Delta_1^s(X_1)$ .

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

$$\beta_i(Y) = Y_i \cup BR_i(Y_{-i}) \text{ and } \beta'_1(Y) = Y_1 \cup BR_1(Y_1), \quad (19)$$

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  $\beta(Y) = [\beta_1(Y) \cup BR_1(Y_1)] \times \beta_2(Y)$ . Similarly, for each integer  $s \geq 1$  let

$$\beta_i^s(Y) = Y_i \cup BR_i^s(Y_{-i}), \beta_1^{s'}(Y) = Y_1 \cup BR_1^s(Y_1) \quad (20)$$

and

$$\beta^s(Y) = [\beta_1^s(Y) \cup \beta_1'^s(Y)] \times \beta_2^s(Y). \quad (21)$$

In the same way, as it is in the proof in [9, *Theorem7.2*], we can show that  $\beta^s(Y) = \beta(Y)$  for all sufficiently large  $s$ .

Consider the process  $R^{m,s,\lambda,0}$ . We shall 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  $R^{m,s,\lambda,0}$ , 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 [9, *Theorem7.2*], there is a positive probability of reaching a state  $h^1$  in which 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

$$\{x^*\} \subseteq \beta(\{x^*\}) \subseteq \beta^{(2)}(\{x^*\}) \subseteq \dots \subseteq \beta^{(j)}(\{x^*\}) \subseteq \dots \quad (22)$$

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

$$\beta^{(j)}(\{x^*\}) = \beta^{(j+1)}(\{x^*\}) = Y^*. \quad (23)$$

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 [9, *Theorem7.2*]. **End of proof.**

**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 have to 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 influence behavior only the non-clever agents in population 1. It means that all mistakes made in population 2 have the same effect for the both processes  $R^{m,s,\lambda,\varepsilon}$  and  $R^{m,s,0,\varepsilon}$ .

Suppose that  $l$  mistakes in a row in population 1 are necessarily 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 that. Hence, if it was less than  $l$  mistakes in population 1, then

the clever agent in role 1 expects an agent in role 2 plays as if they are in recurrent class  $E_p$ . It has to be at least  $l$  mistakes in population 1 to change these expectations of the clever agent. It means that it has to 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 the 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  $E_q$ .

**End of proof.**

**Proof of Lemma 2:** 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  $\mathbf{w}$  such that some best reply of agent in population 2 to a sample from  $\mathbf{w}$  is different from  $1 - x$ . Let  $\mathbf{w}$  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  $1 - x'$ . The number of mistakes in population 2 does not matter in this setting, because both agents look only at the  $x$ -values in  $h^t$ .

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

$$j = \frac{1}{1 + \frac{v(1-\delta)}{v(1-x)}}s.$$

Hence, the least 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 2.

**Proof of Propositions 3 and 4:** It is straightforward to show that the resistances to transiting between the absorbing states  $h_{(A,a)}$  and  $h_{(B,b)}$  are

$$r_{Aa \rightarrow Bb}^{s,s'} = \min \{ \lceil \alpha s \rceil, \lceil \beta s' \rceil, \lceil \beta s \rceil \} \quad (24)$$

and

$$r_{Bb \rightarrow Aa}^{s,s'} = \min \{ \lceil (1 - \alpha)s \rceil, \lceil (1 - \beta)s' \rceil, \lceil (1 - \beta)s \rceil \}, \quad (25)$$

where  $\alpha$  and  $\beta$  are from (12) and (13).

The last terms  $\lceil \beta s \rceil$  and  $\lceil (1 - \beta)s \rceil$  appear because of the clever agents in population 1, who have sample size  $s$  and imitate the behavior of an agent in population 2 in  $2 \times 2$  coordination game.

By Theorem 2,  $h_{(A,a)}$  is stochastically stable if and only if

$$r_{Aa \rightarrow Bb}^{s,s'} \geq r_{Bb \rightarrow Aa}^{s,s'}. \quad (26)$$

Let  $s > s'$ , then we have

$$r_{Aa \rightarrow Bb}^{s,s'} = \min \{ \lceil \alpha s \rceil, \lceil \beta s' \rceil, \lceil \beta s \rceil \} = \min \{ \lceil \alpha s \rceil, \lceil \beta s' \rceil \} \quad (27)$$

and

$$r_{Bb \rightarrow Aa}^{s,s'} = \min \{ \lceil (1 - \alpha)s \rceil, \lceil (1 - \beta)s' \rceil, \lceil (1 - \beta)s \rceil \} = \min \{ \lceil (1 - \alpha)s \rceil, \lceil (1 - \beta)s' \rceil \}, \quad (28)$$

which is exactly the statement of proposition 4.

If  $s < s'$ , then for sufficiently large  $s$

$$r_{Aa \rightarrow Bb}^{s,s'} = \min \{ \lceil \alpha s \rceil, \lceil \beta s' \rceil, \lceil \beta s \rceil \} = \min \{ \lceil \alpha s \rceil, \lceil \beta s \rceil \} = s \min \{ \alpha, \beta \} \quad (29)$$

and

$$r_{Bb \rightarrow Aa}^{s,s'} = \min \{ \lceil (1 - \alpha)s \rceil, \lceil (1 - \beta)s' \rceil, \lceil (1 - \beta)s \rceil \} = s \min \{ (1 - \alpha), (1 - \beta) \}. \quad (30)$$

What does it mean that

$$\min \{\alpha, \beta\} \geq \min \{(1 - \alpha), (1 - \beta)\} \quad (31)$$

from a “risk-dominance” point of view? Suppose that

$$\min \{\alpha, \beta\} = \alpha. \quad (32)$$

Hence

$$\min \{(1 - \alpha), (1 - \beta)\} = 1 - \beta. \quad (33)$$

Find the domain where (31), (32), (33) and  $0 \leq \alpha, \beta \leq 1$  hold. We have the following inequalities:

$$\alpha + \beta \geq 1 \quad (34)$$

and

$$0 \leq \alpha \leq \beta \leq 1. \quad (35)$$

Suppose now that

$$\min \{\alpha, \beta\} = \beta. \quad (36)$$

Hence

$$\min \{(1 - \alpha), (1 - \beta)\} = 1 - \alpha. \quad (37)$$

Find the domain where (31), (36), (37) and  $0 \leq \alpha, \beta \leq 1$  hold. We get the following inequalities:

$$\alpha + \beta \geq 1 \quad (38)$$

and

$$0 \leq \beta \leq \alpha \leq 1. \quad (39)$$

Hence the inequality

$$\min \{\alpha, \beta\} \geq \min \{(1 - \alpha), (1 - \beta)\} \quad (40)$$

is equivalent to the inequalities

$$\alpha + \beta \geq 1, \quad 0 \leq \alpha, \beta \leq 1. \quad (41)$$

By definition, equilibrium  $(A, a)$  is risk-dominant if

$$\alpha\beta \geq (1 - \alpha)(1 - \beta), 0 \leq \alpha, \beta \leq 1, \quad (42)$$

or

$$\alpha + \beta \geq 1, 0 \leq \alpha, \beta \leq 1, \quad (43)$$

which is exactly inequalities (41), and the statement of proposition 3 follows immediately. **End of proof.**

**Proof of proposition 5:** As in the proof of the previous proposition, it is straightforward to show that the resistances of transitions between the absorbing states  $h_{(A,a)}$  and  $h_{(B,b)}$  for sufficiently large  $s$  and  $s'$  are

$$r_{Aa \rightarrow Bb}^{s,s'} = \min \{ \lceil \beta s' \rceil, \lceil \beta s \rceil \} = \beta \min \{ s', s \} \quad (44)$$

and

$$r_{Bb \rightarrow Aa}^{s,s'} = \min \{ \lceil (1 - \beta)s' \rceil, \lceil (1 - \beta)s \rceil \} = (1 - \beta) \min \{ s', s \}. \quad (45)$$

Note that there are only the clever agents in population 1 and therefore the sample of population 2 is irrelevant.

The convention  $h_{(A,a)}$  is stochastically stable if and only if

$$r_{Aa \rightarrow Bb}^{s,s'} \geq r_{Bb \rightarrow Aa}^{s,s'}, \quad (46)$$

or

$$\beta \geq 1 - \beta. \quad (47)$$

This means, see (16), that the convention  $h_{(A,a)}$  is risk-dominant for population 2.

The same logic can be used if the convention  $h_{(B,b)}$  is stochastically stable. Hence, the stochastically stable state(s) of the perturbed process  $R^{m,s,s',1,\varepsilon}$  correspond one to one with the risk-dominant convention(s) of population 2. **End of proof.**



## REFERENCES

- [1] K. Basu and J. Weibull, Strategy subsets closed under rational behavior, *Economics Letters* **36** (1991), 141-146.
- [2] J. Harsanyi and R. Selten, "A General Theory of Equilibrium in Games", Cambridge, MIT Press, 1988.
- [3] M. Kandori, G. Mailath, and R. Rob, Learning, mutation, and long run equilibria in games, *Econometrica* **61** (1993), 29-56.
- [4] M. Saez-Marti and J. Weibull, Clever agents in Young's evolutionary bargaining model, *J. Econ. Theory* **86** (1999), 268-279.
- [5] L. Samuelson, "Evolutionary Games and Equilibrium Selection", Massachusetts, MIT Press, 1997.
- [6] W. Weibull, "Evolutionary Game Theory", Massachusetts, MIT Press, 1995.
- [7] P. Young, The evolution of conventions, *Econometrica* **61** (1993), 57-84.
- [8] P. Young, An evolutionary model of bargaining, *J. Econ. Theory* **59** (1993), 145-168.
- [9] P. Young, "Individual Strategy and Social Structure", Princeton University Press, 1998.