THE VALUE OF (BOUNDED) MEMORY IN A CHANGING WORLD

DANIEL MONTE† AND MAHER SAID‡

AUGUST 5, 2013

ABSTRACT: This paper explores the value of memory in decision making in dynamic environments. We examine the decision problem faced by an agent with bounded memory who receives a sequence of signals from a partially observable Markov decision process. We characterize environments in which the optimal memory consists of only two states. In addition, we show that the marginal value of additional memory states need not be positive, and may even be negative in the absence of free disposal.

KEYWORDS: Bounded memory, Dynamic decision making, Partially observable Markov decision process.

JEL CLASSIFICATION: C61, D81, D83.

†SAO PAULO SCHOOL OF ECONOMICS – FGV, DANIEL.MONTE@FGV.BR

‡OLIN BUSINESS SCHOOL, WASHINGTON UNIVERSITY IN ST. LOUIS, SAID@WUSTL.EDU

This paper supersedes an earlier working paper circulated as "Learning in Hidden Markov Models with Bounded Memory." We thank the editor, an anonymous referee, Dirk Bergemann, Dino Gerardi, Bernardo Guimaraes, Johannes Hörner, Abraham Neyman, and Ben Polak, as well as seminar participants at Yale University and Simon Fraser University for their helpful advice and comments.

1. INTRODUCTION

This paper explores the value of memory in decision making in dynamic environments. In particular, we examine the decision problem faced by an agent with bounded memory who receives a sequence of noisy signals from a partially observable Markov decision process: 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 sufficiently 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] Formally, our decision maker is restricted to a finite number of memory states and must choose both a transition rule and an action rule.[2] Characterizing the optimal behavior of an agent with cognitive limitations in dynamic environments may shed light on the behavioral biases that are present in such settings.[3] Moreover, such characterizations aid in understanding when "simple" heuristics or plans perform well in dynamic environments.[4]

In our first result, we show that if the underlying environment is sufficiently unstable (but still persistent), only two memory states are needed to reproduce the optimal behavior of an unconstrained Bayesian decision maker. This contrasts sharply with static, unchanging environments, where replicating an unconstrained Bayesian decision maker requires an infinite number of memory states—see Hellman and Cover (1970) and Wilson (2004). This suggests that the importance of additional memory stems primarily from its role in relatively stable environments. Even in those environments, however, additional memory need not increase a decision maker's payoff.

To make this point clear, our analysis proceeds by completely characterizing the optimal memory system for a decision maker who is restricted to a small memory of either two or three memory states. We show that, regardless of the uncertainty inherent in the environment or the noisiness of the signals, the optimal two-state memory deterministically uses the last observed signal as a sufficient statistic for decision making. This contrasts with results from the bounded memory

[1] Other recent work in decision problems with limited memory includes Güth and Ludwig (2000); Mullainathan (2002); Wilson (2004); Kocer (2010); Miller and Rozen (2012); and Kaneko and Kline (2013).

[2] Unlike a decision maker with bounded *recall* (see, among others, Lehrer (1988); Aumann and Sorin (1989); or Alós-Ferrer and Shi (2012)) who knows only a finite truncation of history, a decision maker with bounded *memory* has a finite number of states that summarize all her information. Such models have been studied extensively in repeated-game settings: Neyman (1985); Rubinstein (1986); and Kalai and Stanford (1988) are some of the early contributions to this literature, while Romero (2011); Compte and Postlewaite (2012a,b); and Monte (2012) are more recent. Closely related is the literature on "dynastic" games, as in Anderlini and Lagunoff (2005) and Anderlini, Gerardi, and Lagunoff (2008).

[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) also consider a model of dynamic decision making with bounded memory, but focus on the role and value of simplicity and randomization.

literature, starting as early as Hellman and Cover (1970), suggesting that randomization can compensate for memory restrictions. Indeed, Cover and Hellman (1971) show that, in a large class of problems, a two-state memory employing randomization performs arbitrarily better than *any* deterministic memory system. Similarly, Kalai and Solan (2003) show that randomization can lead to payoff improvements over deterministic memory systems, even in highly separable Markovian environments. Our result contributes to this literature by demonstrating that, in certain environments, randomization need not be beneficial.

Finally, we show that the optimal three-state memory involves randomization at the extremal states when the environment is sufficiently persistent relative to the informativeness of signals; this corresponds to the optimal memory system characterized by Hellman and Cover (1970) and Wilson (2004), who studied the optimal bounded memory system in a setting where the underlying state of the world is perfectly persistent. As the degree of instability in the environment increases, however, randomization is no longer optimal. More surprisingly, when the environment is sufficiently unstable (but still persistent), the third memory state becomes redundant—the optimal three-state memory only makes use of two states. Thus, unlike much of the previous literature on decision problems with bounded memory, the optimal memory system may not be irreducible, and the decision maker's optimal expected payoff need not be *strictly* increasing in the bound on memory. Moreover, when restricting attention to irreducible memory systems that make use of all states, the optimal expected payoff may not even be *weakly* monotonic in the number of possible memory states. Thus, the marginal value of additional memory states may be zero or, in some circumstances, may even be negative.

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 $\gamma \in (\frac{1}{2}, 1)$, 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.[5] In particular, we assume that this evolution follows a Markov process with

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

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

where $\alpha \in (0, \frac{1}{2})$. The parameter α measures 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 a sequence of independent flips of a fair coin.

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

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

The use of this payoff criterion allows us to focus on the accuracy of the decision maker's learning and the long-run "correctness" of her actions.[6] 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 her action is the same as the true state of the world. Payoffs are *not* perfectly informative and $\gamma < 1$, however; thus, letting $\delta \in [0, 1]$ denote the long-run proportion of periods in which the "matching" action is taken, the agent's expected utility may be written as

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

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

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

where the inequality follows from the fact that $\frac{1}{2} < \gamma < 1$. Symmetrically, observing a payoff of 1 after l or a payoff of 0 after h is more likely when the true state is L. Thus, we may partition the set of possible action-payoff pairs into a signal space $\mathcal{S} := \{H, L\}$, where $s = H$ represents

[6] With discounting, the optimal bounded memory system will be somewhat present biased, with distortions that are dependent on the decision maker's initial prior. Kocer (2010, Lemma 1) suggests, however, that discounting and the limit of means criterion are "close"—the payoff to the discounted-optimal memory system converges, as the discount rate goes to zero, to the payoff to the limit-of-means-optimal memory system.

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 her expected period-t payoff alone. Since $\gamma > \frac{1}{2}$, her (myopic) action rule, as a function of her 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 first shown by Hellman and Cover (1970), the optimal payoff for a bounded memory agent in a static environment is strictly increasing in her memory size. In our 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 optimal payoffs as a perfectly Bayesian decision maker.

We begin by considering this decision problem from the perspective of a fully Bayesian agent who has no constraints on her 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 $\rho \in [0,1]$, implying that 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. *The decision maker's period-$(t+1)$ beliefs $\rho_{t+1}^s(\rho_t)$ are strictly increasing in her period-t beliefs ρ_t, regardless of the realized signal $s \in \mathcal{S}$.*

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 is strictly positive. □

[7]Therefore, this decision problem is very different from a multi-armed bandit problem and departs from the optimal experimentation literature. See Kocer (2010) for a model of experimentation with bounded memory.

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

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

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

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

PROOF. Note that the belief revision process has a "long-run upper bound" given by the the fixed point of $\rho^H_{t+1}(\cdot)$. The equation $\rho = \rho^H_{t+1}(\rho)$ has only one non-negative solution $\bar{\rho}$, where

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

Because $0 < \alpha < \frac{1}{2} < \gamma < 1$, we have

$$\frac{\sqrt{\alpha^2 + (2\gamma - 1)^2(1 - 2\alpha)} - \alpha}{2(2\gamma - 1)} > \frac{\sqrt{\alpha^2} - \alpha}{2(2\gamma - 1)} = 0,$$

so $\bar{\rho} > \frac{1}{2}$. Likewise,

$$\begin{aligned}\frac{\sqrt{\alpha^2 + (2\gamma - 1)^2(1 - 2\alpha)} - \alpha}{2(2\gamma - 1)} < \frac{\sqrt{\alpha^2 + (2\gamma - 1)^2} - \alpha}{2(2\gamma - 1)} &< \frac{\sqrt{\alpha^2 + (2\gamma - 1)^2 + 2\alpha(2\gamma - 1)} - \alpha}{2(2\gamma - 1)} \\ &= \frac{\sqrt{(\alpha + (2\gamma - 1))^2} - \alpha}{2(2\gamma - 1)} = \frac{1}{2},\end{aligned}$$

so $\bar{\rho} < 1$. Moreover, Lemma 1 implies that $\rho^H_{t+1}(\rho) > \bar{\rho}$ if, and only if, $\rho > \bar{\rho}$; thus, a period-t belief ρ_t can only be larger than this upper bound if the initial belief ρ_0 is greater than $\bar{\rho}$ 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 equation $\rho = \rho^L_{t+1}(\rho)$ has only a single solution $\underline{\rho}$ that is smaller than one, where

$$\underline{\rho} := \frac{(2\gamma - 1) + \alpha - \sqrt{\alpha^2 + (2\gamma - 1)^2(1 - 2\alpha)}}{2(2\gamma - 1)} = 1 - \bar{\rho}.$$

Moreover, Lemma 1 implies that $\rho^L_{t+1}(\rho) < \underline{\rho}$ if, and only if, $\rho < \underline{\rho}$; thus, a period-t belief ρ_t can only be smaller than this lower bound if the initial belief ρ_0 is less than $1 - \bar{\rho}$ 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

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

this is the number of L signals sufficient for beliefs to fall below $\bar{\rho}$, regardless of how high the initial belief is. (Equivalently, it is the number of H signals sufficient for beliefs to go above the boundary $1 - \bar{\rho}$, 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 a one-bit memory suffices 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. *If α and γ are such that $\alpha \geq \gamma(1-\gamma)$, then a decision maker with only two memory states has the same optimal expected payoff as an unconstrained perfectly Bayesian decision maker.*

PROOF. Note first that $\rho^H_{t+1}(0) = \alpha > 0$ and that (as shown in Lemma 1) ρ^H_{t+1} is strictly increasing. Since $\bar{\rho}$ (defined in Equation (1)) is the unique fixed point of ρ^H_{t+1}, it must be the case that $(\rho - \rho^H_{t+1}(\rho))(\rho-\bar{\rho}) \geq 0$ for all $\rho \in [0,1]$, with equality only when $\rho = \bar{\rho}$. In addition, note that when $\alpha \geq \gamma(1-\gamma)$,

$$\gamma - \rho^H_{t+1}(\gamma) = \gamma\frac{\gamma^2+(1-\gamma)^2}{\gamma^2+(1-\gamma)^2} - \frac{\gamma^2(1-\alpha)+(1-\gamma)^2\alpha}{\gamma^2+(1-\gamma)^2} = \frac{(\alpha-\gamma(1-\gamma))(2\gamma-1)}{\gamma^2+(1-\gamma)^2} \geq 0.$$

Therefore, we must have $\bar{\rho} \leq \gamma$ whenever $\alpha \geq \gamma(1-\gamma)$

In addition, notice that $\rho^L_{t+1}(\gamma) = \frac{1}{2}$. 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 $\rho_t \in \left[\frac{1}{2}, \gamma\right]$, 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.

Because Bayesian updating is symmetric in this environment and $\rho^H_{t+1}(\rho) = 1 - \rho^L_{t+1}(1-\rho)$, an analogous 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 $\rho_t \in [1-\gamma, \frac{1}{2}]$, 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 $\bar{t} \in \mathbb{N}$ are such that $\rho_t \in [1-\gamma, \gamma]$, the signal in period $t \geq \bar{t}$ is a sufficient statistic for a Bayesian agent's decision in period $t+1$. Since Lemma 2 implies that $\bar{t} < \infty$ 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. □

This result is intuitive: if the underlying Markov process is sufficiently unstable, then information about the past is not useful. Indeed, in the case where $\alpha = \frac{1}{2}$, so the state of the world in any period is determined by an independent coin toss, it is obvious that history is entirely uninformative. However, the result above shows that this can also be the case when the environment is very persistent and α is arbitrarily small.

In particular, 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 may be a binding constraint.

4. Bounded Memory

We now consider the optimization problem faced by a decision maker in our environment endowed with a (bounded) memory system.

Definition. *A* memory system *is a tuple* $(\mathcal{M}, \varphi, \varphi_0, a)$*, where* $\mathcal{M}$ *is a finite set of memory states;* $\varphi : \mathcal{M} \times \mathcal{S} \to \Delta\mathcal{M}$ *is the memory transition rule;* $\varphi_0 \in \Delta\mathcal{M}$ *is the initial distribution over memory states; and* $a : \mathcal{M} \to A$ *is the action rule.*

For notational convenience, we will use $\varphi^s_{m,m'}$ to denote $\varphi(m,s)(m')$, where $\varphi(m,s)$ is the probability distribution governing transitions after observing signal $s \in \mathcal{S}$ while in state $m \in \mathcal{M}$. In addition, note that since actions affect neither state transitions nor information generation, the decision problem is noninteractive; Kalai and Solan (2003, Theorem 1) then implies that the restriction to deterministic action rules is without loss of generality.

Notice that the combination of state transitions and memory transitions generate a Markov process on an "extended" state space $\widehat{\Omega} := \mathcal{M} \times \Omega$. In principle, such a process may admit several recurrent communicating classes and multiple stationary distributions. We show, however, that it is without loss of generality to restrict attention to memory transition rules that generate a *unique* recurrent communicating class (and hence a unique stationary distribution).

Lemma 3. *Fix any memory system* $(\mathcal{M}, \varphi, \varphi_0, a)$ *with expected payoff* U*. There exists a memory system* $(\mathcal{M}, \varphi', \varphi'_0, a')$ *with expected payoff* $U' \geq U$ *that admits a unique recurrent communicating class and unique stationary distribution on* $\widehat{\Omega}$*.*

Proof. Note first that any recurrent communicating class $\mathcal{R} \subseteq \widehat{\Omega}$ may be written as $\mathcal{R} = M \times \Omega$, where $M \subseteq \mathcal{M}$; that is, $(m, H) \in \mathcal{R}$ for some $m \in \mathcal{M}$ if, and only if, $(m, L) \in \mathcal{R}$. To see why this is true, note that

$$\Pr(\hat{\omega}_t = (m', \omega') | \hat{\omega}_{t-1} = (m, \omega)) = \sum_{s \in \mathcal{S}} \Pr(\omega_t = \omega' | \omega_{t-1} = \omega) \Pr(s_t = s | \omega_t = \omega') \varphi^s_{m,m'}$$

for any $(m, \omega), (m', \omega') \in \widehat{\Omega}$. Recall that both signals occur with positive probability in both underlying states (since $\gamma < 1$), and that both underlying states may occur in any period with positive probability (since $\alpha > 0$). Therefore, $\Pr(\hat{\omega}_t = (m', \omega') | \hat{\omega}_{t-1} = (m, \omega)) > 0$ if, and only if, $\varphi^s_{m,m'} > 0$ for some $s \in \mathcal{S}$. Thus, it is the memory transition rule φ alone which determines whether states in $\widehat{\Omega}$ communicate or not. Since these memory transitions are independent of the underlying state, it must be the case that $(m, H) \in \mathcal{R}$ if, and only if, $(m, L) \in \mathcal{R}$.

Now notice that, since both $\mathcal{M}$ and Ω are finite, Stokey and Lucas (1989, Theorem 11.1) implies that we may partition the extended state space $\widehat{\Omega}$ into $k \geq 1$ recurrent communicating classes $\{\mathcal{R}_1, \ldots, \mathcal{R}_k\}$ and a transient set $\mathcal{T}$. The observation above immediately implies that this partition induces a partition on the memory $\mathcal{M}$; abusing notation slightly, we therefore write $m \in \mathcal{R}_i$ or $m \in \mathcal{T}$ whenever $(m, \omega) \in \mathcal{R}_i$ or $(m, \omega) \in \mathcal{T}$, respectively.

For all $i = 1, \ldots, k$, denote by u_i the decision maker's payoff (under action rule a) conditional on her starting in a memory state $m \in \mathcal{R}_i$. (Because payoffs are evaluated according to the limit of means and each recurrent communicating class has a unique stationary distribution, u_i is constant across all $m' \in \mathcal{R}_i$.) The decision maker's payoff is then

$$U = \sum_{m \in \mathcal{M}} \sum_{i=1}^{k} \varphi_0(m) P(\mathcal{R}_i | m) u_i,$$

where $P(\mathcal{R}_i | m)$ denotes the probability that any state in R_i is reached from initial state m. Since $\varphi_0 \in \Delta\mathcal{M}$ and $\sum_{i=1}^{k} P(\mathcal{R}_i | m) = 1$ for all $m \in \Omega$, the decision maker's payoff U is a convex combination of the payoffs $\{u_1, \ldots, u_k\}$. In particular, this implies that $U \leq \max\{u_1, \ldots, u_k\}$.

We now define an alternative memory transition rule $\varphi' : \mathcal{M} \times \mathcal{S} \to \Delta\mathcal{M}$, where as before we use $\varphi'^{s}_{m,m'}$ to denote $\varphi'(m, s)(m')$. In particular, we fix any $i^* \in \arg\max_{i=1,\ldots,k}\{u_i\}$ and let $N := |\{m : (m, \omega) \in \mathcal{R}_{i^*} \text{ for some } \omega \in \Omega\}|$. Then define, for all $m, m' \in \mathcal{M}$ and all $s \in \mathcal{S}$,

$$\varphi'^{s}_{m,m'} := \begin{cases} \varphi^{s}_{m,m'} & \text{if } m \in \mathcal{R}_{i^*}, \\ 1/N & \text{if } m \notin \mathcal{R}_{i^*}, m' \in \mathcal{R}_{i^*}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, φ' replicates the transitions of φ within the recurrent communicating class $\mathcal{R}_{i^*}$, and transitions uniformly at random into $\mathcal{R}_{i^*}$ from any memory state outside of it. This implies that $\widehat{\Omega}$ can be partitioned into a single recurrent communicating class $\mathcal{R}' = \mathcal{R}_{i^*}$ and a transient set $\mathcal{T}' = \widehat{\Omega} \setminus \mathcal{R}'$.

Moreover, since transitions within $\mathcal{R}'$ under φ' are the same as those under φ, $\mathcal{R}'$ has the same stationary distribution as $\mathcal{R}_{i^*}$, and hence (under the same action rule a) the same payoff u_{i^*}. Finally, since there is only a single recurrent communicating class, Stokey and Lucas (1989, Theorem 11.2) implies that the transition rule φ' induces that same (unique) stationary distribution. Thus, for any initial distribution $\varphi'_0 \in \Delta\mathcal{M}$, the decision maker's payoff is now $U' = u_{i^*} \geq U$. □

As profits are evaluated by the limit of means criterion, the initial conditions of the memory system are relevant only insofar as they influence the long-run distribution on the extended state space $\widehat{\Omega}$. Given Lemma 3 above, however, we are free to consider memory transition rules that generate a unique stationary distribution $\mu \in \Delta\widehat{\Omega}$, where μ_i denotes the mass on state $i \in \widehat{\Omega}$; therefore, we simply assume that the initial memory state is chosen uniformly at random. Note that the marginals of the steady-state distribution μ must agree with those generated by the underlying stochastic processes; in particular, we must have

$$\sum_{m \in \mathcal{M}} \mu_{(m,H)} = \sum_{m \in \mathcal{M}} \mu_{(m,L)} = \frac{1}{2}. \tag{2}$$

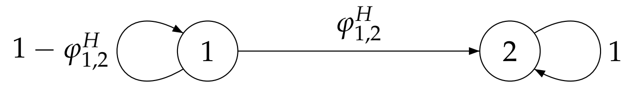

FIGURE 1. A generic symmetric and monotone two-state memory.

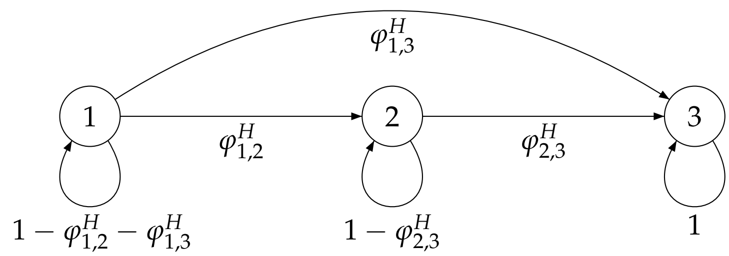

FIGURE 2. A generic symmetric and monotone three-state memory.

Moreover, a steady state must satisfy the standard stationarity condition: for all $(m,\omega) \in \widehat{\Omega}$,

$$\mu_{(m,\omega)} = \sum_{m' \in \mathcal{M}} \sum_{\omega' \in \Omega} \mu_{(m',\omega')} \left(\sum_{s \in \mathcal{S}} \Pr(\omega_t = \omega | \omega_{t-1} = \omega') \Pr(s_t = s | \omega_t = \omega) \varphi^s_{m',m} \right). \tag{3}$$

It is useful to note that, when the decision maker takes the same action in all memory states (that is, when $a(m) = a(m')$ for all $m, m' \in \mathcal{M}$), her 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 $\frac{1}{2}\gamma + \frac{1}{2}(1-\gamma) = \frac{1}{2}$. This is, of course, also the payoff resulting from a single-state memory. With this benchmark in mind, we will restrict attention to memory systems that use both actions. Moreover, we focus on *symmetric* and *montone* memory systems:

DEFINITION. *An n-state memory system is* symmetric *if* $\varphi^L_{j,k} = \varphi^H_{n+1-j,n+1-k}$ *for all* $j,k = 1,\ldots,n$. *An n-state memory system is* monotone *if* $\varphi^L_{j,k} = \varphi^H_{k,j} = 0$ *for all* $1 \leq j < k \leq n$.

Symmetry of the memory system is a natural restriction given the underlying symmetry in the problem.[8] Clearly, such memory systems induce a symmetric stationary distribution μ with

$$\mu_{(k,L)} = \mu_{(|\mathcal{M}|-k+1,H)} \text{ for all } k = 1,\ldots,|\mathcal{M}|.$$

Monotonicity implies that the decision maker never transitions to a "higher" state after a low signal or to a "lower" state after a high signal. Since high signals increase posterior beliefs (and low signals decrease them), monotonicity corresponds to the natural ordering of memory states where higher states are associated with greater posterior beliefs that the true state is H. Figures 1 and 2 present a visual representation of the transition probabilities in generic symmetric and monotone two- and three-state memory systems, respectively.

Thus, the decision maker's optimization problem is to

$$\max_{\varphi,a} \left\{ \sum_{m:a(m)=h} \Big(\gamma\mu_{(m,H)} + (1-\gamma)\mu_{(m,L)} \Big) + \sum_{m:a(m)=l} \Big((1-\gamma)\mu_{(m,H)} + \gamma\mu_{(m,L)} \Big) \right\},$$

[8]Recall that Bayesian updating in this environment is symmetric, with $\rho^H_{t+1}(\rho) + \rho^L_{t+1}(1-\rho) = 1$ for all $\rho \in [0,1]$.

subject to the constraint that μ is the (endogenously determined) steady state of the process induced by φ on $\widehat{\Omega}$. Given the steady state distribution μ, determining the optimal action in each state is trivial: the decision maker should set $a(m) = h$ whenever

$$\gamma\mu_{(m,H)} + (1-\gamma)\mu_{(m,L)} \geq (1-\gamma)\mu_{(m,H)} + \gamma\mu_{(m,L)},$$

and $a(m) = l$ when this inequality is reversed. This implies that the optimal action is h whenever the posterior belief in memory state m is $\frac{\mu_{(m,H)}}{\mu_{(m,H)}+\mu_{(m,L)}} > \frac{1}{2}$, and l when this posterior is less than $\frac{1}{2}$. (If the posterior belief in some state m is *exactly* $\frac{1}{2}$, then both $a(m) = h$ and $a(m) = l$ are optimal.) Since we consider symmetric and monotone memory systems, the ordering of the states then immediately implies that the optimal action rule is

$$a(m) = \begin{cases} l & \text{if } m \leq (|\mathcal{M}|+1)/2, \\ h & \text{otherwise.} \end{cases}$$

We begin by characterizing the optimal symmetric and monotone two-state memory.

THEOREM 2. *For any $\alpha \in (0, \frac{1}{2})$ and $\gamma \in (\frac{1}{2}, 1)$, the optimal symmetric and monotone two-state memory is given by $\varphi^H_{1,2} = 1$.*

PROOF. The proof may be found in the appendix. □

Thus, regardless of the instability of the underlying environment or the informativeness of payoff signals, the optimal symmetric two-state memory is deterministic: the decision maker simply chooses actions based solely on the most recent signal.

With three memory states, the situation is somewhat more subtle. Recall from Theorem 1 that using only two of the three memory states allows the decision maker to achieve the expected payoff of a perfect Bayesian whenever $\alpha \geq \gamma(1-\gamma)$. However, this condition is only sufficient, but not necessary: even when $\alpha < \gamma(1-\gamma)$, using only two memory states may be superior to irreducibly using all three states. Moreover, if the underlying state of the world is sufficiently persistent (when α is small relative to the noisiness of the payoff signals in a sense to be formally defined), then the optimal three-state memory system may involve randomization. As before, we consider only memory systems that are both symmetric and monotone.

THEOREM 3. *For any $\alpha \in (0, \frac{1}{2})$ and $\gamma \in (\frac{1}{2}, 1)$, the optimal symmetric and monotone three-state memory system is given by*

(a) $\varphi^H_{1,2} = \sqrt{\frac{2\alpha}{(1-2\alpha)\gamma(1-\gamma)}}$, $\varphi^H_{1,3} = 0$, *and* $\varphi^H_{2,3} = 1$ *if* $\frac{\alpha}{1-2\alpha} < \frac{\gamma(1-\gamma)}{2}$;
(b) $\varphi^H_{1,2} = 1$, $\varphi^H_{1,3} = 0$, *and* $\varphi^H_{2,3} = 1$ *if* $\frac{\gamma(1-\gamma)}{2} \leq \frac{\alpha}{1-2\alpha} < \gamma(1-\gamma)$; *or*
(c) $\varphi^H_{1,2} = 0$, $\varphi^H_{1,3} = 1$, *and* $\varphi^H_{2,3} = 1$ *if* $\frac{\alpha}{1-2\alpha} \geq \gamma(1-\gamma)$.

PROOF. The proof may be found in the appendix. □

Unlike the case of a two-state memory, the optimal three-state memory depends on the features of the underlying environment. Figure 3 presents the three possibilities. In region (a), where

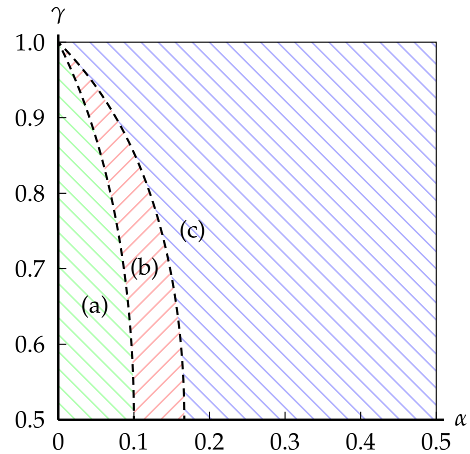

FIGURE 3. Classification of optimal three-state memory systems.

$\frac{\alpha}{1-2\alpha} < \frac{\gamma(1-\gamma)}{2}$, the state of the world is very persistent relative to the informativeness of signals. Therefore, the optimal memory requires multiple (in expectation) signals contradicting an extremal state in order to shift to the state with intermediate beliefs—as in Wilson (2004), the optimal memory leaves the extremal states only stochastically. In our model, however, this is in order to account for the noisiness of signals relative to instability of the underlying state. Note that as α approaches zero, the probability of departing the extremal states also approaches zero; the more stable the underlying environment, the larger the expected number of contradictory signals required to leave an extremal state. On the other hand, when α increases and we enter region (b), the optimal three-state memory becomes deterministic, stepping "linearly" through the memory states. This reflects the fact that the greater variability in the underlying state necessitates additional responsiveness to signals. Finally, in region (c), the environment is sufficiently unstable that the optimal memory only makes use of two state—the memory "jumps" from one extremal state to the other, skipping the middle state entirely. To understand the rationale for skipping the intermediate memory state, consider an increase in $\varphi^H_{1,2}$ to a small $\epsilon > 0$ (and hence a commensurate decrease in $\varphi^H_{1,3}$ to $1-\epsilon$). This change keeps the probability of departing memory state 1 after a high signal unchanged; however, this change *decreases* the arrival rate *into* state 1 after low signals since a strictly positive fraction of time is spent in the intermediate memory state. The net effect of these changes is to slow the response time to contradictory signals observed while in the extremal states. When α is large relative to the informativeness of signals, this dampened response rate has an overall negative effect on the steady-state probability of matching actions to the underlying state of the world.

It is crucial to note that region (c) in Figure 3 is larger than the region described by Theorem 1. In particular, when $\alpha < \gamma(1-\gamma)$, the previous period's signal alone is not a sufficient statistic for the decision of a fully rational perfect Bayesian. Therefore, using only two memory states leads to an expected payoff strictly less than the Bayesian benchmark. However, when

$$\alpha < \gamma(1-\gamma) < \frac{\alpha}{1-2\alpha},$$

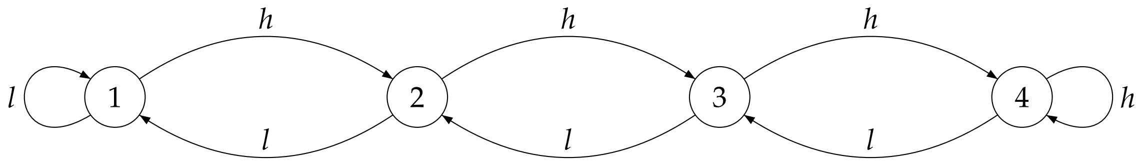

FIGURE 4. A deterministic and irreducible four-state memory system.

a decision maker limited to three memory states *optimally* makes use of only two states. Therefore, while the decision maker's expected payoff is nondecreasing in the number of possible memory states (since increasing the number of states relaxes a constraint in her optimal memory choice problem), this payoff need not be (strictly) increasing. In other words, the "shadow price" of an additional memory state may be zero. Therefore, even in settings where Theorem 1 does not apply and bounded memory is a binding constraint, a decision maker may not be willing to invest in an additional memory state, regardless of how small the cost of such an investment.

This observation suggests that ruling out memory systems with "redundant" memory states by restricting attention to irreducible memory systems (as is frequently done in the literature) is *not* without loss of generality.

DEFINITION. *A memory system is* irreducible *if its induced Markov process admits exactly one communicating class with no transient states.*

Note that an irreducible memory system generates a unique recurrent communicating class encompassing the *entire* state space, thereby precluding states that are essentially unused (such as, for example, the "middle state" appearing in part (c) of Theorem 3). Within the natural class of symmetric and monotone memory systems, the value of memory need not be monotonic once we impose the additional restriction of irreducibility.

THEOREM 4. *There exists an open set $\mathcal{O} \subset (0, \frac{1}{2}) \times (\frac{1}{2}, 1)$ of parameters such that, for all $(\alpha, \gamma) \in \mathcal{O}$, the payoff of the optimal irreducible, monotone, and symmetric memory system is nonmonotonic in the number of memory states.*

PROOF. Recall from Theorem 3 that, when $\gamma(1-\gamma) \leq \alpha/(1-2\alpha)$ (region (c) in Figure 3), the optimal three-state memory is *not* irreducible, but instead makes use of only the two extremal states; indeed, the optimal three-state memory in this region replicates the (irreducible) optimal two-state memory. Therefore, whenever $\gamma(1-\gamma) < \alpha/(1-2\alpha)$, the optimal irreducible three-state memory performs *strictly* worse than the optimal irreducible two-state memory.

Now consider the four-state memory system depicted in Figure 4. This irreducible memory system deterministically transitions to a "higher" state after high signals, and to a "lower" state after low signals. Lemma 4 (in the appendix) shows that this memory system generates an expected payoff of

$$U_4 := \frac{1 - \alpha^2 - 3\gamma + 4\alpha\gamma + 3\gamma^2 - 4\alpha\gamma^2}{1 + \alpha(1-2\alpha) - 2(1-2\alpha)\gamma(1-\gamma)}.$$

Meanwhile, Theorem 2 shows that the optimal two-state memory is irreducible, and Equation (4) implies that it yields an expected payoff of $U_2 := 1 - 2\gamma(1-\gamma)$. Therefore, we may write $U_4 - U_2$

as

$$\frac{1-\alpha^2-(3-4\alpha)\gamma(1-\gamma)}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}-\frac{(1-2\gamma(1-\gamma))\,(1+\alpha(1-2\alpha)-2\gamma(1-\gamma)(1-2\alpha))}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}$$
$$=\frac{\alpha-\alpha^2-\gamma-2\alpha\gamma+4\alpha^2\gamma+5\gamma^2-6\alpha\gamma^2-4\alpha^2\gamma^2-8\gamma^3+16\alpha\gamma^3+4\gamma^4-8\alpha\gamma^4}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}$$
$$=\frac{(2\gamma-1)^2\left(\alpha(2\gamma(1-\gamma)+1)-\gamma(1-\gamma)-\alpha^2\right)}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}=\frac{(2\gamma-1)^2\,(\alpha(1-\alpha)-(1-2\alpha)\gamma(1-\gamma))}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}.$$

Note, however, that both $(2\gamma-1)^2>0$ and $1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma))>0$ since $\alpha\in(0,\frac{1}{2})$ and $\gamma>\frac{1}{2}$. Therefore, $U_4>U_2$ if, and only if, $\alpha(1-\alpha)-(1-2\alpha)\gamma(1-\gamma)>0$, or, equivalently, when $\alpha(1-\alpha)/(1-2\alpha)<\gamma(1-\gamma)$.

Of course, the four-state memory system in Figure 4 need not be optimal (either globally or within the class of irreducible memory systems); thus, U_4 is only a lower bound on the payoff of the optimal irreducible four-state memory. Combining this fact with the payoff comparisons above, we may conclude that the optimal irreducible four-state memory yields a (strictly) greater payoff than the optimal irreducible two-state memory, which in turn yields a (strictly) greater payoff than the optimal irreducible three-state memory, for all $(\alpha,\gamma)\in\mathcal{O}$, where

$$\mathcal{O}:=\left\{(\alpha,\gamma)\middle|\frac{\alpha(1-\alpha)}{1-2\alpha}<\gamma(1-\gamma)<\frac{\alpha}{1-2\alpha}\right\}.$$

Therefore, the value of memory (within the class of irreducible memory systems) is *not* monotonic in the number of memory states for all $(\alpha,\gamma)\in\mathcal{O}$. □

Thus, the marginal value of an additional memory state, when restricting attention to irreducible memory systems (where there is no "free disposal" of individual states), can be strictly negative—even though the incremental payoff from adding *multiple* memory states may be strictly positive. Therefore, investment decisions in additional memory must consider the costs and benefits of acquiring multiple states at once, as a naive marginal analysis alone may not suffice.

5. Conclusion

We have shown that, in a dynamic environment where the state of the world is imperfectly persistent and signals are noisy but informative, the marginal value of additional memory may be zero or even negative. In particular, when the environment is sufficiently unstable, a decision maker needs only two memory states in order to perfectly replicate the behavior of an unboundedly rational Bayesian, and additional memory states are of no extra value. On the other hand, when the environment is relatively stable, a decision maker with bounded memory achieves a lower payoff than her unbounded Bayesian counterpart; in these cases, more memory may be of some value to the decision maker.[9] However, we have shown that there are a non-negligible subset of such environments in which a decision maker optimally leaves some memory resources

[9] Quantifying the loss from bounded memory (relative to an unbounded Bayesian decision maker) is certainly a natural avenue for further inquiry. Such an attempt is complicated, however, by the difficulty of analytically characterizing the general solution to a partially observable Markov decision problem such as our own, and is thus beyond the scope of the present work.

unused—without "free disposal" of memory states, the decision maker may be made worse off with greater memory resources.

In addition to their independent interest, our results have implications for other work in economics. For instance, we have characterized some dynamic environments in which the optimal memory is deterministic. This suggests that, in a changing world, a decision maker with bounded memory may exhibit relatively large swings in beliefs and behavior. Likewise, the payoff rankings for different memory sizes suggest that, even if the cost of additional memory is arbitrarily small but positive, smaller memory systems may be more likely than larger ones; this reinforces the focus of, for instance, Compte and Postlewaite (2012a,b) on relatively simple models of "mental states" and De Grauwe (2011) or Romero (2011) on "simple heuristics."

APPENDIX

PROOF OF THEOREM 2. Note that symmetry implies $\mu_{(1,L)} = \mu_{(2,H)}$ and $\mu_{(2,L)} = \mu_{(1,H)}$; therefore, the decision maker solves

$$\max_{\varphi^H_{1,2} \in [0,1]} \left\{ 2 \left(\gamma \mu_{(2,H)} + (1-\gamma)\mu_{(1,H)} \right) \right\}.$$

We may write the steady-state condition in Equation (3) for state $(2, H)$ as

$$\begin{aligned}
\mu_{(2,H)} &= (\alpha\mu_{(1,L)} + (1-\alpha)\mu_{(1,H)})(\gamma\varphi^H_{1,2} + (1-\gamma)\varphi^L_{1,2}) \\
&\quad + (\alpha\mu_{(2,L)} + (1-\alpha)\mu_{(2,H)})(\gamma\varphi^H_{2,2} + (1-\gamma)\varphi^L_{2,2}) \\
&= (\alpha\mu_{(2,H)} + (1-\alpha)\mu_{(1,H)})(\gamma\varphi^H_{1,2} + (1-\gamma)\varphi^H_{2,1}) \\
&\quad + (\alpha\mu_{(1,H)} + (1-\alpha)\mu_{(2,H)})(\gamma\varphi^H_{2,2} + (1-\gamma)\varphi^H_{1,1}),
\end{aligned}$$

where the second equality follows from symmetry. Recalling that $\varphi^H_{1,1} = 1 - \varphi^H_{1,2}$, and that monotonicity implies $\varphi^H_{2,1} = 0$, this may be written as

$$\begin{aligned}
\mu_{(2,H)} &= (\alpha\mu_{(2,H)} + (1-\alpha)\mu_{(1,H)})(\gamma\varphi^H_{1,2}) \\
&\quad + (\alpha\mu_{(1,H)} + (1-\alpha)\mu_{(2,H)})(\gamma\varphi^H_{2,2} + (1-\gamma)(1-\varphi^H_{1,2})) \\
&= \mu_{(1,H)} \left(\alpha\gamma + \alpha(1-\gamma)(1-\varphi^H_{1,2}) + (1-\alpha)\gamma\varphi^H_{1,2} \right) \\
&\quad + \mu_{(2,H)} \left(\alpha\gamma\varphi^H_{1,2} + (1-\alpha)\gamma + (1-\alpha)(1-\gamma)(1-\varphi^H_{1,2}) \right).
\end{aligned}$$

Combining this expression with the observation from Equation (2) that $\mu_{(1,H)} = \frac{1}{2} - \mu_{(2,H)}$, we can then solve for $\mu_{(2,H)}$. In particular, we must have

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

With this in hand, we may write the decision maker's payoff as

$$U_2(\varphi^H_{1,2}) = (1-\gamma) + (2\gamma - 1)\frac{\alpha + (\gamma-\alpha)\varphi^H_{1,2}}{2\alpha + (1-2\alpha)\varphi^H_{1,2}}. \tag{4}$$

Differentiating with respect to $\varphi^H_{1,2}$ yields

$$\begin{aligned}
U_2'(\varphi^H_{1,2}) &= (2\gamma-1)\frac{(2\alpha + (1-2\alpha)\varphi^H_{1,2})(\gamma-\alpha) - (\alpha + (\gamma-\alpha)\varphi^H_{1,2})(1-2\alpha)}{(2\alpha + (1-2\alpha)\varphi^H_{1,2})^2} \\
&= \alpha \left(\frac{2\gamma - 1}{2\alpha + (1-2\alpha)\varphi^H_{1,2}} \right)^2.
\end{aligned}$$

Since $\alpha \in (0, \frac{1}{2})$ and $\gamma \in (\frac{1}{2}, 1)$, this expression is strictly positive for all $\varphi^H_{1,2} \in [0, 1]$; therefore, the maximum is achieved when $\varphi^H_{1,2} = 1$, yielding a payoff of $U_2(1) = 1 - 2\gamma(1-\gamma)$. □

PROOF OF THEOREM 3. Notice first that symmetry implies that $\mu_{(1,L)} = \mu_{(3,H)}$, $\mu_{(2,L)} = \mu_{(2,H)}$, and $\mu_{(3,L)} = \mu_{(1,H)}$. Thus, $\frac{\mu_{(2,H)}}{\mu_{(2,L)}+\mu_{(2,H)}} = \frac{1}{2}$, implying that the expected payoff, conditional on being in state 2, is $\frac{1}{2}\gamma + \frac{1}{2}(1-\gamma) = \frac{1}{2}$. Therefore, the agent solves

$$\max_{\varphi^H_{1,2},\varphi^H_{1,3},\varphi^H_{2,3}} \left\{ 2\left(\gamma\mu_{(3,H)} + \frac{1}{2}\mu_{(2,H)} + (1-\gamma)\mu_{(1,H)} \right) \right\}$$
$$\text{s.t. } 0 \le \varphi^H_{1,2}, \varphi^H_{1,3}, \varphi^H_{2,3} \le 1$$
$$\text{and } \varphi^H_{1,2} + \varphi^H_{1,3} \le 1.$$

We begin by writing the steady-state condition for states $(1,H)$ and $(3,H)$ from Equation (3) as

$$\begin{aligned}
\mu_{(1,H)} &= (\alpha\mu_{(1,L)} + (1-\alpha)\mu_{(1,H)})(\gamma\varphi^H_{1,1} + (1-\gamma)\varphi^L_{1,1}) \\
&\quad + (\alpha\mu_{(2,L)} + (1-\alpha)\mu_{(2,H)})(\gamma\varphi^H_{2,1} + (1-\gamma)\varphi^L_{2,1}) \\
&\quad + (\alpha\mu_{(3,L)} + (1-\alpha)\mu_{(3,H)})(\gamma\varphi^H_{3,1} + (1-\gamma)\varphi^L_{3,1}) \text{ and} \\
\mu_{(3,H)} &= (\alpha\mu_{(1,L)} + (1-\alpha)\mu_{(1,H)})(\gamma\varphi^H_{1,3} + (1-\gamma)\varphi^L_{1,3}) \\
&\quad + (\alpha\mu_{(2,L)} + (1-\alpha)\mu_{(2,H)})(\gamma\varphi^H_{2,3} + (1-\gamma)\varphi^L_{2,3}) \\
&\quad + (\alpha\mu_{(3,L)} + (1-\alpha)\mu_{(3,H)})(\gamma\varphi^H_{3,3} + (1-\gamma)\varphi^L_{3,3}).
\end{aligned}$$

Imposing symmetry and monotonicity, these may be written as

$$\begin{aligned}
\mu_{(1,H)} &= (\alpha\mu_{(3,H)} + (1-\alpha)\mu_{(1,H)})(\gamma(1-\varphi^H_{1,2}-\varphi^H_{1,3}) + (1-\gamma)) + \mu_{(2,H)}(1-\gamma)\varphi^H_{2,3} \\
&\qquad + (\alpha\mu_{(1,H)} + (1-\alpha)\mu_{(3,H)})((1-\gamma)\varphi^H_{1,3}) \\
&= \mu_{(1,H)}\left(1-\alpha-(1-\alpha)\gamma\varphi^H_{1,2} - (\gamma-\alpha)\varphi^H_{1,3}\right) + \mu_{(2,H)}(1-\gamma)\varphi^H_{2,3} \\
&\qquad + \mu_{(3,H)}\left(\alpha - \alpha\gamma\varphi^H_{1,2} + (1-\alpha-\gamma)\varphi^H_{1,3}\right) \text{ and} \\
\mu_{(3,H)} &= (\alpha\mu_{(3,H)} + (1-\alpha)\mu_{(1,H)})(\gamma\varphi^H_{1,3}) + \mu_{(2,H)}\gamma\varphi^H_{2,3} \\
&\qquad + (\alpha\mu_{(1,H)} + (1-\alpha)\mu_{(3,H)})(\gamma + (1-\gamma)(1-\varphi^H_{1,2}-\varphi^H_{1,3})) \\
&= \mu_{(1,H)}\left(\alpha - \alpha(1-\gamma)\varphi^H_{1,2} + (\gamma-\alpha)\varphi^H_{1,3}\right) + \mu_{(2,H)}\gamma\varphi^H_{2,3} \\
&\qquad + \mu_{(3,H)}\left(1-\alpha-(1-\alpha)(1-\gamma)\varphi^H_{1,2} - (1-\alpha-\gamma)\varphi^H_{1,3}\right).
\end{aligned}$$

Combining the two equations above with the observation in Equation (2) that

$$\mu_{(2,H)} = \frac{1}{2} - \mu_{(1,H)} - \mu_{(3,H)},$$

we can solve for $\mu_{(1,H)}$ and $\mu_{(3,H)}$. In particular, we have

$$\begin{aligned}
\mu_{(1,H)} &= \frac{1}{2}\frac{(\alpha + (\varphi^H_{1,2}+\varphi^H_{1,3})(1-\gamma-\alpha) - \varphi^H_{1,2}(1-2\alpha)\gamma(1-\gamma))\varphi^H_{2,3}}{\alpha(\varphi^H_{1,2}+2\varphi^H_{2,3}) + (1-2\alpha)((\varphi^H_{1,2}+\varphi^H_{1,3})\varphi^H_{2,3} + \varphi^H_{1,2}(\varphi^H_{1,2}+2(\varphi^H_{1,3}-\varphi^H_{2,3}))\gamma(1-\gamma))}, \\
\mu_{(3,H)} &= \frac{1}{2}\frac{(\alpha + (\varphi^H_{1,2}+\varphi^H_{1,3})(\gamma-\alpha) - \varphi^H_{1,2}(1-2\alpha)\gamma(1-\gamma))\varphi^H_{2,3}}{\alpha(\varphi^H_{1,2}+2\varphi^H_{2,3}) + (1-2\alpha)((\varphi^H_{1,2}+\varphi^H_{1,3})\varphi^H_{2,3} + \varphi^H_{1,2}(\varphi^H_{1,2}+2(\varphi^H_{1,3}-\varphi^H_{2,3}))\gamma(1-\gamma))}.
\end{aligned}$$

Furthermore, note that the decision maker's expected payoff is

$$2\left(\gamma\mu_{(3,H)} + \frac{1}{2}\mu_{(2,H)} + (1-\gamma)\mu_{(1,H)}\right) = \frac{1}{2} + (2\gamma-1)(\mu_{(3,H)} - \mu_{(1,H)}),$$

where we have again substituted for $\mu_{(2,H)}$ using Equation (2). This implies that the decision maker maximizes

$$U_3(\varphi^H_{1,2}, \varphi^H_{1,3}, \varphi^H_{2,3}) :=$$
$$\frac{1}{2} + \frac{1}{2}\frac{(2\gamma-1)^2(\varphi^H_{1,2}+\varphi^H_{1,3})\varphi^H_{2,3}}{\alpha(\varphi^H_{1,2}+2\varphi^H_{2,3}) + (1-2\alpha)((\varphi^H_{1,2}+\varphi^H_{1,3})\varphi^H_{2,3} + \varphi^H_{1,2}(\varphi^H_{1,2}+2(\varphi^H_{1,3}-\varphi^H_{2,3}))\gamma(1-\gamma))}. \tag{5}$$

Note first that, if $\varphi^H_{1,2} = 0$ (and, by symmetry, $\varphi^L_{3,2} = 0$), then the "middle" memory state (state 2) is effectively redundant—the memory system only makes use of the two extremal states. Applying Theorem 2, the optimal memory, conditional on $\varphi^H_{1,2} = 0$, must have $\varphi^H_{1,3} = 1$. As in the optimal two-state memory from Theorem 2, this memory yields an expected payoff of

$$U_3(0, 1, \varphi^H_{2,3}) = 1 - 2\gamma(1-\gamma).$$

Clearly, the value of $\varphi^H_{2,3}$ is irrelevant in this case. However, in order to ensure that there is only a single recurrent communicating class, we simply set $\varphi^H_{2,3} = 1$ when $\varphi^H_{1,2} = 0$.

Suppose instead that $\varphi^H_{1,2} > 0$. Then differentiating the payoff in Equation (5) with respect to $\varphi^H_{2,3}$ yields

$$\frac{\partial U_3(\varphi^H_{1,2}, \varphi^H_{1,3}, \varphi^H_{2,3})}{\partial \varphi^H_{2,3}} =$$
$$\frac{\varphi^H_{1,2}(\varphi^H_{1,2}+\varphi^H_{1,3})(\alpha + (1-2\alpha)(\varphi^H_{1,3}+2\varphi^H_{2,3})\gamma(1-\gamma))(2\gamma-1)^2}{2\left(\alpha(\varphi^H_{1,2}+2\varphi^H_{2,3}) + (1-2\alpha)((\varphi^H_{1,2}+\varphi^H_{1,3})\varphi^H_{2,3} + \varphi^H_{1,2}(\varphi^H_{1,2}+2(\varphi^H_{1,3}-\varphi^H_{2,3}))\gamma(1-\gamma))\right)^2}.$$

Clearly, the denominator is positive. Moreover, $\varphi^H_{1,2} > 0$ implies that the numerator is positive. Thus, it is without loss of generality to set $\varphi^H_{2,3} = 1$ whenever $\varphi^H_{1,2} > 0$.

With this in mind, we consider two cases. We first assume that $\varphi^H_{1,2} > 0$ and $\varphi^H_{1,2} + \varphi^H_{1,3} = 1$. In this case, the decision maker's payoff is $U_3(\varphi^H_{1,2}, 1-\varphi^H_{1,2}, 1)$. Note, however, that

$$\frac{\partial^2 U_3(\varphi^H_{1,2}, 1-\varphi^H_{1,3}, 1)}{\partial(\varphi^H_{1,3})^2} = \frac{(2\gamma-1)^2\left(\kappa + \alpha^2 - 3\alpha\kappa\varphi^H_{1,2} + (\kappa\varphi^H_{1,2})^2\right)}{\left(1+\alpha\varphi^H_{1,2} - \kappa(\varphi^H_{1,2})^2\right)^3},$$

where we define $\kappa := (1-2\alpha)\gamma(1-\gamma)$. Since $\alpha \in (0, \frac{1}{2})$ and $\gamma \in (\frac{1}{2}, 1)$, we must have $\kappa \in (0, 1)$. Thus, $\varphi^H_{1,2} \in [0, 1]$ implies that

$$1 + \alpha\varphi^H_{1,2} - \kappa(\varphi^H_{1,2})^2 > 1 - \kappa > 0.$$

In addition, we can write

$$\kappa + \alpha^2 - 3\alpha\kappa\varphi^H_{1,2} + (\kappa\varphi^H_{1,2})^2 = (2\kappa\varphi^H_{1,2} - \alpha)(\kappa\varphi^H_{1,2} - \alpha) + \kappa.$$

Note that $(2\kappa\varphi^H_{1,2} - \alpha)(\kappa\varphi^H_{1,2} - \alpha)$ is negative if, and only if, $2\kappa\varphi^H_{1,2} - \alpha > 0 > \kappa\varphi^H_{1,2} - \alpha$. But $2\kappa\varphi^H_{1,2} - \alpha < 2\kappa$ and $\kappa\varphi^H_{1,2} - \alpha > -\alpha$, implying that

$$(2\kappa\varphi^H_{1,2} - \alpha)(\kappa\varphi^H_{1,2} - \alpha) + \kappa > (2\kappa)(-\alpha) + \kappa = (1-2\alpha)\kappa > 0.$$

Thus, $U_3(\varphi^H_{1,2}, 1-\varphi^H_{1,2}, 1)$ is a convex function of $\varphi^H_{1,2}$, and is therefore maximized either when $\varphi^H_{1,2} = 0$ or $\varphi^H_{1,2} = 1$. The decision maker's expected payoff in each of these cases is

$$U_3(0,1,1) = 1 - 2\gamma(1-\gamma) \text{ and } U_3(1,0,1) = \frac{2+\alpha-4\gamma(1-\gamma)-\kappa}{2(1+\alpha-\kappa)}.$$

Then we have

$$\begin{aligned} U_3(1,0,1) - U_3(0,1,1) &= \frac{(2+\alpha-4\gamma(1-\gamma)-\kappa) - 2(1+\alpha-\kappa)(1-2\gamma(1-\gamma))}{2(1+\alpha-\kappa)} \\ &= -\frac{(\alpha-\gamma+2\alpha\gamma+\gamma^2-2\alpha\gamma^2)(2\gamma-1)^2}{2(1+\alpha-\kappa)} = \frac{(\kappa-\alpha)(2\gamma-1)^2}{2(1+\alpha-\kappa)}. \end{aligned}$$

Recalling the definition of κ, we may then conclude that $U_3(1,0,1) > U_3(0,1,1)$ if, and only if,

$$\frac{\alpha}{1-2\alpha} < \gamma(1-\gamma).$$

Turning to our second case, suppose that $\varphi^H_{1,2} > 0$ and $\varphi^H_{1,2} + \varphi^H_{1,3} < 1$. In addition, assume that $\varphi^H_{1,3} > 0$. Therefore, the first-order conditions for both $\varphi^H_{1,2}$ and $\varphi^H_{1,3}$ must hold; that is, we have

$$\frac{\partial U_3(\varphi^H_{1,2}, \varphi^H_{1,3}, 1)}{\partial \varphi^H_{1,2}} = \frac{(2\gamma-1)^2\left((2-\varphi^H_{1,3})(\alpha+\kappa\varphi^H_{1,3}) - \kappa(\varphi^H_{1,2}+\varphi^H_{1,3})^2\right)}{2\left(\varphi^H_{1,2}+\varphi^H_{1,3}+(2-\varphi^H_{1,2}-2\varphi^H_{1,3})(\alpha-\kappa\varphi^H_{1,2})\right)^2} = 0 \text{ and}$$

$$\frac{\partial U_3(\varphi^H_{1,2}, \varphi^H_{1,3}, 1)}{\partial \varphi^H_{1,3}} = \frac{(2\gamma-1)^2(2+\varphi^H_{1,2})(\alpha-\kappa\varphi^H_{1,2})}{2\left(\varphi^H_{1,2}+\varphi^H_{1,3}+(2-\varphi^H_{1,2}-2\varphi^H_{1,3})(\alpha-\kappa\varphi^H_{1,2})\right)^2} = 0.$$

Solving these two equations yields

$$\varphi^H_{1,2} = \frac{\alpha}{\kappa} \text{ and } \varphi^H_{2,3} = 1 - \frac{\alpha}{2\kappa}.$$

Note, however, that these two expressions sum to *more* than 1, a contradiction. Thus, we must have $\varphi^H_{1,3} = 0$, and only the first of the FOCs above can hold. This implies that

$$\varphi^H_{1,2} = \sqrt{\frac{2\alpha}{\kappa}}.$$

(Of course, this is less than 1 if, and only if, $\frac{2\alpha}{(1-2\alpha)} < \gamma(1-\gamma)$; otherwise, we are at the corner solution where $\varphi^H_{1,2} = 1$.) Note, however, that

$$\begin{aligned} U_3(\varphi^H_{1,2}, 0, 1) - U_3(0,1,1) &= \frac{2\varphi^H_{1,2}(1-2\gamma(1-\gamma)+(2-\varphi^H_{1,2})(\varphi^H_{1,2}-\kappa\varphi^H_{1,2})}{2(\varphi^H_{1,2}+(2-\varphi^H_{1,2})(\alpha-\kappa\varphi^H_{1,2}))} - (1-2\gamma(1-\gamma)) \\ &= \frac{(\varphi^H_{1,2}-2)(\alpha-\kappa\varphi^H_{1,2})(2\gamma-1)^2}{2(\varphi^H_{1,2}+(2-\varphi^H_{1,2})(\alpha-\kappa\varphi^H_{1,2}))} > 0 \end{aligned}$$

if, and only if, $\alpha < \kappa\varphi^H_{1,2}$. Since $\varphi^H_{1,2} = \sqrt{2\alpha/\kappa}$, this implies that $U_3(\sqrt{2\alpha/\kappa}, 0, 1) > U_3(0,1,1)$ if, and only if, $\frac{\alpha}{2(1-2\alpha)} < \gamma(1-\gamma)$. □

Lemma 4. *The expected payoff of the four-state memory system depicted in Figure 4 is*

$$U_4 := \frac{1-\alpha^2-3\gamma+4\alpha\gamma+3\gamma^2-4\alpha\gamma^2}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}.$$

Proof. Note that we can write the Equation (3) steady-state condition for states $(1,H)$, $(2,H)$, and $(3,H)$, for the case of the four-state memory in Figure 4, as

$$\begin{aligned}
\mu_{(1,H)} &= \alpha(1-\gamma)\mu_{(1,L)} + (1-\alpha)(1-\gamma)\mu_{(1,H)} + \alpha(1-\gamma)\mu_{(2,L)} + (1-\alpha)(1-\gamma)\mu_{(2,H)},\\
\mu_{(2,H)} &= \alpha\gamma\mu_{(1,L)} + (1-\alpha)\gamma\mu_{(1,H)} + \alpha(1-\gamma)\mu_{(3,L)} + (1-\alpha)(1-\gamma)\mu_{(3,H)}, \text{ and}\\
\mu_{(3,H)} &= \alpha\gamma\mu_{(2,L)} + (1-\alpha)\gamma\mu_{(2,H)} + \alpha(1-\gamma)\mu_{(4,L)} + (1-\alpha)(1-\gamma)\mu_{(4,H)},
\end{aligned}$$

where we have made use of the fact that the memory transition rule is given by

$$\varphi^s_{m,m'} := \begin{cases} 1 & \text{if } (m,m',s) = (1,2,H),(2,3,H),(3,4,H),(4,4,H),\\ 1 & \text{if } (m,m',s) = (1,1,L),(2,1,L),(3,2,L),(4,3,L),\\ 0 & \text{otherwise.}\end{cases}$$

Symmetry also implies that $\mu_{(1,L)} = \mu_{(4,H)}$, $\mu_{(2,L)} = \mu_{(3,H)}$, $\mu_{(3,L)} = \mu_{(2,H)}$, and $\mu_{(4,L)} = \mu_{(1,H)}$; therefore, we may write

$$\begin{aligned}
\mu_{(1,H)} &= (1-\alpha)(1-\gamma)\mu_{(1,H)} + (1-\alpha)(1-\gamma)\mu_{(2,H)} + \alpha(1-\gamma)\mu_{(3,H)} + \alpha(1-\gamma)\mu_{(4,H)},\\
\mu_{(2,H)} &= (1-\alpha)\gamma\mu_{(1,H)} + \alpha(1-\gamma)\mu_{(2,H)} + (1-\alpha)(1-\gamma)\mu_{(3,H)} + \alpha\gamma\mu_{(4,H)}, \text{ and}\\
\mu_{(3,H)} &= \alpha(1-\gamma)\mu_{(1,H)} + (1-\alpha)\gamma\mu_{(2,H)} + \alpha\gamma\mu_{(3,H)} + (1-\alpha)(1-\gamma)\mu_{(4,H)}.
\end{aligned}$$

In addition, recall from Equation (2) that $\mu_{(1,H)} + \mu_{(2,H)} + \mu_{(3,H)} + \mu_{(4,H)} = \frac{1}{2}$. Solving this system of four equations in four unknowns yields the stationary distribution of this memory system, which is given by

$$\begin{aligned}
\mu_{(1,H)} &= \frac{(1-\gamma)\left(1-\alpha-(2-\alpha-\gamma)(1-2\alpha)\gamma\right)}{2\left(1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)\right)},\\
\mu_{(2,H)} &= \frac{\alpha(1-\alpha)+(1-\alpha-\gamma)(1-2\alpha)\gamma(1-\gamma)}{2\left(1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)\right)},\\
\mu_{(3,H)} &= \frac{\alpha(1-\alpha)+(\gamma-\alpha)(1-2\alpha)\gamma(1-\gamma)}{2\left(1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)\right)}, \text{ and}\\
\mu_{(4,H)} &= \frac{\gamma\left(2\alpha(1-\alpha)+(\gamma-\alpha)(1-2\alpha)\gamma\right)}{2\left(1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)\right)}.
\end{aligned}$$

Therefore, the expected payoff of this memory system is

$$\begin{aligned}
U_4 &:= \gamma(\mu_{(1,L)}+\mu_{(2,L)}+\mu_{(3,H)}+\mu_{(4,H)}) + (1-\gamma)(\mu_{(1,H)}+\mu_{(2,H)}+\mu_{(3,L)}+\mu_{(4,L)})\\
&= 2\gamma(\mu_{(3,H)}+\mu_{(4,H)}) + 2(1-\gamma)(\mu_{(1,H)}+\mu_{(2,H)})\\
&= \frac{\alpha\gamma+\alpha\gamma^2-2\alpha\gamma^3-\alpha^2\gamma+\gamma^3}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)} + \frac{1-\alpha^2-3\gamma+3\alpha\gamma+\alpha^2\gamma+3\gamma^2-5\alpha\gamma^2-\gamma^3+2\alpha\gamma^3}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}\\
&= \frac{1-\alpha^2-(3-4\alpha)\gamma(1-\gamma)}{1+\alpha(1-2\alpha)-2(1-2\alpha)\gamma(1-\gamma)}.
\end{aligned}$$

□

References

Alós-Ferrer, C., and F. Shi (2012): "Imitation with Asymmetric Memory," *Economic Theory*, 49(1), 193–215.

Anderlini, L., D. Gerardi, and R. Lagunoff (2008): "A "Super" Folk Theorem for Dynastic Repeated Games," *Economic Theory*, 37(3), 357–394.

Anderlini, L., and R. Lagunoff (2005): "Communication in Dynastic Repeated Games: 'Whitewashes' and 'Coverups'," *Economic Theory*, 26(2), 265–299.

Aumann, R. J., and S. Sorin (1989): "Cooperation and Bounded Recall," *Games and Economic Behavior*, 1(1), 5–39.

Compte, O., and A. Postlewaite (2012a): "Belief Formation," Unpublished manuscript, University of Pennsylvania.

——— (2012b): "Plausible 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.

De Grauwe, P. (2011): "Animal Spirits and Monetary Policy," *Economic Theory*, 47(2–3), 423–457.

Güth, S., and S. Ludwig (2000): "How Helpful is a Long Memory on Financial Markets?," *Economic Theory*, 16(1), 107–134.

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.

Kalai, E., and W. Stanford (1988): "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56(2), 397–410.

Kaneko, M., and J. J. Kline (2013): "Partial Memories, Inductively Derived Views, and Their Interactions with Behavior," *Economic Theory*, 53(1), 27–59.

Kocer, Y. (2010): "Endogenous Learning with Bounded Memory," Unpublished manuscript, Princeton University.

Lehrer, E. (1988): "Repeated Games with Stationary Bounded Recall Strategies," *Journal of Economic Theory*, 46(1), 130–144.

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 (2012): "Optimally Empty Promises and Endogenous Supervision," Unpublished manuscript, Yale University.

Monte, D. (2012): "Learning with Bounded Memory in Games," Unpublished manuscript, Sao Paulo School of Economics.

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

Neyman, A. (1985): "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma," *Economics Letters*, 19(3), 227—229.

Romero, J. (2011): "Finite Automata in Undiscounted Repeated Games with Private Monitoring," Unpublished manuscript, Purdue University.

RUBINSTEIN, A. (1986): "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 39(1), 83–96.

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

STOKEY, N. L., AND R. E. LUCAS, JR. (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge.

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