Linear Learning in Changing Environments*

Tilman Klumpp†
Indiana University

July 2005

Abstract

The decision maker receives signals imperfectly correlated with an unobservable state variable and must take actions whose payoffs depend on the state. The state randomly changes over time. In this environment, we examine the performance of simple linear updating rules relative to Bayesian learning. We show that a range of parameters exists for which linear learning results in exactly the same decisions as Bayesian learning, although not in the same beliefs. Outside this parameter range, we use simulations to demonstrate that the consumption level attainable under the optimal linear rule is virtually indistinguishable from the one attainable under Bayes' rule, although the respective decisions will not always be identical. These results suggest that simple rules of thumb can have an advantage over Bayesian updating when more complex calculations are more costly to perform than less complex ones. We demonstrate the implications of such an advantage in an evolutionary model where agents "learn to learn."

Keywords: Learning, Changing Environments, Bounded Rationality, Exponential Smoothing Rule, Evolution of Learning Rules.

JEL code: D81, D83.

*I thank Bob Becker, Phil Curry, Arthur Robson, Xuejuan Su, the editor Cars Hommes, three anonymous referees, as well as seminar participants at the University of Western Ontario for their helpful comments.

†Department of Economics, Indiana University. Address: Wylie Hall 105, 100 S. Woodlawn, Bloomington, IN 47405. E-mail: tiklumpp@indiana.edu.

1 Introduction

Bayesian updating is optimal if the decision maker adheres to the expected utility axioms (Luce and Raiffa (1957)). It is well-documented, however, that information processing and decision making by real agents often relies on simple heuristics rather than Bayes' rule. Kahnemann and Tversky (1982), for example, contains a wealth of examples that show that real-life decision makers rarely adhere to Bayes' rule when making probabilistic judgments. Can we then conclude that these agents rarely make correct decisions? The present paper explores the relationship between Bayesian learning, non-Bayesian learning, and decision making in the context of changing environments.

The problem of learning in changing environments has attracted considerable attention by economists. Kiefer (1989), Balvers and Cosimano (1990), and Keller and Rady (1999), extend Rothschild's (1974) model of monopoly with unknown demand, but allow for time-varying states. Nyarko and Olson (1991), Balvers and Cosimano (1993), Rustichini and Wolinky (1995), and Beck and Wieland (2002), present similar problems in more abstract terms; for example, Rustichini and Wolinsky examine learning and experimentation in changing environments in a two-armed bandit framework.[1] The main theme from this literature is the intuitive result that with time-varying and unobservable states, one must "unlearn" past information at some rate in order to learn about the current state. Bayesian learning is then recency biased: Different pieces of information are no longer interchangeable, and new information is treated as more important than old information. With a fixed but unobservable state, on the other hand, even old information is relevant for today's decisions.[2]

Previous work on learning in changing environments has mostly, though not exclusively, focused on Bayesian updating to model learning. In this paper, we measure the performance of non-Bayesian learning rules for decision making in changing environments. Non-Bayesian rules are incorrect from a decision-theoretic perspective, but often computationally less complex than Bayes' rule. We examine a class of especially simple updating processes, known as *exponential smoothing* rules, which generate beliefs as convex combinations of previously observed signals, with declining weights for signals that have been observed further in the past. Thus, only linear operations are needed to compute beliefs, while Bayes' rule requires non-linear operations. In time series econometrics, for example, AR(1) processes are just exponentially weighted averages of past error terms. Since weights given to past observations decline over time, exponential smoothing rules are recency biased just like Bayes' rule, and it appears they could be a substitute for Bayesian learning in a changing environment where unlearning of past signals is crucial. Our aim is to quantify the extent to which linear updating leads to incorrect decisions and inferior

[1] The mentioned papers are only an incomplete list of the relevant literature. See also Sobel (2000) for a survey of learning models in economics. Moreover, the engineering, finance, and psychology literatures contain works on learning in general, and learning in changing environments in particular, which we do not review here.

[2] A related result is that if obtaining information is costly, experimentation may not cede even in the long run when the environment is subject to changes over time. (In contrast, with a fixed environment experimentation will eventually cede, namely when enough has been already learned for the benefit of additional information to fall below the cost of acquiring it (Berry and Fristedt (1985)). An application of this idea to social learning is Moscarini et al. (1998). The authors show that with changing states, only temporary herding and informational cascades can arise.

payoff streams relative to Bayesian updating.

To this end, we utilize a simple symmetric model. Consider an environment that can be in K possible states. In each period a signal is observed that can take on K possible values, where the probability that the signal is the same as the state is sufficiently large to be informative. Prior to observing each signal, one of K possible actions must be taken. Assume that the decision maker obtains a reward of one if she takes the action that matches the signal, and zero otherwise. The value of the state can switch in each period with some probability, but just like the state itself the occurance of a switch is unobservable. We assume that there are no costs or other tradeoffs associated with observing information, so issues of experimentation and optimal control do not arise.[3] The stream of observable signals will hence result in a sequence of posterior Bayesian beliefs about the likelihood of each state, and it is optimal to take whichever action maximizes expected payoffs in this period. This action will be identified with the state that receives the highest weight in current beliefs. Notice that any learning rule that generates beliefs which result in the same utility maximizing action as Bayesian beliefs will result in correct decisions. Thus, if the action space is discrete, it is possible that many different learning rules perform similarly well for decision making.

We consider three performance measures for comparing linear learning and Bayesian updating. First, we examine the likelihood that a linear rule results in an incorrect decision relative to Bayes' rule. Second, we look at two relative consumption measures. The first of these is the average long-run consumption obtained through the linear rule as a fraction of consumption attainable under Bayes' rule. The second relative consumption measure takes as an "outside value" not zero, but the average consumption attainable by an agent who does not process information in any way but simply takes a random action in every period (or, which is equivalent in our model, takes the same action in every period). Thus, it computes the difference of this value and Bayesian payoffs and then measures how much of this difference can be obtained by using a linear updating rule.

We identify a parameter region for which every linear rule that unlearns sufficiently fast yields the same decisions and hence the same payoffs as does Bayes' rule. In this case, the only decision relevant information is the last observed signal. Outside this parameter region, things are less clear and we resort to simulations to find the optimal linear rule and examine its performance. In all of the cases which we consider, the optimal linear rule leads to wrong decisions in less than three percent of all periods when there are only two possible states (the fraction can be higher when there are more than two states). This number increases quickly, however, when moving away from the optimal rule. Strikingly, though, when looking at relative consumption a different picture emerges. The average payoff obtained through the optimal linear rule is virtually indistinguishable from the Bayesian payoff. The reason lies in the fact that even if wrong decisions are made, those occur mostly at times when Bayesian beliefs are diffuse and place high weight on more than one state. Thus, incorrect decisions will typically not be very costly in terms of consumption. Moreover, we show that a large set of suboptimally parameterized linear rules (i.e. those that would yield incorrect decisions much more frequently than the optimal

[3] An example where such an assumption may be justified are (perfectly competitive) financial markets: Traders can observe each security's return whether they decide to buy it or not, and a purchase has no effect on the information obtained.

linear rule) can still perform well in terms of consumption. Akerlof and Yellen (1985) show that relatively large deviations from rationality may entail only small changes in the value of the objective function, and are therefore not very costly. This result is derived in a model with continuous choices and a differentiable objective, so that a first-order change in the choice variables away from the optimizer leads to only second-order changes in the objective (which is flat around the optimum). In our model, the flatness of the payoff function around the optimal linear rule makes it possible that even suboptimal linear rules perform well. It is, however, not the reason why linear rules perform well compared to Bayes' Rule in the first place. The reason for this effect is rather that simple updating procedures can often lead to correct choices if there are only a finite number of choices available. In particular, a wrong belief can lead to a correct decision because the belief space will be partitioned into regions where different choices are optimal. So even if the decision maker holds an incorrect belief, as long as it falls into the right region the choice will not be affected, and we demonstrate that with linear updating this is the case most of the time.

The properties of linear forecasts have been studied extensively, but in very different frameworks. Linear forecasts can be self-enforcing in models with many decision makers. Hommes and Sorger (1998) study how linear procedures can predict an underlying non-linear price process in equilibrium. They develop the concept of a *consistent expectations equilibrium* and demonstrate that the use of simple linear procedures can result in statistically correct and thus self-fulfilling expectations even when the feedback from expectations to actual price realizations is non-linear.

Another strand of literature is based on the *Kalman filter* (Kalman (1960)), a linear technique for tracking an unobservable state variable in $\mathbb{R}^n$ (or a subset of $\mathbb{R}^n$) with normally distributed system shocks and measurement errors. Since the filter maintains the first two moments of Bayesian beliefs, optimal learning is linear if the first and second moments are all one requires. In a similar Gaussian setup, Muth (1960) shows that among all linear estimators the exponential smoothing rule is the one that minimizes the variance of the forecast error. In this environment, exponential smoothing is a special case of the Kalman filter with time-invariant weights. The objective of the current paper is quite different: In our setup the state variable takes on *categorical*, instead of numerical, values. Therefore, no natural metric exists on the state space and no moments can be computed, as required to implement the Kalman filter. We hence examine the exponential smoothing rule not as a special case of the Kalman filter, but as a boundedly rational updating procedure which computes beliefs as a convex combination of two probability distributions: The current prior distribution, and a degenerate distributrion that puts all weight on the last observed signal. The best linear forecast is typically worse than the Bayesian one in terms of error probability and average payoffs—though, as we argue in the quantitative part of the paper, the difference is often not very big. In the Kalman/Muth setup, on the other hand, the linearly computed moments coincide precisely with the moments of the Bayesian belief, so that under an appropriate objective function an agent who uses a linear forecasts obtains precisely the same outcomes as a Bayesian agent.

The accuracy of linear updating rules in changing environments has interesting implications regarding the tradeoff between accuracy and ease of computation. Simon (1955, 1959) advanced the ideas of bounded rationality and satisificing behavior—that optimal

behavior often requires complex computations, and that less complex procedures that perform reasonably well might be adopted instead by humans. The idea obviously can be applied to the case of changing environments, with linear learning being a surrogate for non-linear Bayesian learning. Our results suggest that linear updating may actually win the accuracy vs. complexity tradeoff for many applications. That is, the relative disadvantage in terms of accuracy is surprisingly small (or does not even exist) so that it may well be worthwhile sacrificing accuracy and saving a computational cost in return. In Section 6 we explicitely introduce such a computation cost and show how evolutionary dynamics may select against Bayesian decision makers. This effect may also play an important role in several other applications, including consumer choice in markets with varying product characteristics, and experimentation in changing environments. (Section 7 contains a more detailed discussion of these applications.)

The remainder of the paper is organized as follows. Section 2 contains the basic model. In Section 3 we devise a set of measures to assess the relative performance of linear learning compared to Bayesian learning. In Section 4 we investigate the relationship between accuracy and learning speed of linear rules. Section 5 contains our simulation results as well as a statistical treatment of the discernibility of Bayesian and non-Bayesian payoff streams in changing environments. Section 6 describes a stylized evolutionary model and demonstrates how non-Bayesian rules can have a selection advantage. Section 7 concludes with a discussion of some potential applications of our results. All proofs are contained in Appendix A. Finally, in Appendix B we extend the model to continuous actions and risk-averse preferences and present additional simulation results.

2 Learning in Changing Environments

Time is discrete and indexed by $t = 1, 2, \ldots$. The set of states of the world as well as its cardinality is $K = \{1, \ldots, K\}$, $K \geq 2$. The realization of the state at time t is $\omega_t \in K$. At $t = 1$, an initial state ω_1 is drawn uniformly from K: $Pr[\omega_1 = k] = 1/K$ $\forall k$. Thereafter, the state evolves according to a symmetric switching process with arrival rate $\varepsilon \in (0, 1 - \frac{1}{K})$. With probability $1 - \varepsilon$, $\omega_{t+1} = \omega_t$, and with probability ε, ω_{t+1} is uniformly redrawn from $K \backslash \omega_t$. Let $\hat{\varepsilon} = \frac{\varepsilon}{K-1}$ be the probability that the state switches from k to some given $k' \neq k$. The assumption that $\varepsilon < 1 - \frac{1}{K}$ ensures $1 - \varepsilon > \hat{\varepsilon}$, and thus for any k and $k' \neq k$, $Pr[\omega_{t+1} = k | \omega_t = k] > Pr[\omega_{t+1} = k' | \omega_t = k]$.

An agent observes a signal $m_t \in K$ in each period. The distribution of m_t depends on ω_t in the following way:

$$Pr[m_t = k | \omega_t] = \begin{cases} p & \text{if } \omega_t = k, \\ \hat{p} & \text{if } \omega_t \neq k, \end{cases}$$

where $p \in (\frac{1}{K}, 1)$ and $\hat{p} = \frac{1-p}{K-1}$. The assumption $p > \frac{1}{K}$ ensures that $\hat{p} < p$, and thus for any given k and $k' \neq k$, $Pr[\omega_t = k | m_t = k] > Pr[\omega_t = k | m_t = k']$.

Before observing m_t, the agent must take an action $d_t \in \{1, \ldots, K\}$. The payoff in period t is

$$u_t(d_t, m_t) = \begin{cases} 1 & \text{if } d_t = m_t, \\ 0 & \text{if } d_t \neq m_t. \end{cases}$$

With this specification of payoffs, the outcome of the agent's decision depends on ω_t, but only probabilistically, so that payoff information does not reveal the state. Thus, uncertainty about the state stems from two sources: The fact that signals are imprecise, and the possibility that regime switches occur from time to time. The literature contains some models of changing environments in which the first type of uncertainty is either ruled out or reduced. For example, in the social learning model of Moscarini et al. (1998) the payoff the decision maker observes reveals the state perfectly ex post. In the demand-learning model of Rustichini and Wolinsky (1995), the state is learned perfectly ex post for some actions. In Section 4 we discuss the implications of such an assumption, were it made in the current model.

If the state was observable, then $d_t = \omega_t$ would maximize expected payoff; however only the signals are observable. The agent remembers her past signals and uses them to obtain a statistical inference about the value of ω_t. Let $\sigma_t^k = Pr[\omega_t = k|\{m_l\}_{l<t}]$ be the agent's belief that at the beginning of period t the state is k, and let $\sigma_t = (\sigma_t^1, \ldots, \sigma_t^K) \in \Delta_K$, where Δ_K is the unit simplex in $\mathbb{R}^K$. An updating rule $U : \Delta_K \times K \to \Delta_K$ maps current beliefs σ_t and the current signal m_t into next period's beliefs: $\sigma_{t+1} = U(\sigma_t, m_t)$.

2.1 Bayesian learning

One particular such updating rule is Bayes' rule, which utilizes all relevant information in the "statistically correct" way in order to draw inferences about the unobservable state variable. We denote the Bayesian updating rule by $\mathcal{B}$, i.e. $\sigma_{t+1} = \mathcal{B}(\sigma_t, m_t)$. For changing environments, $\mathcal{B}$ is given by

$$\begin{aligned}\mathcal{B}^k(\sigma, m) &= (1-\varepsilon)Pr[\omega = k|\sigma, m] + \frac{\varepsilon}{K-1}Pr[\omega \neq k|\sigma, m] \\ &= (1-\varepsilon)\frac{\sigma^k Pr[m|\omega = k]}{\sum_{k' \in K} \sigma^{k'} Pr[m|\omega = k']} + \frac{\varepsilon}{K-1}\frac{\sum_{k' \neq k} \sigma^{k'} Pr[m|\omega = k']}{\sum_{k' \in K} \sigma^{k'} Pr[m|\omega = k']}, \quad (1)\end{aligned}$$

where the superscript k indicates the k^{th} component. For $m = k$, (1) can be expressed as

$$\mathcal{B}^k(\sigma, m) = \frac{(1-\varepsilon)\sigma^m p + \hat{\varepsilon}(1-\sigma^m)\hat{p}}{\sigma^m p + (1-\sigma^m)\hat{p}}, \quad (2)$$

and similarly

$$\mathcal{B}^k(\sigma, m) = \frac{(1-\varepsilon)\sigma^k \hat{p} + \hat{\varepsilon}\left[\sigma^m p + (1-\sigma^m-\sigma^k)\hat{p}\right]}{\sigma^m p + (1-\sigma^m)\hat{p}} \quad (3)$$

if $m \neq k$. Note that by our assumptions on p and ε (resp. $\hat{p}$ and $\hat{\varepsilon}$), $\mathcal{B}^k(\sigma, k) > \mathcal{B}^k(\sigma, m)$ for all σ, k, and $m \neq k$. Instead of using the recursive formulations in (1)–(3), beliefs at time t can also be written as a function of initial beliefs σ_1 and the stream of signals $m_1, \ldots, m_{t-1}$ only:

$$\sigma_t = \mathcal{B}(\mathcal{B}(\mathcal{B}(\ldots \mathcal{B}(\sigma_1, m_1) \ldots), m_{t-2}), m_{t-1}).$$

The process by which σ_t evolves is stationary; hence the value of initial beliefs σ_1 will not matter in the long run. Because the sequence of state variables $\{w_t\}_{t=1,2,\ldots}$ is a Markov

process, the sequence of Bayesian beliefs $\{\sigma_t\}_{t=1,2,...}$ itself satisfies the Markov property. (A detailed proof of this fact is in the Appendix.)

An updating rule U is said to be *commuting* if observations are temporally interchangeable, i.e. if $U(U(\sigma, m'), m) = U(U(\sigma, m), m')$ for all m, m' (see Bush and Mosteller (1955)). The Bayesian rule $\mathcal{B}$ is clearly not commuting. In fact, it is easy to show that for all σ and all $m \neq k$ we have

$$\mathcal{B}^k(\mathcal{B}(\sigma, m), k) > \mathcal{B}^k(\mathcal{B}(\sigma, k), m). \tag{4}$$

Recall that observing signal k is indicative of state k. (4) therefore implies that observing signal k later rather than earlier increases the posterior likelihood of state k. Thus, $\mathcal{B}$ is a *recency biased* updating rule, in that signals have a greater impact on beliefs the more recently they have been observed. This feature is an immediate consequence of the changing environments assumption.

Another implication of changing environments is that Bayesian beliefs are bounded away from $\partial\Delta_K$, the boundary of Δ_K. For $0 < \underline{\sigma} < 1/K$, let $\Delta_K(\underline{\sigma}) = \{\sigma \in \Delta_K : \sigma^k \geq \underline{\sigma}\ \forall k\}$. The agent can never be too certain about the value of the state, as it may have changed with probability ε since the last signal was obtained:

Lemma 1. *There exists a tight bound $\hat{\varepsilon} < \underline{\sigma} < \frac{1}{K}$, decreasing in p and increasing in ε, such that if $E \subseteq \Delta_K$ is an ergodic set of Bayesian beliefs, then $E \subseteq \Delta_K(\underline{\sigma})$.*

Since $\underline{\sigma}$ is a lower bound on the probability attached to each possible state, a corresponding upper bound is given by $\overline{\sigma} = 1 - (K-1)\underline{\sigma}$.[4] It is important to note that the existence of these bounds is not driven by the fact that the observed signals are noisy (i.e. $p \neq 1$): If $\varepsilon = 0$, then even with noisy signals the law of large numbers implies that in the long run the agent becomes perfectly informed about the state.

2.2 Linear learning

There are other, non-Bayesian, updating rules that can be used to process the information contained in the stream of signals $\{m_t\}$. One such rule, called the *exponential smoothing* rule, is denoted $\sigma_{t+1} = \mathcal{L}(\sigma_t, m_t)$ and is given by

$$\mathcal{L}(\sigma, m) = (1-\alpha)\sigma + \alpha i(m), \tag{5}$$

where $\alpha \in (0,1)$ and $i(m) = (0, \ldots, 0, 1, 0, \ldots, 0)$, with only the m^{th} entry being non-zero. $i(m)$ can be interpreted as the information contained in signal m. Thus, updated beliefs are simply a convex combination of prior beliefs and new information. More specifically, σ_t is a linear combination of σ_1 and $i(m_1), \ldots, i(m_{t-1})$:

$$\begin{aligned} \sigma_t &= \mathcal{L}(\mathcal{L}(\mathcal{L}(\ldots\mathcal{L}(\sigma_1, m_1)\ldots), m_{t-2}), m_{t-1}) \\ &= \alpha\sum_{l=0}^{t-2}(1-\alpha)^l i(m_{t-1-l}) + (1-\alpha)^{t-1}\sigma_1. \end{aligned}$$

Notice that the rule $\mathcal{L}$ takes as prior beliefs those that were created by itself, not the Bayesian beliefs. As with Bayes' rule, however, σ_t evolves in a stationary way, so that no

[4] We show in the proof of Lemma 1 that also this bound is tight.

assumption on initial beliefs is needed. Sometimes it will be convenient to use the notation $\mathcal{L}_\alpha$ to denote the α-parameterized linear rule.

The linear rule $\mathcal{L}$ has no direct relation to the underlying switching regimes model. Notice, however, that just like the Bayesian operator $\mathcal{B}$, the linear rule $\mathcal{L}$ is recency biased: $\mathcal{L}^k(\mathcal{L}(\sigma, m), k) > \mathcal{L}^k(\mathcal{L}(\sigma, k), m)$ for all σ and all $m \neq k$. The single parameter in (5) is α. This parameter has various intuitive interpretations. Notice that the belief $\tilde{\sigma}_t^k$ can be regarded as the stock of past information indicative of state k available at time t, that is, $i^k(m_\tau)$ for $\tau < t$. The value of α is then simply the rate at which information depreciates: $i^k(m_t)$ will be "worth" $(1-\alpha)i^k(m_t)$ at time $t+1$, $(1-\alpha)^2 i^k(m_t)$ at $t+2$, and so forth. The higher the depreciation rate, the faster new information is unlearned.

Another convenient characterization of learning speed is the *half life* of a signal. The half life is the time after which a signal's weight in current beliefs has decayed from its initial level α to $\alpha/2$, and is calculated as $\lambda = -\frac{\ln 2}{\ln(1-\alpha)}$. A higher value of α corresponds to a smaller value of λ. There is no obvious way of deriving similar measures for informational depreciation or learning speeds for the non-linear Bayesian updating rule $\mathcal{B}$.[5]

3 Relative Performance Measures

If the decision maker's goal is to maximize expected utility, then Bayes' rule is the optimal updating procedure by which to obtain the probabilities that enter the expected utility functional (see, for instance, Luce and Raiffa (1957), p. 313). Using $\mathcal{L}$ instead of $\mathcal{B}$ will result in errornous inferences relative to Bayes' rule. Therefore, a decision maker who employs $\mathcal{L}$ will obtain payoffs which, on average, are less than what is possible by using $\mathcal{B}$.

This paper is concerned with quantifying the extent to which this error in beliefs leads to incorrect decisions and suboptimal consumption levels. Note that $\mathcal{L}$ shares with $\mathcal{B}$ the characteristic that it is recency biased, an important feature for learning in changing environments. At the same time, the exponential smoothing rule is computationally less complex than Bayes' rule. Whereas $\mathcal{B}$ is a non-linear operator on beliefs, $\mathcal{L}$ is a linear operator. How costly is it, then, to use a simple but incorrect rule of thumb for learning in a changing world, as long as it is recency biased?

Let σ_t be Baysian beliefs at time t (i.e. beliefs generated by $\mathcal{B}$), and $\tilde{\sigma}_t$ be beliefs generated by the updating rule $\mathcal{L}$. An agent who does not observe ω_t and who uses $\mathcal{B}$ to update her beliefs will take action $d_t = \arg\max_k \sigma_t^k$. An agent who uses $\mathcal{L}$ takes action $\tilde{d}_t = \arg\max_k \tilde{\sigma}_t^k$. (We may safely ignore the case that $\arg\max_k \sigma_t^k$ or $\arg\max_k \tilde{\sigma}_t^k$ contain more than one element.)

3.1 Error probability

A first performance measure is the probability with which the non-Bayesian agent makes a decision different from the Bayesian agent. Even though these two types of agents hold different beliefs, as long as σ_t and $\tilde{\sigma}_t$ give the highest weight to the same state,

[5] Ideally, one would like to measure how much of Bayesian beliefs at time t can be attributed to a specific piece of information, say m_τ for $\tau < t$. However, the magnitude of change in σ_t that is caused by a change in m_τ depend on σ_τ as well as the sequence of signals $m_{\tau+1}, \ldots, m_{t-1}$. The additive separability of $\mathcal{L}$, on the other hand, eliminates this problem for beliefs obtained from exponential smoothing rules.

d_t and $\tilde{d}_t$ will be the same. Thus, the non-Bayesian agent makes an error at time t if $\arg\max_k \sigma_t^k \neq \arg\max_k \tilde{\sigma}_t^k$. Setting $\rho_t = 1$ if $d_t \neq \tilde{d}_t$, and $\rho_t = 0$ otherwise, we define the error probability of using the learning rule $\mathcal{L}_\alpha$ as

$$D(\alpha) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \rho_t \geq 0. \tag{6}$$

The closer $D(\alpha)$ is to zero, the less likely it is that a non-Bayesian agent makes the "wrong" decision. The use of the limit of averages for time aggregation means we are considering long-run averages, and is appropriate in an environment that is changing but yields a stationary stream of signals and decision. It makes the implicit assumption that the life span of the agent is long relative to the length of the model period.

3.2 Relative consumption

A second group of performance measures concerns the consumption level attainable through the exponential smoothing rule relative to the one attainable by Bayes' rule. As before, we use the limit of averages for time aggregation. The long-run average payoff for the Bayesian agent is

$$v_B = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} u_t(d_t, m_t).$$

The long-run average payoff for an agent who uses updating rule $\mathcal{L}_\alpha$ is

$$v_L(\alpha) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} u_t(\tilde{d}_t, m_t).$$

v_B and $v_L(\alpha)$ are well defined for both types of updating rules. Let

$$C(\alpha) = \frac{v_L(\alpha)}{v_B} \leq 1 \tag{7}$$

be the relative payoff the non-Bayesian agent obtains, measured as a fraction of the Bayesian agent's payoff. The closer $C(\alpha)$ is to one, the smaller the disadvantage of using a non-Bayesian updating rule.

Imagine now an agent who does not use any of the available information and simply chooses a random action in each period, or—which is equivalent for expected payoffs—chooses the same action in each and every period. This agent obtains payoff

$$v_0 = \frac{1}{K} p + \frac{K-1}{K} \hat{p} = \frac{1}{K}.$$

Linear learning is an updating procedure that is somewhere between perfectly rational Bayesian learning and no information processing at all. Therefore, a natural measure for the performance of linear updating is

$$C^*(\alpha) = \frac{v_L(\alpha) - v_0}{v_B - v_0} \leq 1. \tag{8}$$

The denominator in (8) is the additional consumption a perfectly Bayesian learner gets over an unintelligent decision maker. $C^*(\alpha)$ then expresses which fraction of this consumption gain can be had with linear learning, i.e. using $\mathcal{L}_\alpha$ instead of $\mathcal{B}$.

3.3 The relation between errors and consumption

Error probability is related to, but not not fundamentally tied to, measures of economic well-being. Neither it is fundamentally tied to reproductive fitness in an evolutionary framework. Consumption, on the other hand, is fundamental for both economic well-being and reproductive fitness. Error probability is included as a performance measure here since, as we we shall discuss later, the mistakes an agent makes when he learns too fast or too slow shed light on how consumption under linear learning differs from consumption under Bayesian learning. Minimizing errors is instrumental for maximizing consumption, but in a subtle way. Note that if $D(\alpha)$ is close to zero, then $C(\alpha)$ and $C^*(\alpha)$ will necessarily be close to one. However, the converse is not true. Even if the non-Bayesian agent frequently makes wrong decisions, these errors may not be too costly if they occur mostly at times when Bayesian beliefs σ_t place high weight on more than one state. For example, consider $K = 2$. If $\sigma_t^1 = 0.51$ and $\tilde{\sigma}_t^1 = 0.49$, then $d_t = 1$ and $\tilde{d}_t = 2$. The "correct" belief σ_t is diffuse (close to 1/2), and $\tilde{d}_t = 2$ yields only a slightly lower expected payoff than the optimal decision $d_t = 1$. The fact that the non-Bayesian agent chooses a sub-optimal actions is almost without cost in this case. We shall later see that the errors made by non-Bayesian agents are indeed of this type.

4 Informational Depreciation and Optimal Decisions

A decision maker who employs linear rule $\mathcal{L}_\alpha$ instead of $\mathcal{B}$ needs to find a value of the parameter α that optimizes learning performance, as measured by criteria such as error probability or relative consumption. Choosing either too high or too low an informational depreciation rate α leads to more frequent errors and ultimately to inferior consumption streams.

4.1 A sufficient condition for $D(\alpha) = 0$

Under certain conditions, we can identify linear rules that perform exactly as well as Bayesian updating, as the following result shows:

Proposition 2. *$C(\alpha) = 1 = C^*(\alpha) = 1$ and $D(\alpha) = 0$ for all $\alpha \geq \frac{1}{2}$ if the following condition holds:*

$$\varepsilon \geq \varepsilon^* \equiv \frac{p(1-p)(K-1)}{p(K-2)+1}. \tag{9}$$

It is straightforward to verify that $0 < \varepsilon^* < 1 - \frac{1}{K}$, so the region identified in Proposition 2 is always non-empty. If the observed signals are sufficiently precise, or equivalently if the state fluctuation rate is sufficiently high, then every linear rule with $\alpha \geq \frac{1}{2}$ will result in exactly the same decisions as Bayes' rule. Decisions are then based on the last available signal only: The agent simply sets $d_t = m_{t-1}$. In fact, *any* updating procedure T such that $T^k(\sigma, k) > 1/2$ yields these decisions. For perfectly informative signals ($p = 1$) such rules are optimal regardless of ε, since in this case $\varepsilon^* = 0$. As mentioned in Section 2, some papers in the literature consider such informative signals. If the last observed signal reveals the last state, and as long as there is sufficient persistence in the state evolution,

i.e. $\varepsilon < 1 - 1/K$ which we assumed, the decision maker simply needs to take the action that coincides with the last state.

When $p < 1$ and the environment is relatively stable ($\varepsilon < \varepsilon^*$), such a rule of thumb may no longer be successful: Although with a more stable environment the information contained in the last signal becomes more important, there is the countervailing effect that signals from two, three or more periods ago also provide more useful information. A partial converse of Proposition 2 is also true: If $\varepsilon < \varepsilon^*$, then $D(\alpha) > 0$ for $\alpha \geq \frac{1}{2}$. However, even in this case, $\mathcal{L}_\alpha$ with $\alpha \geq \frac{1}{2}$ may be the optimal linear rule, i.e. $D(\alpha) > D(1/2)$ for all $\alpha < 1/2$. This will typically happen if ε is only slightly below ε^*. If ε is sufficiently small, then we will show below through numerical simulations that the optimal linear rule will have a longer half life of information.

Inspection of (9) reveals an interesting relationship between p and ε^*. For $K = 2$, there is a tradeoff between less informative signals (lower p) and higher fluctuation rates (higher ε). When decreasing p, ε^* increases in order for $\mathcal{L}_{1/2}$ to remain optimal: If signals become less informative, there is an added benefit in using more than just the last available signal for learning. An increase in the fluctuation rate can offset this effect. Interestingly, however, this intuition does not hold when $K > 2$. Here, ε^* first increases as p decreases, but as p approaches the non-informative value $1/K$, ε^* starts to decrease.

4.2 The optimal depreciation rate: An example

For the cases not covered by Proposition 2, we numerically search for the values of α that minimize $D(\alpha)$ and maximize $C(\alpha)$ (resp. $C^*(\alpha)$). Unlike in the Proposition, these values will typically be interior solutions, i.e. $0 < \alpha < \frac{1}{2}$. Moving away from these solutions results in learning that is either too fast or too slow, and more prone to certain types of errors. In Section 5, we describe our simulation results in detail. Here, we illustrate some of the effects of too small or too large values of α by means of an example.

The underlying model parameters are set to $K = 2$, $\varepsilon = .05$ and $p = .75$. Figure 1 plots simulated values of beliefs for 16 model periods.[6] After period 3, there is a state switch from $\omega_3 = 2$ to $\omega_4 = 1$; this switch is of course not directly observable by the decision maker. The σ_t line graphs Bayesian beliefs that $\omega_t = 1$. Except for periods 9, 11, and 15, $m_t = \omega_t$. Observe that after the switch, it takes the Bayesian decision maker 3 observations of 1-signals to become sufficiently confident that $\omega_t = 1$ to take action $d_t = 1$. Hence, $d_t = 1$ for $t \geq 7$.

The $\tilde{\sigma}_t$ lines, on the other hand, depict beliefs derived from the linear updating rules $\mathcal{L}_\alpha$, with α taking on three different values. In the first of the three graphs, $\alpha = .22$, which corresponds to a half life of $\lambda = 2.79$ model periods. We see that in all 16 periods, $\tilde{\sigma}_t$ and σ_t lie on the same side of $1/2$, and updating rule $\mathcal{L}_{.22}$ leads to the same decisions as does updating rule $\mathcal{B}$. The value $\alpha = .22$ is in fact the value that minimizes $D(\alpha)$ for the given model parameters, as we shall later see. With this value, the error probability is $D(.22) = 2\%$, so on average only one in 50 periods will be such that σ_t and $\tilde{\sigma}_t$ lie on different sides of $1/2$.

For $\alpha = .05$ and $\alpha = .45$ we have calculated error probabilities of $D(.05) = 20\%$ and $D(.45) = 12.5\%$. Let us examine the nature of errors made when using these values. If

[6] The periods are labelled 1 through 16, but the graph is taken from the middle of a long simulation run and period 1 is *not* the first simulation period.

$\alpha = .05$ ($\lambda = 13.51$), as in the middle graph, linear learning will be "too slow." It takes the linear learner 5 model periods to realize the change in states, so that in periods 7 and 8 a statistically incorrect decision is taken, indicated by circles on the $\tilde{\sigma}_t$ line. If $\alpha = .45$ ($\lambda = 1.16$), on the other hand, linear learning will be "too fast." It only takes 2 periods to adjust beliefs from below 1/2 to above 1/2. As a consequence, action $\tilde{d}_6 = 1$ is taken in period 6. Notice that even though this decision turns out to be ex-post optimal since $\omega_6 = 1$, this fact is irrelevant since ω_t is not observed. Ex-ante, the Bayesian decision $d_6 = 2$ maximizes expected payoffs. One may wonder why it wouldn't be beneficial to learn quickly in order to adjust to state changes more rapidly. The cost at which such quick adjustments come can be seen when looking at periods 11–12 in the bottom graph, for example: There, the occurance of $m_{11} = 2$ in period 11 leads to a sharp drop of beliefs below 1/2 in period 12. If learning is too fast, then random occurances of signals indicative of states other than the true state are wrongly being taken as evidence of a state change. Interestingly, learning that is too slow is not immune to this problem, as can be seen in periods 10 and 12 in the middle graph: The fact that slow learning induces beliefs close to 1/2 makes $\mathcal{L}_{.05}$ prone to incorrect decision as well, as a single signal can be enough to move beliefs from above to below 1/2.

The relation between learning speed, errors, and consumption identified in above has an interesting counterpart in evolutionary economics. Ben-Porath et al. (1993) examine the relation between mutation rates and growth rates in changing environments. There, evolutionary dynamics replace learning and decision making, and mutation rates replace learning speed. If types specialize in reproduction in certain states of nature that fluctuate over time, there exist optimal mutation rates that maximize the population growth.

5 Simulation Results

5.1 The two-state case

We first describe our simulation results for the case $K = 2$. With a switching probability of ε the average duration of a state is ε^{-1}, so when simulating T model periods we should expect to see about εT state changes. Unless noted otherwise, we set $T = 10,000/\varepsilon$, which means that we simulate approximately 10,000 state changes. For a given parameter combination (ε, p), we measure $D(\alpha)$, $C(\alpha)$, and $C^*(\alpha)$ by simulating T model periods, where α is taken from the grid $\{.01, .02, \ldots, .50\}$. All linear updating rules with $\alpha \geq .5$ are equivalent for decision making, so there is no need to consider values of α beyond .50.

We consider two environments, a relatively stable one ($\varepsilon = 0.05$) and a more turbulent one ($\varepsilon = 0.15$). We further examine three values for the signal precision, $p \in \{0.65, 0.75, 0.85\}$.[7] Figure 2 depicts the error probability $D(\alpha)$ and the relative consumption measures $C(\alpha)$ and $C^*(\alpha)$ for all six parameter combinations; the top panel is for the case $\varepsilon = 0.05$ and the bottom panel for the case 0.15. In each of the graphs, the informational depreciation rate α is on the the horizontal axes.

We first observe that linear updating performs almost as well as Bayesian updating: In each of the cases we do get a unique value for α that minimizes $D(\alpha)$, and one that

[7] The model was simulated with essentially the same conclusions for several other parameter values as well.

maximizes $C(\alpha)$ and $C^*(\alpha)$. Although they typically coincide, there are cases when they do not: For example, when $\varepsilon = 0.15$ and $p = 0.75$, then $D(\alpha)$ is minimized at $\alpha = 0.37$, but $C(\alpha)$ and $C^*(\alpha)$ are maximized at $\alpha = 0.35$. The minimized value for D ranges from 0 ($\varepsilon = 0.15$, $p = 0.85$) to 2.09% ($\varepsilon = 0.05$, $p = 0.65$). Looking at the relative consumption measures C and C^*, however, we see that it is possible to make their value virtually indistinguishable from 1. Our simulations resulted in maximized values for C (C^*) between 99.84% (99.58%) for (.05, .75) and 100.02% (100.37%) for (.15, .65). The latter case is theoretically impossible, but only in the limit for an infinite number of observations. Below we show that such cases arise rather frequently for finite simulations. The reason that C and C^* can get closer to 1 than D gets to zero has been discussed in Section 3.3 already: Not all errors are equally costly, and when α is set to minimize $D(\alpha)$, then it turns out that the few remaining errors still made occur when the Bayesian decision maker would almost be indifferent between her available actions. The only case in which $D(\alpha) = 0$ is (.15, .85); this is predicted by Proposition 2. In this case, $\alpha = 0.5$ is a corner solution, and the optimal learning rule is the one that always takes the action that equal the last observed signal. In all other cases we get an interior solution for α. Further examining the graphs of D, C, and C^* we see that the short-memory rule $\mathcal{L}_{0.5}$ would have resulted in significant losses. Thus, there is considerable value in using more than only the last available signal for decision making. However, as the simulations show the decision maker need not go "all the way" and use Bayesian updating in order to reap nearly all the payoffs that a Bayesian agent would obtain.

An important observation, especially regarding the consumption measures, is that there is relatively little variation around the optimum, indicating that even a suboptimally parameterized linear updating rule can perform very well. For example, for $\varepsilon = .05$ and $p = .85$, we calculated a Bayesian payoff of approximately $v_B = 0.76$, so that Bayesian updating yields a 51% increase in payoffs over $v_0 = 0.5$. 90% or more of this inrease can be obtained by any linear updating rule $\mathcal{L}_\alpha$ with $\alpha \in [0.15, 0.49]$ (that is, a half life between 1.03 and 4.27 model periods).

Finally, the graphs contain information on how the optimal learning speed adjusts when the fundamental parameters p and ε change. As p increases, the non-Bayesian decision maker should increase α. This is intuitive: With a low depreciation rate α, $\mathcal{L}_\alpha$ puts relatively high weight on a large number of signals. The more available information is used in the current forecast $\tilde{\sigma}_t$, the longer it takes to pick up possible state changes. When p becomes larger, this tradeoff changes and one can decrease the weights of past signals and increase those of recent ones. When p is held fixed and ε is increased, α increases. Also this response is intuitive, as a higher switching probability diminishes the informational value of past signals relative to recent ones.

5.2 More than two states

To examine the effect of an increase in the state space, we look at the cases $K = 3$, $K = 5$, and $K = 10$. We perform two exercises. First, we fix $\varepsilon = .05$ and $p = .75$; our results are reported in the first group of graphs in Figure 3. For $K = 3$, we find optimized values $D(.29) = 2.70\%$, $C(.24) = C^*(.24) = 100\%$ (two decimals accuracy). As before, C and C^* are rather flat around the optimum. For $K = 5$ and $K = 10$, C and C^* decrease, but only slightly. For example, when $K = 10$, we have $C(.29) = 99.29\%$ and $C^*(.29) = 99.16\%$ at

the optimum.

The parameterizations used in the previous paragraph may not be the most appropriate when comparing different numbers of states. Observe that the graphs for the three different values of K in the top panel of Figure 3 are almost identical, except that the C^* curve moves closer to the C curve as K increases. The reason for this effect is that when p is fixed, $\hat{p}$ decreases in K and as a consequence, it turns out that v_B stays constant at approximately .65. Since $v_0 = 1/K$ becomes smaller, the relative advantage of Bayesian updating over no information processing grows as K increases. Hence, an increase in K that makes the updating problem more difficult should be accompanied by a decrease in p.

For this reason, in our second exercise we still fix $\varepsilon = .05$ but adjust p so that $p/\hat{p} = 3$ remains constant. Thus, regardless of K it will be three times as likely to observe a signal that equals the true state than one that equals any *given* other state. For $p/\hat{p} = 3$ to hold we need to set $p = 3/(K+2)$. The second group of graphs in Figure 3 thus contains the results for $p = 3/5$ ($K = 3$), $p = 3/7$ ($K = 5$), and $p = 1/4$ ($K = 10$). We see that as K increases, the optimal value of α decreases: The decision maker utilizes more of his past information. The minimized value of $D(\alpha)$ rises from 3.71% ($K = 3$) to 9.16% ($K = 10$), while the maximized value of $C^*(\alpha)$ falls slightly from 99.02% to 98.19%. Nonetheless, even when $K = 10$ is it true that many suboptimal linear rules perform relatively well: For $0.03 \leq \alpha \leq 0.12$ the non-Bayesian gets 90% or more of the Bayesian payoff increase over v_0 (a half life between 5.42 and 22.76 model periods).

5.3 Statistical analysis

While graphical inspection of the numerically obtained functions C and C^* suggests that there is almost no difference between the consumption levels under Bayesian updating and an optimal linear rule, we now quantify this statement statistically. We consider only the two-state case here. If $C(\alpha) < 1$, then simulation of T model periods reveals this fact at any desired level of significance with probability one as $T \to \infty$. The statement made earlier that $C(\alpha)$ is virtually indistinguishable from one therefore applies to the observation of a *finite* number of model periods. We hence consider the following experiment: A statistician observes the payoff streams of two decision makers, one using $\mathcal{B}$ and one using $\mathcal{L}$ as an updating procedure, over a finite number of T periods. How likely is it that he will conclude that $C(\alpha) < 1$ at a given level of significance? To answer this question, we again simulate the model for $K = 2$ and various values for ε and p, using the corresponding value α that was found to maximize $C(\alpha)$ in the previous simulations. Using the so generated dataset of Bayesian and non-Bayesian payoffs, we test whether $v_L(\alpha) < v_B$. This process is repeated, and by examining the frequency of various statistical conclusions over these simulations we obtain an estimate of the probability that $C(\alpha) < 1$ at given levels of significance. The closer this probability is to the significance level, the less discernible are Bayesian and non-Bayesian payoffs.

We employ difference of means tests for both independent and paired samples. The first test is appropriate when payoff information is available for two agents who observe independent signals, which means that we need to draw the sequences of state variables and signals twice, once for the Bayesian and once for the non-Bayesian. Given the so

simulated payoff sequences, we compute the z-statistic

$$z^i = (\overline{u}_L - \overline{u}_B)\left(s_L^2/T + s_B^2/T\right)^{-\frac{1}{2}}, \tag{10}$$

where $\overline{u}_L$ and s_L^2 are the sample mean and variance of the non-Bayesian's payoffs, and $\overline{u}_B$ and s_B^2 are the sample mean and variance of the Bayesian's payoffs. A paired sample test, on the other hand, is appropriate when the two agents observe the same signals based on the same state realizations. As Figure 1 reveals, for a well-parameterized linear updating rule Bayesian and non-Bayesian decisions are identical most of the time, so that the associated payoffs are highly correlated. Treating these payoffs as independent would severely overestimate the variability of one stream relative to the other and thus artificially compress (10). Instead we take the sequence of period-by-period payoff differences $\delta_t \equiv u_t(\tilde{d}_t, m_t) - u_t(d_t, m_t)$, and then compute the z-statistic

$$z^p = \overline{\delta}\left(s^2/T\right)^{-\frac{1}{2}}, \tag{11}$$

where $\overline{\delta}$ and s^2 are the mean and variance of the set of observations $\{\delta_t\}$.

If T is large then under the null hypothesis ($C(\alpha) = 1$) the random variables z^i and z^p follow the standard normal distribution. We used $T = 10,000$, which is conservative for our purposes as the number of observations obtainable in experimental or field studies in economics or other disciplines is typically much smaller. We repeated these simulations 4,000 times to obtain empirical distributions of z^i and z^p. We performed this exercise for each of the parameterizations from the previous section. The only exception is ($\varepsilon = 0.15$, $p = 0.85$) where we conclude from Proposition 2 that the distribution of z^i is standard normal, and that of z^p degenerate with all mass at zero. The results are given in Table 1.[8]

Consider the independent case first. For $\varepsilon = 0.05$ and $p = 0.65$, almost one half of the time (47.35%) the linear rule ex-post outperforms Bayes' rule. Bayes' rule outperforms the linear rule in the remaining cases; however, the payoff difference is not statistically significant at the 25%-level (10%-level) in 23.83% (29.53%) of all cases. At the 5%-level the probability that the linear updating rule leads to payoffs significantly lower than the Bayesian payoffs is 8.25%. Keep in mind, however, that the likelihood that a *Bayesian* agent obtains payoffs that are lower than another Bayesian agent's at the 5%-level of significance is exactly 5% (i.e. the probability of making a type-I error), so the chance of rejecting $C(\alpha) = 1$) at the 5%-level is only 3.25 percentage points higher than it would be under the null hypothesis. With paired samples it is less often the case that the linear rule does better than Bayesian updating, so that we observe the correct sign of the test statistic more often. However, the degree of discernibility at given significance levels can be higher or lower than with an independent sample.

When we increase the signal precision the Bayesian and non-Bayesian payoffs typically become more discernible: The z-statistic has the wrong sign less often than before, and is significant with the correct sign more often than before. The reason for this effect is that

[8] Smaller values for T makes the non-Bayesian payoff less discernible from the Bayesian payoff. However, the changes are modest: For example, when $T = 100$, the probability of rejecting the null at the 5%-level for $\varepsilon = 0.05$ and $p = 0.75$ reduces from 9.33% to 8.71%, suggesting that the results reported are fairly robust within a wide range of possible values for T.

Distribution of z-statistic for			incorrect sign	not significant	significant at				
ε	p	α		at 25%	25%	10%	5%	2.5%	1%
Independent sample test									
0.05	0.65	0.14	47.35	23.13	29.53	14.95	8.25	4.90	2.73
	0.75	0.21	46.98	21.83	31.20	16.15	9.33	5.88	3.15
	0.85	0.31	44.70	21.68	33.63	18.05	11.43	7.08	4.05
0.15	0.65	0.32	49.53	25.13	25.35	11.33	6.38	3.15	1.30
	0.75	0.38	50.00	22.80	27.20	12.13	6.58	3.58	1.43
	0.85	≥ 0.5	50.00	25.00	25.00	10.00	5.00	2.50	1.00
Paired sample test									
0.05	0.65	0.14	41.90	26.83	31.28	12.85	6.65	3.50	1.78
	0.75	0.21	35.93	27.90	36.18	17.38	10.25	5.78	2.43
	0.85	0.31	16.43	24.35	59.23	35.55	22.93	14.30	7.75
0.15	0.65	0.32	47.48	26.23	26.30	10.73	5.15	2.70	1.10
	0.75	0.38	47.45	26.95	25.60	9.65	4.70	2.38	0.78
	0.85	≥ 0.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 1: Rejection frequencies of $C(\alpha) = 1$ in percent

The table shows the probabilities with which the null hypothesis $C(\alpha) = 1$ is rejected in a one-tailed test against $C(\alpha) < 1$. The underlying empirical distributions of the test statistics are based on a sample sizes of 4,000 simulations, each containing 10,000 model periods.

with a higher value of p, there is less noise in the data and decision account for a larger fraction of the payoff differences relative to noise. However, the argument that supported Proposition 2 suggests that as ε or p become high the linear updating rule results in the same decisions as Bayes' rule. Hence there will be a region in the parameter space were a further decrease in noise will actually lead to a lesser degree of discernibility between the Bayesian and non-Bayesian payoffs. This effect is especially strong when the linear agent bases his beliefs on the same signals as the Bayesian agent, as can be seen in the last three rows of the table.

6 Evolution of Learning Rules: A Simple Example

In this section, we explore the tradeoff between computation costs and accuracy in an evolutionary setting. Using a specific learning rule could either be a hardwired behavioral trait, or the conscious choice made by the individual. In the former case, biological evolution selects those individuals with "good" learning rules, whereas in the latter case evolution is a metaphor for a learning process by which good rules are selected. It is then possible that the use of simple procedures such as linear learning is a stable outcome of such a Darwinian process if there is a biological cost of performing more complex operations. The biological cost can be interpreted literally (a more powerful brain requires

more energy) or in terms of the opportunity cost if biological resources must be allocated to many different tasks. Using linear updating procedures enables the organism to save on this cost for a loss in accuracy. We have shown that this loss can be modest, so an organism which develops the capability of Bayesian updating may be selected *against* on the basis that the gain in accurate forecasts is too small to be justified by the induced cost. This is certainly true for any cost differential when the environment falls into the parameter range identified in Proposition 2. We now substantiate the claim for other environments by examining a simple and highly stylized example of Darwinian evolution in changing environments.

We consider the case $K = 2$, $\varepsilon = 0.05$, and $p = 0.75$, for which linear rules are less accurate than Bayesian updating. The optimal linear rule in this case is $\alpha = 0.22$. We examine an initial population that consists of 50% Bayesian learners, and 50% non-Bayesians. The non-Bayesians are divided evenly into 50 groups using linear learning rules with $\alpha = 0.01, 0.02, \ldots, 0.50$. Note that this includes agents who always align their action with the last observed signal, namely those with $\alpha = 0.5$. All agents observe the same stream of signals for computational ease. Bayesian updators pay a cost $c = 0.001$ per period, which is about 0.15% of the expected Bayesian payoff per period ($v_B \approx 0.65$).[9] There are no mutations, and each selection cylce consists of 50 model periods. We employ a very simple replicator dynamic which adjusts the proportion of agents in the population proportionally to the agents' performance since the last cycle. That is, if L is the number of types, $f(l)$ is the current fraction of type l in the population, and $u(l) \geq 0$ is the payoff for type l in the current selection-cycle (including any costs involved with a particular updating procedure), then the fraction of type-l agents entering the next cycle is adjusted to

$$f'(l) = \frac{f(l)u(l)}{\sum_{m=1}^{L} f(m)u(m)}.$$

We simulate 250,000 model periods, grouped into 5,000 selection cycles à 50 model periods. Allowing 50 periods between selection stages has several effects. First, it gives each agent the opportunity to accumulate a positive payoff sum (otherwise a type that obtains a zero payoff would go extinct immediately). Alternatively one could simply adjust the "base payoff" the agent gets from taking an incorrect action to a positive number, or make the replicator dynamic less responsive to payoff differences. Secondly, averaging payoffs over 50 periods per cycle reduces the effects of the unrealistic assumption that all agents in the population observe the same signal. Finally, it reduces the volatility of both the Bayesian and the non-Bayesian payoffs across selection steps and hence reduces the impact of the randomness in the underlying stochastic process on outcomes. With short cycles, single payoff draws tend to have rather persistent effects.[10]

Notice that not all linear learning rules are equally good: For the cost advantage

[9] The results are robust for other intitial distributions as well. Furthermore, one can introduce a considerable cost for linear updating with $\alpha < 0.5$ relative to the "short-memory rule" $\alpha = 0.5$ without altering the conclusion, as long as the Bayesians pay the additional cost c.

[10] Selection cycles which are too long are not desirable, either, as they may bias the results in favor of linear rules. The reason is that these rules can involve a more volatile consumption patterns than Bayes' rule, and since a long selection cycle smooths out this volatility too much it tends to diminish the so-called *geometric mean fitness effect* (see, for example, Bergstrom and Godfrey-Smith [1998], and references cited therein). We found 50 model periods to be a good compromise.

of linear rules to translate into an evolutionary advantage, it is necessary that the non-Bayesian agents use linear rules with the parameter α set to the optimal value ($\alpha = 0.22$ in this example) or close to it. The question then arises how such good parameter values can be found by the linear learners who, per definition, are boundedly rational. If there was a realistic computational set-up cost associated with finding the correct value for α it would very likely wash out any subsequent cost savings from using the linear rule. Our model does not presume that the non-Bayesian agents already know the optimal rule, or that they must compute it beforehand. Rather, we present a model that allows agents to "learn how to learn" in a natural evolutionary sense: By including a wide range of linear updators (corresponding to many different values of α) in the initial population, good linear updating rules will be selected through the replicator dynamics in the long run. This means that linear updators who learn too slowly or too quickly will typically be at a disadvantage due to their innacurate predictions, while Bayesian learners will be at a disadvantage due to their higher computational costs. We are therefore interested not only in the question whether Bayesian learning will be selected against by evolution, but also in whether suboptimal linear rules will face such a fate. If they do, then in the long run our model will be populated mostly by agents who use well-parameterized linear rules.

Figure 4 plots the fraction of Bayesian learners and those linear learners with rules $\alpha \in [0.17, 0.27]$. We see that initially the fraction of Bayesians increases: This is due to the fact that the parameter α is evenly distributed among the linear learners, so that relative to the population average the Bayesians do well. However, within the group of linear learners, those with an α-value close to 0.22 increase their frequency, and those with learning speeds that are too high or too low decrease their frequency. This means that over time the linear learners *as a group* become better, and the advantage of the Bayesians *relative to the whole population* becomes smaller. Hence there is a point after which the fraction of linear learners (now consisting almost exclusively of those with learning speeds close to the optimal one) increases dramatically; in the simulation this occurs after about 800 selection stages. After approximately 4,500 stages, the Bayesian's are virtually extinct.[11]

The example serves the purpose of demonstrating the possibility that linear rules have an evolutionary advantage over Bayesian updating. More complicated evolutionary models can obviously be considered, but those are beyond the scope of this work. The evolution of learning rules is studied in detail in several other papers, often in a context where expectations and actual observations feed back into each other in equilibrium. On the

[11] By introducing positive mutation rates, it is possible that Bayesians and non-Bayesian's coexist, as the "outflow" of Bayesian agents due to selection pressure can be compensated by an "inflow" due to random mutations. For example, if the replicator dynamic is adjusted to

$$f'(l) = (1-\mu)\frac{f(l)u(l)}{\sum_{m=1}^{L} f(m)u(m)} + \mu/T,$$

(i.e. a fraction m of the population mutates uniformly across the type space every period), a steady state fraction of at least m/T individuals of each type can be maintained. We found that with $m = 0.1\%$, a fraction of 0.78% of Bayesians coexists with linear agents in the population. The range of linear rules that are used in the population becomes larger as well: For example, while Figure 4 shows that the agents with $\alpha \in [0.17, 0.27]$ make up virutally the entire population when $m = 0$, with the mutation rate those agents account for only about 85% of the population.

computational side, a growing literature on artificial financial markets deals with evolutionary models in which different forecasting and/or trading rules (e.g. "fundamentalist" vs. "chartist" strategies) are used in the population and selected based on their fitness relative to each other (see LeBaron [2005] for a survey of agent-based financial markets). A theoretical treatment of the rather complicated dynamic phenomena in such feedback systems is Brock and Hommes (1997), who consider a model of linear demand and supply with two types of agents—those with naive (i.e. adaptive) expectations about future prices, and those with rational expectations that incorporate the fact that there are also naive agents (see also Hommes [2005] for a survey of the theoretical literature on dynamic heterogeneous agent models). Similar to the evolutionary setup described at the beginning of this section, the use of the rational predictor is costly so that a similar tradeoff between ease of computation and accuracy exists. In the *adaptive rational equilibria* of the model, however, the associated dynamics are more complicated than the simple replicator dynamics studied above, and in fact often chaotic. The reason is that when there is a large fraction of naive agents, prices exhibit cobweb-type fluctuations, which makes the rational predictor more attractive. Conversely, a large fraction of rational agents results in stable prices that can be accurately predicted by naive expectations.

7 Discussion: Applications and Extensions

This paper demonstrated the surprising accuracy of linear updating rules for decision making in a dynamic discrete choice model with changing environments. There are several areas to which our results could be applied. We will discuss two of them below.

7.1 Consumer choice

Exponential smoothing procedures have been used to study consumer choice in the marketing literature. Guadagni and Little (1983) model brand loyalty formation as an exponentially weighted average of past purchasing decisions. The so constructed loyalty variable was used as an explanatory input in a random utility model of consumer choice. (Other inputs were general brand characteristics for grocery items, marketing data, and individual purchase data from a supermarket scanner panel dataset.) While this construction is neither optimal in the Bayesian sense, nor in any obvious way more realistic than other boundedly rational rules, it was found to be of significant explanatory power. Moreover, the simulated evolution of out-of-sample market share was in close agreement with actual purchasing decisions. In light of our results, the appeal of such reduced forms is that they are often as predictive as a fully specified structural version, for example Bayesian approaches or switching regime models in econometrics, or structural models of consumers' belief formation (e.g. Erdem and Keane (1996)). There are several reasons why a simple weighted average of past information can be so predictive. Suppose brand characteristics change over time and consumers try to learn these characteristics through repeated purchases and some updating procedure. Since simple updating rules such as exponential smoothing lead to mostly the same decisions as Bayesian updating, it is possible that consumers in fact use these simple procedures. However, since the correct Bayesian updating rule would be almost observationally equivalent to some linear process,

identifying the process by which consumers actually update their expectations may be difficult. Alternatively, suppose that consumers tastes change over time, resulting in changes in their purchasing decisions. The econometrician must now draw inference about an unobservable time-varying state, and a simple linear specification would be capable of accurately picking up these state changes.

7.2 Experimentation

One obvious extension of this paper is to study linear updating in a framework of experimentation. For instance, in a two-armed bandit model, one may think of one arm as yielding certain payoffs, and the other as having uncertain payoffs. In the uncertain arm, a high payoff that dominates the certain arm is obtained with some probability, and a low payoff that is dominated by the certain arm is obtained with the remaining probability. The probability of obtaining a high payoff is the state of the bandit, and changes over time. Suppose there are only two states: One in which the probability of a high payoff is sufficiently large for the uncertain arm to have a higher expected payoff than the certain arm, and one in which it is so low that the certain arm has the advantage. It is intuitive that any reasonable decision procedure should visit the uncertain arm from time to time, since even if one is confident that the bandit is in the bad state, by pulling the certain arm all the time one learns nothing about a possible state switch. The question, left to future research, is then whether an appropriately parameterized linear rule can perform similarly well compared to the Bayesian rule.

7.3 Fluctuations in the rate of change

Finally, while the model presented in this paper concerns changing environments, the rate of change is assumed to be constant. This may not hold in reality, however. Consider, for example, the current average length of fashion or product cycles and compare it to what it was, say, in the 18th century, or even a few decades ago. An obvious extension of our model is then to allow not only the state variable ω_t to fluctuate over time, but also in the parameter ε itself. With this assumption, agents need to form beliefs about ω_t *and* the current value of ε; so that both Bayesian learning and linear learning would be more complicated. In particular, such a formulation could give rise to more complex evolutionary dynamics. If one were still to consider the simple linear rules $\mathcal{L}_\alpha$ in such a world (i.e. rules that only learn the state ω_t), it might still be possible for them to prevail if their associated compuational cost is very low. However, the cost advantage relative to higher-order rules—even non-Bayesian ones—which also try to learn the current rate of change, would have to be more pronounced. The reason can be seen from Figure 4: For the linear rules to grow in the population, it is necessary that within the pool of linear agents there is sufficient concentration around the optimal rule. In the example, this took about 800 selection cycles. With time-varying values for ε, the optimal value for α will change; thus it becomes more difficult to reach a level of concentration that would put higher-order rules at a realitive disadvantage.

Appendix A: Proofs

The Markov Property for the Bayesian Updating Rule

Here we show formally that the Bayesian updating rule in our setup satisfies the Markov property, that is, the belief σ_{t+1} depends only on σ_t and m_t, but not on m_τ, $\tau < t$. Notice that

$$Pr[\omega_{t+1} = k|\omega_t] = \begin{cases} 1-\varepsilon & \text{if } \omega_t = k, \\ \hat{\varepsilon} & \text{if } \omega_t \neq k, \end{cases} \quad \forall \omega_{t-1}, \omega_{t-2}, \ldots.$$

In other words, the transition probabilities at time t depend only on ω_t, but not on ω_{t-1}, etc. If the value of ω_t is not known, but we are given a probability η^k that $\omega_t = k$, then the probability distribution of the state at $t+1$ can be expressed as a compound probability, i.e.

$$Pr[\omega_{t+1} = k|Pr[\omega_t = k] = \eta^k] = \eta^k(1-\varepsilon) + (1-\eta^k)\hat{\varepsilon}. \tag{12}$$

(Obviously, for (12) to hold, it is necessary that the evolution of ω_t follows a Markov process.) Since the decision maker has access to all signals $m_t, m_{t-1}, \ldots, m_1$, the relevant probability η^k is given by

$$\eta^k = Pr[\omega_t = k|m_t, m_{t-1}, \ldots, m_1],$$

and we can rewrite (12) as

$$\begin{aligned} Pr[\omega_{t+1} = k|Pr[\omega_t = k] = \eta^k] &= Pr[\omega_t = k|m_t, m_{t-1}, \ldots, m_1](1-\varepsilon) \\ &\quad +(1 - Pr[\omega_t = k|m_t, m_{t-1}, \ldots, m_1])\hat{\varepsilon} \\ &= Pr[\omega_{t+1} = k|m_t, m_{t-1}, \ldots, m_1]. \end{aligned} \tag{13}$$

Now let us examine η^k in more detail. Suppose $\sigma_t^k = Pr[\omega_t = k|m_{t-1}, \ldots, m_1]$, and consider the case $m_t = k$. Regardless of how σ_t^k is computed, we can use the standard version of Bayes' rule (i.e. for non-changing environments) to write

$$\begin{aligned} \eta^k = Pr[\omega_t = k|m_t, \sigma_t] &= \frac{\sigma_t^k Pr[m_t = k|\omega_t = k]}{\sigma_t^k Pr[m_t = k|\omega_t = k] + (1-\sigma_t^k)Pr[m_t = k|\omega_t \neq k]} \\ &= \frac{\sigma_t^k p}{\sigma_t^k p + (1-\sigma_t^k)\hat{p}}. \end{aligned}$$

(See, for example, Gelman et al. [1995, p. 11] for an explanation of this "data adding" feature of Bayes' rule.) Plugging this expression back into (13), we get

$$\begin{aligned} Pr[\omega_{t+1} = k|m_t = k, m_{t-1}, \ldots, m_1] &= \eta^k(1-\varepsilon) + (1-\eta^k)\hat{\varepsilon} \\ &= \frac{(1-\varepsilon)\sigma^k p + \hat{\varepsilon}(1-\sigma^k)\hat{p}}{\sigma^k p + (1-\sigma^k)\hat{p}}, \end{aligned}$$

which is identical to (2). The case $m_t = k' \neq k$ is similar. We thus conclude that in order to compute σ_{t+1} one only needs to know m_t and σ_t. Since the argument is independent of t, it follows that the sequence $\{\sigma_t\}_{t=1,2,\ldots}$ satisfies the Markov property.

Proof of Lemma 1

The proof proceeds in a number of steps. In step 1, we collect some technical properties of the operator (2). We then establish the existence of two bounds, $\underline{\sigma}$ and $\overline{\sigma}$, such that if $\sigma_t^k \in [\underline{\sigma}, \overline{\sigma}]$ for some k and t, then $\sigma_{t'}^k \in [\underline{\sigma}, \overline{\sigma}]$ for all $t' > t$. This is done in step 2. Next, in step 3, we show that no belief $\sigma \notin \Delta_K(\underline{\sigma})$ can be in an ergodic set and that the bound $\underline{\sigma}$ is tight. In step 4 we show that $\hat{\varepsilon} < \underline{\sigma} < 1/K$. Finally we establish the comparative properties of $\underline{\sigma}$ in step 5.

Step 1. We prove the following properties of the updating operator $\mathcal{B}^k(\sigma, k)$, defined in (2):

(B1) $\mathcal{B}^k(\sigma, k)$ is continuous, strictly increasing, and strictly concave in σ^k.

(B2) $\mathcal{B}^k(\sigma, k)$ is strictly increasing in p.

(B3) If $\sigma^k > \frac{1}{K}$, then $\mathcal{B}^k(\sigma, k)$ is strictly decreasing in ε.

For (B1), continuity is obvious. To show the other properties, note that any function of the form

$$f(x) = \frac{ax + b(1-x)}{cx + d(1-x)}, \quad x \in [0,1],$$

increases in x if $\frac{a}{c} > \frac{b}{d}$, and provided this is the case, it will be concave in x if $c > d$. Now substitute

$$a = (1-\varepsilon)p, \; b = \hat{\varepsilon}\hat{p}, \; c = p, \; d = \hat{p}$$

to get (2) as a function of σ^k. Since $1 - \varepsilon > \hat{\varepsilon}$ we have $\frac{a}{c} = \frac{b}{d}$, and since $p > \hat{p}$ we also have $c > d$. Hence $\mathcal{B}^k(\sigma, k)$ is strictly increasing and strictly concave in σ^k. For (B2), proceed similarly by setting

$$= (1-\varepsilon)\sigma^k, \; b = \frac{\hat{\varepsilon}(1-\sigma^k)}{K-1}, \; c = \sigma^k, \; d = \frac{1-\sigma^k}{K-1}$$

to get (2) as a function of p. As before, since $1 - \varepsilon > \hat{\varepsilon}$ we have $\frac{a}{c} = \frac{b}{d}$. Hence $\mathcal{B}^k(\sigma, k)$ is strictly increasing in p. Finally, to show (B3), differentiate (2) with respect to ε to get

$$\frac{\partial \mathcal{B}^k(\sigma, k)}{\partial \varepsilon} = \frac{1}{\sigma^k p + (1-\sigma^k)\hat{p}} p \left(\frac{1-\sigma^k}{(K-1)^2} - \sigma^k \right).$$

If $\sigma^k > 1/K$ then

$$\left(\frac{1-\sigma^k}{(K-1)^2} - \sigma^k \right) < \frac{1-\frac{1}{K}}{(K-1)^2} - \frac{1}{K} = \frac{1}{K(K-1)} - \frac{1}{K} < 0,$$

and thus $\frac{\partial \mathcal{B}^k(\sigma,k)}{\partial \varepsilon} < 0$.

Step 2. Fix any $k \in K$. Suppose first that a string of signals indicative of state $\omega = k$ is observed ($m_1, m_2, \ldots = k$), so $\sigma_1^k < \sigma_2^k, \ldots$. Observe that (2) is independent of σ^m for $m \neq k$ and satisfies $\mathcal{B}^k(\sigma, k) > 0$ if $\sigma^k = 0$ and $\mathcal{B}^k(\sigma, k) < 1$ if $\sigma^k = 1$. Together with (B1), there exists a unique value $\overline{\sigma} \in (0, 1)$ such that

$$\overline{\sigma} = \mathcal{B}^k(\sigma, k) \quad \forall \sigma \text{ s.t.} \sigma^k = \overline{\sigma}, \tag{14}$$

and such that $\sigma_t^k \to \overline{\sigma}$. Furthermore, if $\sigma_t^k < \overline{\sigma}$ then $\sigma_{t'}^k < \overline{\sigma}$ $\forall t' > t$. Now suppose that a string of signals indicative of states other than k is observed ($m_1, m_2, \ldots \neq k$), so $\sigma_1^k > \sigma_2^k > \ldots$. Observe that unlike (2), (3) is not independent of σ^m for $m \neq k$. However, to minimize (3) it is sufficient to replace σ^m by $\overline{\sigma}$, which is equivalent to assuming $m_1 = m_2 = \ldots = m$ for some fixed state m (so that $\sigma_t^m \to \overline{\sigma}$). After rearranging terms we have

$$\mathcal{B}^k(\sigma, m) = \hat{\varepsilon} + (1 - \varepsilon - \hat{\varepsilon}) \frac{\sigma^k \hat{p}}{\overline{\sigma} p + (1 - \overline{\sigma}) \hat{p}}.$$

Now $\mathcal{B}^k(\sigma, m)$ is independent of σ^m for $m \neq k$. It is also continuous, strictly increasing, and strictly convex in σ^k, and satisfies $\mathcal{B}^k(\sigma, m) > 0$ if $\sigma^k = 0$ and $\mathcal{B}^k(\sigma, m) < 1$ if $\sigma^k = 1$. Hence there exists a unique $\underline{\sigma} \in (0, 1)$ such that $\underline{\sigma} = \mathcal{B}^k(\sigma, m)$ for all σ with $\sigma^k = \underline{\sigma}$, and such that $\sigma_t^k \to \underline{\sigma}$ $\forall k \neq m$. This implies that $\overline{\sigma} + (K - 1)\underline{\sigma} = 1$. Furthermore, if $\sigma_t^k > \underline{\sigma}$ then $\sigma_{t'}^k > \underline{\sigma}$ $\forall t' > t$.

Step 3. First, consider a finite sequence of signals $m_1 = \ldots = m_T = 1$, $m_{T+1} = \ldots, m_{2T} = 2$, etc., until $m_{(K-1)T+1} = \ldots, m_{KT} = K$ for some finite T. If T is chosen large enough, then $s_T^1 > \underline{\sigma}$ and $\sigma_T^2, \ldots, \sigma_T^K < \overline{\sigma}$. Thus $\forall t > T$, $\sigma_t^1 > \underline{\sigma}$ and $\sigma_t^k < \overline{\sigma}$ $(k \neq 1)$. Similarly $\forall t > 2T$, $\sigma_t^2 > \underline{\sigma}$ and $\sigma_t^k < \overline{\sigma}$ $(k \neq 2)$. Continuing in this fashion, we obtain $\sigma_t^k \in [\underline{\sigma}, \overline{\sigma}]$ for all k and all $t > KT$. Since the so constructed sequence $m_1, \ldots, m_{KT}$ has a positive probability, any ergodic set E of Bayesian beliefs must be a subset of $\Delta_K(\underline{\sigma})$. Next, take the sequence of signal $m_1 = m_2 = \ldots = m_T = k$ for a fixed state k. Then, as $T \to \infty$, $\sigma_T^k \to \overline{\sigma}$ and $\sigma_T^{k'} \to \underline{\sigma}$ for $k' \neq k$, so that the corners of $\Delta_K(\overline{\sigma})$ are limit points of any ergodic set of Bayesian beliefs. The bound $\underline{\sigma}$ is therefore tight.

Step 4. Since $\overline{\sigma} + (K - 1)\underline{\sigma} = 1$ and $\overline{\sigma} > \underline{\sigma}$ it must be that $\underline{\sigma} < 1/K$. (Notice that this implies $\overline{\sigma} > \frac{1}{K}$.) Further, since $\overline{\sigma}$ solves the implicit relation (14), we have

$$\frac{(1 - \varepsilon)\overline{\sigma} p + \hat{\varepsilon}(1 - \overline{\sigma})\hat{p}}{\overline{\sigma} p + (1 - \overline{\sigma})\hat{p}} = \overline{\sigma}. \tag{15}$$

Multiplying (15) by the denominator of the left-hand side and rearranging, we get

$$Ap = B \frac{1 - p}{K - 1}, \tag{16}$$

where $A = \overline{\sigma}(1 - \varepsilon - \overline{\sigma})$ and $B = (1 - \overline{\sigma})(\overline{\sigma} - \varepsilon)$. B is obviously positive, so A must be positive as well. $A > 0$ implies $\overline{\sigma} < 1 - \varepsilon$, and thus $1 - (K - 1)\underline{\sigma} < 1 - (K - 1)\hat{\varepsilon}$, which shows that $\underline{\sigma} > \hat{\varepsilon}$.

Step 5. By (B2) the right-hand side of the implicit relation (14) incrases in p. Similarly, by (B3) it decreases in ε since $\overline{\sigma} > 1/K$ as established in Step 4. By (B1), the right-hand side of (14) is also concave in σ^k, so that as a function of σ^k, $\mathcal{B}^k(\sigma, k)$ crosses the 45°-line from above. Thus, in response to an increase in p, $\overline{\sigma}$ must increase for (14) to continue to hold. Likewise, in response to an increase in ε, $\overline{\sigma}$ must decrease for it to continue to hold. Therefore $\partial\overline{\sigma}/\partial p > 0$ and $\partial\overline{\sigma}/\partial\varepsilon < 0$, which is equivalent to $\partial\underline{\sigma}/\partial p < 0$ and $\partial\underline{\sigma}/\partial\varepsilon > 0$. □

Proof of Proposition 2

Let $\sigma^* = i(k')\overline{\sigma} + (\mathbf{1} - i(k'))\underline{\sigma}$ for some $k' \in K$. That is, σ^* is a belief which puts maximal weight $\overline{\sigma}$ on the state k' and minimal weight k' on all other states. Now take some state $k'' \neq k'$. Suppose that signal $m = k''$ is observed and that

$$\mathcal{B}^{k'}(\sigma^*, k'') < \mathcal{B}^{k''}(\sigma^*, k''), \tag{17}$$

so $k'' = \arg\max_\kappa \mathcal{B}^\kappa(\sigma^*, k'')$. Since $\mathcal{B}^k(\sigma, k)$ is increasing in σ^k for all k, it follows that if (17) holds, then $k'' = \arg\max_\kappa \mathcal{B}^\kappa(\sigma, k'')$ for all σ. Therefore $d_t = k''$ if and only if $m_{t-1} = k''$. Since the argument can be made for any pair k', k'', it follows that if (17) holds for some k', k'' then it holds for all such pairs, and hence $d_t = m_{t-1}$. Now consider linear updating rules: If $\alpha > \frac{1}{2}$, then $\mathcal{L}^k(\tilde{\sigma}_t, k) > \frac{1}{2}$ and $\mathcal{L}^k(\tilde{\sigma}_t, k') < \frac{1}{2}$ if $k \neq k'$. Hence $\tilde{d}_t = k$ if and only if $m_{t-1} = k$. Thus, we conclude that when (17) holds and $\alpha > \frac{1}{2}$, decisions based on $\mathcal{B}$ and decisions based on $\mathcal{L}_\alpha$ will be the same in every period, and therefore $D(\alpha) = 0$ and $C(\alpha) = C^*(\alpha) = 1$.

We now verify that if $\varepsilon > \varepsilon^*$, (17) holds. The strategy is as follows: It can be shown that if $\varepsilon = \varepsilon^*$, $\overline{\sigma} = p$ (this is done in step 1 below). This implies that $\underline{\sigma} = \frac{1-p}{K-1} = \hat{p}$, and since $\underline{\sigma}$ increases in ε by Lemma 1, $\varepsilon > \varepsilon^*$ implies $\underline{\sigma} > \hat{p}$ and $\overline{\sigma} < p$. It can also be shown (step 2 below) that when σ^* is constructed as above, using $\overline{\sigma} = p$ and $\underline{\sigma} = \hat{p}$, (17) becomes an equality for all values of ϵ. Since $\mathcal{B}^k(\sigma, m)$ increases in σ^k, it follows that if $\varepsilon > \varepsilon^*$ (and thus $\overline{\sigma} > p$ and $\underline{\sigma} < \hat{p}$) then (17) holds with the desired strict inequality.

These steps establish the result for $\varepsilon > \varepsilon^*$ and $\alpha > \frac{1}{2}$. To show that the Lemma also holds when $\varepsilon = \varepsilon^*$ and/or $\alpha = \frac{1}{2}$, note that when $\varepsilon = \varepsilon^*$ then (17) is an equality. This means that for those (and only those) beliefs that put maximal weight ($\overline{\sigma} = p$) on some state k and minimal weight ($\underline{\sigma} = 1 - p$) on all others, the Bayesian decision maker will be indifferent between between actions k and k' after observing signal $k' \neq k$. Likewise, a non-Bayesian decision maker who uses $\alpha = \frac{1}{2}$ will be indifferent between actions k and k' if and only if he holds beliefs that put maximal weight (one) on some state k and minimal weight (zero) on all others and observes signal $k' \neq k$. But these limit beliefs can be reached only after observing an infinite sequence of identical signals k. Therefore, these cases do not affect the conclusion that $D(\alpha) = 0$, and thus $C(\alpha) = C^*(\alpha) = 1$.

We now give the detailed algebra to establish steps 1 and 2.

Step 1. Lemma 1 shows that $\overline{\sigma}$ is unique; to show that $\overline{\sigma} = p$ when

$$\varepsilon = \varepsilon^* = \frac{p(1-p)(K-1)}{p(K-2)+1}$$

it hence suffices to verify that $\mathcal{B}^k(\sigma, k) = p$ for σ with $\sigma^k = p$. Using the definition of $\hat{p}$, the condition $\mathcal{B}^k(\sigma, k) = p$ can be written as

$$p\left(p^2 + \frac{1}{K-1}(1-p)^2\right) = (1-\varepsilon^*)p^2 + \hat{\varepsilon^*}\frac{1}{K-1}(1-p)^2. \tag{18}$$

Note that

$$1 - \varepsilon^* = \frac{p^2(K-1) + (1-p)}{p(K-2)+1} \quad \text{and} \quad \hat{\varepsilon^*} = \frac{\varepsilon^*}{K-1} = \frac{p(1-p)}{p(K-2)+1}.$$

Using these values in (18) and multiplying both sides by $p(K-2)+1$ we get

$$(p(K-2)+1)p\left(p^2 + \frac{1}{K-1}(1-p)^2\right) = \left(p^2(K-1) + (1-p)\right)p^2 + p\frac{1}{K-1}(1-p)^3.$$

Eliminating common terms from both sides and rearranging, we obtain

$$\frac{1}{K-1}\left[(1-p)^2 - (1-p)^3\right] = \left[p^3 + p(1-p) - p^2\right] - \frac{K-2}{K-1}p(1-p)^2.$$

Since $\left[(1-p)^2 - (1-p)^3\right] = \left[p^3 + p(1-p) - p^2\right] = p(1-p)^2$, (18) can be further simplified to

$$\frac{1}{K-1} = 1 - \frac{K-2}{K-1},$$

which is obviously true.

Step 2. Without loss of generality let $k' = 1$ and $k'' = 2$, and put $\sigma^* = (p, \hat{p}, \ldots, \hat{p})$. We then get the following updated beliefs after observing signal $k'' = 2$:

$$\mathcal{B}^1(\sigma^*, 2) = \frac{(1-\varepsilon)p\hat{p} + \hat{\varepsilon}(p\hat{p} + (1-p-\hat{p})\hat{p})}{p\hat{p} + (1-\hat{p})\hat{p}}$$

and

$$\mathcal{B}^2(\sigma^*, 2) = \frac{(1-\varepsilon)\hat{p}p + \hat{\varepsilon}(1-\hat{p})\hat{p}}{\hat{p}p + (1-\hat{p})\hat{p}}.$$

Using the definition of $\hat{p}$ we have $1-p-\hat{p} = (K-2)\hat{p}$. For $\mathcal{B}^1(\sigma^*, 2) = \mathcal{B}^2(\sigma^*, 2)$ we hence need

$$\frac{(1-\varepsilon)p\hat{p} + \hat{\varepsilon}(p\hat{p} + (K-2)\hat{p}^2)}{p\hat{p} + (1-\hat{p})\hat{p}} = \frac{(1-\varepsilon)\hat{p}p + \hat{\varepsilon}(1-\hat{p})\hat{p}}{p\hat{p} + (1-\hat{p})\hat{p}},$$

which, by eliminating common terms from both sides, can be simplified to

$$p + (K-2)\hat{p} = 1 - \hat{p}.$$

This condition is true since $p = 1 - (K-1)\hat{p}$. □

Appendix B: Continuous Actions

The basic model presented in this paper assumed the discreteness of the state, signal, and action space. As we argued, in such a discrete environment it may not matter "how far off" an agent's beliefs are from the Bayesian probability distribution, as long as the highest weight is put on the correct state. If the agent has a richer set of actions available, however, the precise magnitude of these probabilities may impact an agents' performace more directly. In this appendix we provide an example of how the model can be extended to a continuous action space. The example leads to similar results as the basic model, although the linear learner's performance is not quite as striking.

The extended model

Consider an environment where the structure of states and signals is exactly the same as in the orignal model of Section 2. Assume, however, that instead of taking an action $d_t \in \{1, \ldots, K\}$ in each period, the agent must now choose an *allocation* $d_t \in \Delta_K$, where Δ_K is the unit simplex in $\mathbb{R}^K$. The payoff the agent obtains in period t is

$$\pi_t(d_t, m_t) = \sum_{k=1}^{K} d_t^k u_t(k, m_t),$$

where d_t^k is the kth component of the vector d^k. (Note that $\pi_t(d_t, m_t)$ is a random variable, since $u_t(k, m_t)$ is random.) We can think of the original model as a special case where the agent is restricted to choosing among the vertices of Δ_K. An interpretation of this extension is that the agent has one unit of time available in each period, which must be devoted to K different activities. One activity is the "correct" one, in that it yields a probabilistically higher payoff than the others, but the agent is not sure about which activity this is.

If the decision maker were risk neutral, he would always set d_t to have all weight on the activity that has the greatest probability in the agent's vector of beliefs. In this case, the extended model would become the same again as our basic model. We thus assume that the decision maker is risk-averse; in particular we assume he maximizes the expectation of $\ln(\pi_t(d_t, m_t))$. If σ_t denotes the decision maker's belief, then it is straightforward to show that he maximizes his expected utility by setting

$$d_t^k = \sigma_t^k p + (1 - \sigma_t^k)\hat{p}, \tag{19}$$

i.e. he allocates resources proportional to the belief that a given alternative yields the high payoff.[12]

We will allow the linear learner to use a rule

$$\mathcal{M}_{\alpha,\gamma}(\sigma, m) = \gamma \mathbf{1} + (1 - K\gamma)\mathcal{L}_\alpha(\sigma, m),$$

[12] Concave utility is clearly needed in order to make the agent choose actions in the interior of Δ_K. However, the question may arise why the agent should be risk-averse, or alternatively why he has a consumption smoothing objective, if we are ultimately interested in long-run performance measures of different learning procedures. We do not address this question here: The objective of this section is merely to show that models with continuous actions can lead to similar results as the basic model.

where $\gamma \in [0, 1/K)$ and $\alpha \in (0, 1)$. This is still a linear rule, with $\mathcal{M}_{\alpha,0}$ corresponding to the exponential smoothing rule $\mathcal{L}_\alpha$. The difference is that if γ is positive, linear beliefs are "compressed" around the barycenter of Δ_K. In the basic model, such a specification is unnecessary: For a given value α, all rules $\mathcal{M}_{\alpha,\gamma}$ and $\mathcal{L}_\alpha$ are equivalent since the transformation from $\mathcal{L}_\alpha$ to $\mathcal{M}_{\alpha,\gamma}$ is order-preserving. With a continuous choice set and concave utility, however, the choice in (19) depends not only on which element of σ is the highest, but also on the magnitude of beliefs. Thus, we now consider the family of linear updating rules $\{\mathcal{M}_{\alpha,\gamma}\}_{\alpha\in(0,1),\gamma\in(0,1/K)}$.

Performance measures

We also have to make some changes in how we define performance measures for the extended model. The measure D as defined in Section 3 would always be 1 and ought to be replaced by the decision error variance.[13] Furthermore, it is unclear how informative the measures C and C^* are in the present context, as they are defined in terms of consumption but the decision maker maximizes expected utility of consumption. It may seem natural to define these measures as "utility ratios" instead. Since risk preferences are preserved under positive affine transformations of the utility function, C and C^* should be invariant to such changes. The measure C, however, is clearly not (adding a constant to the utility function changes C), and thus becomes meaningless. In the measure C^*, however, one can replace u_t with $\ln(\pi_t)$, since affine transformations of the utility function will be cancelled from the ratio of differences. Let us therefore define

$$v_B = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \ln(\pi_t(d_t, m_t)), \quad v_L(\alpha, \gamma) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \ln(\pi_t(\tilde{d}_t, m_t)), \quad v_0 = \ln \frac{1}{K},$$

and let

$$C^*(\alpha, \gamma) = \frac{v_L(\alpha, \gamma) - v_0}{v_B - v_0} \leq 1.$$

Notice that as the outside utility level w_0 we have taken the utility of an agent who does not learn and maintains diffuse beliefs that put equal weight on all K states, and thus allocates the same share to all K activities.

Simulation results

We only consider the two-state case. For the same values for ε and p as used in Section 5, we simulated $200,000$ periods of the extended model. For each of these six simulations, Figure 5 contains a graphical depiction of the measure $C^*(\alpha, \gamma)$ as a contour plot in (α, γ)-space. The three curves contained in each graph correspond to $C^*(\alpha, \gamma) = 0.7$, 0.8, and

[13] We define the error variance as

$$D(\alpha, \gamma) = \frac{1}{2} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} (\tilde{d}_t^k - d_t^k)^2,$$

where $\tilde{d}$ means the linear learner's allocation decision and d the Bayesian's decision. The fraction $\frac{1}{2}$ normalizes the maximal Euclidean distance between two allocations to unity. With $K = 2$, for example the decision to allocate all resources to activity A has distance one from the decision to allocate all resources to activity B.

0.9. The highest point in the plot is represented as a white dot in the center of the contour plot.

As in the basic model, a decision maker who employs a linear rule can do fairly well: The value of C^* that can be achieved with a well-parameterized linear rule is more than 90% in all six cases, and more than 98% in the three cases for $\varepsilon = 0.15$. Furthermore, for $\varepsilon = 0.15$ it appears that, just like in the basic model, a rather large set of linear rules yields $C^*(\alpha, \gamma) \geq 0.9$. The numbers reported here are not quite as striking as before, but the reason for this is simple: The requirements for a "good" linear rule are more demanding in the current setup as when there are only two actions. Not only do we have to make sure that the linear beliefs put a majority of the weight on the same state as the Bayesian beliefs most of the time, but also that we track the magnitude of Bayesian beliefs as closely as possible.

Compared to the discrete setup, the optimal learning speed α for the linear rule is higher in all six cases. The optimal value for the parameter γ is zero for $\varepsilon = 0.05$, so that the linear rue is an exponential smoothing rule as used before. When $\varepsilon = 0.15$, however, it is optimal to set γ to a value between 0.16 and 0.22, so that linear agent's beliefs will be more compressed around $\frac{1}{2}$. Why a higher value of ε leads to such a compression of the optimal linear rule can be seen from Lemma 1. In the correct Bayesian beliefs, which the linear rule must mimic as closely as possible, a minimum weight $\underline{\sigma}$ is put on each state, which increases in ε. The exponential smoothing rule, on the other hand, puts no such minimum weight on each state. As a result, beliefs that are formed by a linear rule with $\gamma = 0$ can move closer to the corners of the belief space than Bayesian beliefs. If ε is small, this does not matter much. If ε gets larger, however, it becomes important to put a similar bound on linear beliefs, and increasing the value of γ does just that.

Finally, how large are the errors in the linear learner's decisions? In the simulations, the minimized value of $D(\alpha, \gamma)$ ranged from 0.0018 (for $(\varepsilon = 0.05, p = 0.85)$ to 4.29×10^{-6} (for $(\varepsilon = 0.15, p = 0.65)$, and the location of the minimum was also the maximer of $C^*(\alpha, \gamma)$. Thus, the standard error of the optimal linear learner's decision is between 0.0425 and 0.0021, with the maximal feasible distance between two allocation decisions being unity.

References

[1] Akerlof, G. and J. Yellen (1985): "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?" *American Economic Review*, **75**, 708–720.

[2] Balvers, R. and T. Cosimano (1990): "Actively Learning about Demand and the Dynamics of Price Adjustment," *Economic Journal*, **100**, 882–898.

[3] Balvers, R. and T. Cosimano (1993): "Periodic Learning about a Hidden State Variable," *Journal of Economic Dynamics and Control*, **17**, 805–828.

[4] Beck, G. and V. Wieland (2002): "Learning and Control in a Changing Environment," *Journal of Economic Dynamics and Control*, **26**, 1359–1377.

[5] Ben-Porath, E., E. Dekel, and A. Rustichini (1993): "On the Relationship between Mutation Rates and Growth Rates in a Changing Environment," *Games and Economic Behavior*, **5**, 576–603.

[6] Bergstrom, C. amd P. Godfrey-Smith (1998): "On the Evolution of Behavioral Heterogeneity in Individuals and Populations," *Biology and Philosophy*, **13**, 205–231.

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

[8] Brock, W. and C. Hommes (1997): "A Rational Route to Randomness," *Econometrica*, **65**, 1059–1095.

[9] Bush, R. and F. Mosteller (1955): *Stochastic Models for Learning*, Wiley, New York.

[10] Erdem, T. and M. Keane (1996): "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, **15**, 1–20.

[11] Gelman, A., J. Carlin, H. Stern, and D. Rubin (1995): *Bayesian Data Analysis*, Chapman & Hall Texts in Statistical Science.

[12] Guadagni, P. and J. Little (1983): "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, **2**, 203–238.

[13] Hommes, C. and G. Sorger (1998): "Consistent Expectations Equilibria," *Macroeconomic Dynamics*, **2**, 287–321.

[14] Hommes, C. (2005): "Heterogeneous Agents Models in Economics and Finance," forthcoming in: K. Judd and L. Tesfatsion: *Handbook of Computational Economics II: Agent-Based Computational Economics*.

[15] Kahneman, D. and A. Tversky (1982): *Intuitive Prediction: Biases and Corrective Procedures*, in: D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgement under Uncertainty: Heuristics and Biases*, Cambridge University Press, London.

[16] Kalman, R. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME-Journal of Basic Engineering*, **82**, 35–45.

[17] Keller, G. and S. Rady (1999): "Optimal Learning in a Changing Environment," *Review of Economic Studies*, **66**, 475–508.

[18] Kiefer, N. (1989): "A Dynamic Model of Optimal Learning with Obsolence of Information," CAE Working Paper 89-14, Cornell University.

[19] LeBaron, B. (2005): "Agent-Based Computational Finance," forthcoming in: K. Judd and L. Tesfatsion: *Handbook of Computational Economics II: Agent-Based Computational Economics*.

[20] Luce, D. and H. Raiffa (1957): *Games and Decisions*, Dover Publications, New York.

[21] Moscarini, G., M. Ottaviani, and L. Smith (1998): "Social Learning in a Changing World," *Economic Theory*, **11**, 657–665.

[22] Muth, J. (1960): "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association*, **55**, 299–306.

[23] Nyarko, Y. and L. Olson (1991): "Optimal Growth with Unobservable Resources and Learning," *Journal of Economic Behavior and Organization*, **29**, 465–491.

[24] Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," Journal of Economic Theory, **9**, 185–202.

[25] Rustichini, A. and A. Wolinsky (1995): "Learning about Variable Demand in the Long Run," *Journal of Economic Dynamics and Control*, **19**, 1283–1292.

[26] Simon, H. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, **69**, 99–118.

[27] Simon, H. (1959): "Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review*, **49**, 253–283.

[28] Sobel, J. (2000): "Economists' Models of Learning," *Journal of Economic Theory*, **94**, 241–261.

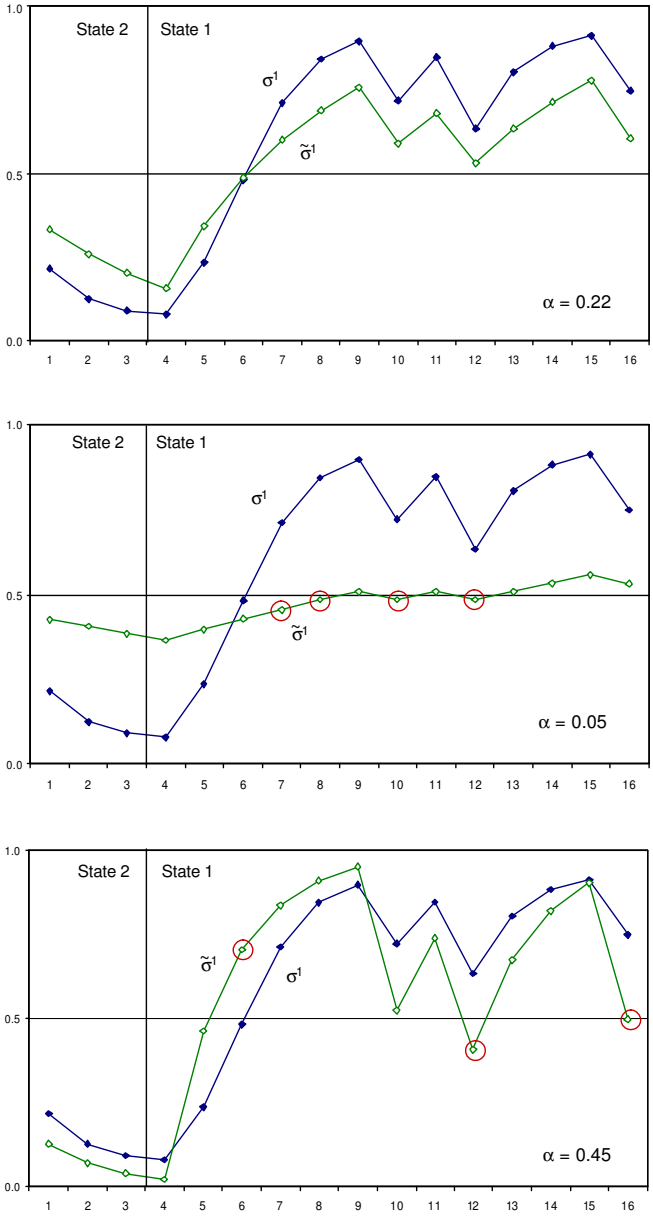

Figure 1: Optimal, Slow, and Fast Learning Speeds.

The top graph shows the evolution of Bayesian and linear beliefs when the linear learning speed is optimal. The middle and bottom graph show the corresponding beliefs for learning speeds which are too slow (α is small) or too fast (α is large).

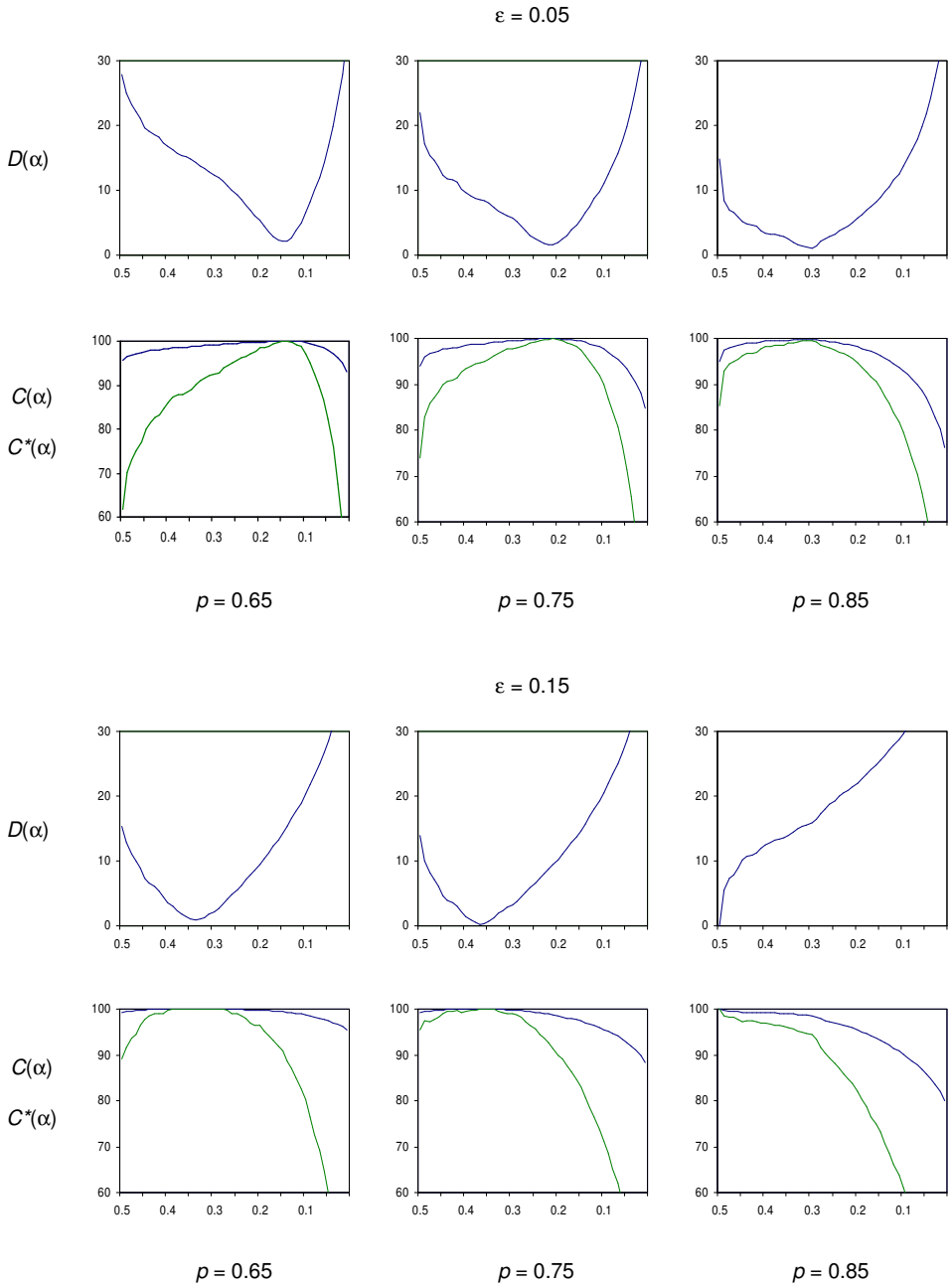

Figure 2: Relative performance measures: Two states

The graphs show the error probability $D(\alpha)$, and the relative consumption measures $C(\alpha)$ and $C^*(\alpha)$, for various values of the signal precision p. The top panel of graphs depicts a relatively stable environment ($\varepsilon = 0.05$), and the bottom panel a more unstable one ($\varepsilon = 0.15$).

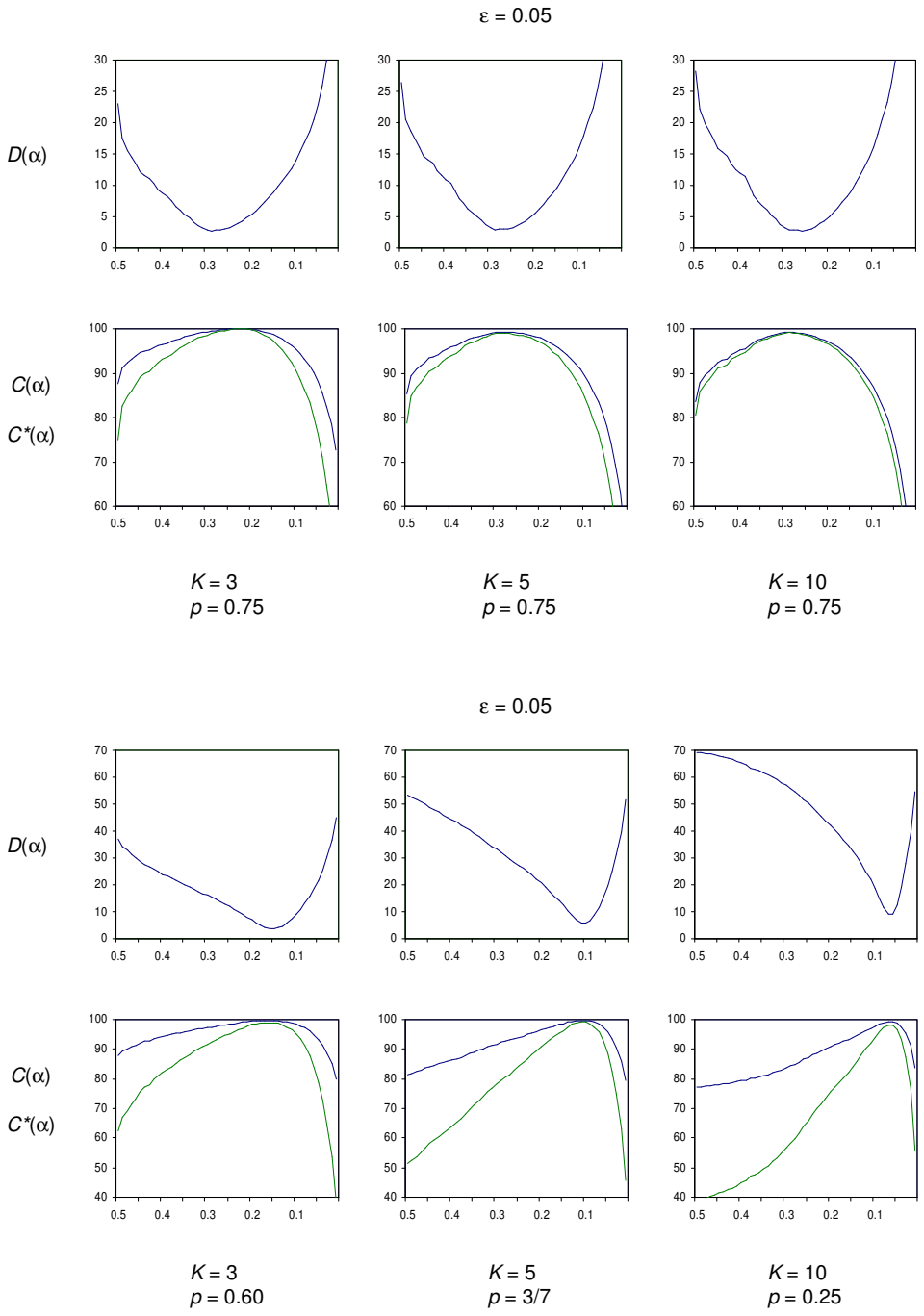

Figure 3: Relative performance measures: More than two states

The top panel shows the measures $D(\alpha)$, $C(\alpha)$, and $C^*(\alpha)$ when there are 3, 5, or 10 possible states, and $\varepsilon = 0.05$, $p = 0.75$ remains fixed. The bottom panel depicts the same measures, but p is adjusted so the $p/\hat{p} = 3$ remains constant.

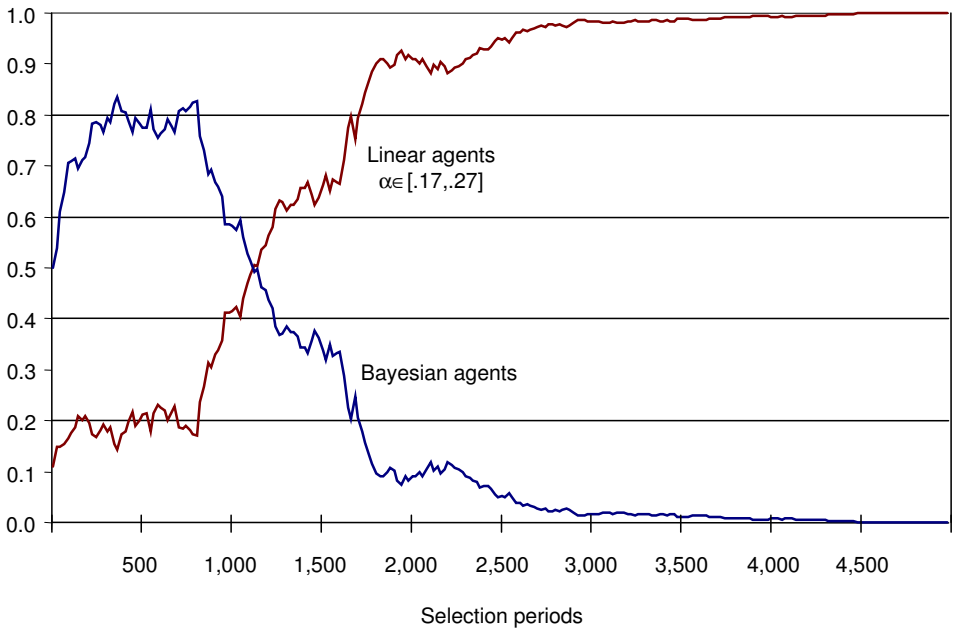

Figure 4: Evolution of learning rules

The graph shows the fraction of Bayesian agents and linear agents with $\alpha \in [0.17, 0.27]$ for the case $K = 2$, $\varepsilon = 0.05$, and $p = 0.75$. Bayes' rule has a cost of 0.0015 relative to the payoff that is obtained from selecting the correct action.

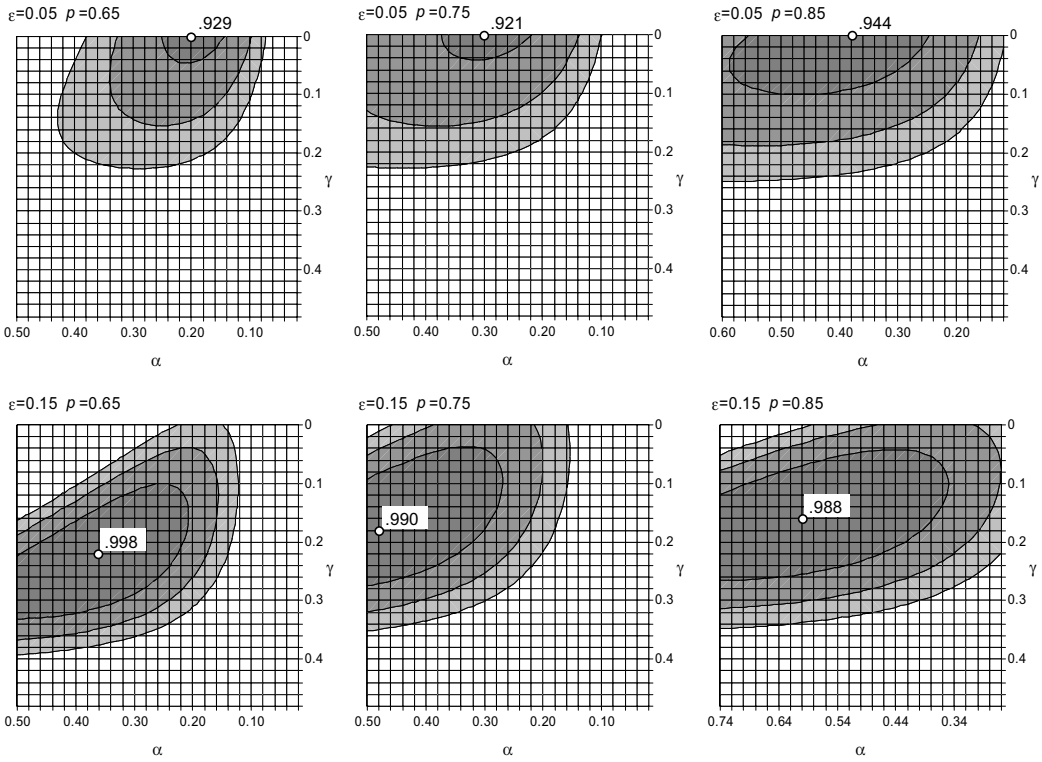

Figure 5: Performance of linear rules with continuous actions

The graphs show the measure $C^*(\alpha, \gamma)$ in the extended model with continuous actions and logarithmic utility. The graphs are for the case $K = 2$ and the same parameterizations as in Figure 2. The three contour lines correspond to $C^*(\alpha, \gamma) =$ 0.7, 0.8, 0.9 (outside to center).