'Imitate Best' vs 'Imitate Best Average'[1]

Karl H. Schlag[2]

March, 1996

[1]Financial support from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged.

[2]Abt. Wirtschaftstheorie III, Department of Economics, University of Bonn, Adenauerallee 24-26, 53113 Bonn, Germany.

Abstract

We consider an infiinite population of identical individuals playing a two-armed bandit. Individuals learn by sampling a fixed number of last pulls of other individuals.

Under idiosyncratic noise and any sample size, individuals eventually learn the efficient action if each of them *imitate the best* individual in their sample and all actions are initially played, *'imitating the best average'* fails to have this property if there is too much noise.

Under general noise, *'imitating the best average'* leads the population close to the efficient action (maximizing expected payoffs) for large samples provided that initially both actions are being played, whereas *'imitate the best'* may drive out the efficient action.

1 Introduction

In this paper we analyze the ability of various social learning behaviors to lead a population to the efficient action. This analysis takes place in a multi-armed bandit setting. Individuals must repeatedly choose from a finite set of actions, each action generating an uncertain payoff. The payoff distributions generated by each action are not known by the individuals. Individuals acquire information about the bandit by observing the realizations of a finite number of other individuals facing the same bandit. We concentrate on simple behavioral rules, rules of thumb that can be used in many different situations. Especially we only consider behavior that does not depend on observations prior to the last time the individual realized a payoff.

A simple rule that is based on one observation is the so-called *proportional imitation rule*. This rule specifies to imitate actions that perform better with probability proportional to how much better they perform. Schlag ([7]) shows that this rule has the following property called *strictly improving*. Faced with any bandit that generates bounded payoffs the average payoff in a large population in which each individual uses this rule increases over time, and eventually, each individual plays the action that maximizes the expected payoff provided that all actions are initially played. In fact, the proportional imitation rule is the best strictly improving rule; we say that it *dominates* any other strictly improving rule in all bandits that generate bounded payoffs since it implements the maximal increase in average payoffs among such rules.

The requirement of leading to the efficient action in any bandit is very strong and comes at a cost. Strictly improving rules only exist for bandits that generate payoffs in a given bounded range. Moreover, the individual needs a randomizing device to implement this rule. In this paper we will analyze the performance of various deterministic behavioral rules found in the literature. We will weaken the condition of strictly improving by analyzing the performance of a given behavior when facing a smaller set of bandits, namely *bandits with idiosyncratic noise*. In these bandits, which we also refer to as *idiosyncratic bandits*, the payoff generated by an action differs

from its expected payoff by a noise term that is independent of the action.

Given only one observation, *'imitate if better'* (Malawski [5], Ellison and Fudenberg [4]) is the obvious candidate for a deterministic rule. Following this imitative rule, an individual switches to an action if he observes that this action achieved a higher payoff. This rule is strictly improving when we restrict attention to bandits with idiosyncratic noise, a property we call *idiosyncratic strictly improving*. In fact, 'imitate if better' dominates the proportional imitation rule in any idiosyncratic bandit that generates bounded payoffs. However, there is no best idiosyncratic strictly improving rule based on one observation, especially there are rules based on one observation that dominate 'imitate if better' in some idiosyncratic bandits.

In the literature there are two deterministic rules for sample sizes of at least two, 'imitate the best' (Axelrod [1]) and 'imitate the best average' (Bruch [3], Ellison and Fudenberg [4]).

Under the rule *'imitate the best'* an individual chooses the action that achieved the highest payoff among the individuals that he last sampled. We show that this rule is idiosyncratic strictly improving for any sample size. However, its performance is not monotonic in the sample size. In a given idiosyncratic bandit, the sample must be larger than a certain size before 'imitate the best' dominates 'imitate if better'. Moreover, this lower bound on the sample size can be arbitrarily large. How does 'imitate the best' perform in more general bandits where noise is not necessarily idiosyncratic? Here it turns out that this rule is too greedy. We show that the individuals adapt in the long run the action that can generate the highest payoffs, which of course might easily be the action that generates the lowest expected payoffs, provided noise is not idiosyncratic.

The other deterministic rule for general sample sizes is the rule *'imitate the best average'*. This rule prescribes to adapt the action that yielded the highest average payoff in the observed sample. In idiosyncratic bandits 'imitate the best average' is not able to lead the population to everyone playing the efficient action. As noted by Ellison and Fudenberg [4], this rule has a tendency to preserve diversity in the population. This happens when there is too much noise or when the sample is too large. However, 'imitate the best

average' leads the population to a state arbitrarily close to everyone using the efficient action, provided that the sample is sufficiently large. This fact holds for arbitrary bandits.

To summarize, we find that 'imitate the best' is a sensible rule in idiosyncratic bandits but that it can perform badly under general noise. On the other hand, 'imitate the best average' always leads most of the population to playing the efficient action when the sample is sufficiently large. However, in the discussion we argue that such rules that are unable to completely eliminate bad actions make them inferior to the proportional imitation rule when rules are under evolutionary selection pressure.

2 The Setting

Consider the following dynamic process of choosing actions, sampling and updating in an infinite population. In a sequence of rounds, a selected number of individuals must each choose one of two actions. Choosing the action i yields an uncertain payoff drawn from a probability distribution P_i where P_i is either atomic or a piecewise continuous density with finite expectation and variance. Payoffs are realized independently of all other events. Let π_i denote the expected payoff generated by choosing action i, i.e., $\pi_i = \int uP_i(u)\,du$, $i \in A$. Then the tuple $\langle A, (P_i)_{i\in A}\rangle$ describes a multi-armed bandit or a game against nature. The set of all multi-armed bandits with action set A that generate payoffs in $\Pi \subseteq \mathbb{R}$ will be denoted by $\mathcal{G}(A, \Pi)$. We will restrict some of our attention to *multi-armed bandits with idiosyncratic noise*, i.e., bandits where the distribution of the noise component $P_i - \pi_i$ does not depend on i. Under such circumstances, let $Z \equiv P_i - \pi_i$ and let the set of all multi-armed bandits with idiosyncratic noise based on the action set A that generate payoffs in $\Pi \subseteq \mathbb{R}$ be denoted by $\mathcal{G}^i(A, \Pi)$. If $\pi_i \in \arg\max\{\pi_j, j \in A\}$ we call the action i an *efficient* action.

In each round, the proportion α of the individuals are selected to choose an action, $0 < \alpha \leq 1$[1]. Before realizing a payoff, an individual sequentially

[1] Alternatively, assume that individuals choose an action in each round after which the

and independently meets (or samples) N other individuals from the population; N is finite and fixed. From each of these individuals she observes the action used and the payoff received in her last pull, i.e., when she last chose an action. We assume that sampling is random according to the current population shares. Formally this means the following. Let $\Delta(A)$ be the set of probability distributions on A which is identified with the set of population shares allocated to each action, $\Delta^{\circ}(A)$ will be the interior configurations, i.e., $\Delta^{\circ}(A) = \{x \in \Delta(A) : x_i > 0 \text{ for all } i \in A\}$. Let $x_i \in [0,1]$ be the proportion of individuals currently using action i, i.e., $x = (x_1, .., x_n) \in \Delta(A)$ and x is called the current state. Then the probability of selecting an individual using action i is equal to x_i. Given this notation, the average expected payoff of the population in state x, $\bar{\pi}(x)$, is given by $\bar{\pi}(x) = \sum x_i \pi_i$.

The description of how an individual chooses her next action whenever she faces a multi-armed bandit is summarized by a revision behavior. We allow for the individual to choose a randomizing device that generates independent events when they determine which action to choose in the next round. We assume that an individual forgets everything that happened before her last realization of a payoff. Following these assumptions, a *revision behavior* is a function $F : (A \times \mathbb{R})^{N+1} \to \Delta(A)$ where $F(i, u, j_1, v_1, .., j_N, v_N)_r$ is the probability of playing action r after obtaining payoff u with action i and sampling individuals using action $j_s \in A$ that obtained payoff $v_s \in \mathbb{R}$, $s = 1, .., N$. The revision behavior F is called *imitating* if $F(j_1, v_1, .., j_{N+1}, v_{N+1})_r = 0$ for $r \notin \{j_1, .., j_{N+1}\}$.

In the following we will investigate the adjustment of the population shares in a single behavior population (i.e., one in which each individual follows the same behavior). For $a \in A^{N+1}$ and $j \in A$ let

$$F^j(a) = \int_{u \in \mathbb{R}^{N+1}} F\left((a_i, u_i)_{i=1,.,N+1}\right)_j dP_1(u_1) ..dP_{N+1}(u_{N+1}) \ ,$$

then the population state at time $t+1$ given state $x^t \in \Delta(A)$ at time t, denoted by $x^{t+1}(F, x^t)$, is determined by

proportion α are allowed to change their action.

$$x_i^{t+1}\left(F,x^t\right)=\left(1-\alpha\right)x_i^t+\alpha\sum_{a\in A^{N+1}}x_{a_1}^t\cdot...\cdot x_{a_{N+1}}^tF^i\left(a\right)\text{ for }t\in\mathbb{N}.$$

In a given bandit, we call the behavior rule F *payoff increasing* if the average payoffs are non decreasing in each state, i.e., if

$$EIP\left(F,x\right)\equiv\sum\pi_ix_i^{t+1}\left(F,x\right)-\sum\pi_ix_i\geq0$$

for all $x^t\in\Delta\left(A\right)$. A behavioral rule F is called *strictly payoff increasing* if it is payoff increasing and the average payoffs increase strictly (i.e., $EIP\left(F,x\right)>0$) whenever $x_ix_j>0$ for some $i,j\in A$ such that $\pi_i\neq\pi_j$.

It follows that individuals playing the same strictly payoff increasing rule will eventually adapt an efficient action provided that initially all actions are present.

Remark 1 *In a given bandit, the state of a single behavior population based on a strictly payoff increasing rule converges to a state in which only efficient actions are present, provided that initially all actions in A are present, i.e., $x_i^0>0$ for all $i\in A$ implies that there exists $x^*\in\Delta\left(\arg\max\left\{\pi_i,i\in A\right\}\right)$ such that $x^t\left(F\right)\to x^*$ as $t\to\infty$.*

The rule F is called *strictly improving (on Π)* if F is strictly payoff increasing in any multi-armed bandit in $\mathcal{G}\left(A,\Pi\right)$. We call the rule F *idiosyncratic strictly improving (on Π)* if F is strictly payoff increasing in any multi-armed bandit $\left\langle A,P\right\rangle$ with idiosyncratic noise, i.e., $\left\langle A,P\right\rangle\in\mathcal{G}^i\left(A,\Pi\right)$. In the special case of two-armed bandits with idiosyncratic noise it follows that the rules that lead the population to the efficient action are precisely the idiosyncratic strictly improving rules.

Corollary 1 *Assume that $A=\left\{1,2\right\}$. Then a behavioral rule F is (idiosyncratic) strictly improving if and only if in any two-armed bandit (with idiosyncratic noise), $\pi_i>\pi_j$ and $x^0\in\Delta^\circ\left(\left\{1,2\right\}\right)$ implies $x_i^t\left(F\right)\to1$ as $t\to\infty$.*

When comparing two rules F and F' in a bandit $\left\langle A,P\right\rangle$ we say that F *dominates* F' in the bandit $\left\langle A,P\right\rangle$ if $EIP\left(F,x\right)\geq EIP\left(F',x\right)$ in $\left\langle A,P\right\rangle$ for all $x\in\Delta\left(A\right)$.

3 Single Sampling

In the following we assume that each individual is only allowed to sample one other individual before realizing her next payoff (i.e., $N = 1$). *'Imitate if better'* (Malawski [5], Ellison and Fudenberg [4]) is the imitating behavioral rule F_{ib} such that for $i \neq j$, $F_{ib}(i,u,j,v)_j = 1$ if and only if $v > u$. As noted by Schlag [6] we obtain the following result.

Theorem 2 *'Imitate if better' is idiosyncratic strictly improving.*

Proof. If F is imitating then the increase in average payoffs is given by

$$EIP(F,x) = \alpha \sum_{i<j} x_i x_j \left[F^j(i,j) - F^i(j,i)\right](\pi_j - \pi_i). \tag{1}$$

Consider the behavioral rule 'imitate if better'. Then

$$\begin{aligned} F^j_{ib}(i,j) - F^i_{ib}(j,i) &= \int_u g_{ij}(u,u) Z(u)^2 du \\ &\quad + \int_{u<v} \left[g_{ij}(u,v) + g_{ij}(v,u)\right] Z(u) Z(v) dudv \end{aligned} \tag{2}$$

where $g_{ij}(u,v) = F_{ib}(i,\pi_i+u,j,\pi_j+v)_j - F_{ib}(j,\pi_j+v,i,\pi_i+u)_i$. If $u \leq v$ and $\pi_j > \pi_i$ then $g_{ij}(u,u) = 1$,

$$g_{ij}(u,v) + g_{ij}(v,u) = \left\{ \begin{array}{ll} 2 & \\ 1 & \text{if } \pi_i + v \left\{ \begin{array}{l} < \\ = \\ > \end{array} \right\} \pi_j + u \\ 0 & \end{array} \right.$$

and hence $F^j_{ib}(i,j) - F^i_{ib}(j,i) \geq 0$. Together with (1) it follows that F_{ib} is idiosyncratic improving. If the noise is atomic then the first term in (2) is strictly positive and hence F_{ib} is idiosyncratic strictly improving. If the noise term is non-atomic then the second term in (2) is strictly positive since

$$\begin{aligned} \int_{u<v<u+|\pi_j-\pi_i|} \left[g_{ij}(u,v) + g_{ij}(v,u)\right] Z(u) Z(v) dudv &= \\ 2 \int_{u<v<u+|\pi_j-\pi_i|} Z(u) Z(v) dudv &> 0. \end{aligned}$$

■

An alternate behavior that plays an important role when facing bandits that generate bounded payoffs where the noise is not necessarily idiosyncratic, is the proportional imitation rule (see Schlag [7]). Assume that $[\pi^-, \pi^+]$ contains the payoffs generated by a given bandit, i.e., $P_i(x) > 0$ for some $i \in A$ implies $x \in [\pi^-, \pi^+]$. Given $\sigma \in \left(0, \frac{1}{\pi^+ - \pi^-}\right]$ the imitating behavioral rule F_{PIR} is called the *proportional imitation rule with rate* σ if $F_{PIR}(i,u,j,v)_j = \sigma \cdot \max\{0, v-u\}$. Schlag [7] shows that the proportional imitation rule is strictly payoff increasing in any bandit that generates payoffs in $[\pi^-, \pi^+]$, and that 'imitate if better' fails to have this property. However, as stated in Theorem 2, 'imitate if better' is strictly payoff increasing in bandits with bounded payoffs if the noise is idiosyncratic. In fact, our next result states that 'imitate if better' outperforms any proportional imitation rule in bandits with idiosyncratic noise and bounded payoffs.

Theorem 3 *'Imitate if better' dominates any proportional imitation rule in any idiosyncratic bandit that generates bounded payoffs.*

Proof. Let $z^- = \inf\{u : Z(u) > 0\}$ and $z^+ = \sup\{u : Z(u) > 0\}$. Consider $i, j \in A$ such that $\pi_j > \pi_i$. For the proportional imitation rule we obtain $F^j_{PIR}(i,j) - F^i_{PIR}(j,i) = \sigma(\pi_j - \pi_i)$ for some $\sigma \leq \frac{1}{\pi_j - \pi_i + z^+ - z^-}$. Let $n \in \mathbb{N}$ be such that $(n-1)(\pi_j - \pi_i) + z^- \leq z^+$ and $n(\pi_j - \pi_i) + z^- > z^+$. For $i = 1, .., n$ let $x_i = P(z^- + (i-1)(\pi_j - \pi_i) \leq Z < z^- + i(\pi_j - \pi_i))$. Then

$$F^j_{ib}(i,j) - F^i_{ib}(j,i) \geq \sum_{i=1}^{n} (x_i)^2 \geq \frac{1}{n} \geq \frac{\pi_j - \pi_i}{\pi_j - \pi_i + z^+ - z^-} \geq \sigma(\pi_j - \pi_i)$$

which together with (1) proves the statement. ■

Besides the fact that it is dominated by 'imitate if better' in idiosyncratic bandits, the proportional imitation rule has the further disadvantage that it is only defined for bandits that generate bounded payoffs. Consider the following extension. Given $\sigma > 0$ we will call the rule F a *truncated proportional imitation rule based on* σ if F is imitating, $F(i,u,j,v)_j = 0$ if $v < u$ and $F(i,u,j,v)_j = \min\{\sigma(v-u), 1\}$ for $v \geq u$.

Proposition 4 *The truncated proportional imitation rule based on* $\sigma > 0$

is idiosyncratic strictly improving and it is strictly payoff increasing in any multi-armed bandit that generates payoffs in $\left[w, w+\frac{1}{\sigma}\right]$ *for some* $w \in \mathbb{R}$.

Proof. All we must show is that it is idiosyncratic strictly improving. For $i \neq j$ with $\pi_j > \pi_i$ let $g_{ij}(u,v) = F(i, \pi_i + u, j, \pi_j + v)_j - F(j, \pi_j + v, i, \pi_i + u)_i$. Consider $v' \geq u'$. Then $g_{ij}(u',v') + g_{ij}(v',u') = \sigma(\pi_j - \pi_i)$ if $\pi_j + v' - \pi_i - x' < \frac{1}{\sigma}$ and $|\pi_j + u' - \pi_i - v'| < \frac{1}{\sigma}$ and $g_{ij}(u',v') + g_{ij}(v',u') \geq 0$ since $g_{ij}(u',v') = 1$ when $\pi_j + v' - \pi_i - x' > \frac{1}{\sigma}$. ■

Schlag [7] shows that the proportional imitation rule with rate $\sigma = \frac{1}{\pi^+ - \pi^-}$ dominates any other rule that is improving in bandits that generate payoffs in $[\pi^-, \pi^+]$. In the following we will show that there is no rule that has this property in idiosyncratic bandits.

Theorem 5 *Under single sampling there is no rule that dominates all idiosyncratic improving rules based on* $[\pi^-, \pi^+]$ *in all bandits in* $\mathcal{G}^i(A, [\pi^-, \pi^+])$.

For the proof of the above theorem requires the following lemma.

Lemma 6 *Idiosyncratic strictly improving rules for bandits in* $\mathcal{G}^i(A, [\pi^-, \pi^+])$ *are imitating.*

Proof. See arguments made by Schlag ([7], [6]), or calculate directly as follows. Let $a \in A^{N+1}$ and $u \in \mathbb{R}^{N+1}$ be such that $F((a_i, u_i))_r > 0$ for some $r \in B = A \backslash \{a_i, i = 1, .., N+1\}$. We will construct a bandit in which this revision behavior is not payoff increasing. Consider a bandit in $\mathcal{G}^i(A)$ such that $\pi_i = \pi_j > \pi_k$ for $i, j \in B$ and $k \in A \backslash B$ and let Z be such that $Z(u_i) = \frac{1}{N+1} |\{j \in \{1, .., N+1\} : u_j = u_i\}|$. For a state $x \in \Delta^\circ(B)$ it follows from (1) that $EIP(F, x) < 0$. ■

Especially, an individual using an idiosyncratic improving rule that samples only individuals using the same action as he is, will not switch regardless of the payoffs he observes. Ellison and Fudenberg [4] refer to this property as the "must-see" restriction.

Proof. (of Theorem 5) By Lemma 6 idiosyncratic improving rules are imitating. Since 'imitate if better' is idiosyncratic improving and it is the unique

rule that dominates all rules that are payoff increasing when the noise is degenerate, i.e., when $Z(u) = 1$ for some u. Hence, all we must do is to construct an idiosyncratic improving rule that is not dominated by 'imitate if better'. Let F_+ be the imitating rule that behaves like 'imitate if better' except that $F_+(1,-1,2,0)_2 = 0$. In the following we will show that F_+ is idiosyncratic improving.

Let $g(u,v) = F_{ib}(1,\pi_1+u,2,\pi_2+v)_2 - F_{ib}(2,\pi_2+v,1,\pi_1+u)_1$. Assume that $v \geq u$. Then $g(u,v)+g(v,u) = -1$ if and only if $\pi_1+u = -1$, $\pi_2+v = 0$ and $0 < \pi_2-\pi_1 < v-u$, otherwise $g(u,v)+g(v,u) \geq 0$. Hence, we only need to check whether F_+ is payoff increasing in a bandit in which $\pi_2 > \pi_1$ and $Z(u)Z(v) > 0$ where $u = -1-\pi_1$ and $v = -\pi_2$. Given $\lambda = \frac{Z(u)}{Z(u)+Z(v)}$, we obtain $0 \leq \lambda \leq 1$ and

$$\begin{aligned} & F^2(1,2) - F^1(2,1) \\ \geq\ & [Z(u)+Z(v)]^2 \\ & \cdot \left[\lambda^2 g(u,u) + \lambda(1-\lambda)(g(u,v)+g(v,u)) + (1-\lambda)^2 g(v,v)\right] \\ =\ & [Z(u)+Z(v)]^2 \left[\lambda^2 - \lambda(1-\lambda) + (1-\lambda)^2\right] \geq 0 \end{aligned}$$

Following (1), F_+ is payoff increasing in such a bandit and hence it is idiosyncratic improving.

W.l.o.g. assume that $[-2,1] \subseteq [\pi^-,\pi^+]$. Consider now a bandit where $Z(-2) = Z(0) = \frac{1}{2}$, $\pi_1 = 1$ and $\pi_2 = 0$. Then $F_+^1(2,1) = F_{ib}^1(2,1)$ and $F_+^2(1,2) = 0 < F_{ib}^2(1,2)$ and hence $EIP(F_{ib},x) < EIP(F_+,x)$ for $x \in \Delta(A)$ such that $x_1x_2 > 0$. ■

4 Multiple Sampling

In the following we consider the more general situation in which the individual samples $N \geq 2$ other individuals before making her next choice. Here

$$EIP(x,F) = \alpha \sum_{a \in A^{N+1}} x_{a_1} \cdot .. \cdot x_{a_{N+1}} \sum_{j \in A} F^j(a)(\pi_j - \pi_{a_1}).$$

We will consider two extensions of the behavior 'imitate if better' for sampling multiple individuals, namely 'imitate the best' (Axelrod [1]) and 'imitate the

best average' (Bruch [3], Ellison and Fudenberg [4]).

4.1 'Imitate the Best'

'Imitate the best' (Axelrod [1]) is the rule that specifies to adapt the action that achieved the highest payoff in the sample modulo some additional assumptions when there are ties. Given $u \in \mathbb{R}^{N+1}$ and $a \in A^{N+1}$ let $m(u) = \max\{u(i), i = 1, .., N+1\}$ and $b(a,u) = \{a(i) : u(i) = m(u)\}$. We refer to a behavioral rule as the rule *'imitate the best'*, and denote this rule by F_{IB}, if

$$F_{IB}(a(1), u(1), .., a(N+1), u(N+1))_j = \frac{1}{|b(a,u)|}$$

when $j \in b(a,u)$ and $u(1) < m(u)$, and

$$F_{IB}(a(1), u(1), .., a(N+1), u(N+1))_{a(1)} = 1$$

when $u(1) = m(u)$.

Theorem 7 *'Imitate the best' is idiosyncratic strictly improving for any sample size $N \geq 2$.*

Proof. W.l.o.g. we may assume that sampling occurs letting $(N+1)$ individuals sample each other. Fix $a \in A^{N+1}$. For $v \in \mathbb{R}^{N+1}$ let $g_i^r(v)$ be the probability that an individual in sample $(a_j, v_j)_{j=1}^{N+1}$ switches from previously playing i to playing r in the next round. Then the increase in average payoffs given by the switching behavior in this sample is equal to

$$\delta(v) = \sum_{i=1}^{n} \frac{m_i}{N+1} \sum_{j=1}^{n} g_i^j(v) \cdot (\pi_j - \pi_i) \ ,$$

where $m_i = |\{r : a_r = i\}|$.

Fix $u \in \mathbb{R}^{N+1}$. Let $\Pi(N+1)$ be the set of all perturbations of $\{1, .., N+1\}$. The main idea of the proof is that samples $\left(a_i, \pi_{a_i} + u_{\tau(i)}\right)_{i=1}^{N+1}$ occur equally likely for all perturbations $\tau \in \Pi(N+1)$. Let

$$\delta^* = \frac{1}{(N+1)!} \sum_{\tau \in \Pi(N+1)} \delta(u_\tau) \tag{3}$$

where $u_\tau \in \mathbb{R}^{N+1}$ is such that $(u_\tau)_i = u_{\tau(i)}$, $i = 1, .., N+1$. Then the proof will be complete if we can show that $\delta^* \geq 0$.

Let $u^* = \max_i \{u_i\}$. W.l.o.g., we may assume that $\pi_i > \pi_{i+1}$ for $i = 1, .., n-1$ and assume that $m_i > 0$ for all $i \in A$. Let z_i be the fraction of the samples $S = \{(a_i, \pi_{a_i} + u_{\tau_i}) : \tau \in \Pi(N+1)\}$ in which no individual playing action $j < i$ received $\pi_j + u^*$ and at least one individual playing i received $\pi_i + u^*$. By definition, $\sum_{i \in A} z_i = 1$ and in the fraction z_1 of the samples S all individuals switch to playing action 1 in the next round. Consider a perturbation $\tau \in \Pi(N+1)$ in which i is the smallest action such that a player in the sample $(a_i, \pi_{a_i} + u_{\tau_i})_i$ received $\pi_i + u^*$. Considering a worst case in which all individuals playing an action with index below i switch to i we obtain that all individuals adapt the action i in the next round. Consequently

$$\begin{aligned} \delta^* &\geq \sum_{i \in A} z_i \sum_{j=1}^{n} m_j (\pi_i - \pi_j) = (N+1) \sum_{i=1}^{n} \left(z_i - \frac{m_i}{N+1} \right) \pi_i \\ &= (N+1) \sum_{i=1}^{n-1} \left(z_i - \frac{m_i}{N+1} \right) (\pi_i - \pi_n) \end{aligned} \tag{4}$$

Let $\gamma = |\arg\max_i \{u_i\}|$. If $\gamma = 1$ then $z_i = \frac{m_i}{N+1}$, more generally for $\gamma \geq 1$ we obtain $\sum_{i=1}^{j} z_i \geq \sum_{i=1}^{j} \frac{m_i}{N+1}$ for any $j < n$. For $r = 1, .., n-1$, let

$$\delta_r = (N+1) \sum_{i=1}^{r} \left(z_i - \frac{m_i}{N+1} \right) (\pi_i - \pi_n).$$

Especially, (4) implies $\delta^* \geq \delta_{n-1}$. In the following we will show by induction that $\delta_r \geq 0$ for all $r = 1, .., n-1$. $\delta_1 \geq 0$ follows directly since $z_1 \geq \frac{m_1}{N+1}$. Assume that $\delta_{r-1} \geq 0$ for $r < n$. Then

$$\delta_r = \delta_{r-1} + (N+1) \left(z_r - \frac{m_r}{N+1} \right) (\pi_r - \pi_n)$$

and $\delta_r \geq 0$ if $z_r \geq \frac{m_r}{N+1}$. If on the other hand, $z_r < \frac{m_r}{N+1}$ then since

$$\delta_r > \delta_{r-1} + (N+1) \left(z_r - \frac{m_r}{N+1} \right) (\pi_{r-1} - \pi_n)$$

together with

$$0 \leq \sum_{i=1}^{r-1} \left(z_i - \frac{m_i}{N+1} \right) (\pi_{r-1} - \pi_n) \leq \sum_{i=1}^{r-1} \left(z_i - \frac{m_i}{N+1} \right) (\pi_i - \pi_n)$$

we again obtain that $\delta_r \geq 0$. Especially we have shown by induction that $\delta_{n-1} \geq 0$ and hence $\delta^* \geq 0$. Hence 'imitate the best' is idiosyncratic improving. The fact that it is also idiosyncratic strictly improving follows along the same lines as in the proof of Proposition 2. Especially, it follows that

$$rF_{IB}^{j}\left(i_r, j_{N+1-r}\right) - \left(N+1-r\right) F_{IB}^{i}\left(j_{N+1-r}, i_r\right) > 0 \text{ if } \pi_j > \pi_i,$$

where $a = \left(i_r, j_{N+1-r}\right) \in A^{N+1}$ is such that $a_s = i$ for $1 \leq s \leq r$ and $a_s = j$ for $r+1 \leq s \leq N+1$. ■

From the above proof we obtain that 'imitate the best' is idiosyncratic strictly improving regardless of which tie breaking rule is used.

In the following we investigate the performance of 'imitate the best' in relation to the underlying sample size N. In a given idiosyncratic bandit with bounded support we show that using 'imitate the best' with sufficiently large sample size N dominates 'imitate if better'. However the sample size necessary to make this statement true can be arbitrarily large. Especially, there is no sample size N such that 'imitate the best' based on N dominates 'imitate if better' in any idiosyncratic bandit.

Theorem 8 *In a given idiosyncratic two-armed bandit $\langle A, P\rangle \in \mathcal{G}^i\left(A, [\pi^-, \pi^+]\right)$, there exists $N_0 \geq 2$ such that $N \geq N_0$ implies 'imitate the best' based on sample size N dominates 'imitate if better' in $\langle A, P\rangle$. However, N_0 can not be chosen uniformly for all idiosyncratic bandits in $\mathcal{G}^i\left(A, [\pi^-, \pi^+]\right)$.*

Proof. W.l.o.g., assume that $\pi_2 > \pi_1$. Let $u^+ = \sup\left\{u : Z\left(u\right) > 0\right\}$ and let $\mu = P\left(\pi_2 + Z > \pi_1 + u^+\right)$. If $\mu = 1$ then $F_{IB}^{2}\left(1, \cdot, 2, \cdot\right) = 1$ and F_{IB} dominates F_{ib}. Assume that $\mu < 1$, i.e., $P\left(\pi_2 + Z \leq \pi_1 + u^+\right) > 0$. Then

$$1 - \epsilon \equiv F_{ib}^{2}\left(1,2\right) - F_{ib}^{1}\left(2,1\right) < 1.$$

Moreover, $F_{IB}^{2}\left(1_j, 2_{N+1-j}\right) \geq 1 - \left(1-\mu\right)^{N+1-j}$, $F_{IB}^{1}\left(2_{N+1-j}, 1_j\right) \leq 1 - F_{IB}^{2}\left(1_j, 2_{N+1-j}\right)$, and hence,

$$\begin{aligned} & jF_{IB}^{2}\left(1_j, 2_{N+1-j}\right) - \left(N-j+1\right) F_{IB}^{1}\left(2_{N+1-j}, 1_j\right) \\ \geq\ & \left(N+1\right)\left[1 - \left(1-\mu\right)^{N+1-j}\right] - \left(N-j+1\right) \\ =\ & j - \left(N+1\right)\left(1-\mu\right)^{N+1-j}. \end{aligned}$$

In the following we will show that there exists N_0 such that $N \geq N_0$ and $1 \leq j \leq N$ implies

$$j - (N+1)(1-\mu)^{N+1-j} > 1 - \epsilon \tag{5}$$

Since the left hand side in (5) is concave in j, all we must check is that (5) holds for $j = 1$ and $j = N$. When $j = N$ then the left hand side goes to ∞ as N goes to infinity. When $j = 1$ then (5) holds for sufficiently large N since $\lim_{\eta \to \infty} \eta (1-\mu)^{\eta} = 0$. Hence, F_{IB} dominates F_{ib} for sufficiently large N.

We now show that N_0 may be arbitrarily large. Fix N, π_1 and π_2 with $\pi_2 > \pi_1$. Let $\mu > \pi_2 - \pi_1$ and consider a bandit with $Z(\lambda - \mu) = \lambda$ and $Z(\lambda) = 1 - \lambda$ for some $0 < \lambda < 1$. Then

$$F^2_{IB}(1, 2_N) = 1 - F^1_{IB}(2_N, 1) = 1 - (1-\lambda)\lambda^N.$$

and

$$\begin{aligned} F^2_{IB}(1, 2_N) - N F^1_{IB}(2_N, 1) &= 1 - (N+1)(1-\lambda)\lambda^N \\ &< F^2_{ib}(1,2) - F^1_{ib}(2,1) = 1 - 2(1-\lambda)\lambda \end{aligned}$$

if and only if

$$\frac{2}{N+1} < \lambda^{N-1} < 1. \tag{6}$$

Since

$$EIP(x, F) = \alpha \left[x_1 (x_2)^N \left[F^2(1, 2_N) - N F^1(2_N, 1) \right] + o\left((x_1)^2 \right) \right] (\pi_2 - \pi_1)$$

we obtain $EIP(x, F_{ib}) > EIP(x, F_{IB})$ when x_2 is sufficiently large ($x_2 < 1$) and (6) holds. ■

Notice that if $\lambda = 0.75$ then (6) is satisfied only if $N \leq 4$, i.e., F_{ib} is not dominated by F_{IB} unless $N > 5$; if $\lambda = 0.85$ then (6) holds if and only if $N \leq 12$.

In fact, the proof of Theorem 8 shows more. For any $N \geq 2$ 'imitate if better' can be better, at least in the long run, for learning the efficient action than 'imitate the best'. More specifically, for any $N \geq 2$ there exists an idiosyncratic bandit such that eventually more individuals in a single

behavior population based on 'imitate if better' adapt the efficient action compared to the situation in which they are all using the rule 'imitate the best' based on sample size N.

Corollary 9 *Let $N \geq 2$. Then there exists an idiosyncratic two-armed bandit with $\pi_2 > \pi_1$ such that for any interior initial state $\tilde{p} \in \Delta^\circ(\{1,2\})$, given $x^0(F_{IB}) = x^0(F_{ib}) = \tilde{p}$ there exists $T > 0$ such that $x_2^t(F_{IB}) < x_2^t(F_{ib})$ for $t > T$.*

Proof. Consider the bandit from the proof of Theorem 8 based on λ such that $\frac{2}{N+1} < \lambda^{N-1} < 1$. Since both F_{ib} and F_{IB} are strictly improving, eventually $x_2^t(F_{IB})$ and $x_2^t(F_{ib})$ are close to 1 provided that $x_2^0(F_{IB}) = x_2^0(F_{ib}) > 0$. Close to $(0,1)$ the step size $x^{t+1} - x^t$ goes to 0 and hence x^{t+1}can be approximated by its linearization, i.e.,

$$x_2^{t+1} \approx x_2^t(1 + \alpha\beta(F)) \ ,$$

where $\beta(F_{IB}) = F_{IB}^2(1, 2_N) - NF_{IB}^1(2_N, 1)$ and $\beta(F_{ib}) = F_{ib}^2(1,2) - NF_{ib}^1(2,1)$, hence, for $n \in \mathbb{N}$,

$$\ln\left(x_2^{t+n}\right) \approx \ln\left(x_2^t\right) + n\ln(\alpha\beta(F)).$$

By construction of the bandit, $\beta(F_{ib}) > \beta(F_{IB})$ so there exists t_0 such that $x_2^{F_{ib}}(t) > x_2^{F_{IB}}(t)$ for all $t > t_0$. Thus the proof is complete. ■

However, one should not overestimate the strength of the statement of Corollary 9, at least concerning the bandit used for proving the statement. In simulations for double sampling ($N = 2$) we have seen that 'imitate if better' overtakes 'imitate the best' in the above example only after a large proportion of individuals has already adapted the efficient action. Given $x_2(0) = \mu$ let $z(\mu) = x_2^{F_{ib}}(t^*(\mu))$ where $t^*(\mu) = \min\{t \in \mathbb{N} : x_2^{F_{ib}}(t) > x_2^{F_{IB}}(t)$ given $x_2^{F_{ib}}(0) = x_2^{F_{IB}}(0) = \mu\}$. When $N = 2$ and $\lambda = 0.85$, then $z(\cdot) > 0.885$, $z(0.75) > 0.997$ and $z(0.5) > 0.9997$. Hence, starting with equal proportions of the two actions, only the last 0.3% of the individuals adapt the efficient action faster with the rule 'imitate if better'.

In the following we show that 'imitate the best' adapts the action that can generate the highest payoffs (if such an action exists) when the sample

size is sufficiently large. This of course is a bad property when noise is not necessarily idiosyncratic since bandits are easily constructed in which 'imitate the best' drives the efficient action from the population.

Corollary 10 *Consider a two-armed bandit in which action i can achieve payoffs that are higher than any payoffs that are achievable by action j, $i, j \in A = \{1, 2\}$, i.e., there exists x such that $P_i(x) > 0$ and $P_j(x') = 0$ for $x' \geq x$. Then there exists $N' \in \mathbb{N}$ such that $N > N'$ and $x^0 \in \Delta^\circ(\{1, 2\})$ implies $x_i^t(F_{IB}) \to 1$ as $t \to \infty$.*

Proof. Assume that $i = 2$. Let $\mu = P(X_2 > x')$ and $\epsilon = 1$. Then following the proof of Theorem 8 there exists N' such that $N > N'$ implies $jF_{IB}^2(1_j, 2_{N+1-j}) - (N - j + 1) F_{IB}^1(2_{N+1-j}, 1_j) > 0$ for $j = 1, .., N$. Hence, $x_i^t \to 1$ as $t \to \infty$ if $x_2^0 > 0$. ■

4.2 'Imitate the Best Average'

'Imitate the best average' (Bruch [3], Ellison and Fudenberg [4]) is the rule that specifies to adapt the action that achieved the highest average payoff in the sample modulo some additional assumptions when there are ties. Given $u \in \mathbb{R}^{N+1}$ and $a \in A^{N+1}$ let

$$q(a, u) = \arg \max_{j \in \{a(i), i=1,..,N+1\}} \frac{\sum_{i:a(i)=j} u(i)}{|\{i : a(i) = j\}|}.$$

We refer to a behavioral rule as the rule *'imitate the best average'*, and denote this rule by F_{BA}, if

$$F_{BA}(a(1), u(1), .., a(N+1), u(N+1))_j = \frac{1}{|q(a, u)|}$$

when $j \in q(a, u)$ and $a(1) \notin q(a, u)$, and otherwise

$$F_{BA}(a(1), u(1), .., a(N+1), u(N+1))_{a(1)} = 1.$$

It turns out that 'imitate the best average' is not always able to learn which action is the best among those played in the population. This result is independent of the tie breaking rule. This phenomenon occurs, given some

additional conditions, when noise is too large, expected payoffs are too close or when the sample is too big. In the following example we see that 'imitate the best average' fails to be payoff increasing in very simple bandits provided that the magnitude of the noise is sufficiently large.

Example 11 *Fix $N \geq 2$, π_1 and π_2 with $\pi_2 > \pi_1$. Let $\mu > \frac{N}{2}(\pi_2 - \pi_1)$. Consider an idiosyncratic two-armed bandit with idiosyncratic noise Z such that $Z(-\mu) = Z(\mu) = \frac{1}{2}$. In the following we will show the population learns the efficient action, i.e., $x_2^t \to 1$ as $t \to \infty$, only in the trivial situation in which the system starts with everyone playing the efficient action, i.e., if $x^0 = (0,1)$.*

Since $\pi_1 + \mu > \frac{1}{N}[(N-1)(\pi_2 + \mu) + \pi_2 - \mu]$,

$$F_{BA}^2(1, 2_N) = 1 - F_{BA}^1(2_N, 1) = \frac{1}{2} + \left(\frac{1}{2}\right)^{N+1} \leq \frac{5}{8} < \frac{N}{N+1}.$$

When $x \in \Delta^\circ(\{1,2\})$ is such that x_1 is sufficiently small we therefore obtain from

$$EIP(F_{BA}, x) = \alpha\left[F_{BA}^2(1, 2_N) - NF_{BA}^1(2_N, 1)\right] x_1 (x_2)^N (\pi_2 - \pi_1) + o\left((x_1)^2\right) \tag{7}$$

that $EIP(F_{BA}, x) < 0$ and $x_2^{t+1}(F_{BA}, x) < x_2$. Hence, $x_2^t \to 1$ as $t \to \infty$ implies $x_2^0 = 1$.

Next we show that 'imitate the best average' has the tendency to lead the population to equal proportions when expected payoffs are close in two-armed bandits based on non-atomic symmetric idiosyncratic noise.

Theorem 12 *Consider a single behavior population based on 'imitate the best average' for some $N \geq 2$. Consider a two-armed-bandit with non-atomic idiosyncratic noise such that $Z(u) = Z(-u)$ for all u. Then for any $\epsilon < 0.5$ there exists $\delta > 0$ such that $|\pi_1 - \pi_2| < \delta$ and $x_2^0 \in (0,1)$ implies $x^t \to x^*$ as $t \to \infty$ where $x_2^* \in (\epsilon, 1-\epsilon)$, especially, 'imitate the best average' is not payoff increasing in this gamble when $|\pi_1 - \pi_2| < \delta$.*

Especially, in such bandits, $x_2^t \to 0.5$ for $\pi_1 = \pi_2$ and $x_2^0 \in (0,1)$, a fact mentioned by Ellison and Fudenberg [4] (p. 108) for normally distributed noise.

Proof. Let $g_j(\pi_2 - \pi_1) = F_{BA}^2(1_j, 2_{N-j+1})$. Let $\bar{X}_j$ be the average of j independent realizations of the noise and Y_{N-j+1} of $N-j+1$ realizations. Then $g_j(\pi_2 - \pi_1) = P\left(\bar{Y}_{N-j+1} - \bar{X}_j > \pi_1 - \pi_2\right)$. Since $\bar{Y}_{N-j+1} - \bar{X}_j$ is symmetric and non-atomic, g_j is continuous and $g_j(0) = 0.5$. Consider $x_2^0 \in (0,1)$. Since $|A| = 2$ there can be no cycles and hence $x^* = \lim_{t\to\infty} x_2^t$ exists, moreover, $x^*(\pi_2 - \pi_1)$ is as a continuous function of $\pi_2 - \pi_1$. Since $x^*(0) = 0.5$ the proof is complete. ■

Of course the inability of 'imitate the best average' to always lead to the efficient action also holds for larger action sets.

Corollary 13 *When $|A| \geq 2$ then there exists a bandit in which for each interior initial state the population converges to a state in which actions are played that do not achieve maximal expected payoffs.*

Proof. Consider an idiosyncratic bandit in which $\pi_2 = \pi_i$ for all $i > 2$. For such a bandit, for any $t \in \mathbb{N}$ and $\tilde{p} \in \Delta(A)$, $x_1^t|_{x^0=\tilde{p}} = x_1^t|_{x^0=r(\tilde{p})}$ where $r(x) \in \Delta(A)$ such that $r_1 = x_1$, $r_2 = \sum_{i\geq 2} x_i$ and $r_i = 0$ for $i > 2$. Hence, two armed bandits from Theorem 12 can be extended to multi-armed bandits to prove the statement. ■

Even if 'imitate the best average' (based on sample size N) is payoff increasing in a given idiosyncratic bandit, increasing the sample size sufficiently causes this property to fail, provided that noise is sufficiently noisy. This property (also noted by Ellison and Fudenberg [4], p. 104) stands in contrast to the rule 'imitate the best' where it is advantageous to sample many individuals (Theorem 8).

Theorem 14 *Consider an idiosyncratic two-armed bandit in which it is possible for an individual with the worse action to achieve a payoff above the expected payoff of the better action, i.e., $P_j(u) > 0$ for some $u > \pi_i > \pi_j$. Then there exists $N_0 \in \mathbb{N}$ such that for $N > N_0$, 'imitate the best average' is not payoff increasing.*

Proof. Assume that $\pi_2 > \pi_1$. Let $\bar{X}_N$ be the average of N independent realizations of the noise. Let $\epsilon > 0$ be such that $\mu = \int_{u>\pi_i+\epsilon} dP_j(u) > 0$. Then there exists N_0 such that $N > N_0$ implies $P\left(\pi_2 + \bar{X}_N < \pi_1 + \epsilon\right) > 1-\mu$ and $1-(1-\mu)\mu < \frac{N}{N+1}$. Then $F^2_{BA}(1, 2_N) \leq 1-(1-\mu)\mu < \frac{N}{N+1}$ and following (7), $EIP(F_{BA}, x) < 0$ when x_2 is large but strictly below 1. ■

Finally we come to some "good" properties of 'imitate the best average'.

Theorem 15 *In a single behavior population based on 'imitate the best average' the efficient action is never eliminated in an idiosyncratic two-armed bandit when starting in interior initial states.*

Proof. Consider an idiosyncratic two-armed bandit with $\pi_2 > \pi_1$. Since 'imitate the best' is payoff increasing, $NF^2_{IB}(1_N, 2) - F^1_{IB}(2, 1_N) \geq 0$. Hence, the proof is complete once we show that

$$NF^2_{IB}(1_N, 2) - F^1_{IB}(2, 1_N) \leq NF^2_{BA}(1_N, 2) - F^1_{BA}(2, 1_N) \ .$$

If an individual switches according to F_{IB} from playing action 1 to action 2 when only one individual in the sample is playing action 2, then he will also switch with F_{BA}, hence $F^2_{IB}(1_N, 2) \leq F^2_{BA}(1_N, 2)$. Similarly, $F^2_{IB}(2, 1_N) \geq F^2_{BA}(2, 1_N)$ and the proof is complete. ■

Although 'imitate the best average' is not able to completely learn the efficient action (Theorem 14), we show that most individuals adapt the efficient action when the sample size is sufficiently large. Especially, this phenomenon also holds when noise is not idiosyncratic.

Theorem 16 *Let* $A = \{1, 2\}$. *For any* $\epsilon > 0$ *there exists* $N' \in \mathbb{N}$ *such that* $N > N'$, $x^0 \in \Delta^\circ(A)$ *and* $\pi_i > \pi_j$ *implies* $\lim_{t\to\infty} x^t_i(F_{BA}) > 1-\epsilon$.

Ellison and Fudenberg [4] (p. 120) mention this property of 'imitate the best average' in a slightly different setting. Their proof is however incomplete, it lacks the arguments necessary to show that the population does not converge arbitrarily close to the inefficient state.

Proof. Assume that $\pi_2 > \pi_1$. If $P(X_2 = \pi_2) = 1$ then the statement is trivial. Assume therefore that $P(X_2 = \pi_2) < 1$. Let Y_k be independent

realizations of the noise of action 2 $(X_2 - \pi_2)$ and let $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$. We will first show that there exists $\epsilon_1 > 0$ such that $P\left(\bar{Y}_n > 0\right) > \epsilon_1$ for all $n \in \mathbb{N}$. Since $\bar{Y}_n = \sqrt{\frac{VarX_2}{n}}\frac{\sum_{k=1}^{n} Y_k}{\sqrt{nVarX_2}}$, by the central limit theorem we obtain that $\lim_{j\to\infty} P\left(\bar{Y}_n > 0\right) = \frac{1}{2}$. This together with the fact that $P\left(\bar{Y}_n > 0\right) > 0$ for all n implies that there exists an $\epsilon_1 > 0$ with the desired property. Similarly, let $\bar{Z}_m$ be the average of m independent realizations of the noise of action one. Then following the same arguments we obtain that there exists $\epsilon_2 > 0$ such that $P\left(\bar{X}_m \leq 0\right) > \epsilon_2$ for all $m \in \mathbb{N}$.

Let $\gamma_k := kF^2_{BA}\left(1_j, 2_{N-j+1}\right) - (N-k+1)\,F^2_{BA}\left(2_{N-k+1}, 1_k\right)$. Then

$$\begin{aligned} F^2_{BA}\left(1_k, 2_{N-k+1}\right) &= P\left(\bar{Y}_{N-k+1} - \bar{Z}_k > \pi_1 - \pi_2\right) \\ &\geq P\left(\bar{Y}_{N-k+1} - \bar{Z}_k > 0\right) > \epsilon_1\epsilon_2 > 0. \end{aligned}$$

Hence,

$$\gamma_k > (N+1)\,\epsilon_1\epsilon_2 - (N-k+1) > 0$$

if

$$\frac{k}{N+1} \geq 1 - \epsilon_1\epsilon_2.$$

Given $\epsilon > 0$ there exists N^* such that $N > N^*$ and $\epsilon < \frac{k}{N+1} < 1 - \epsilon$ implies that in at least $1-\epsilon$ proportion of the samples $(1_k, 2_{N-k+1})$ action 2 achieves the higher average payoff. Hence, $\gamma_k \geq (1-\epsilon)\,k - \epsilon\,(N+1-k) = k - \epsilon\,(N+1) > 0$ when $N > N^*$.

Setting $\epsilon = \epsilon_1\epsilon_2$ we obtain that $\gamma_k > 0$ for all $\epsilon < \frac{k}{N+1}$. This together with the fact that $\gamma_k \geq kF^1_{BA}\left(2_k, 1_{N-k+1}\right) - (N-k+1)\,F^1_{BA}\left(1_{N-k+1}, 2_k\right)$ implies that $x_2^{t+1} - x_2^t > 0$ when $x_2^t \leq \frac{1}{2}$. Hence, given $x_2^0 > 0$, eventually $x_2^t > \frac{1}{2}$.

The rest of the proof follows the arguments of Ellison and Fudenberg [4] (p. 120). Given $\epsilon' > 0$ let $N' > N^*(\epsilon')$ be such that for $N > N'$ and in any state $x \in \left(\frac{1}{2}, 1-\epsilon'\right)$ at least $(1-\epsilon')$ proportion of the individuals sample so many individuals using each action that average payoffs are close enough to expected payoffs such that they choose the better action 2. Consequently, for $N > N'$, $\lim_{t\to\infty} x_2^t > \frac{1}{2}$ provided that $x_2^0 > 0$ and there is no stationary point in $\left(\frac{1}{2}, 1-\epsilon'\right)$ which completes the proof. ■

4.3 Adapted Proportional Imitation

For sample size two Schlag [6] characterizes a rule that is best at performing better than (i.e., dominating) any proportional imitation rule in all bandits in which payoffs are contained in a pre-specified interval $[\pi^-, \pi^+]$. This so-called adjusted proportional imitation rule is defined as follows. Let $\sigma^* : [\pi^-, \pi^+] \to \mathbb{R}^+$ be the linearly decreasing function such that $\sigma^*(\pi^-) = \frac{2}{\pi^+ - \pi^-}$ and $\sigma^*(\pi^+) = \frac{1}{\pi^+ - \pi^-}$, i.e.,

$$\sigma^*(u) = \frac{1}{\pi^+ - \pi^-} + \frac{\pi^+ - u}{(\pi^+ - \pi^-)^2} \text{ for } u \in \left[\pi^-, \pi^+\right].$$

The behavioral rule $\hat{F}$ is called the *adjusted proportional imitation rule* if $\hat{F}(i, u, \{j, v, k, w\})_j = \frac{1}{2}\sigma^*(w)[v - u]_+$, $\hat{F}(i, u, \{j, v, k, w\})_k = \frac{1}{2}\sigma^*(v)[w - u]_+$ and $\hat{F}(i, u, \{j, v, j, w\})_j = \frac{1}{2}\sigma^*(w)[v - u]_+ + \frac{1}{2}\sigma^*(v)[w - u]_+$, $|\{i, j, k\}| = 3$; where $[u]_+ = u$ if $u \geq 0$ and $[u]_+ = 0$ if $u < 0$.

Calculation shows that

$$\hat{F}^j(i, j, j) - 2\hat{F}^i(j, i, i) = \sigma^*(\pi_j)(\pi_j - \pi_i). \tag{8}$$

We show that the adapted proportional imitation rule and 'imitate the best' (sample size 2) can not be ranked according to dominance in all idiosyncratic gambles. This contrasts our dominance result for sample size one (Theorem 3).

Remark 2 *There are idiosyncratic bandits in which 'imitate the best' dominates the adapted proportional imitation rule and there are idiosyncratic bandits in which the opposite is true.*

Proof. Let $\pi^- = \pi_1 + \lambda - 1$, $\pi^+ = \pi_2 + \lambda$ and $\pi_2 > \pi_1 > \pi_2 - 1$ and consider the bandit in the proof of Theorem 8. Then

$$\hat{F}^2(1, 2, 2) - 2\hat{F}^1(2, 2, 1) < \frac{1}{2} + \frac{\lambda}{4}$$

$$2\hat{F}^2(1, 1, 2) - \hat{F}^1(2, 1, 1) < \frac{3}{4} + \frac{\lambda}{4}$$

and hence, if $\lambda < 0.5$

$$\begin{aligned} \hat{F}^2(1,2,2) - 2\hat{F}^1(2,2,1) &< \frac{5}{8} < F^2_{IB}(1,2,2) - 2F^1_{IB}(2,2,1) \\ 2\hat{F}^2(1,1,2) - \hat{F}^1(2,1,1) &< \frac{7}{8} < 2F^2_{IB}(1,1,2) - F^1_{IB}(2,1,1) \end{aligned}$$

which means that F_{IB} dominates $\hat{F}$.

On the other hand, if $\pi_2 - \pi_1 = 0.95$ and $\lambda = 0.6$ then calculation shows that $\hat{F}$ dominates F_{IB}. ■

5 Discussion

Much of the analysis concerning the rule 'imitate the best average' is based on results and side remarks from a model by Ellison and Fudenberg [4]. They consider the behavioral rule 'imitate the best average' for varying sample sizes in non-stationary bandits with idiosyncratic noise. For the modelling of idiosyncratic noise they restrict attention to the normal distribution. Non-stationarity arises through common shocks, a realization before each round of play determines for the entire population the value of the expected payoffs of each action in that round. Ellison and Fudenberg [4] obtain that efficient learning only tends to occur when each individual receives very little information ([4], p. 95). This would also be the result of the analysis without common shocks (as we preformed above) if we had not considered the alternative rule 'imitate the best'. Ellison and Fudenberg [4] do not justify why they choose to model individuals that follow the rule 'imitate the best average'. Clearly, 'imitate the best' is an equally intuitive behavior, and in fact, it is far more easier to implement.

Our approach is to compare the performance of various behavioral rules before deriving implications for the behavior of society. What sense do results make if they are based on individual behavior that no one would want to follow? Basis for our analysis are the concepts of strictly payoff increasing and of strictly improving. Behind these concepts lies our intuition that evolution will select individual behavior. As such we rely on deriving properties of rules that may survive evolution. Björnerstedt and Schlag [2] construct a model

in which behavior under single sampling (sample size one) is selected via the replicator dynamic. A necessary condition for a single behavior population to be robust against the entry of an alternative behavior is that the incumbent rule is payoff increasing. The underlying intuition is not limited to single sampling nor to the replicator dynamic (see also Schlag [7]). Rules that have a tendency to adapt bad actions are at a disadvantage compared to rules that are payoff increasing in a model in which reproductive success is based on current performance. More specifically, a rule that is not payoff increasing when most of the individuals use the efficient action (like 'imitate the best average' in many bandits) will not survive in the majority of the population when it is entered by the proportional imitation rule (the relevant proofs in [2] do not rely on the fact that there is only single sampling).

Therefore we find that 'imitate the best' performs superior to 'imitate the best average' in idiosyncratic bandits, especially, efficient learning occurs with a very intuitive rule for any sample size.

References

[1] R. M. Axelrod, *The Evolution of Cooperation,* Basic Books, New York, 1984.

[2] J. Björnerstedt and K. H. Schlag, "On the Evolution of Imitative Behavior," Mimeo, University of Bonn.

[3] E. Bruch, "Evolution von Kooperation in Netzwerken," Diplomarbeit, University of Bonn, 1993.

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

[5] M. Malawski, "Some Learning Processes in Population Games," Inaugural-Dissertation, University of Bonn, 1989.

[6] K. H. Schlag, "Which one should I imitate?," University of Bonn, Disc. Paper **B-365**, Bonn, 1996.

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