LEARNING IN HIDDEN MARKOV MODELS WITH BOUNDED MEMORY

DANIEL MONTE[†] AND MAHER SAID[‡]

JUNE 23, 2010

ABSTRACT: This paper explores the role of memory in decision making in dynamic environments. We examine the inference problem faced by an agent with bounded memory who receives a sequence of signals from a hidden Markov model. We show that the optimal symmetric memory rule may be deterministic. This result contrasts sharply with Hellman and Cover (1970) and Wilson (2004) and solves, for the context of a hidden Markov model, an open question posed by Kalai and Solan (2003).

KEYWORDS: Bounded Memory, Hidden Markov Model, Randomization.

JEL CLASSIFICATION: C72, C73, D82, D83.

[†]DEPARTMENT OF ECONOMICS, SIMON FRASER UNIVERSITY, DANIEL_MONTE@SFU.CA

[‡]MICROSOFT RESEARCH NEW ENGLAND AND OLIN BUSINESS SCHOOL, WASHINGTON UNIVERSITY IN ST. LOUIS, SAID@WUSTL.EDU

We would like to acknowledge the valuable advice and input of Dirk Bergemann, Dino Gerardi, Johannes Hörner, Abraham Neyman, and Ben Polak, as well as seminar participants at Yale and Simon Fraser.

1. INTRODUCTION

This paper explores the role of memory in decision making in dynamic environments. In particular, we examine the inference problem faced by an agent with bounded memory who receives a sequence of noisy signals from a hidden Markov model: signals are informative about the state of the world, but this underlying state evolves in a Markovian fashion.

An example may help clarify the basic framework of our setting. Consider a firm that decides in each period whether to produce a particular product or not. Demand for the product may be high or low, but sales are only a stochastic function of demand. Thus, the firm's profits depend on both its decision and on the state of the world: if demand is high, then production yields (on average) a high payoff, whereas if demand is low, production yields (on average) a low payoff. If the state of the world is dynamic but not perfectly observable, how should the firm behave after a negative shock? What about two negative shocks? More generally, how many signals does the firm need to track in order to maximize its profits? We show that when the environment is unstable (but still persistent), only a single period of records is required.

We then study the optimal behavior in such an environment by a decision maker whose memory is exogenously constrained.[1] Similar settings appear in nonlinear filtering problems in finance and macroeconomics, and related phenomena may arise in repeated-game settings where players have bounded memory.[2] Understanding the behavior of an agent with cognitive limitations in hidden Markov models may shed light on the behavioral biases that are present in these environments.[3]

Our main result is to show that the optimal symmetric memory rule for a decision maker with bounded memory may involve only deterministic transition rules. This result is in stark contrast with much of the literature on bounded memory. The main result of Hellman and Cover (1970) and Cover and Hellman (1971) is that the optimal memory in a two-hypothesis testing problem (in a static world that does not switch over time) involves randomization at extremal memory states.[4] Moreover, it was later shown by Wilson (2004) that this optimal random transition rule can explain well-documented behavioral biases such as confirmatory bias, belief polarization, and overconfidence or underconfidence bias. We show that, in a world that changes over time, randomization is not optimal for an agent with only two memory states—even if the state of the world is highly persistent and change is unlikely. Thus, the behavioral biases displayed by a decision maker with bounded memory need not correspond to those described by Wilson.

Interestingly, the optimal symmetric memory rule in the hidden Markov problem does not approach the solution of the two-hypothesis testing problem as the probability of state switching goes to zero. The intuition behind this discontinuity at zero lies in the stationary distribution over the state of the world. In a two-hypothesis testing problem, the stationary distribution of the underlying state of the world is a degenerate distribution that places full mass on either one state or

[1] Other recent papers in a similar spirit include Miller and Rozen (2010); Mullainathan (2002); and Wilson (2004).

[2] See, among others, Cole and Kocherlakota (2005); Compte and Postlewaite (2009); Romero (2010); and Monte (2009).

[3] For broad overviews of related work on bounded rationality and behavioral biases, the curious reader may wish to consult Lipman (1995) or Rubinstein (1998), as well as the references therein.

[4] Kalai and Solan (2003) and Kocer (2009) also develop models of decision making with bounded memory in which optimal behavior requires randomization.

the other. In the hidden Markov model with non-zero transition probabilities, on the other hand, the stationary distribution is unique and independent of the initial state.[5] Thus, in our model, even in the presence of very noisy signals and transition probabilities arbitrarily close to zero, the decision maker will still prefer to switch actions in a deterministic way.

This result solves (for the context of a hidden Markov model) an open question raised by Kalai and Solan. They study the value of randomization when agents are boundedly rational and show that randomization may be optimal even in the most elementary environments when agents may be described by two-state automata (that is, when they are restricted to two memory states). They conclude by posing the following open problem: "When is optimality obtained by a random (rather than deterministic) automaton?" (Kalai and Solan, 2003, p. 263). We solve this open question for the case of a dynamic environment.

2. Model

We consider the following single-agent decision problem. Let $\Omega := \{H, L\}$ denote the set of states of the world, where H represents the "high" state and L represents the "low" state, and let $\rho_0 \in [0,1]$ be the decision maker's *ex ante* belief that the initial state of the world is H. In each period $t \in \mathbb{N}$, the decision maker must take an action $a_t \in A := \{h, l\}$, and her objective is to "match" the state of the world ω_t. In particular, taking the action a_t in state ω_t yields a positive payoff (normalized to one) with probability $\pi(a_t, \omega_t)$, and zero payoff with probability $1 - \pi(a_t, \omega_t)$, where

$$\pi(a, \omega) := \begin{cases} \gamma & \text{if } (a, \omega) \in \{(h, H), (l, L)\}, \\ 1 - \gamma & \text{otherwise.} \end{cases}$$

Thus, if the action matches the state, a payoff of one is received with probability γ; and if the action and state do not "match," then the probability of receiving a positive payoff is $1 - \gamma$. We assume that $\frac{1}{2} < \gamma < 1$, implying that receiving a positive payoff is an informative (but not perfectly so) signal of the underlying state of the world.

We make the additional assumption that the state of the world may change in each period.[6] In particular, we assume that this evolution follows a Markov process with

$$\Pr(\omega_{t+1} = \omega_t) = 1 - \alpha,$$

where $0 < \alpha < \frac{1}{2}$. The parameter α may be viewed as measuring the persistence (or, inversely, the instability) of this process: as α approaches 0, the state of the world is increasingly likely to remain the same from one period to the next, while as α approaches $\frac{1}{2}$, the process governing the state of the world approaches an *i.i.d.* flip of a fair coin in each period.

To summarize, the timing of the problem in each period $t \in \mathbb{N}$ is as follows:

- Nature draws a state of the world $\omega_t \in \Omega$, where $\Pr(\omega_1 = H) = \rho_0$ and, for all $t > 1$, $\Pr(\omega_t = \omega_{t-1}) = 1 - \alpha$.

[5] A similar phenomenon is demonstrated by Battaglini (2005) in a long-term principal-agent contracting model—contracting distortions disappear in the long run when an agent's type evolves according to a Markov process for all non-zero transition probabilities, but the distortions are constant and positive when types are perfectly persistent.

[6] This is the main contrast with the stationary models of, for instance, Hellman and Cover (1970) and Wilson (2004).

- The decision maker takes an action $a_t \in A$.
- A payoff $\pi_t \in \{0, 1\}$ is realized according to the distribution $\pi(a_t, \omega_t)$.
- The decision maker observes the payoff π_t, and we proceed to period $t + 1$.

We assume that the agent evaluates payoffs according to the limit of means criterion. In particular, the decision maker's expected utility can be written as

$$U = \mathbb{E}\left[\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \pi_t\right].$$

Note that if γ were equal to one (that is, if payoffs are perfectly informative about the state of the world), then the agent's payoff is precisely the long-run proportion in which his action is the same as the true state of the world. Payoffs are *not* perfectly informative and $\gamma < 1$, however; thus, letting $\sigma \in [0, 1]$ denote the long-run proportion of periods in which the "matching" action is taken, the agent's expected utility may be written as

$$U = \gamma\sigma + (1 - \gamma)(1 - \sigma).$$

It is helpful to think of the decision maker's payoffs π_t as signals about the underlying state of the world; in particular, we may classify action-payoff pairs as being either a "high" signal or "low" signal. To see why, consider any belief $\rho_t = \Pr(\omega_t = H)$, and notice that

$$\Pr(\omega_t = H|a_t = h, \pi_t = 1) = \Pr(\omega_t = H|a_t = l, \pi_t = 0) = \frac{\rho_t\gamma}{\rho_t\gamma + (1 - \rho_t)(1 - \gamma)};$$

thus, observing a payoff of 1 after taking action h provides exactly the same information as observing a payoff of 0 after taking action l. Moreover, observing either of these two action-payoff pairs is more likely when the true state is H than when it is L, as we have

$$\frac{\Pr(\pi_t = 1|a_t = h, \omega_t = H)}{\Pr(\pi_t = 1|a_t = h, \omega_t = L)} = \frac{\Pr(\pi_t = 0|a_t = l, \omega_t = H)}{\Pr(\pi_t = 0|a_t = l, \omega_t = L)} = \frac{\gamma}{1 - \gamma} > 1$$

and $\gamma > \frac{1}{2}$. Similarly, we have

$$\Pr(\omega_t = L|a_t = l, \pi_t = 1) = \Pr(\omega_t = L|a_t = h, \pi_t = 0) = \frac{(1 - \rho_t)\gamma}{(1 - \rho_t)\gamma + \rho_t(1 - \gamma)}$$

and

$$\frac{\Pr(\pi_t = 1|a_t = l, \omega_t = L)}{\Pr(\pi_t = 1|a_t = l, \omega_t = H)} = \frac{\Pr(\pi_t = 0|a_t = h, \omega_t = L)}{\Pr(\pi_t = 0|a_t = h, \omega_t = H)} = \frac{\gamma}{1 - \gamma} > 1.$$

Thus, we may partition the set of possible action-payoff pairs into a signal space $\mathcal{S} := \{H, L\}$, where $s = H$ represents the "high" action-payoff pairs $\{(h, 1), (l, 0)\}$ and $s = L$ represents the "low" action-payoff pairs $\{(h, 0), (l, 1)\}$.

Finally, notice that the action taken by the agent does not affect either state transitions or information generation—in the language of Kalai and Solan (2003), the decision maker faces a *noninteractive* Markov decision problem.[7] This lack of action-dependent externalities implies that, in each period t, the agent will simply take the action that maximizes his expected period-t payoff alone.

[7]Therefore, this decision problem is very different from a multi-armed bandit problem and departs from the optimal experimentation literature. See Kocer (2009) for a model of experimentation with bounded memory.

Since $\gamma > \frac{1}{2}$, his (myopic) action rule, as a function of his beliefs ρ_t that $\omega_t = H$, is given by

$$a_t^*(\rho_t) := \begin{cases} h & \text{if } \rho_t \geq \frac{1}{2}, \\ l & \text{if } \rho_t < \frac{1}{2}. \end{cases}$$

3. Minimal Memory for Unstable Environments

Intuitively, one would presume that memory is an important and valuable resource in a decision problem. As shown by Cover and Hellman (1971), the optimal payoff for a bounded memory agent in a static environment is strictly increasing in his memory size. And, as shown by Wilson (2004), a finite memory may induce several biases in information processing (which may be alleviated somewhat by increasing the number of memory states).

In a dynamic setting, however, we show that for some parameter ranges, the (not-too-distant) past becomes irrelevant, and the agent's optimal choice of action depends only on the previous period. Specifically, if the environment is sufficiently noisy or unstable, only a minimal memory (one bit, or, equivalently, two memory states) is required in order to achieve the same payoffs as a perfectly Bayesian decision maker.

As a benchmark, we first explore this decision problem from the perspective of a fully Bayesian agent who has no constraints on his memory or computational abilities. Recall that ρ_t denotes the agent's belief that the state of the world is H at the beginning of period t. Then beliefs ρ_{t+1}^s after a signal $s \in \mathcal{S}$, taking into account the possibility of state transitions between periods, are given by

$$\rho_{t+1}^H(\rho_t) = \frac{\rho_t\gamma(1-\alpha) + (1-\rho_t)(1-\gamma)\alpha}{\rho_t\gamma + (1-\rho_t)(1-\gamma)} \text{ and } \rho_{t+1}^L(\rho_t) = \frac{\rho_t(1-\gamma)(1-\alpha) + (1-\rho_t)\gamma\alpha}{\rho_t(1-\gamma) + (1-\rho_t)\gamma}.$$

Notice that $\rho_{t+1}^H(\rho) + \rho_{t+1}^L(1-\rho) = 1$ for all $\rho \in [0,1]$; thus, Bayesian belief revision is fully symmetric. Also, notice that $\rho_{t+1}^s(0) = \alpha$ and $\rho_{t+1}^s(1) = 1-\alpha$ for $s = H, L$—even if the agent is absolutely sure of the state of the world in some period t, there will be uncertainty in the following period about this state due to the underlying Markov process. Moreover, it is useful to note the following result:

Lemma 1 (Monotone belief revision)**.**
The decision maker's period-$(t+1)$ beliefs are strictly increasing in his period-t beliefs, regardless of the realized signal.

Proof. Notice that

$$\frac{\partial\rho_{t+1}^H(\rho)}{\partial\rho} = \frac{\gamma(1-\gamma)(1-2\alpha)}{(\rho\gamma + (1-\rho)(1-\gamma))^2} \text{ and } \frac{\partial\rho_{t+1}^L(\rho)}{\partial\rho} = \frac{\gamma(1-\gamma)(1-2\alpha)}{(\rho(1-\gamma) + (1-\rho)\gamma)^2}.$$

Since $0 < \alpha < \frac{1}{2} < \gamma < 1$ and $\rho \in [0,1]$, each of these two expressions must be strictly positive. □

With this in hand, it is straightforward to show that a Bayesian decision maker's beliefs will converge to a closed and bounded "absorbing" set. In particular, we can pin down the upper and lower bounds on long-run beliefs, as shown in the following result:

LEMMA 2 (Bayesian long-run belief bounds)**.**
Fix any $\epsilon > 0$. For any $0 < \alpha < \frac{1}{2} < \gamma < 1$, there exits a time $\bar{t}_\epsilon \in \mathbb{N}$ and a bound $\rho^* \in \left(\frac{1}{2}, 1\right)$ such that

$$\Pr\left(1 - \rho^* \leq \rho_t \leq \rho^*\right) > 1 - \epsilon \text{ for all } t > \bar{t}_\epsilon,$$

where ρ_t is the decision maker's belief at time t that the state of the world is H. Moreover, if $\rho_t \in [1-\rho^*, \rho^*]$ for any $t \in \mathbb{N}$, then $\rho_{t'} \in [1-\rho^*, \rho^*]$ for all $t' > t$.

PROOF. Note that the belief revision process has a "long-run upper bound" given by the fixed point of $\rho^H_{t+1}(\cdot)$—the point which solves

$$\rho = \frac{\rho\gamma(1-\alpha) + (1-\rho)(1-\gamma)\alpha}{\rho\gamma + (1-\rho)(1-\gamma)}.$$

This equation has two solutions, one of which is always negative (and hence irrelevant), and one which is always less than one. Denoting this latter solution by ρ^*, we have

$$\rho^* = \frac{(2\gamma - 1) - \alpha + \sqrt{\alpha^2 + (2\gamma-1)^2(1-2\alpha)}}{2(2\gamma - 1)}. \tag{1}$$

It is straightforward to verify that, for all $0 < \alpha < \frac{1}{2} < \gamma < 1$, we have

$$\frac{1}{2} < \rho^* < 1.$$

Moreover, Lemma 1 implies that $\rho^H_{t+1}(\rho) > \rho^*$ if, and only if, $\rho > \rho^*$; thus, a period-t belief ρ_t can only be larger than this upper bound if the initial belief ρ_0 is greater than ρ^* and sufficiently few L signals have been observed (which occurs with diminishing probability as t grows).

Similarly, the belief revision process has a "long-run lower bound" given by the fixed point of $\rho^L_{t+1}(\cdot)$—the point which solves

$$\rho = \frac{\rho(1-\gamma)(1-\alpha) + (1-\rho)\gamma\alpha}{\rho(1-\gamma) + (1-\rho)\gamma}.$$

This equation has two solutions, one of which is always greater than one (and hence irrelevant), and one which is always less than one. Denoting this latter solution by ρ_*, we have

$$\rho_* = \frac{(2\gamma - 1) + \alpha - \sqrt{\alpha^2 + (2\gamma-1)^2(1-2\alpha)}}{2(2\gamma - 1)} = 1 - \rho^*. \tag{2}$$

Moreover, Lemma 1 implies that $\rho^L_{t+1}(\rho) < \rho_*$ if, and only if, $\rho < \rho_*$; thus, a period-t belief ρ_t can only be smaller than this lower bound if the initial belief ρ_0 is less than $1 - \rho^*$ and sufficiently few H signals have been observed (which occurs with diminishing probability as t grows).

Finally, let $\bar{k} \in \mathbb{N}$ be such that

$$[\rho^L_{t+1}]^{\bar{k}}(1) < \rho^*;$$

this is the number of L signals sufficient for beliefs to "dip below" ρ^*, regardless of how high the initial belief is. (Equivalently, it is the number of H signals sufficient for beliefs to go above the boundary $1 - \rho^*$, regardless of how low initial beliefs may be.) As we are in a world with noisy signals of the underlying state, it is clear that $\bar{t}_\epsilon \in \mathbb{N}$ can be chosen such that the probability of

observing at least $\bar{k}$ low signals in the first $\bar{t}_\epsilon$ periods is at least $1-\epsilon$. Since each additional period yields another opportunity for a low signals to arrive, we have our desired result. □

With these preliminary results in hand, we can go on to show that only a single bit of memory is required for optimal behavior in certain circumstances—specifically, when the environment is sufficiently unstable or noisy (in a sense we will make precise shortly). This result relies on the fact that, in such environments, Bayesian beliefs are sufficiently responsive to new signals that only the most recent signal is a sufficient statistic determining the optimal action.

THEOREM 1 (Minimal memory for unstable environments)**.**
If α and γ are such that $\alpha \geq \gamma(1-\gamma)$, then a decision maker with only two memory states has the same expected payoff as an unconstrained Bayesian decision maker.

PROOF. Note that, for all $0 < \alpha < \frac{1}{2} < \gamma < 1$,

$$\rho^L_{t+1}(\gamma) = \frac{1}{2}.$$

Moreover, when $\alpha \geq \gamma(1-\gamma)$, it is easy to verify (using Equation (1)) that the upper bound of long-run beliefs ρ^* described by Lemma 2 is no larger than γ; that is,

$$\rho^H_{t+1}(\rho^*) = \rho^* \leq \gamma,$$

with equality only when $\alpha = \gamma(1-\gamma)$. Since belief revision is monotone increasing in current beliefs (as shown in Lemma 1), an application of Lemma 2 implies that, for all $\rho_t \in \left[\frac{1}{2}, \gamma\right]$,

$$1-\gamma \leq \rho^L_{t+1}(\rho_t) \leq \frac{1}{2} \leq \rho^H_{t+1}(\rho_t) \leq \gamma.$$

Thus, if $\alpha \geq \gamma(1-\gamma)$ and $1/2 \leq \rho_t \leq \gamma$, a single L signal is sufficient to convince a standard Bayesian decision maker who is following the optimal action rule a^* to switch from taking action h to taking action l.

It is straightforward to use the symmetry of belief revision to see that a similar property holds when a Bayesian decision maker believes that state L is more likely than state H. In particular, if $\alpha \geq \gamma(1-\gamma)$ and $1-\gamma \leq \rho_t \leq 1/2$, a single H signal is sufficient to convince a Bayesian agent who is following the optimal action rule a^* to switch from taking action l to taking action h.

Thus, when $\alpha \geq \gamma(1-\gamma)$ and beliefs at some time $t^* \in \mathbb{N}$ are such that $\rho_t \in [1-\gamma, \gamma]$, the signal in period $t \geq t^*$ is a sufficient statistic for a Bayesian agent's decision in period $t+1$. Since Lemma 2 implies that $t^* < \infty$ with probability one, this implies that the long-run optimal payoff (under the limit of means criterion) of a Bayesian decision maker is exactly equal to that generated by a two-state automaton that simply chooses the action that matches the previous signal. □

Notice that as γ increases and approaches 1 (that is, as signals become more informative about the true state of the world), the set of values of α such that the conditions of Theorem 1 hold increases. Thus, when signals become more and more informative, a restriction to only two memory states does not harm a decision maker. Thus, memory is most valuable when the decision problem is noisy but not too unstable. Therefore, in the following section, we investigate the more interesting cases where $\alpha < \gamma(1-\gamma)$ and the bound on memory is binding.

4. Optimal Two-State Memory

In this section we show that the optimal symmetric two-state memory is deterministic. Consider a decision maker with only two memory states $\mathcal{M} = \{1, 2\}$. His transition rule is a function $\varphi : \mathcal{M} \times \mathcal{S} \to \Delta\mathcal{M}$, where $\varphi(m, s)$ is the probability distribution governing transitions after observing signal $s \in \mathcal{S}$ while in state $m \in \mathcal{M}$. For notational convenience, we will use $\varphi^s_{m,m'}$ to denote $\varphi(m, s)(m')$. In addition, the agent's action rule is a function $a : \mathcal{M} \to A$. Note that the restriction to deterministic action rules is without loss of generality since the decision problem is noninteractive—actions affect neither state transitions nor information generation.

It is useful to note that, when the decision maker takes the same action in both memory states (that is, when $a(1) = a(2)$), his expected payoff is equal to $\frac{1}{2}$. This is because the long-run distribution of the underlying state of the world puts equal mass on both states. Hence, the action taken will be correct half the time, and incorrect half the time, implying that the expected payoff is

$$\frac{1}{2}\gamma + \frac{1}{2}(1 - \gamma) = \frac{1}{2}. \tag{3}$$

Moreover, this is precisely the maximal payoff possible from using a single-state memory or using a memory transition rule with an absorbing memory state (for instance, one in which $\varphi^H_{1,2} = \varphi^L_{1,2} = 0$). With this benchmark in mind, we will restrict attention to memory rules in which the action taken differs across memory states and where the memory states communicate (and we will show that this is, in fact, optimal). Thus, without loss of generality, we will assume that $a(1) = l$ and $a(2) = h$ and that the memory is irreducible.

Notice that the system—both the decision maker's memory and the state of the world—evolve according to a Markov process of an "extended" state space $\widehat{\Omega} := \mathcal{M} \times \Omega$. The combination of state transitions and memory transitions may be summarized by the Markov transition matrix $T = [\tau_{i,j}]$, where $\tau_{i,j}$ is the probability of moving from state $i \in \widehat{\Omega}$ to state $j \in \widehat{\Omega}$, given the memory transition rule φ as well as the parameters α and γ. This stochastic process will induce a stationary distribution $\mu \in \Delta\widehat{\Omega}$, where μ_i denotes the mass on state $i \in \widehat{\Omega}$. Thus, given the action rule in which the agent chooses l in state 1 and h in state 2, the decision maker's problem is to

$$\max_{\varphi} \left\{ \gamma(\mu_{(1,L)} + \mu_{(2,H)}) + (1 - \gamma)(\mu_{(1,H)} + \mu_{(2,L)}) \right\}, \tag{4}$$

subject to the constraints that (1) T is determined by φ; and (2) μ is the (endogenously determined) steady state of the process T on $\widehat{\Omega}$. We will proceed by showing that any memory rule that is irreducible generates a distribution over $\widehat{\Omega}$ that must satisfy certain constraints. We will then show the best possible distribution satisfying these constraints is generated by a deterministic memory transition function. Due to the underlying symmetry of the problem, we will focus on symmetric memory systems—those memory rules for which the transition function satisfies $\varphi^s_{m,m'} = \varphi^{s'}_{m',m}$ and $\varphi^s_{m,m} = \varphi^{s'}_{m',m'}$ for memory states $m \neq m'$ and signals $s \neq s'$. It should be clear that such memory rules induce a symmetric stationary distribution μ with $\mu_{(1,L)} = \mu_{(2,H)}$ and $\mu_{(1,H)} = \mu_{(2,L)}$.

We will denote by λ the relative likelihood of observing $s_{t+1} = H$ given period-t states $\omega_t = H$ and $\omega_t = L$ (or, by symmetry, the relative likelihood of observing $s_{t+1} = L$ given period-t states

$\omega_t = L$ and $\omega_t = H$). Formally, we have

$$\lambda := \frac{\Pr(s_{t+1} = H|\omega_t = H)}{\Pr(s_{t+1} = H|\omega_t = L)} = \frac{\Pr(s_{t+1} = L|\omega_t = L)}{\Pr(s_{t+1} = L|\omega_t = H)} = \frac{(1-\alpha)\gamma + \alpha(1-\gamma)}{(1-\alpha)(1-\gamma) + \alpha\gamma} \tag{5}$$

Note that, since signals are informative and states are persistent, we have $\lambda > 1$. This likelihood ratio will be useful to us in establishing bounds on the behavior of memory transition functions, and in particular on the ratios

$$\delta_1 := \frac{\tau_{(1,H),(1,L)} + \tau_{(1,H),(2,H)}}{\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}} \text{ and } \delta_2 := \frac{\tau_{(2,L),(1,L)} + \tau_{(2,L),(2,H)}}{\tau_{(2,H),(1,H)} + \tau_{(2,H),(2,L)}}. \tag{6}$$

Recall that we are considering memory systems with $a(1) = l$ and $a(2) = h$. Then δ_1 is the ratio between two rates: first, the rate at which the decision maker switches from incorrectly taking action l to correctly matching the underlying state; and second, the rate at which the decision maker switches from correctly taking action l to incorrectly mismatching the underlying state. Thus, δ_1 is the ratio of "correct" departures from the mismatched memory-world state $(1, H)$ to "incorrect" departures from the matched memory-world state $(1, L)$. Similarly, δ_2 is the ratio of "good" departures from the mismatched memory-world state $(2, L)$ to "bad" departures from the properly matched memory-world state $(2, H)$. We show that δ_1 and δ_2 are bounded above by λ.

LEMMA 3 (Bounds on δ_1 and δ_2)**.**
For any irreducible two-state memory, δ_1 and δ_2 are bounded above by λ:

$$\delta_1 \leq \lambda \text{ and } \delta_2 \leq \lambda.$$

PROOF. Consider first the relation above for δ_1. We have

$$\delta_1 := \frac{\tau_{(1,H),(1,L)} + \tau_{(1,H),(2,H)}}{\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}} = \frac{\alpha\left(\gamma\varphi^H_{1,1} + (1-\gamma)\varphi^L_{1,1}\right) + (1-\alpha)\left(\gamma\varphi^H_{1,2} + (1-\gamma)\varphi^L_{1,2}\right)}{\alpha\left(\gamma\varphi^L_{1,1} + (1-\gamma)\varphi^H_{1,1}\right) + (1-\alpha)\left(\gamma\varphi^L_{1,2} + (1-\gamma)\varphi^H_{1,2}\right)}.$$

Therefore,

$$\frac{\partial\delta_1}{\partial\varphi^H_{1,1}} = -\frac{(2\gamma-1)(1-2\alpha)\left(\alpha\varphi^L_{1,1} + (1-\alpha)(1-\varphi^L_{1,1})\right)}{\left(\alpha\left(\gamma\varphi^L_{1,1} + (1-\gamma)\varphi^H_{1,1}\right) + (1-\alpha)\left(\gamma\varphi^L_{1,2} + (1-\gamma)\varphi^H_{1,2}\right)\right)^2}.$$

Clearly, the denominator is positive. As for the numerator, the first term is positive (since $\gamma > 1/2$), as are the second (since $\alpha < 1/2$) and third (since both $\varphi^L_{1,1}, \alpha \in [0,1]$). Thus, we have $\partial\delta_1/\partial\varphi^H_{1,1} < 0$. Similarly,

$$\frac{\partial\delta_1}{\partial\varphi^L_{1,1}} = \frac{(2\gamma-1)(1-2\alpha)\left(\alpha\varphi^H_{1,1} + (1-\alpha)(1-\varphi^H_{1,1})\right)}{\left(\alpha\left(\gamma\varphi^L_{1,1} + (1-\gamma)\varphi^H_{1,1}\right) + (1-\alpha)\left(\gamma\varphi^L_{1,2} + (1-\gamma)\varphi^H_{1,2}\right)\right)^2}.$$

Again, the denominator is positive. And, as above, the numerator is positive (making use of the fact that $0 < \alpha < 1/2 < \gamma$ and $\varphi^H_{1,1} \in [0,1]$). Therefore, $\partial\delta_1/\partial\varphi^L_{1,1} > 0$. This implies that δ_1 is maximized when $\varphi^H_{1,1} = 0$ and $\varphi^L_{1,1} = 1$. Evaluating the ratio at this point—and noting that $\varphi^H_{1,1} + \varphi^H_{1,2} = \varphi^L_{1,1} + \varphi^L_{1,2} = 1$, since transition probabilities sum to one—yields the desired bound.

As for δ_2, note that

$$\delta_2 := \frac{\tau_{(2,L),(1,L)} + \tau_{(2,L),(2,H)}}{\tau_{(2,H),(1,H)} + \tau_{(2,H),(2,L)}} = \frac{(1-\alpha)\left(\gamma\varphi_{2,1}^L + (1-\gamma)\varphi_{2,1}^H\right) + \alpha\left(\gamma\varphi_{2,2}^L + (1-\gamma)\varphi_{2,2}^H\right)}{(1-\alpha)\left(\gamma\varphi_{2,1}^H + (1-\gamma)\varphi_{2,1}^L\right) + \alpha\left(\gamma\varphi_{2,2}^H + (1-\gamma)\varphi_{2,2}^L\right)}.$$

It is simple to show, as above, that δ_2 is strictly decreasing in $\varphi_{2,1}^H$ and strictly increasing in $\varphi_{2,1}^L$. Maximizing δ_2 with respect to these two variables yields an upper bound of λ. □

These bounds on δ_1 and δ_2 are helpful in understanding the long-run behavior of the Markov process on $\widehat{\Omega}$. In particular, we may use the bounds derived in Lemma 3 to partially characterize the steady-state distribution μ, as the bounds on δ_1 and δ_2 can be used to show that the long-run proportion of periods in which the decision maker correctly matches action to state is bounded above (and away from one).

LEMMA 4 (Bounds on accuracy of learning)**.**
For any irreducible symmetric two-state memory, the stationary distribution μ is such that both

$$\mu_{(1,L)} \leq \lambda\mu_{(1,H)} \text{ and } \mu_{(2,H)} \leq \lambda\mu_{(2,L)}.$$

PROOF. Recall from Lemma 3 that $\delta_1 \leq \lambda$, implying that $\delta_1\mu_{(1,H)} \leq \lambda\mu_{(1,H)}$. Similarly, $\delta_2 \leq \lambda$, implying that $\delta_2\mu_{(2,L)} \leq \lambda\mu_{(2,L)}$. Therefore, we have

$$\begin{aligned}\lambda\mu_{(1,H)}(\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}) + \lambda\mu_{(2,L)}(\tau_{(2,H),(1,H)} + \tau_{(2,H),(2,L)}) \\ \geq \mu_{(1,H)}(\tau_{(1,H),(1,L)} + \tau_{(1,H),(2,H)}) + \mu_{(2,L)}(\tau_{(2,L),(1,L)} + \tau_{(2,L),(2,H)}) \\ = \mu_{(1,L)}(\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}) + \mu_{(2,H)}(\tau_{(2,H),(1,H)} + \tau_{(2,H),(2,L)}),\end{aligned}$$

where the equality follows from the fact that, in a steady state, the probability of transitions from one subset of the state space $\widehat{\Omega}$ into its complement must equal the probability of transition from the complement back into the original subset. With this in hand, simple rearrangement yields

$$(\lambda\mu_{(1,H)} - \mu_{(1,L)})(\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}) \geq (\mu_{(2,H)} - \lambda\mu_{(2,L)})(\tau_{(2,H),(1,H)} + \tau_{(2,H),(2,L)}).$$

So, suppose that $\mu_{(1,L)} > \lambda\mu_{(1,H)}$. This implies that the left-hand side of the equation above must be negative. Hence, we must have $\mu_{(2,H)} < \lambda\mu_{(2,L)}$. Note, however, that the stationary distribution of the underlying state of the world places probability $1/2$ on each of states H and L, implying that

$$\mu_{(1,L)} + \mu_{(2,L)} = \frac{1}{2} = \mu_{(1,H)} + \mu_{(2,H)}.$$

Thus, we have

$$\mu_{(1,L)} > \lambda\left(\frac{1}{2} - \mu_{(2,H)}\right) \text{ and } \mu_{(2,H)} < \lambda\left(\frac{1}{2} - \mu_{(1,L)}\right).$$

Combining these two expressions (and the fact that $\lambda > 1$) yields $\mu_{(1,L)} < \mu_{(2,H)}$, which contradicts the symmetry of the memory rule and the symmetry of the Markov state transitions. Thus, we must have $\mu_{(1,L)} \leq \lambda\mu_{(1,H)}$. By symmetry, this then immediately implies that $\mu_{(2,H)} \leq \lambda\mu_{(1,L)}$. □

Thus, the decision maker cannot be "too" convinced that he is taking the correct action in either memory state, as

$$\frac{\mu_{(1,L)}}{\mu_{(1,H)}} = \frac{\mu_{(2,H)}}{\mu_{(2,L)}} \leq \lambda.$$

The final step in characterizing the optimal symmetric memory involves showing that the bounds described in Lemma 3 and Lemma 4 are, in fact, achievable by a deterministic transition rule. Moreover, we will show that the optimal memory system is unique.

THEOREM 2 (Optimal memory is deterministic)**.**
The optimal symmetric two-state memory is uniquely characterized by the deterministic transition rules

$$\varphi^L_{1,1} = \varphi^L_{2,1} = 1 = \varphi^H_{1,2} = \varphi^H_{2,2}. \tag{7}$$

PROOF. Recall the decision maker's optimization problem from . Since $\sum_{i\in\widehat{\Omega}} \mu_i = 1$, we may rewrite the objective function as

$$\gamma(\mu_{(1,L)} + \mu_{(2,H)}) + (1-\gamma)(\mu_{(1,H)} + \mu_{(2,L)}) = (2\gamma - 1)(\mu_{(1,L)} + \mu_{(2,H)}) + (1-\gamma).$$

As $\gamma > 1/2$, this objective function is increasing in $\mu_{(1,L)} + \mu_{(2,H)}$. Recall, however, the bounds from Lemma 4: $\mu_{(1,L)} \leq \lambda\mu_{(1,H)}$ and $\mu_{(2,H)} \leq \lambda\mu_{(2,L)}$. Summing these two inequalities yields

$$\mu_{(1,L)} + \mu_{(2,H)} \leq \lambda(\mu_{(1,H)} + \mu_{(2,L)}) = \lambda(1 - \mu_{(1,L)} - \mu_{(2,H)}) = \frac{\lambda}{\lambda+1}.$$

Thus, if a symmetric memory system is able to achieve this $\lambda/(\lambda+1)$ bound, it must be optimal.

Straightforward computation of the transition matrix and the associated steady state shows that by setting

$$\varphi^L_{1,1} = \varphi^L_{2,1} = 1 = \varphi^H_{1,2} = \varphi^H_{2,2},$$

we achieve this bound exactly. In particular, we arrive at a steady state in which

$$\mu_{(1,L)} = \lambda\mu_{(1,H)} = \lambda\mu_{(2,L)} = \mu_{(2,H)} = \frac{1}{2}\frac{\lambda}{\lambda+1} = \frac{(1-\alpha)\gamma + \alpha(1-\gamma)}{2}.$$

Thus, there exists a deterministic memory rule that is optimal.

To see that it is unique among the class of symmetric memory systems, we will show that no other symmetric memory system can generate the stationary distribution above. To see this, recall that, in a steady state, the probability of transitions from one subset of the state space $\widehat{\Omega}$ into its complement must equal the probability of transition from the complement back into the original subset. Therefore, we must have

$$\begin{aligned}\mu_{(1,L)}(\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}) + \mu_{(2,H)}(\tau_{(2,H),(1,H)} + \tau_{(2,H),(2,L)}) \\ = \mu_{(1,H)}(\tau_{(1,H),(1,L)} + \tau_{(1,H),(2,H)}) + \mu_{(2,L)}(\tau_{(2,L),(1,L)} + \tau_{(2,L),(2,H)}).\end{aligned}$$

By symmetry, this implies that

$$\mu_{(1,L)}(\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}) = \mu_{(1,H)}(\tau_{(1,H),(1,L)} + \tau_{(1,H),(2,H)}).$$

In addition, the requirement that $\mu_{(1,L)} = \lambda\mu_{(1,H)}$ yields

$$\lambda = \frac{\tau_{(1,H),(1,L)} + \tau_{(1,H),(2,H)}}{\tau_{(1,L),(1,H)} + \tau_{(1,L),(2,L)}} = \delta_1.$$

Thus, the steady state distribution described above is generated by transition rules such that δ_1 achieves the upper bound described in Lemma 3. But recall that the proof of Lemma 3 showed that

$$\frac{\partial\delta_1}{\partial\varphi^L_{1,1}} > 0 > \frac{\partial\delta_1}{\partial\varphi^H_{1,1}} = -\frac{\partial\delta_1}{\partial\varphi^H_{1,2}}.$$

Thus, the unique maximizer of δ_1 (and component of the optimal memory rule) is $\varphi^L_{1,1} = \varphi^H_{1,2} = 1$. By applying a symmetry argument, we obtain $\varphi^H_{2,2} = \varphi^L_{2,1} = 1$. □

Finally, note that the memory rule above yields the decision maker a payoff of

$$\gamma\left(\frac{\lambda}{\lambda+1}\right) + (1-\gamma)\left(1 - \frac{\lambda}{\lambda+1}\right).$$

Since $\lambda > 1$, the fraction of correctly chosen actions $\lambda/(\lambda+1)$ is strictly greater than $1/2$, and hence the expected payoff above is strictly greater than the expected payoff in Equation (3) generated by a single-state memory or a memory that chooses only a single action. Thus, the agent is best off using an irreducible memory with a different action in each state.

5. Conclusion

We have shown that, in a dynamic environment where the state of the world may change from period to period with a small but positive probability, the optimal symmetric memory for an agent with bounded memory is deterministic. This contrasts with the optimal memory in a static setting (where the underlying state of the world is unknown but fixed), where randomization is necessary for optimality. This demonstrates that the optimal memory rule is, somewhat surprisingly, discontinuous in the probability that the world may change over time. This discontinuity stems from the discontinuity in the stationary distribution of the underlying state of the world as the (symmetric) switching probability goes to zero—when the switching probability is positive, the stationary distribution is unique and independent of the initial state, while when the switching probability is zero, the stationary distribution is a degenerate distribution with unit mass on the initial-period state.

This suggests that the behavioral biases characterized by Wilson (2004) are not robust to small perturbations in the environment. In particular, if there is a minimal amount of instability in the environment, then a decision maker with bounded memory is likely to exhibit relatively extreme swings in beliefs and behavior (as opposed to the various forms of inertia described by Wilson. Moreover, as the instability increases (but the environment is still persistent), the behavioral biases diminish and bounded memory ceases to be a costly constraint for decision makers.

Finally, this paper tackles the question posed by Kalai and Solan (2003) regarding the optimality of deterministic versus stochastic memory transition rules. We show that, contrary to their examples and conjectures, a deterministic memory rule may induce a higher payoff than a stochastic memory rule when the underlying state of the world changes in a Markovian fashion.

REFERENCES

BATTAGLINI, M. (2005): "Long-Term Contracting with Markovian Consumers," *American Economic Review*, 95(3), 637–658.

COLE, H. L., AND N. R. KOCHERLAKOTA (2005): "Finite Memory and Imperfect Monitoring," *Games and Economic Behavior*, 53(1), 59–72.

COMPTE, O., AND A. POSTLEWAITE (2009): "Effecting Cooperation," Unpublished manuscript, University of Pennsylvania.

COVER, T., AND M. HELLMAN (1971): "On Memory Saved by Randomization," *Annals of Mathematical Statistics*, 42(3), 1075–1078.

HELLMAN, M., AND T. COVER (1970): "Learning with Finite Memory," *Annals of Mathematical Statistics*, 41(3), 765–782.

KALAI, E., AND E. SOLAN (2003): "Randomization and Simplification in Dynamic Decision-Making," *Journal of Economic Theory*, 111(2), 251–264.

KOCER, Y. (2009): "Endogenous Learning with Bounded Memory," Unpublished manuscript, New York University.

LIPMAN, B. L. (1995): "Information Processing and Bounded Rationality: A Survey," *Canadian Journal of Economics*, 28(1), 42–67.

MILLER, D. A., AND K. ROZEN (2010): "Monitoring with Collective Memory: Forgiveness for Optimally Empty Promises," Unpublished manuscript, Yale University.

MONTE, D. (2009): "Learning with Bounded Memory in Games," Unpublished manuscript, Simon Fraser University.

MULLAINATHAN, S. (2002): "A Memory-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 117(3), 735–774.

ROMERO, J. (2010): "Bounded Rationality in Repeated Games," Unpublished manuscript, California Institute of Technology.

RUBINSTEIN, A. (1998): *Modeling Bounded Rationality*. MIT Press, Cambridge.

WILSON, A. (2004): "Bounded Memory and Biases in Information Processing," Unpublished manuscript, Harvard University.