How to Choose

-

A Boundedly Rational Approach to Repeated Decision Making[1]

Karl H. Schlag[2]

May 2001
Very preliminary

[1]The author would like to thank Dorothea Herreiner and Ed Hopkins for helpful comments.

[2]Economics Department, European University Institute, Via dei Roccettini 9, 50016 San Domenico di Fiesole, Italy, Tel: 0039-055-4685251, email: schlag@iue.it

Abstract

Consider an individual repeatedly facing a decision in which he has two actions to choose from. All the individual knows ex-ante is the bounded interval that contains the random payoffs generated. We propose a distribution free method for selecting among behavioral rules for any given discount factor. Selection of behavior is illustrated for simple learning rules that are based on automata where transitions are linear in payoffs and only occur to "neighboring states".

Keywords: improving, maximizing, reinforcement, bounded rationality.

JEL classification: D81, D83.

1 Introduction

Decision making is arguably the most basic task in economics and sociology. The prevalent theory (von Neumann-Morgenstern 1944, Savage 1972) requires for the decision maker to assess a prior probability to any state that may occur. Consequently, finding optimal choices is extremely intricate (if not unsolvable for the average economic theorist) even in the simplest situations involving repeated decision making. Simon (1955, 1982) was among the first to call for alternative models of economic decision making. A prominent approach to boundedly rational decision making is to specify a more or less arbitrary parametrized functional form for a learning rule that does not rely on prior probability distributions, e.g., the Cross Learning Rule (Cross, 1973, see also Bush and Mosteller, 1955), the Payoff Sum Learning Rule[1] (Erev and Roth, 1998) and the Logit Choice model (Block and Marshak, 1960). General models for selecting behavior have been developed for individuals who can observe behavior of others (e.g., Schlag, 1998) or who can recall own experience in similar decisions (Gilboa and Schmeidler, 1995).

Erev and Roth (1998) emphasize that the Payoff Sum Learning Rule is not the only plausible ones, just that its generalizations are very effective in predicting play in games. Never-the-less, this rule is prone to the criticism of being arbitrary. Even observations of human decision making that have been the basis of the Cross Learning Rule (see Bush and Mosteller, 1955) can only reveal qualitative, not quantitative properties. Small differences can drastically change the predicted long run outcomes (e.g., see Arthur, 1993). Our aim is set up a theory for selecting boundedly rational behavior to avoid such problems. The idea is to select among a class of simple rules the ones with the most appealing theoretical properties. Given an understanding that behavior is not necessarily adjusted to each situation, we search for rules with universal properties. The selected rule should perform well when repeatedly facing any decision that involves two actions where each action yields a payoff in $(0, 1)$. We consider the simplest informational setting and assume that after each choice the individual only observes the random payoff realized by the chosen action.[2]

Two major criteria for evaluating the performance of decision rules can be found in the literature. A rule is *maximizing* (Börgers et al. 1998) if the action with the highest expected payoffs is selected in the long run. A rule is *improving* (Schlag 1998, or absolutely expedient, Börgers and Sarin, 1997) if expected payoffs obtained in the next round are larger than the expected payoffs of the current round conditional on the present state. The first condition has the flavor of an infinitely patient individual while the second seems more appropriate for a myopic individual. Both conditions are required to hold in any decision, unlike the classic

[1] Terminology due to (Selten 1999).

[2] An alternative setting is to assume that a random payoff of each action including those not chosen is observed (see Rustichini, 1999, Easley and Rustichini, 1999).

approach where priors lead to taking weighted averages over possible decisions.

Rustichini (1999) shows that the Payoff Sum Learning Rule is maximizing. Arthur (1993) provides a similar rule with this property. Rustichini's analysis also reveals that the Payoff Sum Learning Rule is not improving. Börgers and Sarin (1997) find that the Cross Learning Rule is improving but that it is not maximizing. Börgers et al. (1998) characterize the set of improving rules and show that these are generalizations of the Cross Learning Rule. They find that any of these rules will select the best action in the long run with arbitrarily high probability given that the adjustment between rounds is sufficiently sluggish.

The present paper uses a similar approach to the one followed in (Schlag, 1998). The innovation is that we weaken objectives by measuring expected payoffs in a given round from the ex-ante viewpoint of the individual before he makes the first choice. We refer to this context by using the term "ex-ante", e.g., *ex-ante improving*. Dominance arguments are used to discriminate among ex-ante improving rules. Finally, a maxmin type of argument is used to compare "ex-ante undominated" rules.

We apply our selection technique to a simple class of rules that we call *linear reinforcement rules* and that are based on automata. There are a countable set of states indexed by the integers. In each state the individual chooses action one with a given probability where this probability is increasing in the index of the state. Transitions between states only occur to the neighboring states and involve linear transition probabilities. The only difference to the basic structure underlying the Cross Learning Rule is that here transition only occurs to a limited set of states.

We first consider selection of rules with either two or four states and then consider rules with a countable set of states. For each class of rules we consider selection both under myopia and under complete patience. In addition, for the rules with only two states we also consider intermediate discount factors.

The rest of the paper is organized as follows. Section two introduces the basic decision problem together with the relevant existing rules from the literature. Section three explains the criteria for selecting rules. In Section four linear reinforcement rules are defined. In Section five linear reinforcement rules for two and four states and various discount parameters are selected. Section six then considers selection when there is a countable number of states. Section seven contains the conclusion and the appendix contains the proof for selection given four states.

2 The Setting

Consider a single individual that repeatedly faces a decision with two actions enumerated 1 and 2. Choice of action i yields a random payoff in a given bounded open interval according to the

payoff distribution P_i. Using a linear transformation we can assume without loss of generality that this interval equals $(0,1)$. The set of such decision problems will be denoted by $\mathcal{D}$.

The above decision is like a two-armed bandit except that there is no prior distribution over the possible payoff distributions P_i. Instead, we assume as in Schlag (1998) that the individual only knows that payoffs are contained in $(0,1)$, i.e., the individual could be facing any decision with this property.

Let $\pi_i = \int x dP_i(x)$ denote the expected payoff of choosing action i. Assume that payoffs realized by choosing action i in round n are independent of previous choices and realizations. We assume that the individual prefers higher expected payoffs and discounts payoffs over time with a discount factor $\delta \in [0,1]$. In this sense we will call action i the *better* (*worse*) action if $\pi_i > \pi_j$ ($\pi_i < \pi_j$) where $\{i,j\} = \{1,2\}$. Our analysis also applies to agents that are not risk neutral by replacing payoffs with von Neumann-Morgenstern utilities.

A *behavioral rule* is the formal description of how an individual makes his choice as a function of his previous experience. Here we model this rule as a triple (S,f,g) where S is a countable set of states, $f : \emptyset \cup S \times \{1,2\} \times (0,1) \to \Delta S$ is the transition function and $g : S \to \Delta\{1,2\}$ is the choice function where ΔZ denotes the set of probability distributions with support Z. The interpretation is as follows. When in state s, $g(s)_i$ is the probability of choosing action i. $f(\emptyset)_s$ is the probability of being in state s in the first round. After choosing action i in state s and receiving a payoff x, $f(s,i,x)_{s'}$ is the probability of being in state s' in the next round. Consequently, $z'(s) = \sum_{i=1}^{2} g(s)_i \int_{x\in(0,1)} \sum_{s'\in S} f(s,i,x)_{s'} g(s') dP_i(x)$ is the expected mixed action chosen in the next round after being in state s. Let $z^{(n)}$ denote the expected action chosen in round n from the standpoint of before the first choice. So $z^{(1)} = \sum_{s\in S} f(\emptyset)_s g(s)$ and $z^{(2)} = \sum_{s\in S} f(\emptyset)_s z'(s)$. Let $z^{(\infty)} = \lim_{n\to\infty} z^{(n)}$ whenever this limit exists.

2.1 Some Examples

In the following we present the most popular boundedly rational decision making rules.

Under the *Cross Learning Rule* (see Börgers and Sarin, 1997), $S = \Delta\{1,2\}$, $g(s) = s$ and $f(s,i,x)_i = x + (1-x)s_i$. Notice that this rule is based on *positive reinforcement* as $f(s,i,x)_i \geq s_i$. Then

$$z'(s)_1 = s_1(\pi_1 + (1-\pi_1)s_1) + (1-s_1)(1-\pi_2)s_1 = s_1 + (\pi_1 - \pi_2)s_1(1-s_1). \qquad (1)$$

Thus, given $f(\emptyset)_{(0.5,0.5)} = 1$ we obtain $z^{(2)} = 0.5 + 0.25(\pi_1 - \pi_2)$.

Under the *Payoff Sum Learning Rule* of Erev and Roth (1998), an individual essentially chooses an action in a given round with probability proportional to the sum of payoffs it has generated in the past. More specifically, let $S_0(i) > 0$ be given and let $S_n(i) - S_0(i)$ be the sum of payoffs obtained from choosing action i up to and including round n. The rule prescribes

to choose action i in round $n+1$ with probability $S_n(i)/(S_n(1)+S_n(2))$. Notice that this rule is also based on positive reinforcement. Consider behavior where the individual randomizes equally likely in the first round, i.e., $z^{(1)} = (0.5, 0.5)$,. Then $S_0(1) = S_0(2) = b$. Consider a decision in which $P_1(1) = 1 - P_1(0) = \lambda$ and $P_2(y) = 1$. Then

$$\begin{aligned} z^{(1)} &= 0.5 \\ z^{(2)} &= 0.5\frac{b}{2b+y} + 0.5\lambda\frac{b+1}{2b+1} + 0.5(1-\lambda)\frac{1}{2} = 0.5\left(\frac{b}{2b+y} + 0.5 + \lambda\frac{1}{2(2b+1)}\right) \end{aligned}$$

and it is easily verified that $z^{(2)} < z^{(1)}$ holds when $\pi_1 = \pi_2$ (i.e., $\lambda = y$). Under the *Logit Rule* (Block and Marshak, 1960) an action is chosen with probability proportional to the exponential transformation of the sum of payoffs it has previously generated.

An alternative rule is to choose each action equally often in the first $2k$ rounds and then to choose the action that yielded the highest sum of payoffs. This rule is used by Osborne and Rubinstein (1998) to analyze boundedly rational play in games. With appropriate randomization this yields $z^{(i)} = (0.5, 0.5)$ for $i \leq 2k$. Consider the same decision as in the previous paragraph and assume $y > 0.5$. For $k = 1$ we obtain $z^{(3)} = \lambda$, for $k = 2$ we obtain $z^{(5)} = \lambda^2$.

3 Selection

For $s \in S$ let $E\pi(s) = g(s)_1 \pi_1 + g(s)_2 \pi_2$ be the expected payoff achieved in state s. We will simplify notation and for $n \in \mathbb{N}$ let $E\pi(n) = z_1^{(n)}\pi_1 + z_2^{(n)}\pi_2$ be the expected payoff in round n. Let $E\pi'(s) = z_1'\pi_1 + z_2'\pi_2$ be the expected payoff in the next round after being in state s. Let $E\pi^\delta = (1-\delta)\sum_{k=1}^{\infty}\delta^{k-1}E\pi(k)$ be the ex-ante future value of expected payoffs where ex-ante refers to the fact that payoffs are evaluated from the perspective before making the first choice. Let $E\pi^\delta(n) = (1-\delta)\sum_{k=n}^{\infty}\delta^{k-n}E\pi(k)$ be the ex-ante future value of expected payoffs for an individual in round n, so $E\pi^\delta(1) = E\pi^\delta$.

Our individual has no prior distribution over the probability distributions underlying the choice of a given action. Never-the-less, a Bayes'ian would specify some subjective distribution and then aim to maximize expected payoffs. Along the lines of Schlag (1998) we choose a distribution free approach.[3] The following definitions may seem a bit intricate but the latter analysis will make them clearer.

In the literature we find two relevant definitions. A decision rule f is called *improving* (or absolutely expedient, Narendra and Thathachar, 1989) if for any decision in $\mathcal{D}$ (i.e., for any payoff distributions P_i that yield payoffs in $(0,1)$) and for any state $s \in S$ that occurs with

[3] The artificial intelligence literature refers to a model free approach.

positive probability when using f,

$$E\pi'(s) \geq E\pi(s) \ , \tag{2}$$

i.e., conditional on the current state the payoffs from are expected to increase in each round. A decision rule f is called *maximizing* (Börgers et al. 1998)[4] if for any decision in $\mathcal{D}$

$$\lim_{n\to\infty} E\pi(n) = \max\{\pi_1, \pi_1\} \text{ a.s.},$$

i.e., almost surely the behavioral rule eventually chooses the better action.

Now we introduce our new definitions. In our setting the two actions are a priori identical which makes it plausible that an individual who is not willing or able to learn about payoffs chooses each action equally likely. Our first definition requires that the individual always does better than to ignore previous experience and to just randomize. In this sense, a behavioral rule f is called an *ex-ante learning rule* for discount factor δ if for any decision in $\mathcal{D}$ and any round n

$$E\pi^{\delta}(n) \geq (\pi_1 + \pi_2)/2.$$

Consider an individual using an ex-ante learning rule for $\delta = 0$. Then $z_1^{(1)} = 0.5$ and consequently, $E\pi^0(2) \geq E\pi^0(1) = (\pi_1 + \pi_2)/2$.

Our next definition concern situations where $\delta < 1$ and requires that ex-ante expected payoffs are monotone increasing. A behavioral rule f is called *ex-ante improving* for discount factor δ if for any decision in $\mathcal{D}$ and any round n,

$$E\pi^{\delta}(n+1) \geq E\pi^{\delta}(n) \ . \tag{3}$$

Notice that an ex-ante improving rule with $z_1^{(1)} = 0.5$ is an ex-ante learning rule.

Consider the special case where $\delta = 0$. Then the definition of ex-ante improving is analogous to the above definition of improving except that here payoffs are calculated ex-ante. It is equivalent to requiring that $z_i^{(n)}$ is monotone increasing in n whenever $\pi_i > \pi_j$. When the individual is myopic, dynamic inconsistency cannot arise. The plan of future behavior made in round n will not be regretted once round $n+1$ arises as decisions made after round n are only guided by payoffs achieved in round $n+1$. This is no longer necessarily true when $\delta > 0$. Now the rule has a sequential nature, anticipating when making decisions before round one the goals the individual will set in later rounds. Notice that when $\delta \in (0,1)$ then (3) is equivalent to

$$E\pi^{\delta}(n) \geq E\pi(n) \ ,$$

[4] This criterion is called "optimality" in (Rusticini, 1999) and in the machine learning literature (e.g. Narendra and Thathachar, 1989).

i.e., for any decision and in any round the individual prefers an ex-ante standpoint to apply the rule to never changing actions again in the future.

Next we introduce a concept of dominance. We say that f *ex-ante dominates* g for discount factor δ if for any decision there exists $n \in \mathbb{N} \cup \{\infty\}$ such that $E\pi_f^\delta(k) = E\pi_g^\delta(k)$ for $k < n$ and $E\pi_f^\delta(n) > E\pi_g^\delta(n)$. Accordingly, f is *ex-ante undominated* if f ex-ante dominates g whenever g ex-ante dominates f.

We immediately obtain the following relationships:

Remark 1 *(i) Consider a rule f that is both ex-ante improving and ex-ante undominated when $\delta = 0$. If g ex-ante dominates f for all sufficiently small δ then $E\pi_f^\delta(1) = E\pi_g^\delta(1)$ holds in any decision whenever δ is sufficiently small.*

(ii) If f is maximizing then f ex-ante dominates any other rule for $\delta = 1$.

Looking back at the rules presented in Section 2.1 we find that only the Cross Learning Rule is ex-ante improving. In fact, it follows immediately from (1) that the Cross Learning Rule is improving (see Börgers and Sarin, 1997).

Ex-ante dominance does not induce a complete order over the set of rules. In order to select unique behavior we add the following criterion for comparing ex-ante undominated rules.

$$\Delta_n(\pi_1, \pi_2) = \max\{\pi_1, \pi_2\} - E\pi^\delta(n)$$

measures the incentives for learning in round $n+1$. Let $\bar{\rho}$ be the smallest upper bound for Δ_n, i.e., $\bar{\rho} = \sup_{\mathcal{D}}\{\Delta_n(\pi_1, \pi_2)\}$. For given $\rho \in (0, \bar{\rho})$ one may choose to compare rules according to the minimal increase in expected payoffs among all decisions relative to the maximal increase Δ_n. For a given decision rule f and $n \geq 1$, let $d_f^{n+1} : (0, \bar{\rho}) \to [0, 1]$ be defined by

$$d_f^{n+1}(\rho) = \inf_{\mathcal{D}}\left\{E\pi^\delta(n+1) - E\pi^\delta(n) \text{ s.t. } \Delta_n = \rho\right\}.$$

We say that f *outperforms g for fixed differences* in round $n+1$ if $d_f^{n+1} \geq d_g^{n+1}$, i.e., if $d_f^{n+1}(\rho) \geq d_g^{n+1}(\rho)$ holds for all $\rho \in (0, \bar{\rho})$. Again this does not yield a complete order but we find that this condition suffices in most applications.

When calculating d_f^1 we now need a benchmark to compare the discounted future payoffs starting round one to. Here we choose $0.5(\pi_1 + \pi_2)$ which means that $\Delta_0 = 0.5|\pi_1 - \pi_2|$. Conditions substantially simplify when $\delta = 1$. Here, $d^1(\rho) = 2\rho \inf\left\{\left|z_1^{(\infty)} - 0.5\right| \text{ s.t. } |\pi_1 - \pi_2| = 2\rho\right\}$. f is *maximizing* (Börgers et al. 1998) if $\pi_i > \pi_j$ implies that $z_i^{(n)}$ converges to one almost surely as n tends to infinity (i.e., $z_i^{(\infty)} = 1$). Notice that any rule f with this property is ex-ante improving in round one and outperforms any other rule for fixed differences in round one. In this case, $d_f^1(\rho) = \rho$.

4 Linear Reinforcement

We will consider a specific class of behavioral rules. There is a countable set of states $S \subset \{s_i, i \in \mathbb{N} \cup -\mathbb{N}\}$ where $s_i \in S$ implies $s_{-i} \in S$. To simplify notation, let $\tau_i = g(s_i)_1$ be the probability that the individual plays action one in state s_i. We add the following assumptions. τ_i is increasing in i and $\tau_i = 1 - \tau_{-i}$ for all $i \in \mathbb{N}$. $f(\emptyset)_{s_0} = 1$ if $s_0 \in S$, otherwise $f(\emptyset)_{s_1} = f(\emptyset)_{s_{-1}} = 0.5$. After being in state s_i the individual transits to a neighboring state in $\{s_{i-1}, s_i, s_{i+1}\}$ using the following linear transition rules: $f(s_i, 1, x)_{i+1} = f(s_{-i}, 2, x)_{-i-1} = \alpha_i x$ and $f(s_i, 1, x)_{i-1} = f(s_{-i}, 2, x)_{-i+1} = \gamma_i (1 - x)$ where $\alpha_i, \gamma_i \in [0, 1]$. Of course, α_i and γ_i only yields switching with positive probability if $\tau_i > 0$. A behavioral rule $(S, f, g) = (S, f, \tau)$ with these properties will be called a *linear reinforcement rule*. We say that (S, f, τ) uses *positive reinforcement* in state s_i after receiving payoff x from choosing action one if $\tau_i > 0$, $\alpha_i x (\tau_{i+1} - \tau_i) + \gamma_i (1 - x)(\tau_{i-1} - \tau_i) \geq 0$ and $\tau_{i-1} = \tau_{i+1} > 0$ implies $\alpha_i x \geq \gamma_i (1 - x)$, i.e., (i) if the player is more likely to choose the same action (here action one) in the next round and (ii) if the neighboring states have the same values of τ then the player will be more likely to switch to a state with a higher index than to one with a lower index. (S, f, τ) uses *negative reinforcement* in state s_i after receiving payoff x from choosing action one the two inequalities in the definition above are reversed. Analogously these terms are defined with respect to behavior after choosing action two. We add the term "maximal" if the associated variables are equal to one. $x_0 \in [0, 1]$ will be called the *aspiration level* of action k in state s_i if all payoffs above/below x_0 of action k receive positive/negative reinforcement in state s_i. Notice that $\gamma_i = 0$ ($\alpha_i = 0$) holds if and only if (S, f, τ) uses positive (negative) reinforcement in state s_i after any choice and payoff realization.

Notice the Cross Learning Rule is very similar to a linear reinforcement rule except for the following. In a linear reinforcement rule, transition is random to one of two given states where payoffs received influence transition probabilities. Under the Cross Learning Rule transition occurs to a unique state where this state depends on the payoff received in the present round. In fact, there is a continuum of possible states in the second round if all payoffs in $(0, 1)$ are in the support of P_1 and P_2. This is one of the complex features of the Cross Learning Rule that has lead us to define and analyze the simpler class of linear reinforcement rules.

We immediately obtain:

Proposition 1 *Linear reinforcement rules are ex-ante learning rules.*

Proof. Let $p_n(i)$ be the probability of being in state s_i in round n. Then $z_1^{(n)} = \sum_{s_i \in S} p_n(i) \tau_i$. If $\pi_1 = \pi_2$ then the symmetry property of linear reinforcement rules implies that $p_n(-i) = p_n(i)$ holds for all n and all $s_i \in S$. Thus, $z_1^{(n)} = 0.5$ when $\pi_1 = \pi_2$. For given π_2, if π_1 is increased

starting at $\pi_1 = \pi_2$ then the transition probabilities to states with higher (lower) indices are increased (decreased). Consequently, $p_n(i) \geq p_n(-i)$ for $i > 0$ which implies $z_1^{(n)} \geq 0.5$ and hence $E\pi(n) \geq 0.5(\pi_1 + \pi_2)$. As this holds for any round n, we obtain the ex-ante learning property. ■

5 Bounded Complexity

The number of states is a reasonable measure of the complexity of a linear reinforcement rule. In this section we illustrate which of the simplest rules with either two or four states is selected.

5.1 Two states

Below we select the rule with maximal negative reinforcement.

Proposition 2 *There is a two state linear reinforcement rule that ex-ante dominates all other two state linear reinforcement rules if* $\delta \leq 7/9 = .\bar{7}$ *and which outperforms all other two state linear reinforcement rules for fixed differences if* $7/9 < \delta \leq 1$. *This rule satisfies* $(\tau_1, \gamma_1) = (1, 1)$, *generates*

$$\begin{aligned} z_1^{(n)} &= 0.5 + 0.5(\pi_1 - \pi_2)\frac{1 - (\pi_1 + \pi_2 - 1)^{n-1}}{2 - \pi_1 - \pi_2} \\ E\pi^{\delta}(1) &= 0.5(\pi_1 + \pi_2) + \delta(\pi_1 - \pi_2)^2 \frac{0.5}{1 + \delta(1 - \pi_1 - \pi_2)} \end{aligned}$$

and satisfies $d^1(\rho) = 2\rho^2\delta/(1 + \delta(1 - 2\rho))$ *for* $0 < \rho < 0.5$ *where* $d^{1\prime\prime}(0) = 4\delta/(1 + \delta)$.

Proof. Let $p_n(1)$ be the probability of being in state s_1 in round n. Let ϕ_i be the probability of leaving state s_i where $i \in \{-1, 1\}$. Using the fact that $p_{n+1}(1) = (1 - \phi_1)p_n(1) + \phi_{-1}(1 - p_n(1))$ it is easily verified by induction that

$$\begin{aligned} p_n(1) &= 0.5 + 0.5(\phi_{-1} - \phi_1)\sum_{i=0}^{n-2}(1 - \phi_{-1} - \phi_1)^i \\ &= 0.5 + 0.5(\phi_{-1} - \phi_1)\frac{1 - (1 - \phi_{-1} - \phi_1)^{n-1}}{\phi_{-1} + \phi_1} \end{aligned}$$

Consequently,

$$\begin{aligned} E\pi^{\delta}(1) &= (1 - \delta)\sum_{n=1}^{\infty}\delta^{n-1}((p_n\tau_1 + (1 - p_n)(1 - \tau_1))\pi_1 + ((1 - p_n)\tau_1 + p_n(1 - \tau_1))\pi_2) \\ &= \tau_1\pi_2 + (1 - \tau_1)\pi_1 + (2\tau_1 - 1)(\pi_1 - \pi_2)(1 - \delta)\sum_{n=1}^{\infty}\delta^{n-1}p_n(1) \end{aligned}$$

$$= 0.5(\pi_1+\pi_2)+\delta(\tau_1-0.5)(\pi_1-\pi_2)\frac{\phi_{-1}-\phi_1}{1-\delta(1-\phi_{-1}-\phi_1)}$$
$$= 0.5(\pi_1+\pi_2)+\delta(\pi_1-\pi_2)^2\frac{(\tau_1-0.5)(\tau_1\gamma_1+(1-\tau_1)\alpha_{-1})}{1-\delta+\delta\tau_1\gamma_1(2-\pi_1-\pi_2)+\delta(1-\tau_1)\alpha_{-1}(\pi_1+\pi_2)}$$

and

$$d^1(\rho)=\delta(\tau_1-0.5)\min\left\{(\pi_1-\pi_2)\frac{\phi_{-1}-\phi_1}{1-\delta(1-\phi_{-1}-\phi_1)}\text{ s.t. }|\pi_1-\pi_2|=2\rho\right\}$$

where we use the fact that $\phi_1=\tau_1\gamma_1(1-\pi_1)+(1-\tau_1)\alpha_{-1}\pi_2$ and $\phi_{-1}=\tau_1\gamma_1(1-\pi_2)+(1-\tau_1)\alpha_{-1}\pi_1$.

Since $E\pi^\delta(1)$ is monotone in γ_1 and α_{-1} we obtain three candidates for ex-ante undominated rules:

(i) $\gamma_1=1$ and $\alpha_{-1}=0$ yields

$$E\pi^\delta(1)=0.5(\pi_1+\pi_2)+\delta(\pi_1-\pi_2)^2\frac{(\tau_1-0.5)\tau_1}{1-\delta+\delta\tau_1(2-\pi_1-\pi_2)}$$

and

$$d^1(\rho)=\frac{4\rho^2\delta(\tau_1-0.5)\tau_1}{1-\delta+\delta\tau_1 2(1-\rho)}$$

where both expressions are monotone increasing in τ_1 so $\tau_1=1$ is best which yields

$$E\pi^\delta(1)=0.5(\pi_1+\pi_2)+\delta(\pi_1-\pi_2)^2\frac{0.5}{1-\delta+\delta(2-\pi_1-\pi_2)} \tag{4}$$

and

$$d^1(\rho)=\frac{2\rho^2\delta}{1+\delta(1-2\rho)}. \tag{5}$$

(ii) $\gamma_1=0$ and $\alpha_{-1}=1$ (and $\delta<1$ or $\tau_1<1$) yields

$$E\pi^\delta(1)=0.5(\pi_1+\pi_2)+\delta(\pi_1-\pi_2)^2\frac{(\tau_1-0.5)(1-\tau_1)}{1-\delta+\delta(1-\tau_1)(\pi_1+\pi_2)} \tag{6}$$

and

$$d^1(\rho)=\frac{4\rho^2\delta(\tau_1-0.5)(1-\tau_1)}{1-\delta+\delta(1-\tau_1)2(1-\rho)}\ . \tag{7}$$

In order to show for which values of δ (4) is always greater than (6), notice that

$$\frac{0.5}{1-\delta+\delta(2-\pi_1-\pi_2)}-\frac{(\tau_1-0.5)(1-\tau_1)}{1-\delta+\delta(1-\tau_1)(\pi_1+\pi_2)}$$

is increasing in $(\pi_1+\pi_2)$ so the worst case is $\pi_1=\pi_2=0$ which yields

$$\frac{0.5}{1+\delta}-\frac{(\tau_1-0.5)(1-\tau_1)}{1-\delta}\geq\frac{0.5}{1+\delta}-\frac{.0625}{1-\delta}=.0625\frac{7-9\delta}{1-\delta^2}$$

and hence (6) is strictly smaller than (4) if $\delta < 7/9$.

It is easily shown that (7) is smaller than (5), strictly if $\delta < 1$. If $\delta = 1$ then $\tau_1 = 1$ is best which yields the same values as (5), however $\tau_1 = 1$ is not feasible here as this means that $\phi_1 = 0$.

(iii) $\alpha_{-1} = \gamma_1 = 1$ yields

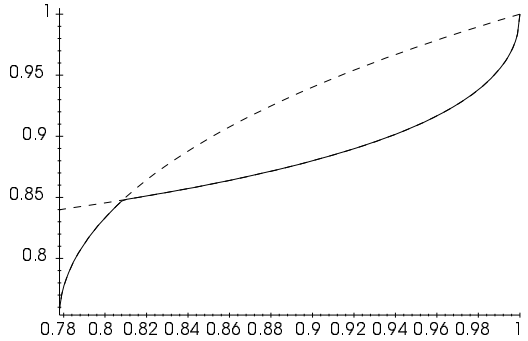
$$E\pi^{\delta}(1) = 0.5(\pi_1 + \pi_2) + \delta(\pi_1 - \pi_2)^2 \frac{\tau_1 - 0.5}{1 - \delta + \delta\tau_1(2 - \pi_1 - \pi_2) + \delta(1 - \tau_1)(\pi_1 + \pi_2)} \tag{8}$$

where it is straightforward to show that (8) is smaller than (4), strictly smaller if $\tau_1 < 1$. Notice that if $\tau_1 = 1$ then the rules in (i) and (iii) have the same behavior. ■

Corollary 3 *Consider* $7/9 < \delta < 1$. *A rule is ex-ante undominated if and only if it satisfies the description in Proposition 2 or it satisfies* $\gamma_1 = 0$, $\alpha_{-1} = 1$ *and*

$$0.75 \leq \tau_1 \leq \min\left\{1 - \sqrt{1-\delta}\frac{1-\sqrt{1-\delta}}{2\delta}, 0.75\left(1 + \frac{\sqrt{\delta - 7/9}}{\sqrt{1+\delta}}\right)\right\}. \tag{9}$$

In particular, undominated linear reinforcement rules with two states have either positive or negative reinforcement in all states.[5] The upper bound in (9) as a function of δ is graphed in the figure below:



Proof. For given $(\pi_1 + \pi_2) > 0$ and δ it is easily shown that

$$\frac{(\tau_1 - 0.5)(1 - \tau_1)}{1 - \delta + \delta(1 - \tau_1)(\pi_1 + \pi_2)} \tag{10}$$

is single peaked in τ_1 on $[0.5, 1]$ with maximum at

$$\tau_1^* = \frac{1}{\delta(\pi_1 + \pi_2)}\left(\delta(\pi_1 + \pi_2) + 1 - \delta - 0.5\sqrt{2(1-\delta)(\delta(\pi_1 + \pi_2) + 2(1-\delta))}\right)$$

[5] The proof also reveals that smaller values of τ_1 perform better for smaller values of $\pi_1 + \pi_2$.

and monotone increasing (decreasing) for $\tau_1 < \tau_1^*$ ($\tau_1 > \tau_1^*$). Moreover, τ_1^* is monotone in $(\pi_1 + \pi_2)$ and takes values in $\left[0.75, 1 - \sqrt{1-\delta}\frac{1-\sqrt{1-\delta}}{2\delta}\right]$. Hence, following (6), a rule is undominated in the class of case (ii) from the proof of Proposition 2 if and only if $0.75 \leq \tau_1 \leq 1 - \sqrt{1-\delta}\frac{1-\sqrt{1-\delta}}{2\delta}$.

Using the same argument as in the proof of the proposition above we obtain that (4) is greater than (6) for all π_1 and π_2 if and only if

$$\frac{0.5}{1+\delta} \geq \frac{(\tau_1 - 0.5)(1-\tau_1)}{1-\delta}$$

which holds if and only if

$$|\tau_1 - 0.75| \leq 0.75\frac{\sqrt{\delta - 7/9}}{\sqrt{1+\delta}}\ .$$

Combining these two arguments completes the proof of the statement. ■

Remark 2 *Consider now $\delta = 1$. Formally there is no rule that is undominated among the rules with positive reinforcement. Following (10) it is best if τ_1 is maximal, however $\tau_1 = 1$ yields $E\pi^1(1) = 0.5(\pi_1 + \pi_2)$ as (10) is not defined for $\delta = \tau_1 = 1$. However, given an exogenous upper bound $\bar{\tau}$ on τ_1 with $\bar{\tau} < 1$ in addition to the negative reinforcement rule of Proposition 2 there is a single alternative undominated rule with $\gamma_1 = 0$, $\alpha_{-1} = 1$ and $\tau_1 = \bar{\tau}$.*

A linear reinforcement rule with two states and $\tau_1 = 1$ is a special case of a pure strategy behavioral rule as defined in Börgers et al. (1998). The only difference to our rule is that they allow for more general switching probabilities between states. Börgers et al. (1998) find that no sequence of pure strategy behavioral rules will find the better action with arbitrarily high probability in each decision. This result is confirmed for a more limited set of behaviors in the proposition above. Since only one of the two rules in parts (i) and (ii) of the proof above are candidates for maximizing the probability of choosing the better action among the improving rules, we obtain

$$\max\left\{\frac{1 - \min\{\pi_1, \pi_2\}}{2 - \pi_1 - \pi_2}, \frac{\max\{\pi_1, \pi_2\}}{\pi_1 + \pi_2}\right\}$$

as an upper bound on the probability of choosing the better action among the improving rules.

Proposition 4 *(i) The rule selected in Proposition 2 fails to be ex-ante improving for any $\delta < 1$ as $E\pi^\delta(2) > E\pi^\delta(3)$ holds whenever $\pi_1 \neq \pi_2$ and $\pi_1 + \pi_2 < 1$.*

(ii) The two state linear reinforcement rule with $(\tau_1, \gamma_1) = (1, 0.5)$ is ex-ante improving for any δ, it dominates any other ex-ante improving rule for $\delta \leq 0.75$ and outperforms all other ex-ante improving rules for fixed differences when $0.75 < \delta \leq 1$. It generates

$$\begin{aligned} z_1^{(n)} &= 0.5 + 0.5(\pi_1 - \pi_2)\frac{1 - (0.5(\pi_1 + \pi_2))^{n-1}}{2 - \pi_1 - \pi_2} \\ E\pi^\delta(1) &= 0.5(\pi_1 + \pi_2) + \delta(\pi_1 - \pi_2)^2\frac{0.5}{2 - \delta(\pi_1 + \pi_2)} \end{aligned}$$

(iii) "Do not switch" is the only two state improving rule.

Proof. It is easily verified that

$$E\pi^{\delta}(2) - E\pi(2) = (\tau_1 - 0.5)(\pi_1 - \pi_2)(\phi_{-1} - \phi_1)\frac{\delta(1 - \phi_{-1} - \phi_1)}{1 - \delta + \delta\phi_{-1} + \delta\phi_1} < 0$$

holds when $\delta > 0$ and $\phi_{-1} + \phi_1 > 1$. Notice that this is true for the rule selected in Proposition 2 when $\pi_1 + \pi_2 < 1$. In order for a linear rule to be ex-ante improving we need that

$$1 \geq \phi_{-1} + \phi_1 = \tau_1\gamma_1(2 - \pi_1 - \pi_2) + (1 - \tau_1)\alpha_{-1}(\pi_1 + \pi_2)$$

is true in any decision with $\pi_1 \neq \pi_2$ which holds if and only if $\tau_1\gamma_1 \leq 0.5$. Notice that $\phi_{-1}+\phi_1 \leq 1$ is also sufficient as this implies that p_n is monotonically increasing (decreasing) whenever $\pi_1 > \pi_2$ ($\pi_1 < \pi_2$).

Using the monotonicity we obtain analogous to case (i) in the proof of Proposition 2 above:

(i') $\tau_1\gamma_1 = 0.5$ and $\alpha_{-1} = 0$ yields

$$E\pi^{\delta}(1) = 0.5(\pi_1 + \pi_2) + 0.5\delta(\pi_1 - \pi_2)^2\frac{\tau_1 - 0.5}{1 - 0.5\delta(\pi_1 + \pi_2)}$$

which is monotone increasing in τ_1 so $\tau_1 = 1$ is best which yields

$$E\pi^{\delta}(1) = 0.5(\pi_1 + \pi_2) + 0.25\delta(\pi_1 - \pi_2)^2\frac{1}{1 - 0.5\delta(\pi_1 + \pi_2)}$$

and

$$d^1(\rho) = \frac{\delta\rho^2}{1 - \delta\rho}. \tag{11}$$

We find that this rule is better than a rule $(\gamma_1, \alpha_{-1}, \tau_1)$ if

$$\frac{0.25}{1 - 0.5\delta(\pi_1 + \pi_2)} - \frac{(\tau_1 - 0.5)(1 - \tau_1)}{1 - \delta + \delta(1 - \tau_1)(\pi_1 + \pi_2)} \geq 0$$

where the left hand side is increasing in $(\pi_1 + \pi_2)$ so entering the worst case $\pi_1 = \pi_2 = 0$ yields

$$0.25 - \frac{(\tau_1 - 0.5)(1 - \tau_1)}{1 - \delta} \geq 0.$$

Setting $\tau_1 = 0.75$ we obtain the worst case which gives us the condition $0.25 - \frac{.0625}{1-\delta} \geq 0$ which implies $\delta \leq 0.75$.

(iii') $\tau_1\gamma_1 = 0.5$, $\alpha_{-1} = 1$ and $\tau_1 < 1$ yields

$$E\pi^{\delta}(1) = 0.5(\pi_1 + \pi_2) + \delta(\pi_1 - \pi_2)^2\frac{(\tau_1 - 0.5)(1.5 - \tau_1)}{1 - \delta(\tau_1 - 0.5)(\pi_1 + \pi_2)}$$

where it is easily verified that the rule in (i') dominates this rule.

Subtracting (7) from (11) and then setting the "worst" case $\rho = \delta = 1$ we find that (11) is always greater than (7). ■

5.2 Four states

Next we consider the case of four states. We skip the analysis of three states as our symmetry condition requires that $\tau_0 = 0.5$ which gives less degree of freedom when constructing good rules. Consider linear reinforcement rules with four states where $f(\emptyset)_1 = f(\emptyset)_{-1} = 0.5$ that are parametrized by τ_1, τ_2 and α_i for $i \in \{-2,-1,1\}$ and γ_j for $j \in \{-1,1,2\}$.

5.2.1 Myopia

Proposition 5 *Consider $\delta = 0$. There is a unique linear reinforcement rule that is ex-ante undominated among the ex-ante improving rules. It satisfies $\tau_1 = \tau_2 = \gamma_1 = \alpha_1 = 1$, $\gamma_2 = 0$ and yields*

$$z_1^{(2k)} = z_1^{(2k+1)} = 0.5 + 0.5(\pi_1 - \pi_2)\left(\frac{1-(1-\pi_1)^k(1-\pi_2)^k}{1-(1-\pi_1)(1-\pi_2)}\right) .$$

Notice that the extreme states s_{-2} and s_2 of this rule selected are absorbing. In round two it yields the same expected payoffs as the two state ex-ante learning rule selected in Proposition 2 and yields larger expected payoffs than the two state ex-ante improving rule selected in Proposition 4.

Proof. For $i > 0$ let $\lambda_i = \tau_i - 0.5$ so that our assumptions on τ_i imply that $0 \leq \lambda_i \leq \lambda_{i+1} \leq 0.5$.

Consider round two. Tedious calculations show

$$z_1^{(2)} = .5 + .5(\pi_1 - \pi_2)\begin{pmatrix} .5(\lambda_2 - \lambda_1)(\alpha_1 + \gamma_{-1}) \\ +\lambda_1(\lambda_2 - \lambda_1)(\alpha_1 - \gamma_{-1}) + (\alpha_{-1} + \gamma_1)\lambda_1 + 2(\gamma_1 - \alpha_{-1})\lambda_1^2 \end{pmatrix} .$$

Assume that (S, f, τ) is ex-ante undominated. Since (S, f, τ) is improving,

$$(\lambda_2 - \lambda_1)(.5(\alpha_1 + \gamma_{-1}) + \lambda_1(\alpha_1 - \gamma_{-1})) + \lambda_1(\alpha_{-1} + \gamma_1 + 2(\gamma_1 - \alpha_{-1})\lambda_1) \geq 0 .$$

If $\lambda_1 < \lambda_2$ then $\alpha_1 = \gamma_{-1} = 1$. If $\lambda_1 > 0$ then $\alpha_{-1} = \gamma_1 = 1$. This yields $z_1^{(2)} = .5 + .5(\pi_1 - \pi_2)(\lambda_1 + \lambda_2)$. If $\lambda_1 = \lambda_2 = 0.5$ then $z_1^{(2)} = .5 + .5(\pi_1 - \pi_2)\gamma_1$ which implies $\gamma_1 = 1$. Comparing these two cases we find that the fact that (S, f, τ) is ex-ante undominated in round two implies $\lambda_1 = \lambda_2 = 0.5$ and $\gamma_1 = 1$. In particular, there are no restrictions on α_{-1}, α_1 and γ_{-1}.

Now consider rounds three and higher and assume $\gamma_1 = 1$. Since $\lambda_1 = 0.5$ means by assumption that $\lambda_i = 0.5$ and $\tau_i = 1$ for all $i > 0$. Hence, for $i > 0$, α_{-i} and γ_{-i} play no more role as action 1 (action 2) is never chosen in states s_{-i} (s_i) for $i > 0$. Let $p_n(i)$ be the probability of being in state s_i in round n. In the following we will use the fact that $z_1^{(n+1)} - z_1^{(n)} = p_n(-1)(1-\pi_2) - p_n(1)(1-\pi_1)$.

For round three we obtain $p_2(1) = 0.5(1-\pi_2) + 0.5(1-\alpha_1)\pi_1$ so that

$$\begin{aligned} z_1^{(3)} &= .5 + .5(\pi_1 - \pi_2)(\alpha_1 + (1-\alpha_1)(\pi_1 + \pi_2)) \\ &= z_1^{(2)} + .5(\alpha_1 - 1)(1 - \pi_1 - \pi_2)(\pi_1 - \pi_2) \ . \end{aligned}$$

This means that ex-ante improving requires $\alpha_1 = 1$ which yields $z_1^{(3)} = z_1^{(2)}$.

Now assume $\alpha_1 = 1$ and consider round four. $p_2(2) = 0.5\pi_1$ and

$$p_3(1) = 0.5(1-\pi_1)(1-\pi_2) + 0.5\pi_1\gamma_2(1-\pi_1) = 0.5(1-\pi_1)(1+\gamma_2\pi_1 - \pi_2)$$

so

$$z_1^{(4)} - z_1^{(3)} = .5(\pi_1 - \pi_2)\left((1-\pi_1)(1-\pi_2) + \gamma_2\left(-1 + 2\pi_1 + 2\pi_2 - \pi_1^2 - \pi_1\pi_2 - \pi_2^2\right)\right) \qquad (12)$$

where $\left(z_1^{(4)} - z_1^{(3)}\right)(\pi_1 - \pi_2) \geq 0$ holds for all γ_2 as it is linear in γ_2 and holds for both $\gamma_2 = 0$ and $\gamma_2 = 1$.

Consider round five. For general γ_2 we find

$$\begin{aligned} p_3(2) &= 0.5(1-\pi_2)\pi_1 + 0.5\pi_1(1 - \gamma_2(1-\pi_1)) \\ p_4(1) &= 0.5(1-\pi_2)(1+\gamma_2\pi_2 - \pi_1)(1-\pi_2) + (0.5(1-\pi_2)\pi_1 + 0.5\pi_1(1-\gamma_2(1-\pi_1)))\gamma_2(1-\pi_1) \\ &= 0.5(1+\gamma_2\pi_2 - \pi_1)(1-\pi_2)^2 + 0.5(2 - \pi_2 - \gamma_2(1-\pi_1))\gamma_2\pi_1(1-\pi_1) \end{aligned}$$

and

$$z_1^{(5)} - z_1^{(4)} = 0.5(\pi_1 - \pi_2)\gamma_2 \left(\begin{array}{c} -1 - \pi_1^2 + 2\pi_1 + 2\pi_2 - \pi_1\pi_2 - \pi_2^2 \\ + \left(1 - \pi_2^3 - 3\pi_2 + 3\pi_2^2 + 3\pi_1\pi_2 - \pi_1\pi_2^2 - 3\pi_1 + 3\pi_1^2 - \pi_2\pi_1^2 - \pi_1^3\right)\gamma_2 \end{array} \right)$$

In order for $\left(z_1^{(5)} - z_1^{(4)}\right)(\pi_1 - \pi_2) \geq 0$ to hold for $\gamma_2 > 0$, the last factor on the right hand side must be non-negative. Checking $\pi_1 = \pi_2 = 0$ we find that $\gamma_2 = 1$ but then checking for $\pi_1 = \pi_2 < 0.5$ we obtain a violation and hence $\gamma_2 = 0$.

Now consider later rounds. For any positive integer k we obtain

$$p_{2k}(1) = 0.5(1-\pi_1)^{k-1}(1-\pi_2)^k, \ p_{2k+1}(1) = 0.5(1-\pi_1)^k(1-\pi_2)^k$$

and hence $z_1^{(2k+1)} = z_1^{(2k)}$ and $z_1^{(2k+2)} - z_1^{(2k+1)} = p_{2k+1}(1)(\pi_1 - \pi_2)$ which means that this rule is ex-ante improving. Using the fact that $p_{2k+1}(2) = 0.5\pi_1(2-\pi_2)\frac{1-(1-\pi_1)^k(1-\pi_2)^k}{1-(1-\pi_1)(1-\pi_2)}$ the value of $z_1^{(2k+1)}$ is easily verified. ■

Next we consider improving rules. Notice that the rule selected in the proposition above is not improving. For a rule to be improving in state s_1 the decrease in the probability of playing action one (due to switching to state s_{-1} which is possible given $\gamma_1 > 0$) has to be offset by an increase in the probability whenever $\pi_1 = \pi_2$. However, such an increase is not possible as $\tau_1 = 1$.

Proposition 6 *Assume $\delta = 0$. The linear reinforcement rule with $\gamma_2 = 0$, $\tau_1 = -\frac{1}{2} + \frac{1}{2}\sqrt{5} \approx .618$, $\tau_2 = \gamma_{-1} = \alpha_{-1} = \gamma_1 = 1$, $\alpha_1 = \frac{1}{2}\left(3 - \sqrt{5}\right) \approx .382$ ex-ante dominates all other improving rules. This rule induces*

$$z_1^{(2)} = 0.5 + \left(\sqrt{5} - 2\right)(\pi_1 - \pi_2) \approx 0.5 + 0.236\,(\pi_1 - \pi_2)$$

Proof. Improving requires $\alpha_{-2} = \gamma_2 = 0$. Let

$$h_1(\pi_1, \pi_2) = \left(\tau_1\alpha_1\pi_1 + (1-\tau_1)\,\gamma_{-1}(1-\pi_2)\right)(\tau_2 - \tau_1) + (\tau_1\gamma_1(1-\pi_1) + (1-\tau_1)\,\alpha_{-1}\pi_2)(1-2\tau_1)$$

then improving in state s_1 implies $h_1(\pi_1, \pi_2) * (\pi_1 - \pi_2) \geq 0$. Consequently,

$$\begin{aligned} h_1(\pi_1, \pi_1) &= \pi_1\left((\tau_2 - \tau_1)\,\tau_1\alpha_1 + (1-\tau_1)(1-2\tau_1)\,\alpha_{-1}\right) \\ &\quad + (1-\pi_1)\left((\tau_2 - \tau_1)(1-\tau_1)\,\gamma_{-1} + \tau_1(1-2\tau_1)\,\gamma_1\right) \\ &= 0 \end{aligned}$$

and hence

$$\begin{aligned} (\tau_2 - \tau_1)(1-\tau_1)\,\gamma_{-1} &= \tau_1(2\tau_1 - 1)\,\gamma_1 \text{ and} \\ (\tau_2 - \tau_1)\,\tau_1\alpha_1 &= (1-\tau_1)(2\tau_1 - 1)\,\alpha_{-1} \end{aligned}$$

which means that $h_1 = (\tau_2 - \tau_1)\left(\tau_1\alpha_1 + (1-\tau_1)\,\gamma_{-1}\right)(\pi_1 - \pi_2)$.

The analogous analysis for state s_{-1} leads to the definition of $h_{-1}(\pi_1, \pi_2)$ where

$$h_{-1}(\pi_1, \pi_2) = ((1-\tau_1)\,\alpha_{-1}\pi_1 + \tau_1\gamma_1(1-\pi_2))(2\tau_1 - 1) + \left((1-\tau_1)\,\gamma_{-1}(1-\pi_1) + \tau_1\alpha_1\pi_2\right)(\tau_1 - 1)$$

Notice that $h_{-1}(\pi_1, \pi_2) = -h_1(\pi_2, \pi_1)$ so there are no additional conditions for state s_{-1}.

Undominated implies $(\tau_2 - \tau_1)\,\alpha_1$ and $(\tau_2 - \tau_1)\,\gamma_{-1}$ maximal for given τ_1. Thus $\tau_2 = 1$ so $\alpha_1 \leq \frac{2\tau_1 - 1}{\tau_1}$ and $\gamma_{-1} \leq \min\left\{\frac{\tau_1(2\tau_1-1)}{(1-\tau_1)^2}, 1\right\}$ where $\frac{\tau_1(2\tau_1-1)}{(1-\tau_1)^2} \leq 1$ holds if and only if $\tau_1 \leq -\frac{1}{2} + \frac{1}{2}\sqrt{5}$.

$\gamma_{-1} = 1$ and $\tau_1 \geq -\frac{1}{2} + \frac{1}{2}\sqrt{5}$ yields $h/(\pi_1 - \pi_2) = (1-\tau_1)\,\tau_1$ which obtains its maximum $\sqrt{5} - 2$ at $\tau_1 = -\frac{1}{2} + \frac{1}{2}\sqrt{5}$.

$\gamma_{-1} = \frac{\tau_1(2\tau_1-1)}{(1-\tau_1)^2}$ and $\tau_1 \leq -\frac{1}{2} + \frac{1}{2}\sqrt{5}$ yields $h/(\pi_1 - \pi_2) = 2\tau_1 - 1$ which obtains its maximum when $\gamma_{-1} = 1$. ■

Interesting results can also be obtained by searching for rules that are undominated among a subset of linear reinforcing rules. The following results are easily verified. Among the set of rules that satisfy $\gamma_i = \alpha_i$, the rule with $\tau_1 = 2/3$, $\gamma_1 = \alpha_1 = 0.5$ and $\gamma_{-1} = \alpha_{-1} = 1$ that yields $z_1^{(2)} = 0.5 + .\bar{2} * (\pi_1 - \pi_2)$ is selected. Among the rules that satisfy $\gamma_{-1} = \gamma_1$ and $\alpha_{-1} = \alpha_1$, the rule with $\alpha = 0$, $\gamma = 1$ and $\tau_1 = -\frac{1}{2} + \frac{1}{2}\sqrt{5}$ where $z_1^{(2)} \approx 0.5 + 0.1459 * (\pi_1 - \pi_2)$ is selected.

5.2.2 Patience

Now consider a patient individual with $\delta = 1$. In the following we will consider for simplicity only rules with $\tau_2 = 1$ so α_{-2} plays no role. We will also restrict attention to small γ_2 and evaluate performance taking the limit as γ_2 tends to zero. We will refer to this limit as "for vanishing γ_2".

Proposition 7 *The unique ex-ante undominated improving rule for fixed differences and vanishing γ_2 is to set $\alpha_{-1} = \alpha_1 = 1$, $\{\gamma_{-1}, \gamma_1\} = \{0, 1\}$ and $\tau_1 = 0.5$ which in the limit as $\gamma_2 \to 0$ yields*

$$z_1^{(\infty)} = \frac{(1-\pi_2)\,\pi_1\,(1+\pi_1-\pi_2)}{(1-\pi_1)\,\pi_2\,(1+\pi_2-\pi_1)+(1-\pi_2)\,\pi_1\,(1+\pi_1-\pi_2)}$$

and $d^1(\rho) = 2\rho^2 \frac{3+4\rho^2}{1+12\rho^2}$ with $d^{1\prime\prime}(0) = 12$.

The proof of this proposition given in the appendix actually reveals many other ex-ante undominated rules. For each of these rules we present the quotients $z_1^{(\infty)}/z_2^{(\infty)}$ and the second derivative of d^1 at $\rho = 0$ as a rough measure of their performance. Each of them satisfies $d^1(0) = 0$, $d^{1\prime}(0) = 0$ which follows from the fact that action one is played in each round with probability 0.5 when $\pi_1 = \pi_2$.

Corollary 8 *The following rules are ex-ante undominated for vanishing γ_2 and satisfy $d^1(0) = d^{1\prime}(0) = 0$:*

(i) $\gamma_{-1} = \gamma_1 > 0$ and $\alpha_{-1} = \alpha_1 = 0$ yields $\frac{(1-\pi_2)^3}{(1-\pi_1)^3}$ ($d^{1\prime\prime}(0) = 6$) (negative reinforcement)

(ii) $\gamma_1 > 0$, $\alpha_1 = 1$ and $\tau_1 = 1$ yields $\frac{\pi_1(1-\pi_2)^2}{\pi_2(1-\pi_1)^2}$ ($d^{1\prime\prime}(0) = 11.657$)

(iii) $\gamma_{-1} = \gamma_1 = 0$, $\alpha_{-1} = \alpha_1 > 0$ and $\tau_1 < 1$ yields $\frac{\pi_1^2(1-\pi_2)}{\pi_2^2(1-\pi_1)}$ ($d^{1\prime\prime}(0) = 11.657$)

For our general analysis we assumed $\tau_2 = 1$ and only consider the limit as γ_2 tends to 0. However, the effect of weakening these assumptions on the above rules is easily analyzed. In fact the above rules remain ex-ante undominated if this is relaxed. It is easily verified (checking case by case) that any of the above rules for $\gamma_2 > 0$ ex-ante dominates any choice of $\gamma_2' > \gamma_2$. Now assume that $\tau_2 < 1$. Then this lowers the probability of choosing action one in state s_2 and raises it by the same amount in state s_{-2}. As $\pi_1 > \pi_2$ implies that state s_2 has higher probability in the stationary distribution than s_{-2} we find that setting $\tilde{\gamma}_2 = \tau_2\gamma_2 + (1-\tau_2)\,\alpha_{-2}$ and $\tilde{\tau}_2 = 1$ yields better performance. Of course, relaxing the assumption $\tau_2 = 1$ introduces additional undominated rules. It is easily verified that $\gamma_i = 0$, $\alpha_i = 1$, $\tau_1 < 1$ (only positive reinforcement) yields for $\tau_2 \to 1$ an undominated rule with relative performance $(\pi_1/\pi_2)^3$.

6 Unbounded Complexity

6.1 An Example

Consider the following linear reinforcement rule with maximal switching probability in each state that we call the *Gradual Confidence Rule*. It is defined by setting $\tau_i = 1$ (and hence $\tau_{-i} = 0$) for $i > 0$, $f(\emptyset)_{-1} = f(\emptyset)_1 = 0.5$ and $\alpha_i = \gamma_i = 1$ for $i > 0$. The behavior of this rule is easily described in words. In each round the individual either plays action one or action two with probability one (there is no mixing in a given state). Play of action i is associated to a level of confidence c where $c = 1, 2, ...$ The individual starts out equally like with each action and with confidence one. With probability equal to the payoff x obtained, the same action is played again and confidence is increased by one. If current confidence is above one then with probability $1-x$ the same action is played again and confidence is decreased by one. If current confidence equals one then with probability $1-x$ the individual keeps $c = 1$ and plays the other action in the next round. Consequently, the aspiration level of each action in each state is 0.5. Calculating ex-ante probabilities of choosing action one in the first seven rounds we find

$$\begin{aligned}
z_1^{(1)} &= 0.5 \\
z_1^{(2)} &= 0.5 + 0.5(\pi_1 - \pi_2) \\
z_1^{(3)} &= z_1^{(2)} \\
z_1^{(4)} &= z_1^{(2)} + 0.5(\pi_1 - \pi_2)(\pi_1(1-\pi_1) + \pi_2(1-\pi_2)) \\
z_1^{(5)} &= z_1^{(4)} \\
z_1^{(6)} &= z_1^{(4)} + (\pi_1 - \pi_2)\left(\pi_1^2(1-\pi_1)^2 + \pi_1\pi_2(1-\pi_1)(1-\pi_2) + \pi_2^2(1-\pi_2)^2\right) \\
z_1^{(7)} &= z_1^{(6)}
\end{aligned}$$

Although the expression for $z_1^{(n)}$ have a regular pattern for $n \leq 7$, this pattern does not continue for rounds $n > 8$ so while we conjecture that $z_1^{(2k+1)} = z_1^{(2k)}$ holds in general a proof is still missing.

6.2 Myopia

In the following we will show that the Gradual Confidence Rule is selected among the linear reinforcement rules for its behavior in the first five rounds.

Proposition 9 *Let (S, f, τ) be a linear reinforcement rule that is ex-ante undominated in the first five rounds. Then the Gradual Confidence Rule ex-ante dominates (S, f, τ) in rounds two, three and five. In round four the Gradual Confidence Rule outperforms (S, f, τ) for fixed differences.*

Proof. The derivation of the terms for the first four rounds is as in the proof of Proposition 5. Here we add some calculations to the fourth round as we wish to select a single γ_2.

The term in (12) with γ_2 cannot be signed as it is negative for $\pi_1 = \pi_2 = 0$ and positive for $\pi_1 = \pi_2 = 0.5$. Given $\pi_1 > \pi_2$, $\Delta_3 = (\pi_1 - \pi_2)\left(1 - z_1^{(3)}\right) = 0.5(\pi_1 - \pi_2)(1 - (\pi_1 - \pi_2))$ which takes values in $[0, 0.125]$ and is symmetric in $(\pi_1 - \pi_2)$ around $(\pi_1 - \pi_2) = 0.5$.

Assume that $\pi_1 = \pi_2 + \varepsilon$ for given $\varepsilon \in (0,1)$. Then

$$\begin{aligned} h(\pi_2, \gamma_2, \varepsilon) &: = z_1^{(4)} - z_1^{(3)} \\ &= .5\varepsilon\left((1 - \pi_2 - \varepsilon)(1 - \pi_2) + \gamma_2\left(-1 + 2\pi_2 + 2\varepsilon + 2\pi_2 - (\pi_2 + \varepsilon)^2 - (\pi_2 + \varepsilon)\pi_2 - \pi_2^2\right)\right) \\ &= 0.5\varepsilon\left(1 - 2\pi_2 + \pi_2^2 - \varepsilon + \varepsilon\pi_2 - \gamma_2 + 4\gamma_2\pi_2 + 2\gamma_2\varepsilon - 3\gamma_2\pi_2^2 - 3\gamma_2\varepsilon\pi_2 - \gamma_2\varepsilon^2\right) \end{aligned}$$

For $\gamma_2 \geq 1/3$, $h(*, \gamma_2, \varepsilon)$ is concave and candidates for minima with respect to π_2 are $0.5\varepsilon(1-\varepsilon)(1 - \gamma_2(1-\varepsilon))$ at $\pi_2 = 0$ and $0.5\varepsilon^2\gamma_2(1-\varepsilon)$ at $\pi_2 = 1 - \varepsilon$ where

$$0.5\varepsilon(1-\varepsilon)(1 - \gamma_2(1-\varepsilon)) \geq 0.5\varepsilon^2\gamma_2(1-\varepsilon)$$

with equality holding if and only if $\gamma_2 = 1$ so that $\min_{\pi_2 \in [0, 1-\varepsilon]} h(\pi_2, \gamma_2, \varepsilon) = 0.5\varepsilon^2\gamma_2(1-\varepsilon)$.

For $\gamma_2 < 1/3$, $h(*, \gamma_2, \varepsilon)$ is convex in π_2 but $\frac{d}{d\pi_2} h(\pi_2, \gamma_2, \varepsilon)|_{\pi_2 = 1-\varepsilon} = -\varepsilon - 2\gamma_2 + 3\gamma_2\varepsilon < 0$ so $\{1 - \varepsilon\} = \arg\min_{\pi_2 \in [0, 1-\varepsilon]} \{h(\pi_2, \gamma_2, \varepsilon)\}$ and $\min_{\pi_2 \in [0, 1-\varepsilon]} h(\pi_2, \gamma_2, \varepsilon) = 0.5\varepsilon^2\gamma_2(1-\varepsilon)$.

Fix $\rho \in [0, 0.125]$. Then there exists $x(\rho) \in [0, 0.5]$ such that $\Delta(\pi_1, \pi_2) = \rho$ if and only if $|\pi_1 - \pi_2| \in \{0.5 - x, 0.5 + x\}$. As $0.5(0.5 - x)^3 \gamma_2(1 - (0.5 - x)) \leq 0.5(0.5 + x)^3 \gamma_2(1 - (0.5 + x))$ the above calculations show that

$$d_f^4(\rho) = \inf\left\{\left(z_1^{(4)} - z_1^{(3)}\right)(\pi_1 - \pi_2) \text{ s.t. } \Delta_3 = \rho\right\} = 0.5(0.5 - x(\rho))^3 \gamma_2(1 - (0.5 - x(\rho))) \tag{13}$$

and hence $\gamma_2 = 1$ maximizes d_f^4.

Assume $\gamma_2 = 1$ and consider round five. Here the calculations are different than for the four state rule as α_2 now enters the picture. We find

$$\begin{aligned} p_3(2) &= 0.5(1 - \pi_2)\pi_1 + 0.5\pi_1^2(1 - \alpha_2) \\ p_4(1) &= 0.5(1 - \pi_2)(1 + \pi_2 - \pi_1)(1 - \pi_2) + \left(0.5(1 - \pi_2)\pi_1 + 0.5\pi_1^2(1 - \alpha_2)\right)(1 - \pi_1) \\ &= 0.5(1 - \pi_2)\left(1 - \pi_1^2 - \pi_2^2 + \pi_1\pi_2\right) + 0.5\pi_1^2(1 - \pi_1)(1 - \alpha_2) \end{aligned}$$

and

$$z_1^{(5)} - z_1^{(4)} = (1 - \alpha_2)(\pi_1 - \pi_2)\left(\pi_1^2 + \pi_1\pi_2 + \pi_2^2 - .5\pi_1 - .5\pi_2 - .5\pi_2^3 - .5\pi_1\pi_2^2 - .5\pi_2\pi_1^2 - .5\pi_1^3\right)$$

where we find that third term in the above expression is negative for small $\pi_1 = \pi_2$. Thus, ex-ante improving implies $\alpha_2 = 1$ and hence $z_1^{(5)} = z_1^{(4)}$. ■

Notice that the Gradual Confidence rule performs identically (in terms of $z^{(n)}$) as the selected myopic four state rule in the first three rounds and is not comparable according to dominance in the fourth round. The reason we choose $\gamma_2 = 1$ was because of our criterion of performance based on fixed differences. Solving for $x(\rho)$ defined in proof above, entering this in (13) and setting $\gamma_2 = 1$ we obtain

$$d^4(\rho) = 0.25\rho\left(1 - \sqrt{1-8\rho}\right)^2$$

whereas the selected myopic four state rule has $\gamma_2 = 0$ and hence $d^4(\rho) = 0$.

6.3 Patience

Next we consider behavioral rules for an individual that is infinitely patient, i.e., he evaluates future payoffs with discount factor 1. Börgers and Sarin (1997) mention that the Cross Learning Rule is not maximizing. As we show below, neither does the Gradual Confidence Rule have this property. On the other hand, Rustichini (1999) claims that the Payoff Sum Learning Rule is maximizing when the payoff distributions have finite support. In the following we construct a maximizing linear reinforcement rule.

Consider linear reinforcement rules and fix $(\tau_i)_i$ such that $\tau_i \in (0,1)$ for all i and $\lim_{i\to\infty}\tau_i = 1$. Let $Ef(s_i)_j = \sum_{k=1}^{2} g(s_i)_k \int f(s_i,k,x)_j \, dP_i(x)$ be the expected switching probability to state s_j. We will say that a linear reinforcement rule has the *symmetric state switching property* if it transits from s_i more likely to s_{i+1} than to s_{i-1} (i.e., $Ef(s_i)_{i+1} \geq Ef(s_i)_{i-1}$) if and only if $\pi_1 \geq \pi_2$ with strict inequalities whenever $\pi_1 > \pi_2$. This means that $\tau_i\alpha_i\pi_1 + (1-\tau_i)\gamma_{-i}(1-\pi_2) \geq (1-\tau_i)\alpha_{-i}\pi_2 + \tau_i\gamma_i(1-\pi_1)$ holds whenever $\pi_1 \geq \pi_2$ which implies $\tau_i\alpha_i = (1-\tau_i)\alpha_{-i}$, $(1-\tau_i)\gamma_{-i} = \tau_i\gamma_i$ and $\alpha_i + \gamma_i > 0$. This implies that the expected index in the next round after being in state s_i equals $Ef(s_i)_{i+1} - Ef(s_i)_{i-1} = \tau_i(\alpha_i + \gamma_i)(\pi_1 - \pi_2)$. The product of $(\pi_1 - \pi_2)$ and the index of the state measures the degree in which the individual has learned which action is better.

Proposition 10 *(i) The Gradual Confidence Rule is not maximizing even if payoffs are restricted to an open subset of* $(0,1)$.

(ii) Any linear reinforcement rule where $\tau_i \in (0,1)$ *for all* i *and* $\lim_{i\to\infty}\tau_i = 1$ *that has the symmetric state switching property is maximizing. None of these rules is improving in each round. Given* $(\tau_i)_i$, *the choice of* $\alpha_{-i} = \gamma_{-i} = 1$ *and* $\alpha_i = \gamma_i = \frac{1-\tau_i}{\tau_i}$ *for* $i > 0$ *maximizes the product of* $(\pi_1 - \pi_2)$ *and the expected state index in any given round among the linear reinforcement rules that have the symmetric switching property.*

Proof. Under the Gradual Confidence Rule, when in the states s_i with $i > 0$ the switching process between states is a random walk with expected increase in the state index of $\pi_1 1 +$

$(1-\pi_1)(-1) = 2\pi_1 - 1$. The theory of random walks tells us that with positive probability the individual will always choose action one if $\pi_1 > 0.5$ even if $\pi_2 > \pi_1$. On the other hand, if $\pi_2 < \pi_1 < 0.5$ then with probability one there will never enter a state with arbitrarily high (or low) index. Thus there is a limit distribution where both actions are chosen with positive probability. This distribution is easily calculated but of no interest here because the existence of this distribution shows that the rule will not select the better action.

With symmetric state switching property, again we have a random walk like process where the expected increase in the state index is $\tau_i(\alpha_i + \gamma_i)(\pi_1 - \pi_2)$. Assume $\pi_1 > \pi_2$. Assume for a moment that $\tau_i(\alpha_i + \gamma_i)$ does not depend on i (which is of course not true). Then we obtain that the random state index tends to infinity. In particular, for any $\varepsilon > 0$ there exists i such that the probability that the state index is always greater than $-i$ is at least $1 - \varepsilon$. Now notice that the probability of never reaching some state s_{-i} does not depend on $\tau_i(\alpha_i + \gamma_i)$ as long as $\tau_i(\alpha_i + \gamma_i) > 0$. Therefore, even when $\tau_i(\alpha_i + \gamma_i)$ depends on i then with a given high probability the process does not reach a state with a sufficiently small index. This means that with probability one the random state index does not tend to minus infinity. On the other hand the random state index cannot remain bounded with high probability as the expected change of the index is strictly positive. Consequently the random state index tends to infinity with probability one.

The change in the expected probability of choosing action one equals

$$\tau_i(\alpha_i\pi_1 + \gamma_i(1-\pi_2))(\tau_{i+1} - \tau_i) + \tau_i(\alpha_i\pi_2 + \gamma_i(1-\pi_1))(\tau_{i-1} - \tau_i)$$

which simplifies to $\tau_i(\tau_{i+1} + \tau_{i-1} - 2\tau_i)(\alpha_i\pi_1 + \gamma_i(1-\pi_1))$ if $\pi_1 = \pi_2$. The fact that $\tau_i \to 1$ as $i \to \infty$ implies that $\tau_{j+1} + \tau_{j-1} - 2\tau_j < 0$ holds for some j. Improving implies that $\alpha_j = \gamma_j = 0$ and hence state s_j is absorbing which however violates the symmetric switching property.

If $i > 0$ then $\tau_i \geq 0.5$ and $\alpha_i, \gamma_i \leq \frac{1-\tau_i}{\tau_i}$ so $\alpha_{-i} = \gamma_{-i} = 1$ and $\alpha_i = \gamma_i = \frac{1-\tau_i}{\tau_i}$ maximizes $\tau_i(\alpha_i + \gamma_i)$ among these rules. ■

7 Conclusion

The present paper is a first very preliminary approach to "optimal" boundedly rational decision making. Analogous to the rational paradigm, methodology is presented for selecting a unique behavior for a given discount factor. We choose a distribution free approach and rely on a maxmin analysis.

Selection is illustrated in a particular class of behavioral rules. Many more results for more general rules can be obtained although proofs are often tedious. We focus in this paper on few

cases as they already bring the relevant issues concerning how to select to light. Moreover, we find the present results interesting.

This is the first paper that presents a theory and analysis of boundedly rational decision making when no additional information about other decisions or other individuals is present. Other approaches add additional information about similar decisions (Gilboa and Schmeidler, 1995) or about the behavior of others (Schlag, 1998).

A related paper classifying boundedly rational decision rules with specific properties without selecting among them is due to Börgers, Morales and Sarin (1998). They characterize improving rules and thus present a first step in selecting for myopic behavior. Their rules are less attractive for completely patient individuals as the best action will only be selected with high probability. Notice that the rule we present in Section 6.3 is intuitively simpler and is maximizing.

References

[1] Arthur (1993), "On Designing Economic Agents that Behave Like Human Agents," *J. Evol. Econ.* **3**, 1-22.

[2] Block and Marshak (1960), "Random Orderings and Stochastic Theories of Responses," in: Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., Mann, H.B. (eds), *Contributions to Probability and Statistics*, Stanford University Press.

[3] Börgers and Sarin (1997), "Learning Through Reinforcement and Replicator Dynamics", *Journal of Economic Theory* **77**, 1-14.

[4] Börgers, Morales and Sarin (1998), "Simple Behavior Rules which lead to Expected Payoff Maximizing Choices", Mimeo, University College London, http://www.ucl.ac.uk/~uctpa01/Papers.htm.

[5] Bush and Mosteller (1955). *Stochastic Models for Learning.* Wiley, New York.

[6] Cross (1973), "A Stochastic Learning Model of Economic Behavior," *Quart. J. Econ.* **87**, 239-266.

[7] Easley and Rustichini (1999), "Choice without Beliefs," *Econometrica* **67**, 1157-84.

[8] Erev and Roth (1998), "Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibrium", *American Economic Review* **88**, 848-881.

[9] Gale, Binmore and Samuelson (1995) "Learning to be Imperfect: the Ultimatum Game", *Games and Economic Behaviour* **8**, 56-90.

[10] Gilboa and Schmeidler (1995) "Case-Based Decision Theory", *Quarterly Journal of Economics* **110**, 605-39.

[11] Narendra and Thathachar (1989) *Learning Automata: An Introduction.* Englewood Cliffs: Prentice Hall.

[12] Osborne and Rubinstein (1998), "Games with Procedurally Rational Players", *American Economic Review* **88**, 834-47.

[13] Rustichini (1999) "Optimal Properties of Stimulus-Response Learning Models". *Games and Economic Behavior* **29**, 244-273.

[14] Savage (1972), *The Foundation of Statistics*, Dover, New York

[15] Schlag (1998), "Why Imitate, and If So, How?", *Journal of Economic Theory* **78**, 130-156.

[16] Selten (1999) "What is Bounded Rationality?" SFB Discussion Paper B-454, University of Bonn.

[17] Simon (1955) "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* **69**, 99-118.

[18] Simon (1982) *Models of Bounded Rationality*, MIT Press.

[19] von Neumann and Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton Univ. Press.

[20] Weibull (1995), *Evolutionary Game Theory*, The MIT press.

A Proof for four states

Assumes that parameters are such that the process is ergodic (e.g., $\alpha_i, \gamma_i > 0$). Let x_i be the probability assigned by the stationary distribution of being in state s_i. Then $(x_2 + \tau_1 x_1 + (1-\tau_1) x_{-1})$ is the probability of choosing action one. As the individual is perfectly patient we only need to consider payoffs in the stationary distribution.

It is easily verified that

$$\begin{aligned} x_{-2} &= \left(\tau_1 \alpha_1^1 \pi_2 + (1-\tau_1)\gamma_{-1}^1 (1-\pi_1)\right)\left((1-\tau_1)\alpha_{-1}^1 \pi_2 + \tau_1 \gamma_1^1 (1-\pi_1)\right)\gamma_2^1 (1-\pi_1)/N, \\ x_{-1} &= \gamma_2^1 (1-\pi_2)\left((1-\tau_1)\alpha_{-1}^1 \pi_2 + \tau_1 \gamma_1^1 (1-\pi_1)\right)\gamma_2^1 (1-\pi_1)/N, \end{aligned}$$

$$
\begin{aligned}
x_1 &= \left(\gamma_2^1\right)^2 (1-\pi_1)(1-\pi_2)\left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_{-1}^1(1-\pi_2)\right)/N, \\
x_2 &= \gamma_2^1(1-\pi_2)\left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_{-1}^1(1-\pi_2)\right)\left(\tau_1\alpha_1^1\pi_1 + (1-\tau_1)\gamma_1^1(1-\pi_2)\right)/N
\end{aligned}
$$

where N is such that $\sum_{i=-2}^{2} x_i = 1$.

$$
\begin{aligned}
& N\left(x_2 + \tau_1 x_1 + (1-\tau_1)x_{-1}\right) \\
= \; & \gamma_2^1(1-\pi_2) \\
& * \begin{pmatrix} \left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_{-1}^1(1-\pi_2)\right)\left(\tau_1\alpha_1^1\pi_1 + (1-\tau_1)\gamma_1^1(1-\pi_2)\right) \\ +\gamma_2^1(1-\pi_1)\left((\tau_1)^2\gamma_{-1}^1(1-\pi_2) + \tau_1(1-\tau_1)\left(\alpha_{-1}^1\pi_1 + \gamma_1^1(1-\pi_1)\right) + (1-\tau_1)^2\alpha_{-1}^1\pi_2\right) \end{pmatrix}.
\end{aligned}
$$

Calculating the analogous expression $N\left(x_{-2} + (1-\tau_1)x_{-1} + \tau_1 x_1\right)$ for action two we obtain the quotient of these two expressions

$$
\begin{aligned}
& \frac{N\left(x_2 + \tau_1 x_1 + (1-\tau_1)x_{-1}\right)}{N\left(x_{-2} + (1-\tau_1)x_{-1} + \tau_1 x_1\right)} \\
= \; & \frac{(1-\pi_2)}{(1-\pi_1)} \\
& \frac{\begin{pmatrix} \left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_{-1}^1(1-\pi_2)\right)\left(\tau_1\alpha_1^1\pi_1 + (1-\tau_1)\gamma_1^1(1-\pi_2)\right) \\ +\gamma_2^1(1-\pi_1)\left((\tau_1)^2\gamma_{-1}^1(1-\pi_2) + \tau_1(1-\tau_1)\left(\alpha_{-1}^1\pi_1 + \gamma_1^1(1-\pi_1)\right) + (1-\tau_1)^2\alpha_{-1}^1\pi_2\right) \end{pmatrix}}{\begin{pmatrix} \left(\tau_1\alpha_1^1\pi_2 + (1-\tau_1)\gamma_{-1}^1(1-\pi_1)\right)\left((1-\tau_1)\alpha_{-1}^1\pi_2 + \tau_1\gamma_1^1(1-\pi_1)\right) \\ +\gamma_2^1(1-\pi_2)\left((\tau_1)^2\gamma_1^1(1-\pi_1) + \tau_1(1-\tau_1)\left(\alpha_{-1}^1\pi_2 + \gamma_{-1}^1(1-\pi_2)\right) + (1-\tau_1)^2\alpha_{-1}^1\pi_1\right) \end{pmatrix}}
\end{aligned}
$$

where $\gamma_2 \to 0$ yields a right hand side of

$$
\frac{(1-\pi_2)\left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_{-1}^1(1-\pi_2)\right)\left(\tau_1\alpha_1^1\pi_1 + (1-\tau_1)\gamma_1^1(1-\pi_2)\right)}{(1-\pi_1)\left(\tau_1\alpha_1^1\pi_2 + (1-\tau_1)\gamma_{-1}^1(1-\pi_1)\right)\left((1-\tau_1)\alpha_{-1}^1\pi_2 + \tau_1\gamma_1^1(1-\pi_1)\right)}
$$

Improving requires that the term above equals one if $\pi_1 = \pi_2$. Our next results states a necessary condition for this to be true.

Lemma 11 *The condition that quotient equals* 1 *for* $\pi_1 = \pi_2$ *requires*

(1) $\gamma_{-1} = \gamma_1$ *which yields*

$$
\frac{(1-\pi_2)}{(1-\pi_1)} \frac{\begin{pmatrix} \left(\tau_1\alpha_1^1\pi_1 + (1-\tau_1)\gamma_1(1-\pi_2)\right)\left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_1(1-\pi_2)\right) \\ +\gamma_2^1(1-\pi_1)\left((\tau_1)^2\gamma_1(1-\pi_2) + \tau_1(1-\tau_1)\left(\alpha_{-1}^1\pi_1 + \gamma_1(1-\pi_1)\right) + (1-\tau_1)^2\alpha_{-1}^1\pi_2\right) \end{pmatrix}}{\begin{pmatrix} \left(\tau_1\alpha_1^1\pi_2 + (1-\tau_1)\gamma_1(1-\pi_1)\right)\left((1-\tau_1)\alpha_{-1}^1\pi_2 + \tau_1\gamma_1^1(1-\pi_1)\right) \\ +\gamma_2^1(1-\pi_2)\left((\tau_1)^2\gamma_1^1(1-\pi_1) + \tau_1(1-\tau_1)\left(\alpha_{-1}^1\pi_2 + \gamma_1(1-\pi_2)\right) + (1-\tau_1)^2\alpha_{-1}^1\pi_1\right) \end{pmatrix}}
$$

and for $\gamma_2 \to 0$

$$
\frac{(1-\pi_2)}{(1-\pi_1)} \frac{\left(\tau_1\alpha_1^1\pi_1 + (1-\tau_1)\gamma_1(1-\pi_2)\right)\left((1-\tau_1)\alpha_{-1}^1\pi_1 + \tau_1\gamma_1(1-\pi_2)\right)}{\left(\tau_1\alpha_1^1\pi_2 + (1-\tau_1)\gamma_1(1-\pi_1)\right)\left((1-\tau_1)\alpha_{-1}^1\pi_2 + \tau_1\gamma_1^1(1-\pi_1)\right)}
$$

or

(2) [$\tau_1 = 0.5$ and $\alpha_{-1} = \alpha_1$] which yields

$$\frac{(1-\pi_2)}{(1-\pi_1)}\frac{\begin{array}{c}(\alpha_1\pi_1+\gamma_{-1}(1-\pi_2))(\alpha_1\pi_1+\gamma_1(1-\pi_2)) \\ +0.5\gamma_2(1-\pi_1)(\gamma_{-1}(1-\pi_2)+\alpha_1\pi_1+\gamma_1(1-\pi_1)+\alpha_1\pi_2)\end{array}}{\begin{array}{c}(\alpha_1\pi_2+\gamma_{-1}(1-\pi_1))(\alpha_1\pi_2+\gamma_1(1-\pi_1)) \\ +0.5\gamma_2(1-\pi_2)(\gamma_1^1(1-\pi_1)+\alpha_1^1\pi_2+\gamma_{-1}^1(1-\pi_2)+\alpha_1^1\pi_1)\end{array}}$$

and for $\gamma_2 \to 0$

$$\frac{(1-\pi_2)}{(1-\pi_1)}\frac{(\alpha_1\pi_1+\gamma_{-1}(1-\pi_2))(\alpha_1\pi_1+\gamma_1(1-\pi_2))}{(\alpha_1\pi_2+\gamma_{-1}(1-\pi_1))(\alpha_1\pi_2+\gamma_1(1-\pi_1))}.$$

Proof. (of Lemma) The requirement is

$$\begin{array}{rc} & \left(\begin{array}{c}((1-\tau_1)\alpha_{-1}^1\pi_1+\tau_1\gamma_{-1}^1(1-\pi_1))(\tau_1\alpha_1^1\pi_1+(1-\tau_1)\gamma_1^1(1-\pi_1)) \\ +\gamma_2^1(1-\pi_1)\left((\tau_1)^2\gamma_{-1}^1(1-\pi_1)+\tau_1(1-\tau_1)\gamma_1^1(1-\pi_1)+(1-\tau_1)\alpha_{-1}^1\pi_1\right)\end{array}\right) \\ = & \left(\begin{array}{c}((1-\tau_1)\alpha_{-1}^1\pi_1+\tau_1\gamma_1^1(1-\pi_1))(\tau_1\alpha_1^1\pi_1+(1-\tau_1)\gamma_{-1}^1(1-\pi_1)) \\ +\gamma_2^1(1-\pi_1)(\tau_1\gamma_1^1(1-\pi_1)+\tau_1(1-\tau_1)\gamma_{-1}^1(1-\pi_1)+(1-\tau_1)\alpha_{-1}^1\pi_1)\end{array}\right)\end{array} \quad (14)$$

so for $\pi_1 = 0$ this means

$$\tau_1\gamma_{-1}^1\gamma_2^1+(1-\tau_1)\gamma_1^1(\gamma_{-1}^1+\gamma_2^1) = \tau_1\gamma_1^1\gamma_2^1+(1-\tau_1)\gamma_{-1}^1(\gamma_1^1+\gamma_2^1)$$

which is equivalent to $(1-2\tau_1)\gamma_2(\gamma_1-\gamma_{-1}) = 0$ so either $\tau_1 = 0.5$ or $\gamma_{-1} = \gamma_1$.

If $\tau_1 = 0.5$ then the requirement is $0 = \pi_1(1-\pi_1)(\alpha_{-1}-\alpha_1)(\gamma_1-\gamma_{-1})$ which is also sufficient if $\alpha_{-1} = \alpha_1$.

On the other hand, $\gamma_{-1} = \gamma_1$ implies that (14) holds with equality. ■

Proposition 12 *Best ex-ante undominated improving rule for fixed difference for* $\gamma_2 \to 0$: $\alpha_{-1} = \alpha_1 = 1$, $\{\gamma_{-1}, \gamma_1\} = \{0, 1\}$ *and* $\tau_1 = 0.5$ *which yields*

$$\frac{(1-\pi_2)\pi_1(1+\pi_1-\pi_2)}{(1-\pi_1)\pi_2(1+\pi_2-\pi_1)+(1-\pi_2)\pi_1(1+\pi_1-\pi_2)}$$

where $d^1(\rho) = 2\rho^2\frac{3+4\rho^2}{1+12\rho^2}$.

Proof. Consider $\gamma_2 \to 0$ and $\gamma_{-1} = \gamma_1$ then we aim to maximize

$$\frac{(1-\pi_2)(\tau_1\alpha_1\pi_1+(1-\tau_1)\gamma_1(1-\pi_2))((1-\tau_1)\alpha_{-1}\pi_1+\tau_1\gamma_1(1-\pi_2))}{(1-\pi_1)(\tau_1\alpha_1\pi_2+(1-\tau_1)\gamma_1(1-\pi_1))((1-\tau_1)\alpha_{-1}\pi_2+\tau_1\gamma_1(1-\pi_1))}$$

subject to $\pi_1 > \pi_2$. Notice that this expression is monotone in α_{-1} and in α_1. Taking a minimum according to π_2 and a maximum according to the parameters shows that α_1 and α_{-1} are on the

boundary. In the following we will calculate this expression for the extreme values of α_{-1} and α_1 for given π_1 and π_2 and then derive $d^1(\rho)$.

$\alpha_{-1} = \alpha_1 = 0$ yields $\frac{(1-\pi_2)^3}{(1-\pi_1)^3}$ where $d^1 = \rho\left(\frac{2}{1+(1-2\rho)^3} - 1\right)$.

$\{\alpha_{-1}, \alpha_1\} = \{0, 1\}$ yields

$$\frac{(\tau_1\pi_1 + (1-\tau_1)\gamma_1(1-\pi_2))(1-\pi_2)^2}{(\tau_1\pi_2 + (1-\tau_1)\gamma_1(1-\pi_1))(1-\pi_1)^2}$$

and hence using linearity the candidates are $\frac{\pi_1(1-\pi_2)^2}{\pi_2(1-\pi_1)^2}$ with

$$d^1 = \frac{\rho}{2}\frac{\left(1-\rho-\sqrt{\rho^2+2}\right)\left(\rho+2-\sqrt{\rho^2+2}\right)^2}{\rho^2+7-5\sqrt{\rho^2+2}} - \rho$$

or $\frac{(1-\pi_2)^3}{(1-\pi_1)^3}$.

$\alpha_{-1} = \alpha_1 = 1$ yields

$$\frac{(1-\pi_2)}{(1-\pi_1)}\frac{(\tau_1\pi_1 + (1-\tau_1)\gamma_1(1-\pi_2))((1-\tau_1)\pi_1 + \tau_1\gamma_1(1-\pi_2))}{(\tau_1\pi_2 + (1-\tau_1)\gamma_1(1-\pi_1))((1-\tau_1)\pi_2 + \tau_1\gamma_1(1-\pi_1))}$$

Candidates for extrema are $\tau_1 = 1$ that yields $\frac{\pi_1(1-\pi_2)^2}{\pi_2(1-\pi_1)^2}$ and $\tau_1 = 0.5$ which yields $\frac{(1-\pi_2)(\pi_1+\gamma_1(1-\pi_2))^2}{(1-\pi_1)(\pi_2+\gamma_1(1-\pi_1))^2}$. Given ρ, candidates are $\gamma_1 = 0$ which yields $\frac{(1-\pi_2)\pi_1^2}{(1-\pi_1)\pi_2^2}$ (same d^1 as $\frac{\pi_1(1-\pi_2)^2}{\pi_2(1-\pi_1)^2}$), $\gamma_1 = 1$ which yields $\frac{(1-\pi_2)(\pi_1+1-\pi_2)^2}{(1-\pi_1)(\pi_2+1-\pi_1)^2}$ ($d^1 = \frac{\rho(1+2\rho)^2}{1-\rho+8\rho^2-4\rho^3} - \rho$) and γ_1, π_2 such that $\frac{d}{d\gamma_1} = 0$ which means $\frac{\pi_1}{\pi_2} = \frac{1-\pi_2}{1-\pi_1}$ and hence $\pi_2 = 0.5(1-2\rho)$. Verifying $\frac{d}{d\pi_2} = 0$ at $\pi_2 = 0.5$ and $\rho = 0$ we find $\gamma_1 = 1/3$ which yields as candidate $\frac{(1-\pi_2)(\pi_1+(1-\pi_2)/3)^2}{(1-\pi_1)(\pi_2+(1-\pi_1)/3)^2}$

Consider $\gamma_2 \to 0$, $\tau_1 = 0.5$ and $\alpha_{-1} = \alpha_1$ which yields

$$\frac{(1-\pi_2)\left(\alpha_1\pi_1 + \gamma_{-1}(1-\pi_2)\right)\left(\alpha_1\pi_1 + \gamma_1(1-\pi_2)\right)}{(1-\pi_1)\left(\alpha_1\pi_2 + \gamma_{-1}(1-\pi_1)\right)\left(\alpha_1\pi_2 + \gamma_1(1-\pi_1)\right)}$$

(1) $\gamma_1 = \gamma_{-1}$ is best if either $\alpha_1 = 0$ or $\alpha_1 = 1$ which we already have above,

(2) $\{\gamma_{-1}, \gamma_1\} = \{0, 1\}$ yields $\frac{\pi_1(\alpha_1\pi_1+1-\pi_2)(1-\pi_2)}{\pi_2(\alpha_1\pi_2+1-\pi_1)(1-\pi_1)}$ so best if $\alpha_1 = 1$ which yields $\frac{(1-\pi_2)\pi_1(1+\pi_1-\pi_2)}{(1-\pi_1)\pi_2(1+\pi_2-\pi_1)}$ ($d^1 = 2\rho^2\frac{3+4\rho^2}{1+12\rho^2}$) or $\alpha_1 = 0$ which yields $\frac{\pi_1(1-\pi_2)^2}{\pi_2(1-\pi_1)^2}$.

So $\alpha_{-1} = \alpha_1 = 1$, $\{\gamma_{-1}, \gamma_1\} = \{0, 1\}$ and $\tau_1 = 0.5$ best as

$$\frac{\pi_1(\pi_1+1-\pi_2)(1-\pi_2)}{\pi_2(\pi_2+1-\pi_1)(1-\pi_1)} \geq \frac{(\pi_1+(1-\pi_2)/3)^2(1-\pi_2)}{(\pi_2+(1-\pi_1)/3)^2(1-\pi_1)},$$

$$\frac{\pi_1(1-\pi_2)^2}{\pi_2(1-\pi_1)^2} \geq \frac{(\pi_1+1-\pi_2)^2(1-\pi_2)}{(\pi_2+1-\pi_1)^2(1-\pi_1)}$$

for $\pi_1 > \pi_2$ and it outperforms all other candidates above in fixed differences. ■

Proof. (of Corollary 8) In the proof of Proposition 7 we derived the candidates for maximizing $z_1^{(\infty)}/z_2^{(\infty)}$ for given π_1 and π_2 with $\pi_1 > \pi_2$. These four rules are described in Proposition 7 and in Corollary 8 (i)-(iii). If one of these rules is not ex-ante undominated then it must be dominated by one of the other four candidate rules. However, we find that none of these four rules dominates one of the others as $\frac{1-\pi_2}{1-\pi_1} > (<) \frac{\pi_1}{\pi_2}$ implies

$$\frac{(1-\pi_2)^3}{(1-\pi_1)^3} > (<) \frac{\pi_1 (1-\pi_2)^2}{\pi_2 (1-\pi_1)^2} > (<) \frac{\pi_1 (\pi_1 + 1 - \pi_2)(1-\pi_2)}{\pi_2 (\pi_2 + 1 - \pi_1)(1-\pi_1)} > (<) \frac{\pi_1^2 (1-\pi_2)}{\pi_2^2 (1-\pi_1)} .$$

■