Imitation and Minimax Regret[1]

Karl H. Schlag[2]

December 3 2004

[1]This first version of this paper was written for an imitation workshop at UCL, London 5.12.2003.

[2]Economics Department, European University Institute, Via della Piazzuola 43, 50133 Florence, Italy, Tel: 0039-055-4685951, email: schlag@iue.it

Abstract

Myopic individuals belonging to an infinite population repeatedly and independently have to choose one of two actions. Between choices each individual is informed about the success of one random other individual. We search among rules with a single round of memory for a rule that attains minimax regret in each round when all others use this rule. Behavior will not be imitative in round two but can be imitative starting round three. A simple linear imitation rule even causes all to learn the best action in the long run. The informational setting is varied to further investigate the role of imitation.

1 Introduction

Imitation is a commonly observed human behavior. This paper aims to add to our understanding of circumstances in which imitation is desirable.

We start by listing a few of the many reasons why imitation may be a good "strategy": (a) to take advantage of being regarded similar to others, (b) to save calculation costs, (c) to profit from superior information of others, and (d) to aggregate information under heterogeneous information. In line with (a), Veblen (1899) describes lower social classes imitating higher ones through adapting their fashion, in the model of Kreps and Wilson (1982) the more flexible chain store imitates the behavior of the chain store that can only act tough. A nice example for (b) is Rogers (1989) who shows in an evolutionary model how an equilibrium proportion of imitators arises in a society when learning the true state is costly. Examples for (c) range from children learning to research and development strategies. Banerjee (1992) and Squintani and Välimäki (2002) are two examples belonging to category (d) that show how intrinsic information in observed choice can lead rational individuals to imitate others. A less rational perspective on (d) is given by Schlag (1998) who shows that individuals who aim to increase average payoffs will choose to imitate.

We will consider the same learning environment as Schlag (1998). However, instead of being concerned with change in average payoffs we will assume that individuals evaluate their decisions in unknown environments according to regret. More specifically they will choose to 'minimax regret'. The presentation of how to model regret will be very detailed as it has not been done before in an environment that is characterized by repeated decision making and learning from others. We will find (i) that aiming to minimax regret can support imitative behavior if individuals are sufficiently experienced and (ii) that such a level of knowledge or experience is reached by myopic individuals after they have faced the decision-problem twice. This research fits in the broader program of showing *how simple rules need not be 'stupid'*. Simplicity will not only be informally associated to imitation but also to linearity of the behavioral rule.

To be more specific, consider the decision-making and learning environment used in Schlag (1998). Individuals belonging to a large (countably infinite) population have to repeatedly and independently choose between two actions. Each action yields a random payoff according to some given stationary but unknown payoff distribution. Between rounds individuals receive information about the success of others. More specifically, after each own choice each individual observes the action chosen and payoff received of a single other random member from the population. We also consider an alternative scenario with entry and exit in which each individual only makes a single choice.

We limit attention to behavioral rules (or strategies) in which choice only

depends on observations made in the previous round. This limitation captures exogenous memory and computational complexity bounds often imposed in models of bounded rationality and in particular in models involving evolution. It is particularly natural in the entry and exit scenario. So a behavioral rule for an individual is the specification of which action to choose in the first round, and for any later round, which action to choose conditional on own choice and payoff in the previous round, as well on observed behavior of the random other in the previous round. In particular, choice may depend on the number of the round. We refer to *imitative behavior* if an individual chooses the same action again whenever the action observed coincides with the own action chosen. In this sense imitation is not conditional on success. A rule is *linear* if transition probabilities are linear in payoffs.

The decision involving the choice between two actions together with the specification of the payoff distribution underlying each action is called the *decision-problem.* So all individuals repeatedly and independently face the same unknown decision-problem. Individuals do not have a prior over the set of possible decision-problems as they do in the formal description of a *two-armed bandit.* Instead they only know that payoffs belong to a given bounded interval that we normalize to $[0, 1]$. The decision-problem that individuals face is a primitive of the model.

Individuals are myopic and only care about payoffs achieved in the next round. We assume that individuals are risk neutral regarding objective probabilities. In other words, if an individual would know the decision-problem and hence would know the underlying probability distributions then she would choose in each round the action that yields the higher expected payoff. For a given decision-problem an action that maximizes the expected payoff will be called a *best action.* However, as stated above, individuals do not know the underlying payoff distributions.

Up to here there are no changes as compared to the setting by Schlag (1998). The difference is that we consider decision-making under *minimax regret* while Schlag (1998) searches for rules that guarantee that average payoffs always increase. First we describe what we mean by regret, then show how we avoid using a prior and finally explain how we incorporate learning.

Regret is associated to choosing a given action in a given decision-problem. There are two basic ways of evaluating the regret in our environment which we differentiate by referring either to 'ex-post-regret' or to 'interim-regret'. This distinction is novel. Ex-post-regret is associated to the stimuli received ex-post when the individual is informed about the payoff the other action would have realized.[1][2] Formally, *ex-post-regret* is measured in a given state by the difference

[1] Foregone payoffs are learned after each choice. This scenario arises naturally in situations in which the individual learns the state of nature ex-post and hence can calculate the payoff that the alternative action would have yielded.

[2] Payoffs of others are assumed not to enter the calculation of own regret.

between the largest payoff the individual could have realized and the actual payoff she realized. Expectations are taken in order to evaluate ex-post-regret ex-ante in a given decision-problem. Interim-regret on the other hand is not associated to stimuli received ex-post. Instead it measures the opportunity loss due to not knowing the characteristics of an unknown random environment, here of the decision-problem. Formally *interim-regret* is calculated in a given decision-problem by the difference between the expected payoffs of the best action and the expected payoffs of the particular action chosen. Consequently, interim-regret measures the extent to which learning is valuable in a given decision problem. When evaluated in a minimax approach (see below) it has the flavor of inducing learning when learning matters.

Regret measures the increase in expected payoffs that would result if the individual would learn more about the environment before she chooses an action. Under ex-post-regret she would learn the specific payoffs realized by each action, under interim-reget she would learn the characteristics (i.e. payoff distributions) of the decision-problem itself. Consequently, interim-regret from choosing a best action is necessarily zero while ex-post-regret is typically strictly positive. Of course, if each arm yields a deterministic payoff then ex-post-regret and interim-regret coincide.

Decision-making under regret typically proceeds by evaluating *minimax regret* (Wald, 1950, for a notable exception see Bell, 1982). Accordingly, the decision-maker chooses the action that minimizes among all actions the maximum among all potential decision-problems of (ex-post- or interim-) regret. In particular, if the individual were to know which decision problem she is facing then she would choose the action that minimizes regret. Under either interim- or ex-post-regret this is equivalent to maximizing expected payoffs. In this sense minimax regret is an alternative to specifying a prior over the set of potential decision problems.

Both interim- and ex-post-regret are consistent with the original definition of minimax regret due to Wald (1950). To derive minimax regret we build on Savage (1951) who showed how to calculate minimax regret via solving for an equilibrium of a zero-sum game. Milnor's (1954) axiomatization of minimax regret and Linhart and Radner's (1989) minimax regret analysis refer to ex-post-regret. On the other hand, minimax regret analyzed by Berry and Fristedt (1985) and Chamberlain (2000) is what we term interim-regret. Notice that the "no regret" literature (e.g. Hart and Mas-Colell, 2000) applies only to infinitely patient individuals but can interpreted as a refinement of interim-regret.

We use the above definitions to analyze choice under minimax regret in round one. However, for later rounds we have to specify how to deal with own previous experience and with behavior of others. We accommodate for previous history by assuming as in Berry and Fristedt (1985) that an individual commits to a behav-

ioral rule for entire play before even making her first choice.[3] Behavior of others is incorporated by using an equilibrium approach where we refer to *equilibrium minimax regret*. Accordingly, each individual calculates minimax regret taking behavior of others as given. We focus on situations where all use the same rule.

We are not aware of other papers that consider minimax regret in an equilibrium approach. However, Linhart and Radner (1989: sec. 3) do consider a very weak form of beliefs as they assume that the buyer anticipates minimal rationality of the seller (no sale below cost) when calculating own ex-post-regret.

Most of our analysis is concerned with interim-regret. We find that equilibrium minimax interim-regret is attained when all individuals use the following linear rule. Choose each action equally likely in the first round. Starting round two, with probability equal to payoff obtained be satisfied with performance of action chosen and choose it again. If not satisfied in round two then switch actions, if not satisfied in round three or later then imitate action of individual observed. Individual behavior thus involves some experimentation in round two and only imitative behavior thereafter. This observation is more generally true for any linear rule that attains equilibrium minimax interim-regret. When all use such a linear rule then average payoffs increase over time and all choose the best action in the long run.[4]

In an infinite population observed play is not influenced by own previous play. However, in a finite population typically own and observed choice are correlated through observing each others behavior in previous rounds. Consequently one may expect that an individual is more reluctant to imitate in finite populations. To uncover the impact of such correlation we consider the extreme case in which the population consists of only two individuals. We find that the linear rule presented above still attains equilibrium minimax interim-regret. Of course the long run convergence result no longer holds as with positive probability both individuals choose the same action in round two and then, due to their imitative behavior starting round three, never change actions again.

So what is the intuition for why imitation emerges in round three but not in round two? Consider the event in which own and observed action coincide and where own and observed payoff equal the lowest possible value zero. If this happens in round one then the individual switches actions in round two as low payoffs are sufficient evidence that the alternative action is possibly better. There is no value in the fact that the observed chose the same action as individuals random-

[3] We also present an alternative entry and exit scenario in which we assume that individuals calculate expected regret prior to entry. Here regret is calculated prior to entering into the population.

[4] So as an aside we obtain that a linear rule that attains minimax interim-regret (when the individual only cares for the current payoff) also attains equilibrium minimax regret if individuals are infinitely patient.

ize equally likely in the first round. However, if our specified event happens in round two (or later) then observing an action contains information about earlier outcomes. The fact that observed and own action coincide reinforces the belief that the own action is best. It turns out that this offsets the negative impression of the own action resulting from the poor performance.

To further investigate the role of information included in the observed action we consider an alternative scenario and assume that the individual observes after round two the success of an inexperienced individual who has chosen each action equally likely. In this alternative setting we find that behavior is no longer imitative. This conclusion remains true even if the observed individual is marginally experienced. So we find that imitation is not justified when there is only some intrinsic information. It requires sufficient information.

Results on ex-post-regret are less insightful. While the rule presented above also attains equilibrium minimax ex-post-regret both for infinite populations and for populations with only two individuals, a much simpler rule that specifies to never switch actions after round two also has this property. In particular, the minimax value of ex-post-regret remains constant after round two.

We proceed as follows. In Section 2 we introduce the environment and the set of behavioral rules. In Section 3 we present the two different concepts of regret together with the definitions of minimax regret and equilibrium minimax regret. Section 4 contains the main body on selecting behavior based on equilibrium minimax interim-regret with separate subsections on the infinite population, on the setting with only two individuals and on alternative scenarios to investigate the role of intrinsic information in observed behavior. Section 5 contains the material on ex-post-regret. In Section 6 we conclude and in the appendix we elaborate on why it is not sensible to define minimax regret conditional on observed histories.

2 Decision Problems and Rules

There are two actions A and B. Letters C, C', D and D' will be used to denote generic actions from $\{A, B\}$. A *decision-problem* ψ is defined as a tuple $\psi = (P_A, P_B)$ where payoffs realized by choosing action C are distributed according to P_C. We assume that P_A and P_B are independent and that only payoffs in $[0, 1]$ can be achieved. ψ is called a *Bernoulli decision-problem* if $P_C(0) + P_C(1) = 1$. Let $\pi_C := \pi_C(\psi) := \int_0^1 x dP_C(x)$ be the expected payoff achieved when choosing C in decision-problem ψ.

Consider a countably infinite population of individuals and a single decision-problem ψ unknown to the individuals. Assume that each individual repeatedly and independently faces ψ. Choices are synchronous so we speak of *rounds* where round n refers to the $n-th$ choice of each of the individuals. After making

her choice in round n an individual observes the $n-th$ choice and payoff of another randomly selected individual. A behavioral rule is the description of what an individual chooses in each round as a function of her previous observations. To reflect use of simple rules we limit attention to rules that do not condition on choices made or payoffs yielded prior to the previous round. Accordingly, a *behavioral rule* f can be described as follows. $f = \left(f^{(n)}\right)_{n=1}^{\infty}$ is such that $f^{(1)} \in \Delta\{A, B\}$ and $f^{(n)} : (\{A, B\} \times [0,1])^2 \to \Delta\{A, B\}$ for $n \geq 2$ where $f_D^{(1)}$ denotes the probability of choosing action D in the first round and $f^{(n)}(C, x, C', y)_D$ describes the probability of choosing action D in round $n \geq 2$ after choosing action C and achieving payoff x in round $n-1$ and observing an individual who chose action C' and obtained payoff y in round $n-1$. $f^{(n)}$ is called the rule for round n.

Our analysis will also apply to the following alternative scenario in which each individuals only makes a single choice and is then replaced by a new individual. The entering individual observes choice and payoff of the replaced, observes choice and payoff of one random other and then chooses an action in the next round. We loosely say that an individual does not switch actions when we mean that she adapts the action of the individual replaced. In this alternative setting an individual entering in round n selects a rule $f^{(n)}$ for round n. Here it is particularly natural to assume that the rule $f^{(n)}$ for round n conditions on actions and payoffs in round $n-1$ only. Of course there is no information about behavior of others upon entry in the first round. For later reference we refer to this setting as the *entry and exit scenario* while the previous will be called the *longevity scenario*. Both settings are taken directly from Schlag (1998).

We now introduce some possible attributes of rules. For $n \geq 2$ we say that the rule f is *imitative* in round n (or the rule $f^{(n)}$ for round n is imitative) if $f^{(n)}(C, x, C, y)_C = 1$. Notice that as there are only two actions, imitation is equivalent to not changing actions whenever own and observed actions coincide. We say that the rule f is *linear* in round $n \geq 2$ if $f^{(n)}(C, x, C', y)_D$ is a linear function of x and y for all $C, C', D \in \{A, B\}$. We say that the rule f is *quasi-linear* in round n if $f^{(n)}(C, x, C, y)_D$ and $\left(f^{(n)}(A, x, B, y)_B - f^{(n)}(B, y, A, x)_A\right)$ are linear in x and y (for $x, y \in [0,1]$). So linear rules are quasi-linear. We call f_L the *linearization of* f if $f_L^{(1)} = f^{(1)}$, $f_L^{(n)}$ is linear in all rounds $n \geq 2$ and $f_L^{(n)}(C, x, C', y) = f^{(n)}(C, x, C', y)$ holds whenever $x, y \in \{0, 1\}$ and $n \geq 2$. Recall that a function $g : [0,1]^2 \to [0,1]$ is linear if $g(x, y) \equiv \sum_{v,w \in \{0,1\}} x^v (1-x)^{1-v} y^w (1-y)^{1-w} g(v, w)$. Rules that do not depend on how the actions are labelled will be called symmetric. Formally, we say that the rule $f^{(n)}$ for round n is *symmetric* if $n = 1$ implies $f_A^{(1)} = \frac{1}{2}$ and if $n \geq 2$ implies $f^{(n)}(C, x, C', y)_D = f^{(n)}(\rho(C), x, \rho(C'), y)_{\rho(D)}$ where $\rho(A) := B$, $\rho(B) := A$. Accordingly, we call the rule f symmetric if $f^{(n)}$ is symmetric for all $n \geq 1$. Given a rule f we will define the rule f_S by setting

$f_{SA}^{(1)} = 1 - f_A^{(1)}$ and $f_S^{(n)}(C, x, C', y)_D = f^{(n)}(\rho(C), x, \rho(C'), y)_{\rho(D)}$ for $n \geq 2$. Thus, $\frac{1}{2}f + \frac{1}{2}f_S$ is a symmetric rule.

Next we present some examples of symmetric rules for round $n \geq 2$. The *simple reinforcement rule* is the linear rule $f^{(n)}$ for round n defined by $f^{(n)}(C, x, D, y)_C = x$ so this rule is not imitative. The imitative quasi-linear rule $f^{(n)}$ for round n that satisfies $f^{(n)}(C, x, D, y)_D = \max\{0, y - x\}$ is called the *proportional imitation rule*. The *proportional reviewing rule* $f^{(n)}$ and the *proportional observation rule* $\tilde{f}^{(n)}$ are the linear imitative rules for round n defined by $f^{(n)}(C, x, D, y)_D = 1 - x$ and $\tilde{f}^{(n)}(C, x, D, y)_D = y$ for $C \neq D$. The imitation rules presented above are taken from Schlag (1998) and Schlag (1999). Learning according to the proportional reviewing rule is also used by Bjoernerstedt and Weibull (1996) and by Gale et al. (1995).

3 Myopic Decision-Making under Regret

Individuals are assumed to only care about payoffs achieved in the next round. However, they do not know which decision-problem they are facing, and unlike a Bayesian decision-maker, our individuals have no (*subjective*) prior over the set of decision-problems they may be facing. Notice that the set of decision-problems that an individual may be facing is uncountable. It is not easy to write up an example of a prior that is not too restrictive, let alone update this prior once information has been gathered. Moreover, optimal behavior has to be calculated every time one finds oneself in a similar environment. We assume that individuals follow a distribution free approach, one that does not depend on a prior. The individual selects a single rule to use in many situations.

One approach to decision-making without priors is to search for a rule that has some property in all decision-problems (e.g. see Schlag (1998) and Boergers et al. (2004)). An alternative is to specify for each decision-problem some measure of performance of a rule and then to evaluate each rule according to the decision-problem in which it performs worst. Decision-making under maximin falls in this category (Savage, 1951, Gilboa & Schmeidler, 1989) and uses expected payoff as the measure of performance. To be more specific, a rule satisfies the maximin criterion if it maximizes among all rules the minimum expected payoff among all decision-problems. This common selection criterion does not yield sensible results in our setting as our set of decision-problems has only little structure. For any rule and in any round the minimum (expected) payoff is obtained in the decision-problem in which both actions yield payoff 0 with certainty. In this 'worst case' decision-problem all rules perform equally (bad) in each round and there is no means to discriminate between different rules.[5]

[5]The same problem is commented on by Linhart & Radner (1989, p. 154) in a different

Instead, we will evaluate performance in a given decision-problem by the difference between what could be achieved under perfect information and what is actually achieved in terms of expected payoffs. This measurement criteria is referred to as regret (Wald, 1950). We present two alternative ways how to evaluate what could be achieved and distinguish accordingly between interim-regret and ex-post-regret. We focus first on choice in round 1 and later then present how to deal with own history as well as with observations of others.

3.1 Interim-Regret

We use the term "interim-regret" to refer to the notion of regret used by Robbins (1952), Berry and Fristedt (1985) and Schlag (2003) among others. We add the prefix "choice" to differentiate this concept from the notion of ex-post-regret defined below. While the original term is not completely changed the prefix highlights that the reader should not think of regret as an emotion felt after observing payoffs of alternative actions ex-post. Regret that is associated to such an emotion is captured by ex-post-regret. Interim-regret is instead a more theoretical construction that measures the effectivity of learning. We choose to introduce interim-regret first as we believe that this concept is a more meaningful in our environment.

To start, uncertainty implicit in the decision-problem is treated as *objective* uncertainty and individuals are assumed to be risk neutral whenever facing only objective uncertainty. Hence, if an individual would know that she is facing decision-problem ψ then she would choose $C \in \arg\max \{\pi_A(\psi), \pi_B(\psi)\}$.[6] Accordingly we call $\arg\max_{C\in\{A,B\}} \pi_C(\psi)$ the set of *best actions* of the decision-problem ψ.

Interim-regret from choosing a specific action is evaluated in a given random environment (with objective uncertainty) before the uncertainty is resolved. It measures the difference between what an individual could obtain if the parameters of the environment are known and what she actually (objectively) obtains (in expectation). In our setting the environment is given by the decision-problem with the probability distributions P_A and P_B. So if the parameters of the environment are known then the individual can achieve $\max \{\pi_A(\psi), \pi_B(\psi)\}$. Here the individual actually obtains $\sum_{C\in\{A,B\}} q_C^{(1)} \pi_C(\psi)$ where $q_C^{(1)}$ is the probability that the given individual chooses action C in round 1. Notice what the individual actually obtains is measured in objective terms using the underlying decision-problem. To summarize, interim-regret in round 1 of an individual facing decision-problem ψ

setting.

[6] More generally one can assume that individuals are identical and maximize expected utility where utilities are contained in a bounded interval. Then payoffs can be identified with utilities and it is as if individuals are risk neutral.

is given by

$$r^{(1)}\left(\psi, q^{(1)}\right) = \max\left\{\pi_A\left(\psi\right), \pi_B\left(\psi\right)\right\} - \sum_{C\in\{A,B\}} q_C^{(1)}\pi_C\left(\psi\right) \ . \tag{1}$$

In particular in a given decision-problem interim-regret is minimized (it equals zero) if the action chosen is a best action e.g. when both actions yield the same expected payoff.

3.2 Ex-post-Regret

We use the term 'ex-post-regret' to refer to the notion of regret used for instance by Milnor (1954) and by Linhart and Radner (1989). Ex-post-regret (from making an irrevocable choice) can be associated to an emotion that arises after uncertainty is resolved when an individual compares the payoff realized by the action she chose to the payoff that would have been realized by the alternative action. Thus, the individual has to learn ex-post about the payoff that the alternative action would have yielded (foregone payoffs are observable). To remain in the original framework we assume that the individual does not base her regret on the payoffs she observes that others have achieved. Such payoffs are only used for learning what to choose in the next round. Ex-post-regret is measured ex-post by the difference between the maximal payoff that was realized by some action and the own payoff realized. More specifically, in state $\omega\left(x,y\right)$ in which action A yields payoff x and action B yields payoff y *ex-post-regret* equals $\max\left\{0, y-x\right\}$ if action A was chosen and equals $\max\left\{0, x-y\right\}$ if action B was chosen. In order to incorporate ex-post-regret into decision-making, ex-post-regret will be measured in a given decision-problem by taking expectations over the ex-post-regret that will arise once uncertainty is resolved. Thus, ex-post-regret in round 1 of an individual facing decision-problem ψ is given by

$$R^{(1)}\left(\psi, q^{(1)}\right) = \int_0^1\int_0^1 \left(q_A^{(1)}\max\left\{0, y-x\right\} + q_B^{(1)}\max\left\{0, x-y\right\}\right) dP_A\left(x\right) dP_B\left(y\right) \ . \tag{2}$$

Notice that we use the fact that arms are assumed to be independent, an assumption not needed in the interim-regret setting.

Ex-post-regret can also be interpreted as the (expected) difference between payoff achievable when parameters of the environment are known and the actual payoff. In this formulation the environment does not only refer to the decision-problem itself but also to the state associated to the specific payoffs realized by each action. More specifically, ex-post-regret from choosing A in state $\omega\left(x,y\right)$ equals $\max\left\{x,y\right\} - x$ which is the same definition as presented above. In other words, ex-post-regret is the difference between what is achieved if all moves of

nature are observable and if they are not. Taking expectations we obtain

$$R^{(1)}\left(\psi, q^{(1)}\right) = \int_0^1 \int_0^1 \max\{x, y\}\, dP_A(x)\, dP_B(y) - \sum_{C\in\{A,B\}} q_C^{(1)} \pi_C(\psi) \ .$$

So the definitions of interim-regret and ex-post-regret differ in what perfect information means when we state that the individual calculates what she could achieve under perfect information. Under interim-regret perfect information means knowing the decision-problem while under ex-post-regret it also means to know the specific payoffs realized by each action. Clearly ex-post-regret is always larger than interim-regret where the two coincide if both arms yield a deterministic payoff.

Similar to interim-regret we find for any given decision-problem that ex-post-regret is minimized if a best action is chosen. However, in contrast to interim-regret, ex-post-regret is typically strictly positive even when choosing a best action (unless one action always yields higher payoffs than the alternative action). In other words, knowledge of the decision-problem cannot keep an individual from experiencing (strictly positive) ex-post-regret unless the decision-problem is particularly simple.

3.3 Minimax Regret

Decision-making under regret typically follows a lexicographic approach (Wald, 1950, for an exception see Bell, 1982). Choice in round one is formally a separate single decision problem for each individual and not a game among the individuals as each individual only cares about payoffs in the next round, as there is no history yet and as payoffs do not depend on what others do. So we can apply the standard lexicographic approach in which the performance of a choice is measured according to the maximum regret it can generate among all decision-problems and where the individual then chooses the (mixed) action that performs best. Formally, the rule $f^{(1)}$ for round one attains *minimax interim-regret in round one* if $f^{(1)} \in \arg\min_{q\in\Delta\{A,B\}} \sup_{\psi} r^{(1)}(\psi, q)$. *Minimax ex-post-regret in round one* is defined analogously.

For all later rounds we have to adjust the above definition as behavior can be conditioned on histories and as individuals influence each other through their observed play. Formally we can adapt the definitions of interim-regret and ex-post-regret and simply replace the index 1 by n in (1) and (2) to define $r^{(n)}\left(\psi, q^{(n)}\right)$ and $R^{(n)}\left(\psi, q^{(n)}\right)$. However, the open question is how to define $q^{(n)}$ in rounds $n > 1$?

Consider first the longevity scenario presented in Section 2 in which individuals face the decision-problem infinitely often. The first thought is to define regret

in round n conditional on the history experienced up to round n. On closer look this does not make sense. It would be as if nature chooses a different decision-problem conditional on which history will occur. In the appendix we explore this in more detail and show how this definition does not yield meaningful results. An alternative is to calculate regret ex-ante before making any choice in the first round and thus to incorporate learning that takes place while facing the decision-problem. This is the approach followed by Berry and Fristedt (1985). It is as if an individual commits to a rule (that prescribes choices in each round based on previous history) before making the first choice. Regret in round n is calculated conditional on each possible history up to round n and then the histories are summed up taking into account the probability that each of them arises in the underlying decision-problem. This is equivalent to calculating the probability $q_C^{(n)}$ of choosing action C in round n from an ex-ante perspective before making a choice in round one and then entering $q^{(n)}$ into the definition of $r^{(n)}$ and $R^{(n)}$. To be more specific, let individuals be indexed by α with $\alpha = 1, 2, ...$ and let f_α denote the rule used by individual α. For each individual α we define $q^{(n)} = q^{(n)}(\alpha)$ recursively as follows. For round one, $q^{(1)}(\alpha) \equiv f_\alpha^{(1)}$. For any later round $n \geq 2$ assume that the proportion of individuals choosing action C in round $n-1$ is well-defined and denote this by $z_C^{(n-1)}$. So $z_C^{(n-1)} = \lim_{N\to\infty} \frac{1}{N} \sum_{\alpha'=1}^{N} q_C^{(n-1)}(\alpha')$, in particular $z^{(n-1)}$ is independent of $q_C^{(n-1)}(\alpha)$ for any α. Note that when all individuals except for perhaps one use the same rule then $z_C^{(n-1)}$ is well defined and $z_C^{(n-1)} = q_C^{(n-1)}(\alpha)$ for almost all α. Then

$$q_D^{(n)}(\alpha) = \sum_{C,C' \in \{A,B\}} q_C^{(n-1)}(\alpha) \cdot z_{C'}^{(n-1)} \cdot \int_0^1 \int_0^1 f_\alpha^{(n)}(C, x, C', y)_D \, dP_C(x) \, dP_{C'}(y) \tag{3}$$

where we use the fact that there are infinitely many individuals so that $z_C^{(n-1)}$ also denotes the average choice of C among all individuals except for individual α.

Using the above definition of $q^{(n)}(\alpha)$ to describe regret in round n we now extend the definition of minimax regret to accommodate the fact that individuals are learning from others facing the same situation. For this we need some more notation. First we replace the index α in $q^{(n)}(\alpha)$ by the profile of rules used and write $q^{(n)}\left(f_\alpha, (f_{\alpha'})_{\alpha' \neq \alpha}\right)$. Next we allow for variations of the rule used by individual α in order to be able to analyze her choice in round n. Let $f \backslash \tilde{f}^{(n)}$ be the rule that behaves like f in rounds $m = 1, .., n-1$ and like $\tilde{f}$ in round n, i.e. $\left(f \backslash \tilde{f}^{(n)}\right)^{(m)} = f^{(m)}$ for $m \neq n$ and $\left(f \backslash \tilde{f}^{(n)}\right)^{(n)} = \tilde{f}^{(n)}$. Combining these two definitions we say that $(f_\alpha)_\alpha$ attains *equilibrium minimax interim-regret* if $f_\alpha^{(n)} \in \arg\min_{\tilde{f}^{(n)}} \sup_\psi r^{(n)}\left(\psi, q^{(n)}\left(f_\alpha \backslash \tilde{f}^{(n)}, (f_{\alpha'})_{\alpha' \neq \alpha}\right)\right)$ holds for all n and all

α. In the special case where all individuals use the same rule f we will abuse notation and say that f attains equilibrium minimax interim-regret when in fact $(f)_{\alpha=1}^{\infty}$ does. Definitions for ex-post-regret are analogous.

Consider now the entry and exit scenario from Section 2 in which individuals only make a single choice and are then replaced. Analogous to the longevity scenario it does not makes sense to assume that individuals choose their rule conditional on regret calculated right before they make their choice, i.e. after they have entered and have observed behavior of others. Instead we assume as above that each individual commits to a rule before she enters. So maximum regret is minimized before entering. This is reminiscent of the approach in Schlag (1998) where individuals calculate the expected change in payoffs before entering.

For round one we set $q^{(1)}(\alpha) \equiv f_{\alpha}^{(1)}$ as above. Behavior in later rounds is defined again recursively. Consider round $n \geq 2$. All information about the state in which an individual enters is included in the decision-problem ψ, the round n and the distribution of choices $z_C^{(n-1)}$. Consequently,

$$q_D^{(n)}(\alpha) = \sum_{C,C' \in \{A,B\}} z_C^{(n-1)} \cdot z_{C'}^{(n-1)} \cdot \int_0^1 \int_0^1 f_{\alpha}^{(n)}(C, x, C', y)_D \, dP_C(x) \, dP_{C'}(y) \ . \quad (4)$$

Equilibrium minimax regret is then defined as in the longevity scenario.

Notice that if individual α never switches actions between rounds $n-1$ and n so $f_{\alpha}^{(n)}(C, \cdot, \cdot, \cdot)_C = 1$ then $q^{(n)}(\alpha) = q^{(n-1)}(\alpha)$. This leads us to the following simple result.

Remark 1 *In either the longevity or entry and exit scenario, if $(f_{\alpha})_{\alpha}$ attains equilibrium minimax (interim- or ex-post-) regret then the maximum value of regret for an individual is non increasing over time.*

Below we will see examples for both types of regret where regret does not strictly decrease.

In the rest of the paper we will only consider equilibria in which all individuals use the same rule. Here the definitions of equilibrium minimax interim-regret and ex-post-regret do not depend on whether we consider the longevity or the entry and exit scenario. This is because (3) and (4) yield the same value of $q_D^{(n)}(\alpha)$ if all individuals except for possibly α are using the same rule. As almost all use the same rule we usually can drop the individual index α. We will also use the fact that $z_D^{(n)} = q_D^{(n)}$, i.e. the ex-ante probability that an individual chooses action D in round n equals the proportion of individuals choosing action D in round n.

4 Equilibrium Minimax Interim-Regret

4.1 Preliminaries

We now show how to recursively find a symmetric linear rule that attains equilibrium minimax interim-regret. Assume that all individuals are using the same symmetric linear rule up to round $n-1$. Consider a specific individual and note that behavior of others in round n does not influence payoffs of this individual. For each Bernoulli decision-problem ψ and each symmetric linear rule g for round n calculate choice $q^{(n)}$ of this individual in round n and use $q^{(n)}$ to calculate interim-regret $r^{(n)}(\psi, g)$. Find an equilibrium (ψ^*, g^*) of the zero-sum game in which the individual chooses a symmetric linear rule g^* for round n in order to minimize $r^{(n)}(\psi, g)$ and nature chooses a Bernoulli decision-problem ψ^* in order to maximize $r^{(n)}(\psi, g)$. Part (i) of the lemma below then states that g^* attains minimax interim-regret in round n. Assume then that all individuals use g^* in round n and continue to round $n+1$.

To simplify notation throughout this subsection, fix n, α and $(f_{\alpha'})_{\alpha' \neq \alpha}$, let g be a generic rule for round n and let $r(\psi, g) := r^{(n)}\left(\psi, q^{(n)}\left(f_\alpha \backslash g^{(n)}, (f_{\alpha'})_{\alpha' \neq \alpha}\right)\right)$. So the claim stated in part (i) of the lemma below is that $g^* \in \arg\min_g \sup_\psi r(\psi, g)$. Part (ii) of the lemma states that any symmetric rule that attains minimax regret will be an equilibrium strategy for the individual in the above game. Part (iii) shows that there is always a symmetric linear rule that attains minimax regret if an equilibrium of the fictitious zero-sum game exists. This justifies that our initial search for a candidate rule to symmetric linear rules.

Lemma 2 *Let ψ^* be a Bernoulli decision-problem and let g^* be a symmetric linear rule. Assume that*

$$r(\psi, g^*) \leq r(\psi^*, g^*) \leq r(\psi^*, g) \tag{5}$$

holds for any Bernoulli decision-problem ψ and for any symmetric linear rule g. Then

(i) $g^* \in \arg\min_g \sup_\psi r(\psi, g)$,

(ii) if $\tilde{g}$ is symmetric with $\tilde{g} \in \arg\min_g \sup_\psi r(\psi, g)$ then $r(\psi, \tilde{g}) \leq r(\psi^*, g^*) = r(\psi^*, \tilde{g}) \leq r(\psi^*, g)$ for all ψ and all symmetric g, and

(iii) if $\tilde{g} \in \arg\min_g \sup_\psi r(\psi, g)$ then $\left\{\tilde{g}_L, \frac{1}{2}\tilde{g} + \frac{1}{2}\tilde{g}_S\right\} \subseteq \arg\min_g \sup_\psi r(\psi, g)$.

The possibility to derive minimax interim-regret behavior by finding an equilibrium of a zero-sum game was first pointed out by Savage (1951). We presented a simpler condition by restricting attention in (5) to symmetric linear rules and to Bernoulli decision-problems. In the original formulation, the individual would choose among all rules for round n and nature would choose among all priors over

the set of decision-problems. The connection being that if (ψ^*, g^*) solves (5) then an equilibrium of the game with enlarged strategy spaces is given if the individual chooses g^* and nature chooses ψ^* and ψ_S^* each with probability $\frac{1}{2}$ where ψ_S^* is defined by taking ψ^* and interchanging the labels of the two actions. In other words, g^* is a best response among all rules for round n to an appropriate prior over the set of decision-problems.

(5) is a very useful sufficient condition for finding minimax interim-regret behavior. The trick of being able to limit attention to linear rules was first used by Schlag (2003). This simplifies analysis greatly as linear rules are only characterized by very few parameters. Notice however that such an equilibrium characterization need not exist. In other settings we have seen that minimax interim-regret can exist even if the associated zero-sum game does not have an equilibrium. In this paper we will not deal with existence issues as we are able to provide explicit solutions for all cases we are interested in.

Proof. Let $\ominus_0$ be the set of Bernoulli decision-problems, let $\mathcal{L}$ be the set of linear rules and let $\mathcal{S}$ be the set of symmetric rules. As g^* is quasi-linear we obtain $\max_{\psi\in\ominus_0} r(\psi, g^*) = \max_{\psi} r(\psi, g^*)$. Since ψ^* is a Bernoulli decision-problem we obtain $\min_{g\in\mathcal{S}} r(\psi^*, g) = \min_{g\in\mathcal{L}\cap\mathcal{S}} r(\psi^*, g)$. Let $\rho(A) = B$ and $\rho(B) = A$. Given ψ let ψ_S be the decision-problem that results when interchanging the labels of the actions, i.e. $P_C(\psi_S) = P_{\rho(C)}(\psi)$ for $C \in \{A, B\}$. Given g let g_S be such that $g_S(C, x, C', y)_D = g(\rho(C), x, \rho(C'), y)_{\rho(D)}$. Then $r(\psi, g_S) = r(\psi_S, g)$ and $\frac{1}{2}g + \frac{1}{2}g_S$ is a symmetric rule. Thus

$$\begin{aligned}
2r(\psi^*, g^*) &= 2\max_{\psi\in\ominus_0} r(\psi, g^*) = 2\max_{\psi} r(\psi, g^*) \geq 2\inf_g \sup_\psi r(\psi, g) \\
&\geq \inf_g \sup_\psi (r(\psi, g) + r(\psi_S, g)) \geq \sup_\psi \inf_g (r(\psi, g) + r(\psi_S, g)) \\
&\geq \inf_g (r(\psi^*, g) + r(\psi_S^*, g)) = \inf_g (r(\psi^*, g) + r(\psi^*, g_S)) \\
&= 2\inf_g r\left(\psi^*, \frac{1}{2}g + \frac{1}{2}g_S\right) = 2\min_{g\in\mathcal{S}} r(\psi^*, g) \\
&= 2\min_{g\in\mathcal{L}\cap\mathcal{S}} r(\psi^*, g) = 2r(\psi^*, g^*)
\end{aligned}$$

using the general inequality that $\inf\sup \geq \sup\inf$. This proves part (i).

For part (ii) assume that $\tilde{g} \in \mathcal{S}\cap\arg\min_g \sup_\psi r(\psi, g)$. Then the claim follows from

$$r(\psi^*, g^*) \leq r(\psi^*, \tilde{g}_L) = r(\psi^*, \tilde{g}) \leq \sup_\psi r(\psi, \tilde{g}) = \sup_\psi r(\psi, g^*) = r(\psi^*, g^*) \ .$$

Part (iii) follows from part (i) and (ii) and from the fact that our calculations above show that

$$\inf_g \sup_\psi r\left(\psi, \frac{1}{2}g + \frac{1}{2}g_S\right) = \frac{1}{2}\inf_g \sup_\psi (r(\psi, g) + r(\psi, g_S)) = \inf_g \sup_\psi r(\psi, g) \ .$$

■

4.2 Equilibria

In the following we characterize symmetric quasi-linear rules that attain equilibrium minimax interim-regret. In particular we find that one such rule employs the simple reinforcement rule in round two and the proportional reviewing rule in all later rounds. In the following we refer to this rule as our *favorite rule* as intuitively it can be argued to be the simplest rule.

Proposition 3 *(a) A symmetric quasi-linear rule f attains equilibrium minimax interim-regret if and only if (i) $f_A^{(1)} = \frac{1}{2}$, (ii) $f^{(2)}(C,0,C,0)_C = 0$, $\sum_{C\in\{A,B\}} \left(f^{(2)}(C,0,C,1)_C + f^{(2)}(C,1,C,0)_C\right) = 2$ and $f^{(2)}(C,1,C,1)_C = 1$, (iii) $f^{(m)}(C,x,C,y)_C = 1$ for all $x,y \in \{0,1\}$ and $m \geq 3$, and (iv) $f^{(n)}(C,0,C',1)_{C'} = 1$ and $f^{(n)}(C,1,C',0)_{C'} = 0$ for all $C \neq C'$ and $n \geq 2$.*

(b) If all individuals use the same symmetric quasi-linear rule that attains equilibrium minimax interim-regret then

$$q_A^{(1)} = \frac{1}{2},\ q_A^{(2)} = \frac{1}{2}(1+\pi_A-\pi_B) \text{ and } q_A^{(n)} = q_A^{(n-1)} + q_A^{(n-1)} q_B^{(n-1)} (\pi_A - \pi_B) \text{ for } n \geq 3. \tag{6}$$

(c) Any rule f that attains equilibrium minimax interim-regret must satisfy (i) - (iv) in statement (a) above.

We limit our characterization in (a) to rules that are both symmetric and quasi-linear for the following reasons. These two properties are natural attributes for describing that a rule is simple. We want to understand which properties of our "favorite" rule involving simple reinforcement in round two and proportional reviewing later are important for obtaining equilibrium minimax interim-regret. We want to contrast our scenario to selection by Schlag (1998). Last but not least, a more complete characterization would be disproportionately involved and thus beyond the scope of the present paper.

Some observations can be made on the necessity of imitative behavior. Consider a rule that attains equilibrium minimax interim-regret. Then this rule is not imitative in round 2 (parts (c) and (ii) above). If the rule is also quasi-linear rule then it is imitative in rounds three and higher (parts (c) and (iii) above). However, it is conceivable that this rule is neither imitative nor quasi-linear in round three or even later. The only constraint is that it behaves like an imitative rule starting round three whenever facing a Bernoulli decision-problem.

The set of imitative quasi-linear rules for round n (with $n \geq 3$) satisfying (iv) is precisely the set of rules selected by Schlag (1998) as being "dominant" among the "improving" rules. This set includes the proportional reviewing rule, the

proportional observation rule and the proportional imitation rule. In particular we find that the possibility to observe the payoff of the sampled individual is not necessary to attain equilibrium minimax regret as the proportional reviewing rule does not rely on this information.

Notice that while play in the population that evolves according to (6) is not autonomous, it otherwise resembles a discrete version of the replicator dynamics (Taylor, 1979). In particular, the proportion of individuals playing a best action increases over time and approaches unity in the long run.

Proof. Consider parts (a) and (b). Given Lemma 2 it is sufficient to provide for each round n a Bernoulli decision-problem $\psi^{(n)} = \left(P_A^{(n)}, P_B^{(n)}\right)$ such $f^{(n)}$ and $\psi^{(n)}$ satisfy (5) for any Bernoulli decision-problem ψ and any symmetric linear rule g for round n. Notice that we can restrict attention to decision-problems in which $\pi_A \geq \pi_B$ and that we only need to specify $\pi_C^{(n)} := \pi_C\left(\psi^{(n)}\right)$ for $C \in \{A, B\}$.

Notice that $f^{(1)}$ is the unique symmetric rule for round 1. If we set $\pi_A^{(1)} = 1$ and $\pi_B^{(1)} = 0$ then it is immediate that (5) holds for round 1.

For rounds 2 and higher we will use the following prerequisites. Define the switching probabilities $\lambda_{xy}^{(n)} = 1 - f^{(n)}(C, x, C, y)_C$ and $\mu_{xy}^{(n)} = f^{(n)}(C, x, C', y)_{C'}$ for $C \neq C'$ and $x, y \in \{0, 1\}$. Then

$$\begin{aligned} q_B^{(n+1)} &= q_B^{(n)} + \left(\left(q_A^{(n)}\right)^2 \pi_A^2 - \left(q_B^{(n)}\right)^2 \pi_B^2\right) \lambda_{11}^{(n+1)} \\ &\quad + \left(\left(q_A^{(n)}\right)^2 \pi_A (1 - \pi_A) - \left(q_B^{(n)}\right)^2 \pi_B (1 - \pi_B)\right) \left(\lambda_{01}^{(n+1)} + \lambda_{10}^{(n+1)}\right) \\ &\quad + \left(\left(q_A^{(n)}\right)^2 (1 - \pi_A)^2 - \left(q_B^{(n)}\right)^2 (1 - \pi_B)^2\right) \lambda_{00}^{(n+1)} \\ &\quad + q_A^{(n)} q_B^{(n)} (\pi_A - \pi_B) \left(\mu_{10}^{(n+1)} - \mu_{01}^{(n+1)}\right) \end{aligned}$$

and $r^{(n+1)} = \pi_A - \left(q_A^{(n+1)} \pi_A + q_B^{(n+1)} \pi_B\right) = (\pi_A - \pi_B) q_B^{(n+1)}$ as $\pi_A \geq \pi_B$ holds by assumption. Notice that $r^{(n+1)}$ is linear in $\lambda_{xy}^{(n+1)}$ and in $\mu_{xy}^{(n+1)}$.

Now consider choice in round 2 given $q_A^{(1)} = \frac{1}{2}$. Set $\pi_A^{(2)} = \frac{3}{4}$ and $\pi_B^{(2)} = \frac{1}{4}$. Then $r^{(2)}\left(\psi^{(2)}, (f)_\alpha\right) = \frac{1}{4} + \frac{1}{16}\left(\lambda_{11}^{(2)} - \lambda_{00}^{(2)} + \mu_{10}^{(2)} - \mu_{01}^{(2)}\right)$ which is minimized by setting $\lambda_{00}^{(2)} = \mu_{01}^{(2)} = 1$ and $\lambda_{11}^{(2)} = \mu_{10}^{(2)} = 0$.

Assume $\lambda_{00}^{(2)} = \mu_{01}^{(2)} = 1$ and $\lambda_{11}^{(2)} = \mu_{10}^{(2)} = 0$. Then

$$q_B^{(2)} = \frac{1}{2}(1 - \pi_A + \pi_B) + \frac{1}{4}(\pi_A (1 - \pi_A) - \pi_B (1 - \pi_B)) \left(\lambda_{01}^{(2)} + \lambda_{10}^{(2)} - 1\right)$$

and it is easily verified that $r^{(2)}$ is maximized over ψ by setting $\pi_A = \pi_A^{(2)} = \frac{3}{4}$ and $\pi_B = \pi_B^{(2)} = \frac{1}{4}$ if and only if $\lambda_{01}^{(2)} + \lambda_{10}^{(2)} = 1$. In particular, $\lambda_{01}^{(2)} + \lambda_{10}^{(2)} = 1$

implies $q_A^{(2)} = \frac{1}{2}(1 + \pi_A - \pi_B)$.

Next consider some round $n \geq 3$ and assume that statement (b) holds up to round $n-1$ so $q_A^{(2)} = \frac{1}{2}(1 + \pi_A - \pi_B)$ and $q_A^{(m+1)} = q_A^{(m)} + q_A^{(m)} q_B^{(m)} (\pi_A - \pi_B)$ for $m \in \{2, .., n-2\}$. Assume first $\mu_{01}^{(n)} = 1$ and $\lambda_{xy}^{(n)} = \mu_{10}^{(n)} = 0$. Then $q_A^{(n)} = q_A^{(n-1)} + q_A^{(n-1)} q_B^{(n-1)} (\pi_A - \pi_B)$. Given the recursive formula for $q_B^{(n)}$ we see that $r^{(n)} = (\pi_A - \pi_B) q_B^{(n)}$ is continuous, bounded and only depends on π_A and π_B only through $(\pi_A - \pi_B)$. Moreover, $r^{(n)} \geq 0$ and $r^{(n)} = 0$ if $(\pi_A, \pi_B) \in \{(0,0), (1,0)\}$. Set $\pi_B^{(n)} = 0$ and choose $\pi_A^{(n)} \in \arg\max_{\pi_A} \left\{ \pi_A q_B^{(n)} \text{ s.t. } \pi_A \in (0,1) \right\}$ and note that our previous arguments show that $\pi_A^{(n)}$ is well defined.

Consider what happens when facing $\psi^{(n)}$. Note that $\pi_A^{(n)} < 1$ implies $q_A^{(2)} \in (0,1)$ which is easily shown to imply $q_A^{(n)} \in (0,1)$. Together with the fact that $\frac{d}{d\lambda_{11}^{(n)}} r^{(n)} = \frac{d}{d\lambda_{11}^{(n)}} \left(q_B^{(n)} - q_B^{(n-1)} \right)$ we obtain that $\frac{d}{d\lambda_{11}^{(n)}} r^{(n)} > 0$. Similarly

$$\frac{d}{d\lambda_{01}^{(n)}} r^{(n)} = \frac{d}{d\lambda_{10}^{(n)}} r^{(n)} > 0 \text{ and } \frac{d}{d\left(\mu_{10}^{(n)} - \mu_{01}^{(n)}\right)} r^{(n)} > 0.$$

All we are left to show is that

$$\frac{d}{d\lambda_{00}^{(n)}} r^{(n)} = \pi_A^{(n)} \left(q_A^{(n)} \left(1 - \pi_A^{(n)} \right) - q_B^{(n)} \right) > 0.$$

This will be done by proving that $q_A^{(m)} (1 - \pi_A) > q_B^{(m)}$ holds for any $\pi_A \in (0,1)$ and any $m = 2, .., n$.

Note that $q_A^{(2)} (1 - \pi_A) > q_B^{(2)}$ holds as

$$\frac{1}{2}(1 + \pi_A)(1 - \pi_A) - \frac{1}{2}(1 - \pi_A) = \frac{1}{2}\pi_A (1 - \pi_A) > 0.$$

Assume that $q_A^{(m)} (1 - \pi_A) > q_B^{(m)}$ is true. Then $q_A^{(m+1)} (1 - \pi_A) > q_B^{(m+1)}$ follows from

$$\begin{aligned} q_A^{(m+1)} (1 - \pi_A) - q_B^{(m+1)} &= q_A^{(m)} \left(1 + \left(1 - q_A^{(m)} \right) \pi_A \right) (1 - \pi_A) - \left(1 - q_A^{(m)} \right) \left(1 - q_A^{(m)} \pi_A \right) \\ &= -1 + q_A^{(m)} (2 - \pi_A) \left((1 + \pi_A) - q_A^{(m)} \pi_A \right) \\ &\geq -1 + (1 + \pi_A) - q_A^{(m)} \pi_A = \pi_A \left(1 - q_A^{(m)} \right) \end{aligned}$$

where the last inequality follows as $q_A^{(m)} (1 - \pi_A) \geq q_B^{(m)}$ implies $(2 - \pi_A) q_A^{(m)} - 1 \geq 0$.

Finally, part (c) regarding (ii)-(iv) follows using Lemma 2 (iii). The statement about (i) in part (c) follows from the fact that $\max_\psi r^{(1)} \left(\psi, q^{(1)} \right) = \max_\psi r^{(1)} \left(\psi, \left(\frac{1}{2}, \frac{1}{2} \right) \right)$ implies $q_A^{(1)} = \frac{1}{2}$. ■

Remark 4 *(Memory) It is natural to ask whether we need to restrict attention to rules with a single round of memory to attain our result. Let us check whether our selected rules satisfy the saddle point characterization in Lemma 2 if we do not restrict memory. For this we only have to consider best responses among the symmetric rules to Bernoulli decision-problems in which* $\pi_B = 0$ *and* $\pi_A \in (0,1)$*. Bayesian updating after history* $((C,0),(D,0,D,0),(C,0,C,0))$ *with* $C \neq D$ *implies that* D *should be chosen in round four. However, any rule that attains equilibrium minimax interim-regret among the rules with a single round of memory specifies to choose* C *in round four (see Proposition 3(c)). Hence, we need this restriction on memory in order to derive minimax interim-regret behavior using condition (5). Extensive arguments along the lines of those in Schlag (2003) can be used to show that (5) also has a solution if rules are not restricted. Only this additional information allows us to conclude that more memory is needed in order to attain equilibrium minimax interim-regret within the unrestricted set of behavioral rules.*

Remark 5 *(Time Preference) Individuals are modelled as being completely myopic as they are only interested in the payoffs of the next round. This approach was taken among other reasons as it is the obvious counterpart to the model in (Schlag, 1998) and as it reflects the standard modelling approach of evolutionary game theory (e.g. see Weibull, 1995). Notice however that our analysis also yields insights in the longevity scenario for the completely opposite setting with infinitely patient individuals that are only concerned with average long run payoffs. If all individuals use the same symmetric quasi-linear rule satisfying (i) - (iv) in the proposition above then the expected proportion of rounds in which any given individual chooses a best action approaches unity in the long run. Thus, the interim-regret of each individual equals zero and hence these rules also attain equilibrium minimax interim-regret when all agents are infinitely patient.*

The intermediate case where future payoffs are discounted is unfortunately both conceptually and computationally too involved for this paper (see Schlag (2003) for the individual learning setting in which individuals do not have the opportunity to observe behavior of others).

4.3 Two Individuals

In the following we assume that there are only two individuals in the population. As above, regret in round n of individual α is determined by $q^{(n)}(\alpha)$. While $q_D^{(1)}(1) = f_D^{(1)}(1)$ as above, the formulae for $q^{(n)}(\alpha)$ have to be adjusted as $z_C^{(n)}$ is no longer necessarily independent of $q^{(n)}(\alpha)$ for $\alpha \in \{1,2\}$ when $n \geq 2$. Let $q_{DD'}^{(n)}$ be the probability that individual one chooses D and individual two chooses D'

in round n. Then $q_D^{(n+1)}(1) = \sum_{D'} q_{DD'}^{(n+1)}$ and $q_{DD'}^{(n)}$ is determined recursively by

$$\begin{aligned} q_{DD'}^{(1)} &= f_D^{(1)}(1) \cdot f_{D'}^{(1)}(2) \\ q_{DD'}^{(n+1)} &= \sum_{C,C'} q_{CC'}^{(n)} \cdot \int_{x,y\in[0,1]} f_1^{(n+1)}(C,x,C',y)_D \cdot f_2^{(n+1)}(C',y,C,x)_{D'} \, dP_C(x) \, dP_{C'}(y) \\ \text{for } n &\geq 1. \end{aligned}$$

From the above we see that the formula for $q^{(n+1)}(\alpha)$ remains the same as in the infinite population setting if own play is independent of previous play in round n, i.e. if $q_{CC'}^{(n)} = q_C^{(n)}(1) \cdot q_{C'}^{(n)}(2)$ for all C, C'. As this independence holds in round 1 it follows that the characterization of play in round 2 does not depend on whether there are two or infinitely many individuals in the population. Our next result shows that our favorite rule from the infinite population setting also attains equilibrium minimax interim-regret when there are only two individuals.

Proposition 6 *Consider a population consisting of two individuals. Let f be a symmetric quasi-linear rule that satisfies (i)-(iv) in Proposition 3 (a) above and (v) $f^{(n)}(C,x,C',x)_{C'} \in \{0,1\}$ for all $x \in \{0,1\}$, $C \neq C'$ and $n \geq 3$. Then f attains equilibrium minimax interim-regret. If all individuals use the same such rule then $q_A^{(1)} = \frac{1}{2}$ and*

$$q_A^{(n)} = \frac{1}{2}(1+\pi_A-\pi_B)\left(1+\frac{1}{2}(1-\pi_A+\pi_B)\frac{1-(\pi_A\pi_B+(1-\pi_A)(1-\pi_B))^{n-2}}{\pi_A+\pi_B-2\pi_A\pi_B}(\pi_A-\pi_B)\right)$$

for $n \geq 2$. In particular,

$$\lim_{n\to\infty} q_A^{(n)} = \frac{1}{2} + \frac{1}{4}(\pi_A-\pi_B)\frac{1+(\pi_A+\pi_B)(2-\pi_A-\pi_B)}{\pi_A+\pi_B-2\pi_A\pi_B}$$

with $\lim_{n\to\infty} q_A^{(n)} \in (0,1)$ if $(\pi_A,\pi_B) \notin \{(0,1),(1,0)\}$ and $\lim_{n\to\infty} \sup_\psi r^{(n)} \approx 0.0674$.

Proof. Consider only symmetric linear rules. Clearly choice in round one is as in the infinite population setting. As $q_{CC'}^{(1)} = q_C^{(1)} q_{C'}^{(1)}$ holds for $C, C' \in \{A,B\}$, choice in round two also does not depend on whether there are two or infinitely many individuals. Behavior induces $q_{CC'}^{(2)} = q_C^{(2)} q_{C'}^{(2)}$ for all $C, C' \in \{A,B\}$ so choice in round three also remains unchanged. However in round 3 we find $q_{AA}^{(3)} > \left(q_A^{(3)}\right)^2$ if $0 < \pi_B < \pi_A < 1$ as

$$q_{AA}^{(3)} = q_{AA}^{(2)} + 2q_{AB}^{(2)}(\pi_A-\pi_B) = \left(q_A^{(2)}\right)^2 + 2q_A^{(2)}q_B^{(2)}(\pi_A-\pi_B)$$

and

$$q_A^{(3)} = q_A^{(2)} + q_{AB}^{(2)}(\pi_A - \pi_B) = q_A^{(2)} + q_A^{(2)} q_B^{(2)}(\pi_A - \pi_B)$$

which implies

$$q_{AA}^{(3)} - \left(q_A^{(3)}\right)^2 = q_A^{(2)}\left(q_B^{(2)}\right)^2 \left(2 - q_A^{(2)}(\pi_A - \pi_B)\right)(\pi_A - \pi_B).$$

So starting choice in round 4 we have to provide new calculations.

Consider round $n+1$ with $n \geq 3$ and assume that all individuals use up to round n the same (symmetric linear) rule that satisfies (i)-(iv) in Proposition 3 (a) and property (v) given above.

First we investigate the best response of nature when $\lambda_{ij}^{(n+1)} = \mu_{10}^{(n+1)} = 0$, $\mu_{01}^{(n+1)} = 1$ and $\mu_{00}^{(n+1)}, \mu_{11}^{(n+1)} \in \{0,1\}$. Then $q_{AB}^{(n)} = q_{BA}^{(n)}$ and $q_B^{(n+1)} = q_B^{(n)} + q_{AB}^{(n)}(\pi_B - \pi_A)$. So if $\pi_A \geq \pi_B$ then

$$r^{(n+1)} = q_B^{(n+1)}(\pi_A - \pi_B) = \left(q_B^{(n)} + q_{AB}^{(n)}(\pi_B - \pi_A)\right)(\pi_A - \pi_B) = r^{(n)} - q_{AB}^{(n)}(\pi_A - \pi_B)^2$$

Using the fact that $q_{AB}^{(m)} = q_{AB}^{(m-1)}(\pi_A\pi_B + (1-\pi_A)(1-\pi_B))$ for $n \geq m \geq 2$, $r^{(2)} = \frac{1}{2}(1 - \pi_A + \pi_B)(\pi_A - \pi_B)$ and $q_{AB}^{(2)} = \frac{1}{4}(1 + \pi_A - \pi_B)(1 - \pi_A + \pi_B)$ we obtain

$$\begin{aligned} r^{(n+1)} &= \frac{1}{2}(1 - \pi_A + \pi_B)(\pi_A - \pi_B) \\ &\quad \cdot \left(1 - \frac{1}{2}(1 + \pi_A - \pi_B)\frac{1 - (\pi_A\pi_B + (1-\pi_A)(1-\pi_B))^{n-1}}{\pi_A + \pi_B - 2\pi_A\pi_B}(\pi_A - \pi_B)\right) \end{aligned}$$

for $n \geq 2$.

Set $x = \frac{1}{2}(\pi_1 + \pi_2)$ and $y = \frac{1}{2}(\pi_1 - \pi_2)$ with $y \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ and $x \in [|y|, \min\{1-y, 1+y\}]$ so $\pi_1 = x + y$ and $\pi_2 = x - y$. For $x \notin \{0,1\}$ and $y \neq 0$ we find

$$\begin{aligned} \frac{d}{dx} r^{(n+1)} &= (1-2y)\, y^2 (2y+1)(1-2x) \\ &\quad \cdot \frac{1 - (2x^2 - 2y^2 + 1 - 2x)^{n-2}(2(n-2)(x - x^2 + y^2) + 1)}{(x - x^2 + y^2)^2} \end{aligned}$$

where it can then be shown that $\frac{1}{1-2x}\frac{d}{dx} r^{(n+1)} \geq 0$ and that $\frac{d}{dx} r^{(n+1)} = 0$ if $x = \frac{1}{2}$ or $y = \frac{1}{2}$ (which implies $x = \frac{1}{2}$). Looking at the curvature and the ranges for x and y we see that $x = \frac{1}{2}$ (and hence $\pi_A = 1 - \pi_B$) holds in any maximum of $r^{(n+1)}$. This is all we wanted to derive for nature's best response behavior.

Now consider the best response for the individual when $\pi_A = 1 - \pi_B \geq \frac{1}{2}$. We find

$$\frac{d}{d\lambda_{00}^{(n+1)}} r^{(n+1)} = \left(q_{AA}^{(n)} (1 - \pi_A)^2 - q_{BB}^{(n)} (1 - \pi_B)^2 \right) (\pi_A - \pi_B) \ .$$

Using the fact that $q_{AA}^{(n)} = q_{AA}^{(n-1)} + \left(1 - q_{AA}^{(n-1)} - q_{BB}^{(n-1)} \right) \pi_A (1 - \pi_B)$ and a similar expression for $q_{BB}^{(n)}$ it is easily verified that

$$q_{AA}^{(n)} (1 - \pi_A)^2 - q_{BB}^{(n)} (1 - \pi_B)^2 = q_{AA}^{(n-1)} (1 - \pi_A)^2 - q_{BB}^{(n-1)} (1 - \pi_B)^2 \ .$$

Since

$$q_{AA}^{(2)} (1 - \pi_A)^2 = \left(\frac{1}{2} (1 + \pi_A - \pi_B) \right)^2 (1 - \pi_A)^2 = q_{BB}^{(2)} (1 - \pi_B)^2$$

we obtain that $\frac{d}{d\lambda_{00}^{(n+1)}} r^{(n+1)} = 0$.

$$\begin{aligned} \frac{d}{d\lambda_{01}^{(n+1)}} r^{(n+1)} &= \frac{d}{d\lambda_{10}^{(n+1)}} r^{(n+1)} = \left(q_{AA}^{(n)} \pi_A (1 - \pi_A) - q_{BB}^{(n)} \pi_B (1 - \pi_B) \right) (\pi_A - \pi_B) \\ &= \pi_A \pi_B \left(q_{AA}^{(n)} - q_{BB}^{(n)} \right) (\pi_A - \pi_B) \geq 0 \end{aligned}$$

as $\pi_A \geq \pi_B$ and $q_{AA}^{(n)} \geq q_{BB}^{(n)}$ where later follows from

$$q_{AA}^{(n)} - q_{AA}^{(n-1)} = \left(1 - q_{AA}^{(n-1)} - q_{BB}^{(n-1)} \right) \pi_A^2 \geq \left(1 - q_{AA}^{(n-1)} - q_{BB}^{(n-1)} \right) \pi_B^2 = q_{BB}^{(n)} - q_{BB}^{(n-1)}$$

and $q_{AA}^{(2)} \geq q_{BB}^{(2)}$. Similarly we can verify

$$\frac{d}{d\lambda_{11}^{(n+1)}} r^{(n+1)} = \left(q_{AA}^{(n)} \pi_A^2 - q_{BB}^{(n)} \pi_B^2 \right) (\pi_A - \pi_B) \geq 0.$$

Note that $r^{(n+1)}$ is independent of $\mu_{00}^{(n+1)}$ and $\mu_{11}^{(n+1)}$ and that

$$\frac{d}{d\mu_{01}^{(n+1)}} r^{(n+1)} = -\frac{d}{d\mu_{10}^{(n+1)}} r^{(n+1)} = q_{AB}^{(n)} \left(\pi_B (1 - \pi_A) - \pi_A (1 - \pi_B) \right) (\pi_A - \pi_B) \leq 0.$$

This shows that (i)-(v) are sufficient to establish the conditions for equilibrium minimax interim-regret in round $n + 1$.

The maximal value of interim-regret in the long run can be computed analytically but as the formula is messy we simply provide the approximate numerical value 6.7442×10^{-2}. ■

Notice that the three imitation rules proportional reviewing, proportional imitation and proportional observation satisfy condition (v). In the following we

briefly discuss why condition (v) has been added. Consider a symmetric quasi-linear rule that satisfies conditions (i) to (iv). In the infinite population setting we found that the choice of $\mu_{00}^{(n)}$ and $\mu_{11}^{(n)}$ do not alter the performance in round n. This is also true with only two individuals as

$$q_B^{(n)} = q_B^{(n-1)} - q_{AB}^{(n-1)} (\pi_A - \pi_B)$$

holds for all $\mu_{00}^{(n)}, \mu_{11}^{(n)} \in [0,1]$ where $\mu_{xx}^{(n)} = f^{(n)} (C, x, C', x)_{C'}$ for $C \neq C'$. However, unlike in the infinite population setting, when there are only two individuals then choice of $\mu_{00}^{(n)}$ and $\mu_{11}^{(n)}$ influences behavior starting round $n+1$. This follows from the fact that

$$q_{AB}^{(n)} = q_{AB}^{(n-1)} \left(\pi_A \pi_B \left(1 - 2\mu_{11}^{(n)} \left(1 - \mu_{11}^{(n)} \right) \right) + (1 - \pi_A)(1 - \pi_B) \left(1 - 2\mu_{00}^{(n)} \left(1 - \mu_{00}^{(n)} \right) \right) \right) .$$

Looking at the two equations above we see that $r^{(n+1)}$ is decreasing in $q_{AB}^{(n)}$ and that $q_{AB}^{(n)}$ is maximized if $\mu_{00}^{(n)}, \mu_{11}^{(n)} \in \{0,1\}$. So if individuals anticipate effects of behavior in round n on interim-regret in round $n+1$ when being indifferent in terms of interim-regret in round n then one can argue that each individual prefers rules with $\mu_{00}^{(n)}, \mu_{11}^{(n)} \in \{0,1\}$ for all $n \geq 2$.

4.4 Alternative Scenarios

In the following we investigate the value of the observed information in our above results. In the first scenario below the individual only learns from own history. In the second she observes the success of a random draw. An alternative interpretation of this second scenario is that the individual observes the success of an inexperienced individual. This is because an individual that minimizes maximum regret will choose each action equally likely in the first round.

4.4.1 Individual Learning

Consider an individual who only learns from own experience. We retain the assumption on limited memory and assume that the rule used by this individual only depends on her success in the previous round. Formally a rule $f = \left(f^{(n)}\right)_n$ is given by $f^{(1)} \in \Delta\{A, B\}$ and $f^{(n)} : \{A, B\} \times [0,1] \to \Delta\{A, B\}$ for $n \geq 2$. This is a pure decision-making context as individuals no longer influence each other. Accordingly we drop the prefix "equilibrium" and refer only to minimax interim-regret.

Proposition 7 *Assume that an individual does not observe behavior of others.*

(a) Then a symmetric linear rule f attains minimax interim-regret if and only if $f_A^{(1)} = \frac{1}{2}$, $f^{(2)} (C, x)_C = x$ and $f^{(n)} (C, x)_C = 1$ for $n \geq 3$.

(b) If f attains minimax interim-regret and $n \geq 3$ then $f^{(n)}(C,x)_C = 1$ for $x \in \{0,1\}$ and $\sup_\psi r^{(n)} = \frac{1}{8}$.

The proof is left to the reader. In order to verify minimax interim-regret of the presented symmetric linear rules, derive for each round the (unique) interim-regret maximizing Bernoulli decision-problem and then verify (5). For part (b) use the following result. If a rule f attains minimax interim-regret then the symmetric linear rule $\left(\frac{1}{2}f + \frac{1}{2}f_S\right)_L$ will also have this property. Its proof is immediate given the one of Lemma 2(iii).[7]

Minimax interim-regret behavior in rounds one and two comes at no surprise as it follows directly from Proposition 3 (a). The interesting implication of (b) is that the individual never switches actions after round two when using a linear rule or when facing a Bernoulli decision-problem. Comparing this to our previous findings we conclude that information about success of others has substantial value when individuals aim to minimax interim-regret.

4.4.2 Sampling Inexperienced

Recall the main setting of this paper. The individual has access to two sources of information when observing behavior of others. The first source of information is contained directly in the payoff observed. The individual learns about the payoff distribution of the action chosen by the individual observed. This source of information is not necessary for attaining equilibrium minimax regret (see properties of the favorite rule involving simple reinforcement and proportion reviewing). The second source of information is contained in the observed choice as this indirectly reveals information about previous success of the two actions. Our intuition is that this second source is essential to support imitative behavior of quasi-linear rules starting round three as found in Propositions 3 and 6. The fact that another learning individual is using the same action is treated as sufficient evidence that the action chosen is best and offsets the negative signal received when observing that the apparently best action yielded twice the lowest payoff.

In the following we investigate the validity of this intuition by maintaining the first source of information while eliminating the second. To do this we assume that the individual observes after each own choice the outcome of a random choice, i.e. an independently realized payoff of an action where each action is observed equally likely. Recall that an individual who aims to minimax interim-regret will choose each action equally likely in the first round. Thus, it is as if the individual observes the success of an inexperienced individual. Formally we calculate $q^{(1)}$ as above and $q^{(n)}$ from (3) by setting $z_C^{(n-1)} = \frac{1}{2}$ when $n \geq 2$.

[7] Note that it does not follow from Lemma 2 (iii) as this lemma presumes the existence of a saddle point.

Proposition 8 *Assume that an individual only observes the outcome of a random choice after each round.*

(a) A symmetric linear rule f attains minimax interim-regret in the first three rounds if and only if $f^{(1)} = \frac{1}{2}$, $f^{(2)}(C,x)_C = x$, $f^{(3)}(C,0,C,0)_C = 0$, $f^{(3)}(C,x,C,y)_C = 1$ for $(x,y) \in \{(0,1),(1,0),(1,1)\}$, $f^{(3)}(C,0,C',1)_{C'} = 1$ and $f^{(3)}(C,x,C',y)_{C'} = 0$ for $(x,y) \in \{(0,0),(1,0),(1,1)\}$ and $C \neq C'$. The maximal value of interim-regret in round three under such a rule is approximately equal to 0.087.

(b) There exists $\phi^* > 0$ such that if $0 \leq \phi < \phi^*$ and $z_A^{(2)} = (1-\phi)\frac{1}{2} + \phi\frac{1}{2}(1+\pi_A-\pi_B)$ then any rule f that attains equilibrium minimax interim-regret satisfies $f^{(3)}(C,0,C,0)_C < 1$ for some $C \in \{A,B\}$ and hence is not imitative in round three.

Part (a) shows which symmetric linear rules attain minimax interim-regret. The important implication of $f^{(3)}(C,0,C,0)_c = 0$ is that no rule that attains minimax interim-regret will be imitative in round three (part (b) for $\phi = 0$). Thus we confirm our intuition that imitation requires learning from others who are also learning. The proof of part (a) follows the same steps as in that of Proposition 7 (a).

In fact, we obtain more and present this in (b). Consider a population with either two or infinitely many individuals and assume that an individual observes in round three the performance of an inexperienced individual with probability $1-\phi$ and of an equally experienced other individual (from round two) with probability ϕ. Following Propositions 3 and 6 the experienced individual chooses action A in round two with probability $(1+\pi_A-\pi_B)$. Part (b) of Proposition 8 then states that there is no rule that exhibits imitative behavior in round three if ϕ is sufficiently small. In words, imitation is only warranted in round three if the observed success contains sufficient information about the value of the action chosen. The proof of part (b) restricted to symmetric linear rules is easily established by using continuity, in particular by verifying that the values π_A and π_B associated to the Bernoulli decision-problem ψ^* used to verify (5) are continuous in ϕ. The extension to non-symmetric rules is then obtained by invoking Lemma 2 (iii).

For completeness we have to mention that the value of minimax interim-regret in round three is the same when observing an inexperienced individual (Proposition 8(a)) as it is when observing another equally experienced individual (Propositions 3 and 6). In fact, the choice probabilities are identical if the underling rule is both symmetric and linear.[8] So one might think that it is not important whether the observed individual is inexperienced or experienced. To investigate this further we consider behavior in round four when observing an inexperienced

[8]The precise value of choice-regret equals $\frac{1}{4}(\pi_A-\pi_B)(2+\pi_A-\pi_B)(1-\pi_A+\pi_B)^2$.

individual. Using basic numerical methods we find that $f^{(4)}$ is the same as $f^{(3)}$ except that $f^{(4)}(C,0,C,0)_C \approx 0.5075$ which implies $\max_\psi r^{(4)} \approx 7.9147 \times 10^{-2}$. For comparison, when observing an experienced individual interim-regret in round four is bounded above by $7.551\,3 \times 10^{-2}$ if there are only two individuals and by 6.6438×10^{-2} if there are infinitely many individuals in the population. Hence, anticipating performance in round four when choosing in round three we conclude that individuals strictly prefer to learn from similarly informed individuals (in round three).

5 Ex-post-Regret

Next we consider selection under ex-post-regret in a population consisting of either two or countably infinitely many individuals.

5.1 Preliminaries

We establish some useful bounds on ex-post-regret. First we show that ex-post-regret is maximal among the Bernoulli decision-problems when all individuals use linear rules.

Lemma 9 *For any decision-problem ψ and linear rule f_α,*

$$\begin{aligned} R^{(n)}\left(\psi, q^{(n)}\left(f_\alpha, (f_{\alpha'})_{\alpha'\neq\alpha}\right)\right) \;\leq\; & q_A^{(n)}\left(f_\alpha, (f_{\alpha'})_{\alpha'\neq\alpha}\right)\cdot \pi_B\left(1-\pi_A\right) \\ & + q_B^{(n)}\left(f_\alpha, (f_{\alpha'})_{\alpha'\neq\alpha}\right)\cdot \pi_A\left(1-\pi_B\right) \end{aligned}$$

where equality holds if ψ is a Bernoulli decision-problem.

Proof. The inequality statement follows from the fact that

$$\begin{aligned} & \pi_B\left(1-\pi_A\right) - \int_0^1\int_0^1 \max\left\{0, y-x\right\} dP_A(x)\, dP_B(y) \\ = \;& \int_0^1 y \int_0^1 (1-x)\, dP_A(x)\, dP_B(y) - \int_0^1\int_0^y (y-x)\, dP_A(x)\, dP_B(y) \\ \geq \;& \int_0^1\int_0^y \left(y(1-x) - (y-x)\right) dP_A(x)\, dP_B(y) \\ = \;& \int_0^1 (1-y)\int_0^y x\, dP_A(x)\, dP_B(y) \geq 0\ . \end{aligned}$$

As f_α is linear, $q_A^{(n)}\left(f_\alpha, (f_{\alpha'})_{\alpha'\neq\alpha}\right)$ only depends on the expected payoff of each action. The equality for Bernoulli decision-problems then follows by direct verification. ■

Let ψ' denote the Bernoulli decision-problem with $\pi_A(\psi') = \pi_B(\psi') = \frac{1}{2}$. Then $R^{(n)}\left(\psi', q^{(n)}\right) = \frac{1}{4}$ and consequently:

Lemma 10 $\sup_\psi R^{(n)}\left(\psi, q^{(n)}\left((f_\alpha)_\alpha\right)\right) \geq \frac{1}{4}$ *for any* $(f_\alpha)_\alpha$.

We use these two lemmata to characterize equilibrium minimax ex-post-regret by presenting rules below that do not yield ex-post-regret above 1/4 in any Bernoulli decision-problem after round one.

5.2 Equilibrium Minimax Ex-Post-Regret

Below we show that observed behavior is not needed in order to attain equilibrium minimax ex-post-regret. Equilibrium minimax ex-post-regret is attained by the simple symmetric linear rule selected under individual learning in Proposition 7 (a): flip a coin in round one, apply the simple reinforcement rule in round two and then choose this action in all later rounds. We also find that reacting appropriately to observed behavior is not harmful as our favorite rule selected under interim-regret also attains equilibrium minimax ex-post-regret.

Proposition 11 *Consider a population with either two or infinitely many individuals. (a) The rule f given by $f_A^{(1)} = \frac{1}{2}$, $f^{(2)}(C, x, C', y)_C = x$ and $f^{(n)}(C, x, C', y)_C = 1$ for $n \geq 3$ attains equilibrium minimax ex-post-regret. (b) Any quasi-linear rule satisfying conditions (i)-(iv) in Proposition 3 attains equilibrium minimax ex-post-regret. (c) If the combination of rules $(f_\alpha)_\alpha$ attains equilibrium minimax ex-post-regret then the maximal value of ex-post-regret is equal to $\frac{1}{2}$ in round one and equal to $\frac{1}{4}$ in all later rounds.*

Proof. Proof of (b) for an infinite population. Let f satisfy the stated assumptions. Let ψ_0 be a Bernoulli decision-problem.

Consider round 1. Then $R^{(1)}\left(\psi_0, q^{(1)}\right) = q_A^{(1)}\pi_B(1-\pi_A) + q_B^{(1)}\pi_A(1-\pi_B)$. Evaluating $(\pi_A, \pi_B) = (0,1)$ and $(\pi_A, \pi_B) = (1,0)$ we find that $R^{(1)} \geq 1/2$. Setting $q_A^{(1)} = 1/2$ it is easily verified that $\max_{\psi_0} R^{(1)}\left(\psi_0, q^{(1)}\right) = \frac{1}{2}$ which completes the proof for round 1.

Given Lemma 9 and Lemma 10 all we need to show for rounds $n \geq 2$ is that $\sup_{\psi_0 \in \ominus_0} R^{(n)}\left(\psi_0, q^{(n)}\left((f)_\alpha\right)\right) \leq \frac{1}{4}$.

Consider round 2. Since $q_A^{(2)} = \frac{1}{2}(1 + \pi_A - \pi_B)$ we obtain $R^{(2)} = \frac{1}{2}(\pi_A(1-\pi_A) + \pi_B(1-\pi_B))$ which attains its maximum $\frac{1}{4}$.

Now consider any round $n \geq 3$. Then $R^{(n)} = q_B^{(n)}(\pi_A - \pi_B) + \pi_B(1-\pi_A) = R^{(n-1)} - q_A^{(n-1)} q_B^{(n-1)}(\pi_A - \pi_B)^2$ so $R^{(n)} \leq R^{(n-1)}$ with $R^{(n)} = R^{(n-1)}$ if $\pi_A = \pi_B$. Hence, $\sup_{\pi_A, \pi_B} R^{(n)} \leq \sup_{\pi_A, \pi_B} R^{(2)} = \frac{1}{4}$.

The arguments for a population consisting of only two individuals are straightforward and are thus omitted.

Proof of parts (a) and (c) are immediate given the above and given Lemma 10. ■

Remark 12 *Analogous to Remark 5, consider briefly the case where individuals are infinitely patient and hence only care for long run average payoffs. Previous analysis in this section shows that the rule specified in Proposition 11 (a) and (b) also attains equilibrium minimax ex-post-regret when all individuals are infinitely patient. However, it is not plausible that an infinitely patient individual chooses not to change actions after round two. Ex-post-regret alone seems to be an inappropriate selection criterion in our setting.*

6 Conclusion and Outlook

We show how imitation can result from concern for regret. We use the term imitation to indicate that an individual either chooses the same action again or switches to the action observed. So imitation is in this sense not conditional on payoffs. As we consider a model with two actions only, imitation is characterized by never switching when the observed individual chooses the same action. We find that a very simple rule attains equilibrium minimax regret. Behavior when own action differs from observed action does not depend on how often the decision-problem has been faced. Experience only matters when own action coincides with observed action where behavior in round 2 is sensitive to payoffs while in later rounds it is imitative.

Notice that behavior starting round 3 is the same as selected by Schlag (1998) under a very different approach. We briefly compare the major findings in the two papers in terms of reasons to imitate, the role of linearity and in terms of long run behavior.

Schlag (1998) shows that imitative behavior is necessary if individuals prefer that average payoffs always weakly increase. The individual does not switch to an unobserved action as she fears that she could be in the situation in which she is the only one switching away from the best action. In the present paper we show that imitative behavior is sufficient, a particular imitative behavior will minimize maximum interim-regret after round two. Myopic learning after round one creates sufficient experience in the population so that it is very likely that own action is best when own and observed action coincide in round two. The proof pins down behavior when payoffs are either zero or one. Alternative rules that are not imitative after round one are easily constructed by perturbing behavior when observing intermediary payoffs.

Only linear rules can always make average payoffs weakly increase in the model of Schlag (1998). In our paper we find linearity to be one possible way of achieving the goal of minimax regret but not necessarily the only one.

In Schlag (1998) all learn to choose a best action in the long run only if initially some choose a best action. Choice in the initial round cannot be investigated. In the infinite population setting of this paper, all individuals learn to choose a best action in the long run. This long run efficiency of imitation stands in contrast to the inefficient herding outcomes discussed for instance in Banerjee (1992) and Squintani and Välimäki (2002).

The complexity and novelty of the equilibrium approach to minimax regret led us to limit attention to decision-problems with two actions only. In future research we intend to investigate decision-problems with more actions and to also include asymmetric information and correlated payoff realizations (aggregate shocks).

References

[1] Banerjee, A. (1992), "A Simple Model of Herd Behavior," *Quart. J. Econ.* **107**, 797–817.

[2] Bell, D.E (1982), "Regret in Decision Making under Uncertainty," *Oper. Res.* **30**, 961–81.

[3] Berry, D.A. and B. Fristedt (1985), *Bandit Problems: Sequential Allocation of Experiments*, Chapman-Hall, London.

[4] Björnerstedt, J. and J. Weibull (1996), Nash Equilibrium and Evolution by Imitation, in *The Rational Foundations of Economic Behaviour* (K. Arrow and E. Colombatto, eds.). Macmillan.

[5] Börgers, T., A. Morales and R. Sarin (2004) "Expedient and Monotone Learning Rules," *Econometrica* **72**, 383–405

[6] (with Antonio Morales and Rajiv Sarin).

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

[8] Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with a Non-Unique Prior," *J. Math. Econ.* **18**, 141–53.

[9] Hart, S. and A. Mas-Colell (2000), "A Simple Adaptive Procedure Leading to Corellated Equilibrium," *Econometrica* **68(5)**, 1127–50.

[10] Kreps, D. and R. Wilson (1982), "Reputation and Imperfect Information," *J. Econ. Theory* **27**, 253–279.

[11] Linhart, P. B. and R. Radner (1989), "Minimax-Regret Strategies for Bargaining over Several Variables," *J. Econ. Theory* **48**, 152–178.

[12] Milnor, J. (1954), Games against Nature, in *Decision Processes* (R.M. Thrall, C.H. Coombs, and R.L. Davis, eds.), Wiley and Chapman & Hall, New York and London.

[13] Robbins, H. (1952), "Some Aspects of the Sequential Design of Experiments," *Bull. Amer. Math. Soc.* **58(5)**, 527–35.

[14] Rogers, A. (1989), "Does Biology Constrain Culture?" *Amer. Anthrop.* **90**, 819–831.

[15] Savage, L. J. (1951), "The Theory of Statistical Decision," *J. Amer. Stat. Assoc.* **46(253)**, 55–67.

[16] Schlag, K. H. (1998), "Why Imitate, and if so, How? A Boundedly Rational Approach to Multi-Armed Bandits," *J. Econ. Theory* **78(1)**, 130–156.

[17] Schlag, K. H. (1999), "Which One Should I Imitate?" *J. Math. Econ.* **31**, 493–522.

[18] Schlag, K. H. (2003), *How to Minimize Maximum Regret under Repeated Decision-Making*, Mimeo, European University Institute, http://www.iue.it/Personal/Schlag/papers/regret7.pdf.

[19] Squintani, F. and J. Välimäki (2002), "Imitation and Experimentation in Changing Contests," *J. Econ. Theory* **104**, 376–404.

[20] Taylor, P. (1979), "Evoluationarily Stable Strategies with Two Types of Players," *J. Applied Prob.* **16**, 76–83.

[21] Veblen, T. (1899), *The Theory of the Leisure Class, An Economic Study of Institutions,* The Macmillan Company, New York.

[22] Wald, A. (1950), *Statistical Decision Functions*, Bronx, Chelsea.

[23] Weibull, J. (1995) *Evolutionary Game Theory*, MIT Press, Cambridge.

A Appendix: Defining Minimax Regret Conditional on Histories

In the following we show why it does not make sense to calculate minimax regret conditional on the history experienced. We provide the argument for interim-regret, the reasoning for ex-post-regret is analogous.

Consider an individual who chose action A and obtained payoff x in round one and after observed someone else who chose B and obtained y in round one. We denote this history by (A, x, B, y). The definition of interim-regret $r^{(2)}$ depends on how we define $q^{(2)} \in \Delta\{A, B\}$. Assume that $q^{(2)}$ is calculated conditional on (A, x, B, y) so for a given rule f we would set $q^{(2)} = f^{(2)}(A, x, B, y)$.

Proceeding then to calculate the supremum of regret under a rule f it only makes sense to consider decision-problems that can generate the history (A, x, B, y). In Propositions 3 and 6 we find that only $q_A^{(1)} = \frac{1}{2}$ attains minimax (interim- or ex-post-) regret in round one. So it follows that (A, x, B, y) can arise under any decision-problem in which the payoff distributions P_A and P_B have full support. Let $\mathcal{D}$ be the set of all decision-problems and let $\mathcal{D}'$ be the subset of decision-problems in which both payoff distributions have full support. Notice that the supremum of interim-regret over all ψ in $\mathcal{D}$ is the same it is over all ψ in $\mathcal{D}'$.

As $q^{(2)}$ is only a function of the given history (A, x, B, y) it does not change when ψ is varied within $\mathcal{D}'$. It could be that $\pi_C \approx 1$ and $\pi_D \approx 0$ for some $C \neq D$. As in round one it then follows that only $q^{(2)} = \frac{1}{2}$ will minimax interim-regret in round two given the behavior of others. Similarly, we obtain $q_A^{(n)} = \frac{1}{2}$ for all $n \geq 3$ regardless of the history experienced.

So a definition of interim-regret (and similarly of ex-post-regret) conditional on past history is not sensible. Notice that a similar issue arises in Schlag (1998) when calculating change in average payoffs.

Notice that the standard Bayesian approach to multi-armed bandits can also be seen has being based on such a commitment ex-ante. Maximizing payoffs (or utility) in each round conditional on the updated prior is equivalent to maximizing payoffs ex-ante in each round. Notice that in the latter case, priors are not updated explicitly.