On the Evolution of Imitative Behavior[1]

Jonas Björnerstedt[2] Karl H. Schlag[3]

Projektbereich B
Discussion Paper No. **B-378**
Juli, 1996

[1]We wish to thank Georg Nöldeke and Jörgen Weibull for helpful comments. Financial support from the Söderberg and Wallander foundations, from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn and from the Industriens Utrednings Institut, Stockholm is gratefully acknowledged.

[2]Department of Economics, Stockholm University, 10691 Stockholm, Sweden, e-mail: JB@ne.su.se - currently visiting Harvard University, e-mail: bjorners@husc.harvard.edu

[3]Abt. Wirtschaftstheorie III, Department of Economics, University of Bonn, Adenauerallee 24-26, 53113 Bonn, Germany, e-mail: schlag@glider.econ3.uni-bonn.de

Abstract

We analyze the evolution of behavioral rules for learning how to play a two-armed bandit. Individuals have no information about the underlying payoff distributions and have limited memory about their own past experience. Instead they must rely on information obtained through observing the performance of other individuals. Evolution is modelled using the replicator dynamic with the revision behaviors as replicators.

We find that evolution favors a special class of imitative rules. These so-called strictly improving rules, that also play an important role in a bounded rational selection approach (Schlag [16]), are found to be neutrally stable when facing any two-armed bandit.

JEL classification numbers: C72, C79.

Keywords: multi-armed bandit, social learning, payoff increasing, improving, proportional imitation rule, replicator dynamic, neutral stability, survival.

1 Introduction

In every day life we are constantly making decisions. Many of our decisions are based on minimal information or on very limited previous experience. Changing environments make it difficult to rely on the previous experience that we do have. Computational abilities and memory capacities of human beings are limited. Consequently there are many situations in which a decision maker will follow a simple behavioral rule.

One type of simple behavior that immediately comes to mind is *imitation*, the act of copying or mimicking the actions of others. Its undeniable presence among behavioral traits of human beings is perhaps the reason why lately, with the growing popularity of Evolutionary Game Theory, many models of social learning have emerged in which individuals select their future actions by imitating others. Most of these models postulate that individuals follow an imitative behavior and concentrate on the implications this has on population behavior (e.g., Björnerstedt and Weibull [2]; Cabrales [3]; Ellison and Fudenberg [5]; Gale et al. [6]; Hofbauer [7], Novak and May [11]), only few models attempt to explain the emergence of imitative behavior first before analyzing its implications (Banerjee [1]; Rogers [12]). In our investigation we follow the later approach.

Concerning the analysis of simple rules, current theory is hard pressed to say what rational behavior would be when information and memory is extremely limited. This motivates our pursuit of an evolutionary approach where such theoretical problems do not arise. In this paper we compare the performance of a wide variety of simple behavioral rules in an evolutionary framework. Our study is very closely related to the model of Schlag [16] who approaches the same questions from a bounded rational viewpoint. The main question we wish to answer is: *will evolution select imitative behavior, and if so, what types of imitative behavior will be selected?*

Our basic finding is that evolution favors a specific form of imitative behavior among the set of behavioral rules that are without memory. This is the same class of behavior - called strictly improving - that performs well from the bounded rational viewpoint of Schlag [16]. Although being simple, a strictly improving rule, when used by the entire population, can prevent mutants that are specialized for a specific environment from taking over. This however does not mean that strictly improving rules always take over the entire population; they are not able to drive out any other type of imitative behavior. Moreover, their imitative nature prevents them from being able to successfully introduce a superior action (e.g., technology) in a society in which the rest of the individuals stubbornly (unknowingly) adhere to an inferior action.

Going more into detail, our investigation takes place in the classical *two-armed bandit* setting (Rothschild [13]). An individual must repeatedly choose one of two actions, each action yielding an uncertain payoff with an unknown stationary distribution. This could be the model of an individual choosing a restaurant or of a manager setting the price of a good. A priori it is assumed that individuals do not know anything about the distribution of the payoffs in the two-armed bandit they face. Instead they must rely on information they receive about the performance of other individuals that are facing the same situation. This occurs according to the following social learning scenario: before making his choice, an individual observes the previous action and payoff achieved by some other (randomly selected) individual.

Which action the individual chooses as a function of his previous experience and observations is called his *revision behavior* or rule. Following our previous motivation and to keep the model simple, we assume that an individual has no recall about how he came to choose his present action. Consequently, a revision behavior is a function of the previous action and previously achieved payoff (if the individual had previous experience) and of the action and payoff observed.

We analyze the evolution of behavior in an infinite population of individuals each facing the same two-armed bandit. In each round a small proportion of individuals reproduce and then die. Reproduction is modelled according to the biological model underlying the standard replicator dynamic (Taylor and Jonker [17]); reproductive fitness of an individual (i.e., the number of offspring) is determined by the payoff last achieved; offspring inherit the revision behavior of their parent and enter the population. Thus the individuals that last chose the action that yields the highest expected payoff produce on average the most offspring. In this sense we will refer to a better / a worse action by the action achieving the higher / the lower expected payoff in the two-armed bandit.

Since the individuals have no memory, the success of individual behavior depends on how other individuals in the population behave. For simplicity and thereby in the spirit of classical Evolutionary Game Theory (that started with Maynard Smith and Price [9]), we restrict our attention to populations in which at most two different revision behaviors exist, a *two behavior contest*. We evaluate the success of individual behavior when it is used by the majority of the population and its success when it is adapted by a minority.

An important limit case will be the situation in which every individual uses the same revision behavior. For a given two-armed bandit, we characterize revision behavior - and call it *strictly payoff increasing* - that, when used by everyone, enables eventually everyone to learn which action maximizes expected payoffs. As previously shown by Schlag [16], a revision behavior has this property in any two-armed bandit - such a rule is called *strictly improving* - if and only if it is based on imitation and its switching behavior satisfies a specific linear relationship. For example, *proportional imitation rules* are strictly improving; these rules specify to imitate the action of the sampled individual if he achieved a higher payoff and to do this with a probability proportional to the difference between the observed payoff and one's own payoff. One of the side results of Schlag [16] is that the probably most intuitive behavior, to switch to the observed action if and only if it achieved a higher payoff (e.g., see Ellison and Fudenberg [5]), is **not** strictly improving.

One of the main results of our paper shows that the evolutionary robustness of a behavior used by the majority is directly related to the ability of the rule to learn which action achieves the higher expected payoff when no other rule is around. Along the lines of Maynard Smith [8], we call a revision behavior *neutrally stable* if its offspring constantly constitute the majority of the population after a one time mutation occurs in which a small proportion of individuals adapt an (arbitrary) alternative behavior. We show that a rule is neutrally stable precisely if it is strictly payoff increasing. Especially, this means that even if all the individuals using the strictly payoff increasing rule initially choose the worse action and the mutants know which action is best, the strictly payoff increasing behavior will still be used by the majority of the individuals in the long run. This occurs because the individuals using the strictly payoff increasing behavior learn that the action adapted by the mutants is better before the offspring of the mutants can take over a substantial proportion of the population.

Especially it follows that only a strictly improving rule is neutrally stable in any two-armed bandit. This has consequences for a slightly modified setting. Only a strictly improving rule, when adapted by everyone, is robust against a one time mutation of behavior and to a one time mutation of the payoff structure of the bandit.

However, an evolutionarily stable behavior fails to exist in our model, i.e., there is no revision behavior that, when used by the majority, is able to drive out any alternative behavior. Any imitative behavior survives when the majority is using a strictly payoff increasing rule, this is even true if the majority of the population knows which action is best. This is because the majority of the population teaches the minority which action is better and thus eliminates selection pressure.

This intuition is also the reason why it is much more difficult for a behavior to survive in the long run when the majority does not take over the role of teaching which action is best (i.e., when the majority does not follow a strictly payoff increasing rule). We show that only strictly payoff increasing rules will never be eliminated from the population. Especially this means that 'imitate if better' will be driven out of some two behavior contests.

Interestingly enough, the fact that a strictly payoff increasing imitation rule always survives does not necessarily mean that it is able to teach the population which action is better. When mutants that are playing the better action and that follow an imitation rule enter a population that is stubbornly playing the worse action then no one plays the better action in the long run. Only non-imitative behavior that has a tendency to play the better action even if none of the observed individuals do, can successfully introduce a better action into a population in which there is a large fraction of stubborn play. It is in this sense that sophisticated behavior (i.e., strictly improving behavior) when adopted by a minority can fail to cause an innovation to spread.

We proceed as follows. In Section 2, we define two-armed bandits, revision behaviors over possible actions in a two-armed bandit, and switching probabilities. In Section 3, we consider the dynamic on the shares using the different actions in the two-armed bandit, given that all individuals in the population follow the same revision behavior. Section 4 extends the analysis to two behavior contests; a population with two revision behaviors present, reproduction determining the relative size of each. Neutral stability and survival of a revision behavior in the two behavior dynamic are analyzed. Section 5 considers an additional class of behavioral rules based on experimentation. In Section 6 the results together with the related literature is discussed. Finally, the appendix contains the more lengthy proofs.

2 Choice and Revision Behavior

We will consider individuals who must repeatedly choose one of two actions, denoted by 1 and 2. After an individual chooses an action $i \in \{1, 2\}$, he receives a payoff $p \in \mathbb{R}$, according to the probability density function P_i that has support on the bounded interval $\left[\pi^L, \pi^H\right]$ and that yields the expected value π_i. The payoff from a given action i is realized independently over time and across individuals. The pair $\langle\{1, 2\}, P = (P_1, P_2)\rangle$ constitutes a *two-armed bandit* (Rothschild [13]). The set of all such two-armed bandits will be denoted by $\mathcal{G}\left(\{1, 2\}, \left[\pi^L, \pi^H\right]\right)$. When $\pi_i > \pi_j$ we will sometimes call i the efficient or better action and j the inferior or worse action.

Individuals belong to an infinite population. The following scenario takes place every τ time units, so-called rounds, where $\tau \in (0, 1)$ is fixed. We initiate the process in a state in which each individual has previously chosen an action at least once and ignore how this first state came to being. Before each round, each individual is independently selected with probability τ. With probability $\mu \in (0, 1)$ a selected individual reproduces and then dies; the offspring who inherit the behavior of their parent enter the population. We will refer to μ as the *reproduction rate*. It will not be necessary to introduce the process governing the number of offspring until Section 4. Selected individuals (or the offspring of the ones that died) are each given the opportunity to choose an action (to "pull").

Before choosing an action, the individual is able to gather information about the performance of actions as follows. The individual samples an individual among those that previously chose an action, i.e., among those that did not enter the population since the last round. He learns the payoff and the associated action that the sampled individual obtained at his last pull. It is assumed that sampling occurs at random, i.e., the probability of observing an individual using a given action who achieved a payoff in a given range is equal to the associated population share of such individuals. This process of obtaining information will be called *random sampling*.

The action the individual chooses after he samples as a function of the information from this sample and from previous observations is called his *revision behavior* or *rule*. We assume that the experience of the individual prior to his last pull does not influence his future behavior, a restriction that can result from limited memory capabilities. Formally, a revision behavior is therefore a function

$$X : \left\{\{\emptyset\} \cup \left(\{1,2\} \times \left[\pi^L, \pi^H\right]\right)\right\} \times \{1,2\} \times \left[\pi^L, \pi^H\right] \to \Delta\left(\{1,2\}\right)$$

where $\Delta\left(\{1,2\}\right)$ is the set of probability distributions on $\{1,2\}$, $X_k\left(i,x,j,y\right)$ is the probability of choosing action k when he last chose action i that yielded payoff x and just sampled an individual who chose action j and received payoff y; $X_k\left(\emptyset,j,y\right)$ is the corresponding probability for an individual that newly entered the population and hence has no previous experience. The realizations of X are assumed to be independent of the randomizations of the other individuals, independent of the probability distributions P_i, and also independent of time.

For a given two-armed bandit $\langle\{1,2\}, P\rangle$ in $\mathcal{G}\left(\{1,2\}, \left[\pi^L, \pi^H\right]\right)$ and a given triplet $i,j,k \in \{1,2\}$, a revision behavior X induces the *switching probabilities* $F_{\emptyset j}^k$ and F_{ij}^k; $F_{\emptyset j}^k$ is the probability that the individual will switch to action k after he samples action j when he newly entered the population and F_{ij}^k is the corresponding probability for when the individual has previous experience and last chose action i, both calculated a priori to the realization of the payoffs. Formally,

$$\begin{aligned} F_{\emptyset j}^k &:= \int_{\pi^L}^{\pi^H} X_k(\emptyset, j, q) P_j(q) \, dq \\ F_{ij}^k &:= \int_{\pi^L}^{\pi^H} \int_{\pi^L}^{\pi^H} X_k(i, p, j, q) P_i(p) P_j(q) \, dp dq. \end{aligned}$$

Notice that this probability is a function of the two-armed bandit and that it is independent of the population shares playing the different actions.

2.1 Examples

Definition 1 *We call a revision behavior X* ***stubborn*** *if the induced choice of an action is independent of the observations, i.e., if there exists $u \in \Delta(\{1,2\})$ such that $X_k(\emptyset, j, \cdot) = X_k(i, \cdot, j, \cdot) = u_k$ for all $i, j, k \in \{1,2\}$. We will say that X stubbornly chooses action i if $u_i = 1$.*

A more versatile class of revision behaviors are those we call imitating. Using an imitating behavior the individual will never switch to an action that he did not just observe, formally,

Definition 2 *a revision behavior X is called* ***imitating*** *if $\forall i, j, k \in \{1,2\}$,*

$$\begin{aligned} X_k(\emptyset, k, \cdot) = 1 \text{ and} \\ X_k(i, \cdot, j, \cdot) = 0 \text{ if } k \notin \{i, j\}. \end{aligned}$$

Due to the different ways in which newly born choose their first action, there is no revision behavior that is both stubborn and imitating. In the following we present some examples of imitating revision behaviors.

- *Reciprocal switching rules* are imitating revision behaviors where

$$X_2(1, p, 2, q) = X_1(2, q, 1, p) \text{ for } p, q \in \left[\pi^L, \pi^H\right].$$

Special cases are the rule *'never switch'* where $X_i(i,\cdot,\cdot,\cdot) = 1$ for $i \in \{1,2\}$ and the rule *'always switch'* (or copy-cat) where $X_j(\cdot,\cdot,j,\cdot) = 1$ for all $j \in \{1,2\}$.

- The *proportional imitation rules* are the imitating revision behaviors that satisfy
$$X_j(i,p,j,q) = \sigma \cdot \max(q-p,0)$$
for any $i,j \in \{1,2\}$, $i \neq j$ and some fixed $0 < \sigma \leq \frac{1}{\pi^H - \pi^L}$. The proportional imitation rule with $\sigma = \frac{1}{\pi^H - \pi^L}$ is the optimal rule in a the closely related model of Schlag [16].

- *'Imitate if better'* is the imitating revision behavior that satisfies
$$X_j(i,p,j,q) = \begin{cases} 1 & q > p \\ 0 & q \leq p \end{cases}$$
for $i \neq j$.

The switching probabilities of a given revision behavior will generally depend on the two-armed bandit at hand. Moreover, in a given two-armed bandit many revision behaviors can induce the same switching probabilities. Calculating the set of revision behaviors that correspond to given switching probabilities is generally difficult. In section 3.2 some general relationships between revision behaviors and switching probabilities will be presented.

3 Adaptive Dynamics of a Single Behavior Population

In this section we analyze how the shares of actions change over time in a single behavior population, a population in which each individual is using the same revision behavior. Since offspring inherit the behavior of their parent, we can analyze the dynamics in single behavior populations without specifying the actual process of reproduction. Let X denote the underlying revision behavior, let F be the associated switching probabilities. The population state at any given time is characterized by the proportion of individuals $x_i = x_i(t)$ using the action i $(i \in \{1,2\})$; hence $x_1, x_2 \geq 0$ and $x_1 + x_2 = 1$. We identify the vector of population shares $x = (x_1, x_2)$ with the probability distribution $x \in \Delta(\{1,2\})$ associated to randomly selecting an individual from the population. According to the scenario given in Section 2 the population shares change over time according to the following adjustment process:

$$\begin{aligned} x_k(t+\tau) &= x_k(t) + \tau(1-\mu) \sum_{i,j\in\{1,2\}} \left[F_{ij}^k x_i(t) x_j(t) - F_{kj}^i x_k(t) x_j(t)\right] \quad (1) \\ &\quad + \tau\mu \sum_{i\in\{1,2\}} \left[x_i(t) F_{\emptyset i}^k - x_k(t) F_{\emptyset k}^i\right], \end{aligned}$$

where $t = r\tau$ for some $r \in \mathbb{N}$ and $k \in \{1,2\}$. In order to simplify our analysis we take the limit in (1) as τ goes to 0 and obtain the following differential equation:

$$\dot{x}_k = (1-\mu) \sum_{i,j\in\{1,2\}} \left[F_{ij}^k x_i x_j - F_{kj}^i x_k x_j\right] + \mu \sum_{i\in\{1,2\}} \left[x_i F_{\emptyset i}^k - x_k F_{\emptyset k}^i\right], k \in \{1,2\}. \quad (2)$$

3.1 The 'Payoff Increasing' Condition

We will now define some properties of switching probabilities and then characterize which switching probabilities satisfy them. The properties relate to the change of the average payoff in a population where each individual uses a revision behavior that induces the given switching probabilities. The average payoff $\bar{\pi} = \bar{\pi}(x)$ in a given population state $x \in \Delta(\{1,2\})$ is given by $\bar{\pi} = x_1\pi_1 + x_2\pi_2$.

Definition 3 *The switching probabilities F are called* ***payoff increasing*** *if, for any $x \in \Delta(\{1,2\})$, the average payoff in a monomorphic population does not decrease over time, i.e., $\frac{d}{dt}\bar{\pi} = \dot{x}_1\pi_1 + \dot{x}_2\pi_2 \geq 0$ where $\dot{x}_k$ defined in (2). A revision behavior X is called* ***payoff increasing*** *if it induces payoff increasing switching probabilities.*

Definition 4 *The switching probabilities F are called* ***strictly payoff increasing*** *if for any $x \in \Delta(\{1,2\})$ with $0 < x_1 < 1$ and if $\pi_1 \neq \pi_2$ then the average payoff in the population strictly increases over time, i.e., $\frac{d}{dt}\bar{\pi} = \dot{x}_1\pi_1 + \dot{x}_2\pi_2 > 0$.*

Notice that every behavioral rule is strictly payoff increasing when $\pi_1 = \pi_2$. The following proposition states that an individual induces switching probabilities that are payoff increasing in two-armed bandits precisely when he does not switch "without reason" away from the efficient action and is more likely to switch to the efficient action when it is sampled than vice versa.

Proposition 1 *In a two-armed bandit $\langle\{1,2\}, P\rangle$, the switching probabilities F are* payoff increasing *if and only if*

$$\pi_i > \pi_j \Rightarrow F^j_{\emptyset i} = F^j_{ii} = 0 \text{ and } F^i_{ji} \geq F^j_{ij} \ , \tag{3}$$

the switching probabilities F are strictly payoff increasing *if and only if they are payoff increasing and*

$$\pi_i > \pi_j \Rightarrow F^i_{\emptyset j} > 0 \text{ or } F^i_{jj} > 0 \text{ or } F^i_{ji} > F^j_{ij} \ . \tag{4}$$

Proof. The statement follows directly from

$$\frac{d}{dt}\bar{\pi} = \left[\begin{array}{c} (1-\mu)\left(F^1_{22}(x_2)^2 + (F^1_{21} - F^2_{12})\,x_1x_2 - F^2_{11}(x_1)^2\right) \\ +\mu\left(F^1_{\emptyset 2}x_2 - F^2_{\emptyset 1}x_1\right) \end{array} \right] \cdot (\pi_1 - \pi_2) \ .$$

■

In two-armed bandits, the underlying rule in a single behavior population is strictly payoff increasing if and only if inferior actions are eliminated when the efficient action is initially present. This result is a direct consequence of Definition 4 and Proposition 1.

Corollary 1 *Consider a two-armed bandit $\langle\{1,2\}, P\rangle$ and a behavioral rule X. Then X is strictly payoff increasing if and only if $\pi_i < \pi_j$ and $x_j(0) > 0$ implies $x_i(t) \to 0$ as $t \to \infty$.*

3.2 Revision Behavior and Switching Behavior

In this section we will quote results of Schlag [16] that concern connections between properties of revision behaviors and their induced switching probabilities. The following proposition completely characterizes a revision behavior that is payoff increasing in any two-armed bandit yielding payoffs in $\left[\pi^L, \pi^H\right]$, a property Schlag [16] refers to as improving.

Definition 5 *A revision behavior based on action set* $\{1,2\}$ *and payoff interval* $\left[\pi^L, \pi^H\right]$ *is called (**strictly**[1]) **improving** if it is (strictly) payoff increasing in any two-armed bandit in* $\mathcal{G}\left(\{1,2\}, \left[\pi^L, \pi^H\right]\right)$.

Proposition 2 *Let* X *be a revision behavior. Then the following statements are equivalent.*

i) X *is improving.*

ii) X *is imitating and in any two-armed bandit in* $\mathcal{G}\left(\{1,2\}, \left[\pi^L, \pi^H\right]\right)$,

$$\pi_i > \pi_j \Rightarrow F^i_{ji} \geq F^j_{ij}.$$

iii) X *is imitating and there exists* $\sigma \in \left[0, \frac{1}{\pi^H - \pi^L}\right]$ *such that for any* $p, q \in \left[\pi^L, \pi^H\right]$,

$$X_2(1,p,2,q) - X_1(2,q,1,p) = \sigma(q-p). \tag{5}$$

Proof. The fact that statements i) and ii) are equivalent follows directly from Proposition 1 and the definitions. Regarding the proof of the equivalence of statement iii) we refer to Schlag [16]. ■

It follows from the above characterization that reciprocal switching rules (e.g., 'never switch', 'always switch') are improving. On the other hand, as noted by Schlag [16], the seemingly intuitive rule 'imitate if better' fails to be improving. Notice that part iii) in the above characterization requires for the revision behavior to react in a smooth manner to small changes in the payoffs, a property the rule 'imitate if better' lacks.

The following result characterizes revision behavior that is strictly payoff increasing for any underlying payoff distribution in the two-armed bandit.

[1] In Schlag [16] the term "non stationary" is used instead.

Proposition 3 *Let X be a revision behavior. Then X is a strictly improving rule if and only if X is improving with underlying $\sigma > 0$ where σ defined by the characterization in Proposition 2.*

The following conclusions are easily derived from the above results: i) proportional imitation rules are strictly improving rules, ii) improving rules that are not strictly improving are precisely the reciprocal switching rules, and iii) strictly improving rules only exist if the payoffs realized by choosing an action are contained in a bounded interval.

Schlag [16] argues that the proportional imitation rule with $\sigma = \frac{1}{\pi^H - \pi^L}$ is the unique most preferred rule for the individual under certain circumstances. As we will see in the following, our model will not be able to select among strictly improving rules, especially the proportional imitation rule will not play a special role in our analysis.

4 Two Behavior Contests

We will now investigate the dynamics of a population in which various revision behaviors compete against each other. In order to simplify the analysis we will restrict attention to populations in which at most two different rules (denoted by X and Y) are present. The resulting dynamic adjustment process will be called a *two behavior contest.* This section develops the dynamic equations that will be needed in the dynamic stability analysis in the next section.

Selection among the revision behaviors is determined through the number of offspring an individual produces before he dies. As stated in Section 2, before any round a given individual is selected and then dies with probability $\tau\mu$. However, before he dies he reproduces as follows. The number of offspring of an individual is assumed to be equal to $h + \lambda\pi' > 0$ where π' is the payoff the individual last realized, $\lambda > 0$ is fixed and h is such that the total number of offspring is equal to the total number of parents, i.e., to the number individuals that were withdrawn; hence, $h = 1 - \lambda\bar{\pi}$ where $\bar{\pi}$ is the average payoff among those withdrawn which is equal to the average payoff in the population and $\lambda \in \left(0, \frac{1}{\pi^H - \pi^L}\right)$. Individuals breed true, i.e., offspring inherit the behavioral rule of their parent. Finally, the offspring randomly take the positions of the individuals that died.

After reproduction is completed, the individuals that were selected and that did not die together with the offspring of the selected individuals that died are given the opportunity to choose an action.

Since the population size is assumed to be infinite, the state of the population is uniquely determined by the proportion of the individuals using a given rule and a given action. At a given point in time t let $x_i = x_i(t)$ denote the proportion of individuals playing action $i = 1, 2$ and using the revision behavior X and y_i denote the corresponding proportions using the behavior Y. By definition, $x_1 + x_2 + y_1 + y_2 = 1$. Since the population size is infinite, the average number of offspring of individuals using action i is $1 + \lambda(\pi_i - \bar{\pi})$ where $\bar{\pi} = \pi_1 \cdot (x_1 + y_1) + \pi_2 \cdot (x_2 + y_2)$. Hence, the proportion of offspring using the behavioral rule F is $x_1 [1 + \lambda(\pi_1 - \bar{\pi})] + x_2 [1 + \lambda(\pi_2 - \bar{\pi})]$.

As in the previous setting without selection, an entering individual samples an individual that was not replaced and then chooses his first action according to his inherited behavioral rule. Hence, the increase in the proportion of individuals using the behavioral rule F and playing action 1 between the present and the next round is given by

$$\begin{aligned}\phi_{F1} &= [x_1 (1 + \lambda(\pi_1 - \bar{\pi})) + x_2 (1 + \lambda(\pi_2 - \bar{\pi}))] \\ &\quad \left[(x_1 + y_1) F^1_{\emptyset 1} + (x_2 + y_2) F^1_{\emptyset 2}\right] - x_1 \\ &= (x_1 y_2 - x_2 y_1) [(x_1 + y_1) \lambda(\pi_1 - \pi_2) - 1] \\ &\quad + \left[F^1_{\emptyset 2} (x_2 + y_2) - F^2_{\emptyset 1} (x_1 + y_1)\right] [(x_1 y_2 - x_2 y_1) \lambda(\pi_1 - \pi_2) + x_1 + x_2]\end{aligned} \tag{6}$$

and similarly, the increase in the proportion of individuals using F and action 2 is given by

$$\begin{aligned}\phi_{F2} &= (x_1 y_2 - x_2 y_1) [(x_2 + y_2) \lambda(\pi_1 - \pi_2) + 1] \\ &\quad + \left[F^2_{\emptyset 1} (x_1 + y_1) - F^1_{\emptyset 2} (x_2 + y_2)\right] [(x_1 y_2 - x_2 y_1) \lambda(\pi_1 - \pi_2) + x_1 + x_2],\end{aligned} \tag{7}$$

especially,

$$\phi_{F1} + \phi_{F2} = (x_1 y_2 - x_2 y_1) \lambda(\pi_1 - \pi_2) \ . \tag{8}$$

If each selected individual is replaced (i.e., $\mu = 1$), the adaptation of x_1 at time $t = r\tau$ ($r \in \mathbb{N}$) is given by

$$x_1(t+\tau) = x_1(t) + \tau\phi_{F1}. \tag{9}$$

If $\mu = 1$, all individuals using X stubbornly play action 1 and all individuals using Y stubbornly play action 2, i.e., $F^1_{\emptyset i} = G^2_{\emptyset i} = 1$, $i = 1, 2$ and $x_2(0) = y_1(0) = 0$ then $x_2(t) = y_2(t) = 0$ and

$$\begin{aligned} x_1(t+\tau) &= x_1(t) + \tau\lambda(\pi_1 - \bar{\pi})x_1 \\ y_2(t+\tau) &= y_2(t) + \tau\lambda(\pi_2 - \bar{\pi})y_2 \end{aligned} \tag{10}$$

which is a discrete version of the replicator dynamic (Taylor and Jonker [17]).

If the reproduction rate $\mu \in (0, 1)$, then combining (9) and (1) we obtain

$$x_1(t+\tau) = x_1(t) + \tau[(1-\mu)\Psi_F + \mu\phi_{F1}]$$

and taking the limit as τ goes to 0,

$$\dot{x}_1 = (1-\mu)\Psi_F + \mu\phi_{F1}, \tag{11}$$

and similarly,

$$\dot{x}_2 = -(1-\mu)\Psi_F + \mu\phi_{F2}, \tag{12}$$

where Ψ_F gives the net increase of individuals using action 1 among those that are not replaced, given by

$$\Psi_F = F^1_{22}(x_2+y_2)x_2 + F^1_{21}(x_1+y_1)x_2 - F^2_{12}(x_2+y_2)x_1 - F^2_{11}(x_1+y_1)x_1\ . \tag{13}$$

Similarly,

$$\dot{y}_1 = (1-\mu)\Psi_G + \mu\phi_{G1} \tag{14}$$

and

$$\dot{y}_2 = -(1-\mu)\Psi_G + \mu\phi_{G2},$$

where ϕ_{G1}, ϕ_{G2} and Ψ_G are derived by replacing x by y and F by G in (6), (7) and (13).

From (8), (11) and (12) it follows that

$$\dot{x}_1 + \dot{x}_2 = \mu\lambda\left(\pi_1 - \pi_2\right)\left(x_1 y_2 - x_2 y_1\right). \tag{15}$$

Consider a population in which both behaviors X and Y are present and $\pi_1 > \pi_2$. Then (15) implies that the proportion of individuals using behavior X grows and the proportion of individuals using Y falls if and only if among the individuals using X the relative proportion using action 1 is greater than among individuals using Y, i.e.,

$$\text{if } \pi_1 > \pi_2 \text{ then } \dot{x}_1 + \dot{x}_2 > 0 \text{ if and only if } \frac{x_1}{x_1 + x_2} > \frac{y_1}{y_1 + y_2}.$$

4.1 Neutral Stability

In this section we investigate the robustness or stability of a single behavior population with respect to mutations. Individuals are characterized by their revision behavior and by the action they chose last. Given a single behavior population with a given distribution of actions we assume that a small proportion of individuals mutate such that they then all follow a different behavior and choose different actions. Since selection pressure on the revision behaviors only takes place as long as the inferior action is present in the population we can not hope for an evolutionarily stable behavior (i.e., one that drives out every mutant behavior; Maynard Smith and Price [9]) to exist. Instead we will analyze revision behaviors that have a slightly weaker property. We will characterize a revision behavior that, when played by the incumbents, is able to prevent mutant behavior from spreading. Mutants replace incumbents according to their relative proportions in the population; if $y^\circ \in \Delta(\{1,2\})$ is the action profile played among the mutants, $x^\circ \in \Delta(\{1,2\})$ are the population shares of the incumbent before entry and δ is the fraction of the original population that is replaced, then $\left((1-\delta)\,x_1^\circ, (1-\delta)\,x_2^\circ, \delta y_1^\circ, \delta y_2^\circ\right)$ is the population state after the mutation. Along the lines of Maynard Smith [8] a behavioral rule will be called neutrally stable at a given distribution of actions if the proportion of any mutant rule stays arbitrarily small provided that the initial proportion of mutants is sufficiently small.

Definition 6 *For a given two-armed bandit and given* $x^{\circ} \in \Delta(\{1,2\})$, *a behavior* X *will be called* ***neutrally stable*** *at* x° *if for any behavior* Y, *any* $y^{\circ} \in \Delta(\{1,2\})$ *and any* $\varepsilon > 0, \exists \delta > 0$ *such that* $0 < \delta' < \delta$ *and* $(x_1(0), x_2(0), y_1(0), y_2(0)) = ((1-\delta')x_1^{\circ}, (1-\delta')x_2^{\circ}, \delta' y_1^{\circ}, \delta' y_2^{\circ})$ *implies*

$$y_1(t) + y_2(t) < \varepsilon \ \forall t \geq 0.$$

In other words, the proportion of a neutrally stable behavior X will not decrease substantially provided that the initial proportion of the entering mutation is sufficiently small. Never-the-less the relative proportion of the *actions* played among the individuals using the neutrally stable behavior may change drastically over time, especially if there are incumbents that are not using the efficient action (see Section 4.2). This fact complicates the proofs of neutral stability since the usual local linearization techniques concern local stability.

In the following we will investigate the class of rules that are neutrally stable when initially the incumbents are using the efficient action. It turns out that any revision behavior that is strictly payoff increasing is neutrally stable. Moreover, being payoff increasing is a necessary condition. Alongside we obtain the intuitive result that after a small mutation individuals stop using the inferior action in the long run.

Proposition 4 *Let* $\langle \{1,2\}, P \rangle$ *be a two-armed bandit such that* $\pi_1 > \pi_2$.

i) If X *is neutrally stable at* $(x_1^{\circ}, x_2^{\circ}) = (1,0)$ *then* X *is payoff increasing.*

ii) If X *is strictly payoff increasing then* X *is neutrally stable at* $(x_1^{\circ}, x_2^{\circ}) = (1,0)$ *and* $x_2(t) + y_2(t)$ *converges to* 0 *as* t *goes to infinity, provided* $x_1(0)$ *is sufficiently large.*

Proof. (see appendix).

An open question remains whether the converse of statement i) is true or not.

4.2 Neutral Stability for Arbitrary Initial States

In this section we wish to characterize revision behaviors that are neutrally stable for any initial state. In Proposition 4 we saw that any strictly payoff increasing rule playing the efficient action is neutrally stable. The question arises whether or not it is also neutrally stable if it must learn which action is the better one. In particular, is a strictly payoff increasing behavior neutrally stable even when all the incumbents are initially playing the inferior action? The following result states that this is true.

Proposition 5 *Let* $\langle \{1, 2\}, P \rangle$ *be a two-armed bandit such that* $\pi_1 > \pi_2$*. Then the behavioral rule* X *is neutrally stable at* x° *for any* $x^\circ \in \Delta(\{1, 2\})$ *if and only if* X *is strictly payoff increasing. Especially, if* X *is strictly payoff increasing then* $x_2(t) + y_2(t)$ *converges to* 0 *as* t *goes to infinity provided that* $x_1(0) + x_2(0)$ *is sufficiently large.*

Proof. (see appendix).

If the neutrally stable incumbents all start out playing the inferior action and the mutant stubbornly plays the efficient action then although the frequency of the mutant behavior increases steadily over time, the incumbents learn which action is better and thus eliminate their selective disadvantage by playing the efficient action too - our result states that this happens before the mutants have taken over a substantial proportion of the population. Therefore, even if all mutants were to know which action is better, arbitrarily long after the mutation occurred the majority of the individuals will still be using the strictly payoff increasing behavior.

With Proposition 5 we can now characterize behaviors that are neutrally stable in any two-armed bandit (Notice that any rule is neutrally stable when $\pi_1 = \pi_2$).

Corollary 2 X *is neutrally stable at any* $x^\circ \in \Delta(\{1, 2\})$ *and for any two-armed bandit* $\langle \{1, 2\}, P \rangle$ *in* $\mathcal{G}\left(\{1, 2\}, \left[\pi^L, \pi^H\right]\right)$ *if and only if* X *is a strictly improving rule.*

Rules that are neutrally stable for any two-armed bandit and for any initial configuration have intuitive appeal in an environment where the payoff distributions in the bandit are subject to rare changes. The notion of an evolutionarily stable strategy was introduced by Maynard Smith and Price [9] as an intuitive concept for a population playing a single strategy to be able to survive rare mutations. As an approximation of very rare mutations, formally their concept only considers a one time mutation. In fact, when there are only a finite number of sufficiently rare mutations, then their concept remains valid (see Schlag [14]). Similarly, when interested in rare changes in the payoffs of a two-armed bandit (more precisely, in changes in the distributions of the payoffs associated to an action in the two-armed bandit), one might consider a one time change as a first approximation. Consider a single behavior population that is subject to a one time mutation of behavior and to a one time change in the payoffs of the two-armed bandit, the mutation and the change of the payoffs must not occur simultaneously. With Corollary 2 it follows easily that strictly improving rules are neutrally stable in this setting. If the two-armed bandit changes before the mutant enters then the property of being neutrally stable at any initial configuration is sufficient to prevent the mutant from taking over a substantial proportion of the population. If the mutant enters before the two-armed bandit changes then Corollary 2 must be applied "twice". In fact, strictly improving rules are the only rules that are neutrally stable in this setting of slowly changing environments.

An alternative justification for searching for a rule that is neutrally stable for any distribution of the payoffs is that it has intuitive appeal when an individual is uncertain about the specifications of the two-armed bandit. Such considerations are the basis of the bounded rational selection approach pursued in the model of Schlag [16].

4.3 Survival

In the previous two subsections we focused on the stability of a single behavior population with respect to the entry of some mutant. In this section we analyze the long run outcomes of two behavior contests. We will say that a revision behavior survives if it does not vanish in the long run for some entry proportions.

Definition 7 *For a given two-armed bandit and given* $x^\circ, y^\circ \in \Delta(\{1,2\})$ *we will say that a behavior* Y *playing* y° ***survives*** *in a two behavior contest with* X *playing* x° *if for every* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that* $(x_1(0), x_2(0), y_1(0), y_2(0)) = ((1-\varepsilon)x_1^\circ, (1-\varepsilon)x_2^\circ, \varepsilon y_1^\circ, \varepsilon y_2^\circ)$ *implies* $y_1(t) + y_2(t) > \delta$ *for all* $t \geq 0$.[2]

At first we compliment Proposition 4 and characterize the class of mutant rules that survive in a two behavior contest with a strictly payoff increasing rule that is playing the efficient action. Mutants that survive under such circumstances never switch away from the efficient action without reason and do not stubbornly adhere to the inferior action when the efficient action is only being played by the incumbent.

Proposition 6 *Consider a two-armed bandit* $\langle\{1,2\}, P\rangle$ *where* $\pi_1 > \pi_2$. *Let* X *be a strictly payoff increasing revision behavior and let* $x^\circ, y^\circ \in \Delta(\{1,2\})$. *Then* Y *playing* y° ***survives*** *in the two behavior contest with* X *playing* x° *if and only if i)* $G_{11}^2 = G_{\emptyset 1}^2 = 0$ *and ii) if* $x_1^\circ > 0 = y_1^\circ$ *then* $G_{21}^1 + G_{22}^1 + G_{\emptyset 2}^1 > 0$.

Proof. (see appendix).

It follows from Proposition 6 that any imitating rule that does not ignore better actions ($F_{ij}^j > 0$ whenever $\pi_j > \pi_i$) will survive in a two behavior contest against any strictly payoff increasing rule. Reproduction eliminates the inferior action from the population at a stronger rate than any rule might be biased towards the inferior action (as long as the efficient action is initially played). Hence one might argue that the incumbent "teaches" the mutant which action is better. Especially, following Propositions 4 and 6, selection pressure is not strong enough for an *evolutionarily stable behavior* to exist, i.e., one that is neutrally stable and where no mutant playing a different rule than the incumbent can survive.

We conjecture that a revision behavior Y that violates the 'if' condition in the above proposition will be driven out when starting at any initial proportion (less than 1) of players using Y. Our proof only shows this statement when the proportion of individuals using the rule Y is sufficiently small. It

[2] In the dynamics systems literature this condition is called *persistence*.

is conceivable that the population state enters a cycle if there are initially a large fraction of individuals using Y.

Above we saw that it is quite easy to survive in a population where the other behavior is strictly payoff increasing; the incumbent rule will show which action is better and "tagging" along is sufficient. It turns out that it is much more difficult to survive in **any** two behavior contest. Only payoff increasing revision behaviors (that do not always stubbornly adhere to the inferior action) are able to survive in any two behavior contest.

Proposition 7 *Let $\mu^* \in (0,1]$ be fixed and consider a two-armed bandit in which $\pi_1 > \pi_2$. Then Y playing y°* ***survives*** *in a two behavior contest with X playing x° for any revision behavior X, any $x^\circ \in \Delta(\{1,2\})$ and any $0 < \mu < \mu^*$ if and only if Y is payoff increasing and $y_1^\circ > 0$ if $G_{21}^1 + G_{22}^1 + G_{\emptyset 2}^1 = 0$.*

Proof. (in the appendix).

In the proof of the above proposition we showed a slightly stronger statement, namely that a behavior Y that is not payoff increasing is driven out by some other behavior X even if initially the *majority* is using Y. Furthermore, notice that we used a stubborn rule X to drive out an imitating but non payoff increasing rule Y.

Corollary 3 *Consider a two-armed bandit in which $\pi_1 > \pi_2$. If X is not payoff increasing then there exists a behavior Y such that $x_1(t) + x_2(t) \to 0$ as $t \to \infty$ when starting in $(x_1(0), x_2(0), y_1(0), y_2(0))$ such that $x_1(0) + x_2(0) < 1$.*

Especially this means that the rule 'imitate if better', even when used by the majority, can be driven out by appropriately choosing mutant behavior and payoff distribution.

Analogue to Section 4.2 one may be interested in rules that survive for any initial actions against any other rule in any two-armed bandit. Such a rule will then always survive if the two-armed bandit is subject to rare changes (see scenario described at the end of Section 4.2). Analog to Corollary 2 we obtain from Proposition 7 the following result:

Corollary 4 *Y playing any $y^\circ \in \Delta(\{1,2\})$ survives in a two behavior contest with any X playing any $x^\circ \in \Delta(\{1,2\})$ for any reproduction rate $0 < \mu < 1$ and for any two-armed bandit $\langle \{1,2\}, P \rangle$ if and only if Y is improving but not the rule 'never switch'.*

Especially we find that copy-cats (i.e., the rule 'always switch') always survive.

Proof. Since y° is arbitrary, Y must be an improving rule such that $G_{ij}^j > 0$ whenever $\pi_j > \pi_i$. These are precisely the improving rules such that $G_j(i,p,j,q) > 0$ whenever $q > p$, a property that any improving rule other than the rule 'never switch' has. ■

In Proposition 5 we saw that a majority of the population using a strictly improving rule causes the entire population to adapt the superior action (provided that it is initially been chosen by someone). In the following we will see that this is no longer necessarily true when only a minority is playing such a rule. In a population in which the majority of the individuals stubbornly choose the inferior action, whether the superior action survives in the population depends on whether the mutant behavior has a tendency to play the superior action even the rest of the world is not, i.e., it depends on the existence of a "genetic" bias towards this action. Especially, a strictly improving behavior will survive but, due to its imitative nature, in the long run it will end up choosing the inferior action.

Proposition 8 *Consider a two-armed bandit with $\pi_1 > \pi_2$. Assume that individuals using X stubbornly choose the inferior action, i.e., $X_2(\cdot) = 1$. If Y is such that $G_{22}^1 + G_{\emptyset 2}^1 = 0$ then X remains in the majority and $y_1 \to 0$ as $t \to \infty$. If Y is biased towards action 1 when no one else is choosing action one, i.e., if $G_{22}^1 + G_{\emptyset 2}^1 > 0$ then $y_1 \to 1$ as $t \to \infty$.*

Proof. Assume that $G_{22}^1 + G_{\emptyset 2}^1 = 0$. Then it follows that

$$\begin{aligned}
\dot{x}_2 &= -\mu\lambda(\pi_1 - \pi_2)\, x_2 y_1 \\
\dot{y}_2 &\geq -(1-\mu)\, G_{21}^1 y_1 y_2 + \mu x_2 y_1 \left[1 + (x_2 + y_2)\lambda(\pi_1 - \pi_2)\right] \\
&\geq -(1-\mu)\, y_1 y_2 + \mu x_2 y_1
\end{aligned}$$

and hence y_2 increases for sufficiently large x_2. Consequently, all we are left to show is that y_2 increases at a larger rate than x_2 decreases. Given $c \in \left(1, \frac{1}{\lambda(\pi_1 - \pi_2)}\right)$,

$$\frac{\dot{y}_2}{\dot{x}_2} < -c < -1$$

if and only if

$$\mu x_2 \left[1 - c\lambda\left(\pi_1 - \pi_2\right)\right] > (1 - \mu)\, y_2 \ ,$$

which holds when x_2 is sufficiently large. Hence, for every unit that x_2 decreases, y_2 increases at least by $c > 1$ units which means that y_1 converges to 0 as t goes to infinity. Especially $x_2(t)$ remains arbitrarily large as long as $x_2(0)$ is sufficiently large.

Assume now that $G^1_{22} + G^1_{\emptyset 2} > 0$. Assume that $y_1 \to 0$ as $t \to \infty$. Since $\dot{x}_2 \leq 0$ there exists $c > 0$ such that $y_2(t) > c$ for all t. However, since $\dot{y}_1 > 0$ when $y_1 = 0$ and $y_2 > 0$, we obtain a contradiction to the fact that $y_1 \to 0$. Since y_1 is bounded away from 0, $x_2 > 0$ implies $\dot{x}_2 < 0$ and hence $x_2 \to 0$ as $t \to \infty$. ■

5 Experimentation

In this section we allow for an alternative mechanism for obtaining information about the two-armed bandit. Instead of sampling an individual may experiment with an action before he chooses his next action. *Experimentation* will mean, in contrast to sampling, to select an action according to some fixed probability distribution (that may depend on the current action) and to realize one payoff with this action. After experimentation, as with sampling, the individual chooses his future action according to some revision behavior (that belongs to the same class of revision behaviors allowed under sampling). Consequently this leads to switching probabilities F^k_{ij}, $i, j, k \in \{1, 2\}$,where F^k_{ij} is independent of j. (Here we assume that entering individuals do not experiment.)

Notice that including experimenting rules in our model requires no new computations since the switching behavior associated to an experimenting rule is already incorporated. Experimenting rules behave just like sampling rules that induce switching probabilities with $F_{ii}^k = F_{ij}^k$ for all $i, j \in \{1, 2\}$. Especially, there is no non-trivial experimenting behavior that is improving (only the degenerate experimenting behavior 'never switch' is improving). Following Proposition 4 and Corollary 2 we see that there is no experimenting behavior that is neutrally stable in all two-armed bandits.

6 Discussion

The Success of Imitation. We consider competition among memory-less rules, thereby including many rules that are not purely imitative. Similar to the bounded rational approach of Schlag [16] we find that individuals' inability to aggregate individual information favors imitative behavior. Here is some intuition behind this result. If no one else plays the efficient action then there is no selection pressure for me to play the efficient action. If someone else has found the efficient action then by observing his performance I will be able to learn that this is the efficient action, provided that I employ the right kind of imitative behavior. On the other hand, by following a behavior that does not rely solely on imitation, I may be punished by experimenting with a suboptimal action that no one else uses and thus create a selective disadvantage for my rule. A more sophisticated behavior would be necessary to be able to always profit from experimenting. Does this mean that the only reason that imitative behavior is selected is because it only competes against stupid rules? Notice that for a given bandit we allow for a mutant rule that (by chance) always plays the efficient action. Even if all the incumbents are playing the worse action before the mutation occurs, as long as they all follow the same strictly improving rule, the mutant can not take over a substantial proportion of the population. Thus, we have shown that individuals that are without any knowledge about the underlying situation, who follow a very simple behavior (e.g., the proportional imitation rule) can hold off extremely knowledgable (or lucky) entrants. We do not want to argue that the assumptions made in our model are typical. Often individuals have more information about the environment and employ more sophisticated behavior based on a richer memory. The objective of this paper is merely to demonstrate that sophisticated simple behavior can already be very effective.

Evolutionary verses Bounded Rational Approach. Although the bounded rational approach of Schlag [16] might be intuitive, bounded rational objectives are always subject to criticism due to their subjective nature (e.g., the uniform prior assumption in [16]). Under the evolutionary approach, once the population dynamic is defined, the criteria for the analysis are easily derived from the dynamical systems literature and classical evolutionary game theory (neutral stability and survival). Criticism can thus focus on the underlying population assumptions. Especially we provide an explicit (although admittingly simplified) story of a changing environment to justify why we are interested in a rule that is neutrally stable in any two armed-bandit (see remark made after Corollary 2). The analogue assumption in the bounded rational setting is mainly justified by the fact that the individual has no knowledge about the underlying payoff distributions in the bandit.

Schlag [16] obtains a unique optimal behavior in both a bounded rational and in a population oriented approach. This behavior is the proportional imitation rule with rate $\sigma = \frac{1}{\pi^H - \pi^L}$ which is of course a specific strictly improving rule. The mentioned population oriented approach, a preliminary analysis to this paper, considers the adaptive dynamics of a single behavior population dealt with in Section 3. Schlag [16] shows that this specific proportional imitation rule generates the highest growth rates among the improving rules. How does this rule perform in our setting? It can be shown that among all improving rules this proportional imitation rule gives up the least proportion of the population to an entering mutant that stubbornly uses the efficient action. We conjecture that this statement is also true for any behavior of the entering mutants. Of course, in a given two-armed bandit, the overall highest growth rate after the mutation is achieved by the lucky rule that stubbornly uses the efficient action.

Related Literature. To our knowledge, this is the first paper that deals analytically with the evolutionary selection of behavioral rules for social learning. In a closely related evolutionary setting Rogers [12] investigates the trade-off between individual and social learning. Individual learning refers to the act of incurring a cost and then learning which action is efficient. Social learning means that the action of a random individual from the population is imitated (payoffs are not observable). In a specific example of a constantly changing environment the evolutionarily stable proportions of individual and social learning are calculated.

In our analysis, an evolutionarily stable rule fails to exist because evolutionary selection pressure sometimes eliminates the inferior action before sufficient selection has occurred among the revision behaviors. The survival of various learning rules that behave similarly is a common phenomenon. In a model of Nelson and Winter [10], profit maximizers survive in the long run along with firms that act like profit maximizers. Schlag [15] studies the evolution of automata (which are in some sense rules) in the repeated Prisoners' Dilemma. A unique stable component is selected where both Tat-for-Tit and automata that imitate Tat-for-Tit when matched against Tat-for-Tit survive.

We restrict attention to two behavior contests to simplify the analysis. An alternate way of handling the typically very complex dynamical systems that arise when analyzing the evolution of learning rules is to employ simulations. Dixon et al. [4] simulate the evolution of rules for playing an infinitely repeated Cournot Duopoly game. In their setup a rule is a linear function that determines future output level based on their opponents output in the last round. In their simulations (the ones without noise) a unique rule survives, it is essentially the joint profit maximizing behavior.

Finally, Banerjee [1] finds that imitative behavior can also be justified on rational grounds, provided that there is the possibility that some agents are more informed than others.

References

[1] A. V. Banerjee, A Simple Model of Herd Behavior, *Quart. J. Econ.* **107** (1992), 797-818.

[2] J. Björnerstedt, and J. Weibull, "Nash Equilibrium and Evolution by Imitation," In K. Arrow and E. Colombatto (eds.), Rationality in Economics, New York: Macmillan (forthcoming), 1993.

[3] A. Cabrales, "Stochastic Replicator Dynamics," mimeo, University of California, San Diego, 1993.

[4] Dixon, H., Wallis, S., and Moss, S. (1995). "Axelrod meets Cournot: Oligopoly and the Evolutionary Metaphor, Part 1," Disc. Paper **95/8**, The University of York.

[5] G. Ellison, and D. Fudenberg, Word-Of-Mouth Communication and Social Learning, *Quart. J. Econ.* **440** (1995), 93-125.

[6] J. Gale, K. G. Binmore, and L. Samuelson, Learning to be Imperfect: the Ultimatum Game, *Games Econ. Beh.* **8** (1995), 56-90.

[7] J. Hofbauer, "Imitation Dynamics for Games," University of Vienna, mimeo, 1995.

[8] Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge University Press, Cambridge.

[9] Maynard Smith, J., and Price, G. R. (1973). "The Logic of Animal Conflict," *Nature* **246**, 15–18.

[10] Nelson, R., and Winter, S. (1982). *An Evolutionary Theory of Economic Change.* Harvard University Press, Cambridge.

[11] Novak, M. A., and R. M. May (1992). "Evolutionary Games and Spatial Chaos," Nature **359**, 826-829.

[12] Rogers, A. (1989). "Does Biology Constrain Culture?," *Amer. Anthropol.* **90** , 819-831.

[13] Rothschild, M. (1974). "A Two-Armed Bandit Theory of Market Pricing," *J. Econ. Theory* **9**, 185-202.

[14] Schlag, K. H. (1992). "Evolutionary Stability in Games with Equivalent Strategies, Mixed Strategy Types and Asymmetries," Disc. Paper **912**, Northwestern University.

[15] Schlag, K. H. (1993). "Dynamic Stability in the Repeated Prisoners' Dilemma Played by Finite Automata," Disc. Paper **B-243**, University of Bonn.

[16] Schlag, K. H. (1996). "Why Imitate, and if so, How? A Bounded Rational Approach to Multi-Armed Bandits," Disc. Paper **B-361**, University of Bonn.

[17] Taylor, P. D., and Jonker, L. (1978). "Evolutionarily Stable Strategies and Game Dynamics," *Math. Biosc.* **40**, 145–156.

[18] Wiggins, S. (1990). *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* Springer-Verlag: New York, Heidelberg, Berlin.

A Proofs

A.1 Neutral Stability

Let F (G) be the switching probabilities related to the revision behavior X (Y) for a given two-armed bandit. Let $a = \frac{x_2}{x_1+x_2}$, $b = \frac{y_2}{y_1+y_2}$ and $c = x_1 + x_2$. We then have

$$\begin{aligned} \dot{a} &= \frac{\dot{x}_2 x_1 - \dot{x}_1 x_2}{c^2} = -\left(1-\mu\right)\Psi_F \frac{1}{c} \\ &+\mu\left[\left(b-a\right)\left(1-c\right) - F^1_{\emptyset 2}\left(ac + b\left(1-c\right)\right)\right]\left[\lambda\left(\pi_1-\pi_2\right)\left(b-a\right)\left(1-c\right)+1\right] \end{aligned}$$

and with

$$\begin{aligned} \Psi_F &= \left[\left(1-a\right)c + \left(1-b\right)\left(1-c\right)\right]\left[F^1_{21}a - F^2_{11}\left(1-a\right)\right]c \\ &+\left[ac + b\left(1-c\right)\right]\left[F^1_{22}a - F^2_{12}\left(1-a\right)\right]c \end{aligned}$$

we obtain

$$\begin{aligned}\dot{a} &= (1-\mu)(1-a)\left[(1-a)c+(1-b)(1-c)\right]F_{11}^{2} \\ &+ (1-\mu)(1-a)\left[ac\left(F_{12}^{2}-F_{21}^{1}\right)+b(1-c)F_{12}^{2}\right] \\ &- (1-\mu)a\left[(ac+b-bc)F_{22}^{1}+(1-b)(1-c)F_{21}^{1}\right] \\ &+\mu\left[\left(bF_{\emptyset 2}^{2}-a\right)(1-c)-F_{\emptyset 2}^{1}ac\right]\left[\lambda(\pi_1-\pi_2)(b-a)(1-c)+1\right] \\ &+\mu F_{\emptyset 1}^{2}\left[1-ac-b(1-c)\right]\left[\lambda(\pi_1-\pi_2)(b-a)(1-c)+1\right] .\end{aligned} \tag{16}$$

Similarly,

$$\dot{c} = \mu\lambda(\pi_1-\pi_2)c(1-c)(b-a).$$

In the following we will establish some lemmata for a two-armed bandit in which $\pi_1 > \pi_2$. The first lemma states that the proportion of individuals playing the inferior action among the individuals using a strictly payoff increasing behavior X decreases as long as sufficiently many individuals are using X.

Lemma 1 *If F is strictly payoff increasing then $\exists$ a continuous decreasing function $\bar{a}:(0,1]\to\mathbb{R}$ such that $\bar{a}(1)=0$ and $\dot{a}<0$ if $\bar{a}(c)\leq a<1$.*

Proof. The first and fifth line in (16) are equal to zero since $F_{\emptyset 1}^{2}=F_{11}^{2}=0$, using the fact that $b(1-c)F_{12}^{2}\leq 1-c$ in the second line and dropping the third line, we obtain

$$\begin{aligned}\dot{a} &\leq (1-\mu)\left[-a^2F_{22}^{1}+(1-a)ac\left(F_{12}^{2}-F_{21}^{1}\right)+(1-a)(1-c)\right] \\ &+\mu\left[2(1-c)-F_{\emptyset 2}^{1}ac\right],\end{aligned}$$

and hence the statement is true. ■

The next lemma states that the proportion of individuals using a strictly payoff increasing behavior X does not decrease much while it decreases when almost all of the individuals are using X.

Lemma 2 *If F is strictly payoff increasing and μ is sufficiently small then there exists a function $s:(0,1)\to(0,1)$ such that $c(0)>1-s(\varepsilon)$ implies $c(t)>1-\varepsilon$ as long as $\dot{c}<0$, i.e., for any $t\in[0,T)$ where $T\in\mathbb{R}_0^{+}\cup\{\infty\}$ satisfies $\dot{c}(t)<0$ $\forall t\in[0,T)$.*

Proof. Let $y := y_1 + y_2$ and assume $\pi_1 > \pi_2$. Consider a trajectory with $0 < c(0) < 1$ and let $T \in \mathbb{R}_0^+ \cup \{\infty\}$ satisfy $\dot{c}(t) < 0 \ \forall t \in [0, T)$. Assume that $F_{21}^1 - F_{12}^2 > 0$, then

$$\dot{x}_1 \geq I(x_2 + y_2)x_1 + J(x_2 y_1 - x_1 y_2), \tag{17}$$

where

$$\begin{aligned} I &= (1-\mu)\left(F_{21}^1 - F_{12}^2\right) > 0 \text{ and} \\ J &= (1-\mu)F_{21}^1 + \mu\left[1 - \lambda(\pi_1 - \pi_2)(x_1 + y_1)\right] > 0. \end{aligned}$$

Note that $\dot{c} < 0$ ($\Leftrightarrow \dot{y} > 0$ - by definition) implies $x_2 y_1 > x_1 y_2$ and hence $\dot{x}_1 > 0$ and $\dot{y} > 0$ for all $t \in [0, T)$.

In the following we will put an upper bound on how much y can grow by time T for given $x_1(0)$ and $y(0)$. Only interested in an upper bound and since $\dot{x}_1 > 0$ it is enough to consider only trajectories with $x_1(0) = 0$.

From (15) and (17) we obtain

$$\frac{\partial x_1}{\partial y} = \frac{\dot{x}_1}{\dot{y}} \geq \frac{I \cdot x_1}{\mu\lambda(\pi_1 - \pi_2)y_1} + \frac{J}{\mu\lambda(\pi_1 - \pi_2)} \geq \alpha\frac{x_1}{y} + \beta,$$

where $\alpha = \frac{I}{\mu(\pi_1 - \pi_2)} > 1$ and $\beta = \mu(1 - \lambda(\pi_1 - \pi_2)) > 0$. The smaller the fraction $\frac{\dot{x}_1}{\dot{y}}$ is, the more y will grow relative to x_1. Hence, in order to find an upper bound on y when $y(0) = s$ and $x_1(0) = 0$ it is sufficient to find an upper bound on $\bar{q} = \bar{q}(s)$ where $p(\bar{q}) = 1$, $p(s) = 0$ and

$$\frac{\partial p}{\partial q} = \alpha\frac{p}{q} + \beta. \tag{18}$$

The solution to (18) given $p(s) = 0$ is

$$p(q) = \frac{\beta q}{\alpha - 1}\left(\left(\frac{q}{s}\right)^{\alpha - 1} - 1\right).$$

Consequently, for any given $\varepsilon > 0$ there exists $s = s(\varepsilon) \in (0, \varepsilon)$ such that $\bar{q}(s) = \varepsilon$. Therefore, $0 < y(0) < s(\varepsilon)$ implies $y(t) \leq \varepsilon$ for all $t < T$ which completes the proof for $F_{21}^1 - F_{12}^2 > 0$.

Assume now that $F_{21}^1 = F_{12}^2$. Since F is strictly payoff increasing, $F_{22}^1 > 0$ or $F_{\emptyset 2}^1 > 0$ must hold. Hence,

$$\dot{x}_2 \leq -(x_2)^2\left[(1-\mu)F_{22}^1 + \mu F_{\emptyset 2}^1\right] - J(x_2 y_1 - x_1 y_2)$$

where $J = (1-\mu) F^1_{21} + \mu \left[1 + \lambda (\pi_1 - \pi_2) F^2_{\emptyset 2} (x_2 + y_2)\right] > 0$. Then $\dot{x}_2 < 0$ and $\dot{y} > 0$ for all $t \in [0, T)$. With $\alpha = -\frac{(1-\mu) F^1_{22} + \mu F^1_{\emptyset 2}}{\mu \lambda (\pi_1 - \pi_2)} < 0$ and $\beta = -\frac{1}{\pi_1 - \pi_2} < 0$ we obtain

$$\frac{\dot{x}_2}{y} \leq \alpha \frac{x_2}{y} + \beta \ .$$

Hence we will search for an upper bound on $\bar{q} = \bar{q}(s)$ where $p(\bar{q}) = 0$, $p(s) = 1 - s$ and $p = p(q)$ satisfies (18). It follows that

$$p(q) = \frac{q}{1-\alpha} \left(\left(\frac{1-\alpha}{s} (1-s) + \beta \right) \left(\frac{q}{s} \right)^{1-\alpha} - \beta \right).$$

Consequently, for any given $\varepsilon > 0$ there exists $s = s(\varepsilon) \in (0, \varepsilon)$ such that $\bar{q}(s) = \varepsilon$ and $p(\bar{q}(s)) = 0$. Therefore, $0 < y(0) < s(\varepsilon)$ implies $y(t) \leq \varepsilon$ for all $t < T$. ■

Next we show that when either a or b is small, along any trajectory there will be at most one switch from c decreasing to c strictly increasing, or vice versa.

Lemma 3 *Let F be payoff increasing. Then there exists $\gamma > 0$ such that either*

i) $\dot{b} > \dot{a}$ whenever $0 < a = b < \gamma$ or

ii) $\dot{b} \leq \dot{a}$ whenever $0 < a = b < \gamma$.

Especially, if $G^2_{11} > 0$ or $G^2_{\emptyset 1} > 0$ then i) holds.

If $0 < c < 1$ then $\dot{c} \geq 0$ if and only if $b \geq a$. Hence, as long as either a or b is less than γ, if i) holds then only a switch from $\dot{c} \leq 0$ to $\dot{c} > 0$ is possible, if ii) holds then only a switch from $\dot{c} > 0$ to $\dot{c} \leq 0$ is possible.

Proof. Consider allocations where $a = b$. Then the proof follows directly using the fact that

$$\dot{a} = (1-\mu) \left[a (1-a) \left(F^2_{12} - F^1_{21}\right) - a^2 F^1_{22}\right] - \mu F^1_{\emptyset 2} a^2 \qquad (19)$$

and

$$\begin{aligned} \dot{b} \;=\; & (1-\mu) \left[(1-a)^2 G^2_{11} + a(1-a) \left(G^2_{12} - G^1_{21}\right) - a^2 G^1_{22}\right] \\ & + \mu \left[G^2_{\emptyset 1} (1-a)^2 - G^1_{\emptyset 2} a^2\right] \ . \end{aligned} \qquad (20)$$

■

Next we show that the inferior action vanishes when the fraction of individuals using the incumbent rule always strictly increases.

Lemma 4 *If F is payoff increasing and $\dot{c}(t) > 0$ for all $t \geq 0$ then $a \to 0$ and $b(1-c) \to 0$ as $t \to \infty$.*

Proof. If $\dot{c}(t) > 0$ for all $t \geq 0$ then $c \to c^*$ as $t \to \infty$ for some $c^* \leq 1$. If $c^* = 1$ then following Lemma 1, $a \to 0$ as $t \to \infty$ and the claim is proven.

Assume that $c^* < 1$. Let (a', b', c^*) be an ω limit. Since $\dot{c} \to 0$ as $t \to \infty$ it follows that $a' = b'$. Assume that $a' > 0$ and let $T_\alpha := \{(1-\alpha) b < a < b\}$. For any $\alpha > 0$ the trajectory eventually stays in $T_\alpha \cup \{b < \alpha\}$ for ever. Moreover, following (19), there exists $\alpha' > 0$ such that $\dot{a} < 0$ in $T_{\alpha'} \cap \left\{a > \frac{a'}{2}\right\}$. This contradicts the fact that (a', b', c^*) is an ω limit and hence $a \to 0$ and $b \to 0$ as $t \to \infty$. ■

Using the above lemmata, we are now able to present the proof of statement ii) in Proposition 4.

Proof. (of statement ii) in Proposition 4) Let X be strictly payoff increasing. Let $\varepsilon > 0$ be given. Consider the dynamic in the space $a := \frac{x_2}{x_1+x_2}, b := \frac{y_2}{y_1+y_2}$ and $c := x_1 + x_2$. Let $\gamma > 0$ be given from Lemma 3.

Assume that $\dot{b} > \dot{a}$ whenever $0 < a = b < \gamma$ as in part i) of Lemma 3. Let $0 < \delta < \varepsilon$ be such that $\bar{a}(1-\delta_1) < \gamma$ where $\bar{a}(\cdot)$ is given by Lemma 1. Consider a trajectory starting with $a(0) = 0$ and $1 - \delta < c(0) < 1$. Then $\dot{c}(t) > 0$ as long as $a(t) < b(t)$. However the later holds for all t, since at any time t either $\dot{a}(t) < 0$ or $a(t) \leq \bar{a}(c) < \gamma$. Hence $c(t) > 1 - \delta$ for all t and hence X is neutrally stable when $a(0) = 0$. Lemma 4 implies that $x_2 + y_2 \to 0$ as $t \to \infty$.

Assume that $\dot{b} \leq \dot{a}$ whenever $0 < a = b < \gamma$ as in part ii) of Lemma 3. Let $\delta' \in (0, \varepsilon)$ be such that $\bar{a}(1-\delta') < \gamma$ and let $\delta = s(\delta')$ ($s(\cdot)$ given by Lemma 2). Notice that $\delta < \delta'$. Consider a trajectory starting with $a(0) = 0$ and $1 > c(0) > 1 - \delta$. Let $T \in \mathbb{R}_0^+ \cup \{\infty\}$ be such that $a(t) < b(t)$ for all $0 \leq t < T$ and $a(T) = b(T)$ if $T < \infty$. Then $\dot{c}(t) > 0$ and hence $c(t) > 1 - \delta > 1 - \varepsilon$ while $t < T$.

If $T = \infty$ then following Lemma 4 the proof is complete.

Now assume that $T < \infty$. Then $c(T) \geq c(0) > 1 - s(\delta')$ and $a(T) < \bar{a}(c(0)) < \gamma$. Moreover, $\dot{c}(t) \leq 0$ and $\dot{b}(t) \leq \dot{a}(t)$ as long as $a(t) < \gamma$. By Lemma 2, $c(T) > 1 - s(\delta')$ implies $c(t) > 1 - \delta'$ as long as $\dot{c} < 0$. However, $c(t) > 1 - \delta'$ implies $a(t) \leq \bar{a}(1 - \delta') < \gamma$ and hence $\dot{c}(t) \leq 0$ and $c(t) > 1 - \delta' > 1 - \varepsilon$ for all $t > T$. Hence $a(t) - b(t) \to 0$ as $t \to \infty$. Since $0 < a = b < \gamma$ implies $\dot{a} < 0$ we obtain $a \to 0$ and $b \to 0$ as $t \to \infty$. ■

The next lemma shows that in the long run the majority of the population will not be playing the efficient action and simultaneously using a non payoff increasing rule.

Lemma 5 *Let X be a revision behavior that is not payoff increasing. Then there exists $\alpha > 0$ such that $0 \leq a \leq \alpha$ and $0.5 \leq c < 1$ implies $\dot{a} > 0$.*

Proof. Setting $a = 0$ in (16) we obtain

$$\begin{aligned} \dot{a} \geq\ & (1-\mu)\left[c + (1-b)(1-c)\right] F_{11}^2 \\ & + (1-\mu)\, b\,(1-c)\, F_{12}^2 \\ & + \mu F_{\emptyset 1}^2 \left[1 - b(1-c)\right]\left[\lambda(\pi_1 - \pi_2)\, b\,(1-c) + 1\right]. \end{aligned} \tag{21}$$

Hence, the proof follows directly the fact that X is not payoff increasing implies $F_{\emptyset 1}^2 > 0$ or $F_{11}^2 > 0$ or $F_{12}^2 > F_{21}^1$. ■

We now prove statement i) of Proposition 4 using Lemma 5.

Proof. (of statement i) in Proposition 4) Let Y be a rule that induces $G_{ij}^1 = G_{\emptyset i}^1 = 1$ for $i, j \in \{1, 2\}$. Let α be given by Lemma 5. Consider a trajectory starting with $0 < b(0) < 1$ and $0.5 < c(0) < 1$. We will show that there exists $t' > 0$ such that $c(t') < 0.5$ which then implies that X is not neutrally stable. Assume that $c(t) > 0.5$ for all t. Since $\dot{b} \leq -b\mu\left[(a-b)\, c\lambda(\pi_1 - \pi_2) + 1\right] < 0$ for $b > 0$, together with Lemma 5 it follows that there exists a finite time T such that $b(t) < \alpha/2$ and $a(t) > \alpha$ for all $t > T$. Hence, $c \to 0$ as $t \to \infty$ which contradicts our assumption that $c(t) > 0.5$ for all t. Consequently, X is not neutrally stable for any x°. ■

A.2 Neutral Stability in Arbitrary Initial States

The following lemma states that while a is decreasing, c can not decrease then increase more than twice along the same trajectory.

Lemma 6 *Let F be payoff increasing. Then there is no trajectory such that $\dot{a} < 0$ for $0 \leq t \leq T_5$, $\dot{c} < 0$ for $t \in [0, T_1) \cup (T_2, T_3) \cup (T_4, T_5]$ and $\dot{c} \geq 0$ for $t \in [T_1, T_2] \cup [T_3, T_4]$ where $0 < T_1$ and $T_i < T_{i+1}$ for $i = 1, .., 4$.*

Proof. Using (19) and (20) it is easily shown that $\dot{a} = \dot{b}$ and $a = b$ has at most two solutions $d' \neq d''$ such that $d', d'' \in (0, 1)$, especially any solution is independent of c. Following the same reasoning as in Lemma 3, using the fact that $\dot{c} < 0$ if and only if $b < a$ and $0 < c < 1$, the statement is proven. ■

This leads us to the proof of Proposition 5.

Proof. (Proposition 5) 'if' statement: In the proof of Proposition 4 we only considered trajectories starting with $a(0) = 0$. However we see that the proof also goes through for $a(0) < \gamma$.

Given the above, all we must now check is initial configurations where $a(0)$ is not small. Following Lemma 1, $\dot{a} < 0$ provided that c is sufficiently large. Using Lemma 2 we know that on sections of the trajectory where c is decreasing, c does not decrease too much. This fact we will use together with Lemma 6 to prove the statement.

Let $\varepsilon > 0$ be given. Let $\delta_1 > 0$ be such that $a(0) < \delta_1$ and $c(0) > 1 - \delta_1$ implies that $c(t) > 1 - \varepsilon$ for all t. Let $0 < \delta' < \delta_1$ be such that $\bar{a}(1 - \delta') < \delta_1$ and let $\delta = s(s(\delta'))$ where s is given by Lemma 2. Consider a trajectory starting with $c(0) > 1 - \delta$. Let T_i be such that $\dot{c} \geq 0$ for $t \in [0, T_1) \cup (T_2, T_3) \cup (T_4, T_5)$ and $\dot{c} < 0$ for $t \in [T_1, T_2] \cup [T_3, T_4]$ where $0 < T_1$ and $T_i < T_{i+1}$ or $T_i = \infty$ for $i = 1, .., 4$. Similar to previous reasoning we have in the intervals where $\dot{c} \geq 0$ that $\dot{a} < 0$ or $a \leq \bar{a}(c)$. In the intervals where $\dot{c} < 0$, $\dot{a} < 0$. Moreover, $c \geq 1 - \delta$ when $t \in [0, T_1)$, $c \geq 1 - s(\delta')$ when $t \in [T_1, T_2]$ and also for $t \in (T_2, T_3)$, $c \geq 1 - \delta'$ when $t \in [T_3, T_4]$. Consequently, for all $T_4 \leq t < T_5$ and hence for all $t < T_5$ we obtain that $c \geq 1 - \delta' > 1 - \delta_1$ and $[\dot{a} < 0$ or $a \leq \bar{a}(c) < \delta_1]$. Hence, following Lemma

6, there exists $0 < T' < T_5$ such that $c(T') > 1 - \delta_1$ and $a(T') < \delta_1$. By definition of δ_1 it follows that $c(t) > 1-\varepsilon$ for all t which completes the proof.

'only if' statement: Proposition 4 implies that X is payoff increasing. Assume that X is not strictly payoff increasing, i.e., $F_{21}^1 = F_{12}^2$ and $F_{11}^2 = F_{22}^1 = F_{\emptyset 1}^2 = F_{\emptyset 2}^1 = 0$. Consider the rule Y that stubbornly plays the efficient action, assume $y_2(0) = 0$. Then $y_2(t) = 0$ for all t and following

$$\begin{aligned} \dot{x}_2 &= -\mu x_2 y_1 [\lambda(\pi_1 - \pi_2) x_2 + 1] \\ \dot{y}_1 &= \mu\lambda(\pi_1 - \pi_2) y_1 x_2 \ , \end{aligned}$$

y_1 increases, x_2 decreases as long as $x_2 > 0$ and

$$\frac{dx_2}{dy_1} = \frac{\dot{x}_2}{\dot{y}_1} = -x_2 - \frac{1}{\pi_1 - \pi_2} \ . \tag{22}$$

Solving for the path (x_1, x_2, y_1, y_2) as a function of y_1 that initiates in $(\beta, 1-\alpha-\beta, \alpha, 0)$ for some $\alpha, \beta > 0$ with $\alpha + \beta \leq 1$, we obtain

$$x_2(y_1) = e^{\alpha - y_1}\left(1 - \alpha - \beta + \frac{1}{\pi_1 - \pi_2}\right) - \frac{1}{\pi_1 - \pi_2} \ .$$

Let $\bar{y}_1$ be such that $x_2(\bar{y}_1) = 0$, then

$$\bar{y}_1(\alpha) = \alpha + \ln((1-\alpha-\beta)\lambda(\pi_1 - \pi_2) + 1) \ .$$

Hence, for very small initial value $y_1(0) = \alpha$, by the time x_2 has vanished, y_1 has grown to be approximately $\ln((1-\beta)\lambda(\pi_1-\pi_2)+1) > 0$ and hence X is not neutrally stable at $(x_1^\circ, x_2^\circ) = (\beta, 1-\beta)$ when $\beta < 1$. ■

A.3 Survival

Proof. (of Proposition 6) Let X be a strictly payoff increasing behavior. If $G_{21}^1 = G_{22}^1 = G_{\emptyset 2}^1 = 0$ and $y_1^\circ(0) = 0$ then $b(t) = 1$ for all t. Then clearly, Y does not survive.

If $G_{11}^2 > 0$ or $G_{\emptyset 1}^2 > 0$ then following Proposition 4, for large enough initial c, a and $b(1-c)$ go to 0 and c stays close to 1 as t goes to ∞. Since $\dot{b} > \beta$ for some $\beta > 0$ when a and b are sufficiently small, b does not go to zero which means that $(1-c)$ goes to zero, hence Y does not survive.

On the other hand, if $G_{11}^2 = G_{\emptyset 1}^2 = 0$ and $G_{21}^1 + G_{22}^1 + G_{\emptyset 2}^1 > 0$ then w.l.o.g. we may assume that initially some players using Y are selecting action 1. Consequently for the 'only if' statement all we must show is that $G_{11}^2 = G_{\emptyset 1}^2 = 0$ and $y_1(0) > 0$ implies that Y survives. Following Proposition 4, for large enough initial c, a goes to 0 and c stays close to 1 as t goes to ∞. When $a = 0$ and $c = 1$ we obtain

$$\dot{b} = -b\left[(1-\mu) G_{21}^1 + \mu(1 - b\lambda(\pi_1 - \pi_2))\right]$$

and hence $\dot{b} < 0$. Hence the trajectory eventually stays close to $(0,0,1)$.

Notice that $M = \{(0,0,c), 0 \leq c \leq 1\}$ is a set of rest points. Moreover, calculating the eigenvalues at $(0,0,1)$ reveals two strictly negative $((1-\mu)(F_{12}^2 - F_{21}^1) - \mu F_{\emptyset 2}^1$ and $-(1-\mu) G_{21}^1 - \mu))$ and one zero eigenvalue. The zero eigenvalue is associated to the eigenvector $(0,0,1)$, in fact the centre manifold (e.g., see Wiggins [18]) associated to this eigenvalue is M. Moreover, trajectories starting in the two dimensional subspace $\{(a,b,1), a,b \in [0,1]\}$ stay in this set. Hence, trajectories starting close to $(0,0,1)$ with $c(0) < 1$ converge to $(0,0,c')$ with $c' < 1$. ■

A.4 Survival among Arbitrary Behavior

We now present the proof of Proposition 7.

Proof. 'if' statement: Let Y be payoff increasing such that $y_1^\circ > 0$ if $G_{21}^1 + G_{22}^1 + G_{\emptyset 2}^1 = 0$. W.l.o.g. we may hence assume that $y_1^\circ > 0$. Consider the trajectories for $c = 1$. We obtain

$$\begin{aligned} \dot{a} &= (1-\mu)\left[(1-a)^2 F_{11}^2 + a(1-a)\left(F_{12}^2 - F_{21}^1\right) - a^2 F_{22}^1\right] \\ &\quad + \mu\left[(1-a) F_{\emptyset 1}^2 - a F_{\emptyset 2}^1\right] \end{aligned}$$

and

$$\begin{aligned} \dot{b} &= (1-\mu)\left[a(1-b) G_{12}^2 - b(1-a) G_{21}^1 - ab G_{22}^1\right] \\ &\quad + \mu\left[a - b - G_{\emptyset 2}^1 a\right]\left[(a-b)(\pi_1 - \pi_2) + 1\right] \end{aligned}$$

and hence the trajectory starting in $(a,b,1)$ leads to $(a^*, b^*, 1)$ such that $b^* \leq a^*$.

Assume that Y does not survive, i.e., the trajectory starting in (a, b, c) converges to $(a^*, b^*, 1)$. If $b^* < a^*$ then $\dot{c} < 0$ in a neighborhood of $(a^*, b^*, 1)$ when $c < 1$ which gives a contradiction. Assume that $b^* = a^*$. Then $(0, 0, 1)$ is an eigenvector to the eigenvalue 0 which implies that c does not converge to 1 (same arguments as used at the end of the proof of Proposition 6).

'only if' statement: Here we will show a bit more than stated in the proposition, namely that if Y is either not payoff increasing or $y_1^\circ = G_{21}^1 = G_{22}^1 = G_{\emptyset 2}^1 = 0$ then Y vanishes in a two behavior contest with some X even if the initial fraction of Y is not small.

If $y_1^\circ = 0$ and $G_{21}^1 = G_{22}^1 = G_{\emptyset 2}^1 = 0$ then $y_1(t) = 0$ for all t and clearly Y will not survive against 'never switch' playing the efficient action, i.e., $x_1^\circ = 1$. In fact, for any initial $0 < c < 1$, $c \to 1$ as $t \to \infty$.

Similarly, if Y induces switching probabilities such that $G_{11}^2 + G_{\emptyset 1}^2 > 0$ then it will not survive against the stubborn rule that always plays the efficient action. When $a = 0$ and b is small then $\dot{b} > \beta$ for some $\beta < 0$. Hence, after finite time T, $a(t) = 0 < \varepsilon < b(t)$ holds for all $t > T$ and hence $0 < c(0) < 1$ implies $c \to 1$ as $t \to \infty$.

Finally, assume that $y_1^\circ > 0$ if $G_{21}^1 = G_{22}^1 = G_{\emptyset 2}^1 = 0$, that $G_{11}^2 = G_{\emptyset 1}^2 = 0$ but that Y is not payoff increasing, i.e., $G_{12}^2 > G_{21}^1$. If $a \geq b$ then

$$\dot{b} \geq b\left[(1-\mu)\left[(1-b)\left(G_{12}^2 - G_{21}^1\right) - aG_{22}^1\right] - \mu G_{\emptyset 2}^1\left((a-b)\,c\,(\pi_1 - \pi_2) + 1\right)\right]$$

and there exists $a^* > 0$, $\mu^* > 0$ and $\beta > 0$ such that $\dot{b} > \beta b$ for $0 \leq b \leq a \leq a^*$ and $0 < \mu < \mu^*$.

Consider now a revision behavior X such that $F_{\emptyset i}^2 = F_{ij}^2 = a^*$ for $i, j \in \{1, 2\}$. Then

$$\dot{a} = (a^* - a)\left[1 - \mu + \mu\left[(b-a)(1-c)(\pi_1 - \pi_2) + 1\right]\right] \; .$$

Consider a trajectory starting with $0 < c(0) < 1$. Since $y_1^\circ > 0$ if $G_{21}^1 = G_{22}^1 = 0$ we may assume that $b(0) > 0$. Since $\dot{b} > \gamma b$ when $0 \leq b \leq a^*$ and $a \to a^*$ as $t \to \infty$, there exists $T > 0$ and $\varepsilon > 0$ such that $b(t) > a(t) + \varepsilon$ for all $t > T$. Hence, $c \to 1$ as $t \to \infty$. ■