Limited Memory, Learning, and Stochastic Choice*

Drew Fudenberg† Giacomo Lanzani‡
Philipp Strack§

August 6, 2026

Abstract

We characterize the long-run behavior of agents who recall only a random subset of their past experiences. When empirical frequencies of actions converge, the limit must be a *stochastic memory equilibrium*. Stochastic memory equilibria are consistent with random utility, and limited memory can rationalize the stochastic choices of random utility models. Limited memory also explains how reminders and the skewness of induced consequences influence behavior, choice overload patterns, and underreaction to large samples. Allowing for recency and rehearsal effects, along with limited memory, explains correlated prediction errors in forecasts of returns.

*We thank Chiara Aina, Larbi Alaoui, Ian Ball, Pierpaolo Battigalli, Modibo Camara, Andrew Caplin, Simone Cerreia-Vioglio, John Conlon, Roberto Corrao, Glenn Ellison, Amanda Friedenberg, Nicola Gennaioli, Duarte Goncalves, Cosmin Ilut, Philippe Jehiel, Jay Lu, Fabio Maccheroni, Peter Maxted, David Miller, Tomasz Strzalecki, Elias Tsakas, Gerelt Tserenjigmid, and Alex Wolitzky for helpful comments, and NSF grant SES-2417162 as well as the Cowles foundation for financial support. We also thank Tsong-Hong Cheng for excellent research support.

†Department of Economics, MIT
‡Department of Economics, UC Berkeley
§Department of Economics, Yale University

1 Introduction

People typically remember only a limited number of their past experiences, and what they remember is stochastic: they might recall some things one week and different things the next. For example, when deciding whether to go to the gym, a person may recall only a few past visits, and a manager deciding whom to promote may recall episodes in which creativity or grit mattered. In both cases, different memories can lead the same person to make different choices in the same decision problem, not because their preferences have changed but because different memories surfaced. Moreover, past memories shape choices, and choices generate future memories. We develop a tractable model of limited and stochastic memory that enables us to analyze its long-run implications. The model unifies established behavioral regularities: stochastic choice, reminders, skewness effects, choice overload, underreaction to large samples, and some forms of the equity premium puzzle.

In our model, agents are naive: they update their beliefs as if the experiences they remember were the only ones that occurred and maximize their expected utility given these beliefs. Because agents recall only a small subset of their experiences, their beliefs and behavior remain stochastic as the number of their experiences grows large. We call an action distribution a *stochastic memory equilibrium* if it is generated by best responses to the distribution of remembered databases that the action distribution induces. We show that such equilibria exist, and that whenever the empirical action frequencies converge, they converge to a stochastic memory equilibrium.

Limited memory provides a learning foundation for stochastic choice. In the long run, behavior is as if utility were redrawn each period, but the randomness comes from memory rather than taste shocks. Conversely, every stochastic choice generated by a random utility model can be obtained as the long-run outcome of learning with limited memory. This connection links random utility models to the objective environment: the distribution of random utilities is constrained by the data-generating process and by the fixed-point requirement that action frequencies be consistent with the memories they generate. In binary-choice problems with affiliated signals, this implies payoff monotonicity: improving an action's objective payoff distribution increases its equilibrium choice frequency.

With normally distributed outcomes and a normal prior, stochastic memory equilibrium yields a mixed probit model that preserves payoff monotonicity and diminishing sensitivity

while adding frequency dependence: actions chosen less frequently are estimated with lower precision. When actions are bundles of desirable features, the model yields elimination by aspects [Tversky, 1972] with weights determined by the probability of recalling occasions on which each aspect mattered.

Finally, limited memory affects how agents evaluate rare events. Because rare events are often absent from finite remembered samples, agents underweight them when probabilities are learned from experience rather than given, as in Hertwig et al. [2004], Fox and Hadar [2006], and Wulff et al. [2018]. This underweighting biases choices toward actions with more frequent moderate payoffs and against actions whose value depends on rare large payoffs. Our model also explains why neutral reminders about past experiences increase use of beneficial actions, such as going to the gym, with stronger effects for those who take the action less frequently [Calzolari and Nardotto, 2017; Augenblick et al., 2026]. Limited memory further accounts for sample-size insensitivity and underreaction to signals, phenomena usually attributed to "underinference" [Phillips and Edwards, 1966]. Our model goes beyond underinference by also explaining why people perceive uncertainty even with very large samples [Kahneman and Tversky, 1972].

Since real-world memory is limited, the perfect memory of standard models is, at best, an approximation. We show that it is a good approximation to the form of limited memory we analyze here: as the expected number of recalled experiences goes to infinity, the stochastic memory equilibria converge to the self-confirming equilibria, which are the limit outcomes with perfect memory.

We then extend the model to capture "rehearsal" and "recency" by allowing experiences that occurred or were remembered in the previous period to be more likely to be recalled, consistent with the evidence in Kandel et al. [2000], Davelaar et al. [2005], and Conlon [2026]. A generalization of our fixed-point condition characterizes the limits of the action distribution: beliefs conditional on the objective history are autocorrelated rather than i.i.d., and the limit action distribution must be consistent with the stationary distribution of the Markov chain of beliefs it induces. We call such distributions ergodic memory equilibria. They extend Mullainathan [2002]'s analysis of rehearsal in income forecasts to the long run and more general functional forms.

Related Work Psychological evidence supports our assumptions that retrieval, rather than storage, is the main memory bottleneck and that retrieval is limited and stochas-

tic (Gershman et al. [2025], Kahana [2012], Shadlen and Shohamy [2016], and Sial et al. [2025]). d'Acremont et al. [2013] provides fMRI evidence that the brain maintains a neural representation of accumulated frequencies in memory-related regions, which are then combined with prior knowledge when forming posterior beliefs. This motivates modeling belief updating as operating on retrieved samples. Reder [2014] and Duncan and Shohamy [2020] document partial or complete unawareness of memory limitations. Kaanders et al. [2022] provides evidence of frequency dependence in active learning problems. This dependence is a general implication of our model; we explicitly characterize its effect for normal-normal and binomial-beta environments.

In a stochastic memory equilibrium, the agent's actions are stochastic because they are conditioned on a random sample of their (endogenous) experiences. Several classes of models derive random choice from randomness in exogenous or endogenous signals. Perhaps the oldest example of this is an optimal stopping problem in which the agent seeks to match a binary action to a binary state while incurring a flow cost for observing a Brownian signal. Fudenberg et al. [2018] extends the analysis of optimal stopping to settings where the agent is uncertain about the payoff difference between the actions; Che and Mierendorff [2019] further extends it to more general signal structures, and Hébert and Woodford [2023] replaces specific signals with a bound on the rate of information flow. Matějka and McKay [2015] and Ke and Villas-Boas [2019] also connect stochastic choice to the objective environment in models in which agents first obtain information and then make a single decision. Lu [2016] and Natenzon [2019] axiomatize stochastic choice arising from Bayesian updating; in these models, the distribution and number of signals are exogenous.

Wilson [2014] and Jehiel and Steiner [2020] study the optimal use of a stochastic memory by an agent who receives a stream of exogenous signals until they stop and take a single action. Osborne and Rubinstein [1998] studies a notion of equilibrium in two-player games where players receive a fixed number of samples of the payoffs of each of their actions against the equilibrium mixed action of the other player and choose the action that maximizes the expected payoff against these empirical distributions. Salant and Cherry [2020] extends this prior-free approach to other statistical inference procedures, focusing on estimating the frequency of other players taking a particular action, again with a fixed sample size.[1] Danenberg and Spiegler [2026] study the equilibrium action distribution when the precision

[1]Koszegi et al. [2021] considers a different form of circularity in the memory process, allowing negative memories to be recalled more frequently when holding more negative beliefs.

of the signals about the payoff to each action depends on the probability the action is played, but does not model the agent's period-by-period decisions and information. Gonçalves [2023] defines an equilibrium concept for agents who make a single decision based on optimal sequential sampling from the equilibrium distribution. In these papers, aggregate behavior is assumed to be described by a fixed-point condition; we instead prove that a specific fixed-point condition characterizes the possible limit frequencies.

Our earlier paper Fudenberg et al. [2024] studies the long-run outcomes of an agent with unlimited but "selective" memory, meaning that they are more likely to remember some experiences than others; in this paper we abstract away from selectivity and let each experience have the same probability of being recalled.[2] Because the agent eventually has an infinite sample, when memory is not selective, the long-run outcome in that model is the same as if the agent had perfect memory. Gottlieb [2014] studies a model of deliberate memory manipulation that induces regret-sensitive preferences and shows that behavior converges to that predicted by expected utility when memory is perfect. Karlan et al. [2016] models the effectiveness of reminders by assuming that agents always take into account their future need for ordinary consumption expenditures but only sometimes account for infrequently-occurring future expenditures, even if they are perfectly deterministic, as with car registration fees or school fees. The paper attributes the randomness to random attention, but it can be viewed as stochastic memory.

The baseline version of our model, without rehearsal, can be interpreted as a social learning model in which, at each period, different agents draw a subset of past agents' experiences and make a decision. Under this interpretation, the closest paper is Banerjee and Fudenberg [2004], but that paper differs in many ways. For example, there agents do make inferences from the prevalence of each action in their database, and later agents do not observe the private signals of their predecessors.

There is an extensive psychology literature documenting various forms of the recency effect and proposing explanations for it, see, e.g., the summaries in Davelaar et al. [2005] and Erev and Haruvy [2016]. There is also extensive evidence of the importance of rehearsal; see, e.g., the textbook by Kandel et al. [2000]. Mullainathan [2002] analyzes the short-run implications of rehearsal in a specific parametric context but does not study its long-run

[2]In particular, with or without selectivity, the unlimited memory of that paper cannot generate any of the patterns of random utility, skewness and reminder effects, persistent uncertainty in the long run, and choice overload we study in Section 4. Conversely, while selective memory is able to rationalize all the outcomes of misspecified learning, limited memory is not.

effects. Bordalo et al. [2023] links limited memory to recurrent errors in inference, while Bordalo et al. [2025] shows how memory can impact choice through matching of the current problems with a category of similar past experiences.

2 The Model

In each period $t \in \mathbb{N}_+$, a single agent chooses an action a from the finite set A, generating an outcome in the finite set Y according to the objective distribution $p_a^* \in \Delta(Y)$. The agent knows that the map from actions to probability distributions over outcomes is fixed, but is uncertain about the outcome distributions each action induces. The agent has a prior μ_0 over data generating processes $p \in \Delta(Y)^A$, where $p_a(y)$ denotes the probability of outcome $y \in Y$ when action a is played under data generating process p.[3] The support of μ_0 is Θ; its elements are the p that the agent initially thinks are possible. We maintain the following assumption throughout:

Assumption 1. The agent is correctly specified, i.e., $p^* \in \Theta$.

We assume correct specification to highlight the issues that arise solely from limited memory, but our results generalize to the case where the agent is misspecified.

Histories, Memory, and Recalled Periods We call action-outcome pairs $(a, y) \in A \times Y$ *experiences*. A period $t \in \mathbb{N}$ *history* is a sequence $h_t = (A \times Y)^t$. We assume that the probability the agent remembers any given past experience at the beginning of period $t+1$ is $m_{t+1} = \min\{1, k/t\}$, with $k > 0$. As we will see, k is the expected number of experiences the agent recalls when their sample is large.

After history $h_t = (a_i, y_i)_{i=1}^t$, the *recalled periods* r_t are a random subset of $\{1, ..., t\}$. We assume that each past experience has an independent probability of being recalled, so[4]

$$\mathbb{P}[r_t = R] = m_{t+1}^{|R|}(1 - m_{t+1})^{t-|R|} \qquad \forall R \subseteq \{1, ..., t\}. \tag{1}$$

For every objective history h_t and set of recalled periods R, the *recalled history* is the subsequence of recalled experiences listed in the order in which they were realized.

[3] Many of our examples identify y with the agent's realized utility, but the model allows different realizations of y to have the same payoff, as when paying a fixed cost to obtain an informative signal.

[4] Section 6 lets experiences that happened or were recalled at $t-1$ be more likely recalled at t.

Beliefs We assume the agent recomputes their beliefs each period based on all of their recalled experiences, as opposed to simply updating their period-t beliefs based on their period-t observation,[5] and that the agent is unaware of their limited memory and naively updates their beliefs as if the experiences they remember are the only ones that have occurred. Thus, the agent's beliefs only depend on the number of times each (a, y) pair is recalled, which we call the agent's *database*. We let $\mathcal{D} = \mathbb{N}^{A\times Y}$ denote the set of these databases, so that the posterior probability of any (measurable) $C \subseteq \Theta$ after database $d \in \mathcal{D}$ is[6]

$$\mu\big(C \mid d\big) = \frac{\int_C \prod_{(a,y)\in A\times Y}(p_a(y))^{d(a,y)} d\mu_0(p)}{\int_\Theta \prod_{(a,y)\in A\times Y}(p_a(y))^{d(a,y)} d\mu_0(p)}. \tag{2}$$

Optimal Policies The agent is myopic, with utility function $u : A\times\int_\Theta \sum_{y\in Y} u(a, y)p_a(y)d\nu(p)$ denote the actions that maximize the agent's current period expected utility when their belief is $\nu \in \Delta(\Theta)$.[7] A *Markovian policy* $\pi : \Delta(\Theta) \to A$ is a measurable function that specifies a pure action for every belief.

The agent uses an *optimal Markovian policy* π, i.e., for every $\nu \in \Delta(\Theta)$, $\pi(\nu) \in BR(\nu)$. Together, true data generating process p^*, prior μ_0, and Markovian policy function π induce a unique probability measure over histories, denoted as $\mathbb{P}_{\mu_0,\pi}$.

Limit Action Frequencies For every t, define the *action frequency* at time t by

$$\alpha_t(a') = \frac{1}{t}\sum_{\tau=1}^{t} \mathbb{1}_{\{a'\}}(a_\tau) \qquad \forall a' \in A.$$

We say that $\alpha \in \Delta(A)$ is a *limit frequency* if there exists an optimal Markovian policy π such that for every $\varepsilon > 0$

$$\mathbb{P}_{\mu_0,\pi}\left[\limsup_{t\to\infty} ||\alpha_t - \alpha||_\infty \leqslant \varepsilon\right] > 0\,.$$

[5] As noted above, there is fMRI evidence that agents re-access memories of their experiences when forming beliefs. Note that if the same data is relevant in many different decision problems, it is more efficient to store the data than all of the potentially relevant posterior beliefs.

[6] If the agent believes their utility function is subject to very unlikely random shocks that can induce each action as the best reply, and dogmatically believes they recall every experience, they attribute unexplained behavior to the utility shock so Bayes rule coincides with equation (2). Heidhues et al. [2023] studies agents who forget their preference shocks and tries to infer them from their own actions.

[7] For every $X \subseteq \mathbb{R}^n$, $\Delta(X)$ denotes the set of Borel probability distributions on X endowed with the topology of weak convergence.

The definition of limit frequency requires that for every ε, there is a strictly positive probability that the empirical frequency remains within the ε-ball around α forever.[8]

3 Stochastic Memory Equilibrium

This section defines stochastic memory equilibrium, proves its existence, and characterizes it as the limit of empirical action frequencies.

For every action distribution α, define $\eta_\alpha \in \Delta(\mathcal{D})$ by

$$\eta_\alpha(d) = \prod_{(a,y)\in A\times Y} \frac{[\alpha(a)p_a^*(y)k]^{d(a,y)}}{d(a,y)!} e^{-k\alpha(a)p_a^*(y)} \qquad \forall d \in \mathcal{D}.$$

This is a product distribution where the marginal distribution for each action-outcome pair (a, y) is Poisson with mean $k\alpha(a)p_a^*(y)$.

We will show that, for large t, η_{α_t} well approximates the distribution of databases. Intuitively, because recall probability is diminishing in t, the remembered outcome distribution for a given (a, y) is analogous to drawing from a binomial with shrinking weights, so it converges to a Poisson distribution by the Poisson limit theorem. A law of large numbers shows that the joint frequency of each action and outcome pair (a, y) converges to $\alpha_t(a)p_a^*(y)$, so the expected number of times a pair (a, y) is recalled is proportional to the frequency of action a, the probability of the outcome given the action $p_a^*(y)$, and the average memory capacity k.

Lemma 1. *For $\mathbb{P}_{\mu_0,\pi}$-almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$, the distance between the distribution of databases at $t+1$ given h_t and η_{α_t} converges to 0 as $t \to \infty$.*[9]

Lemma 1 shows that the limiting database distribution η_α depends only on α and the true outcome distribution p^*. There is no additional history-dependence or path-dependence in the limit, which allows a tractable necessary condition for long-run outcomes. We use this below to define stochastic memory equilibrium as action frequencies consistent with the beliefs they induce.

[8] This covers convergence to α and also the case where no single α has strictly positive probability of being the limit.

[9] Note that Lemma 1 implies that as the number of the agent's observations grows to infinity, the probability they recall nothing at all is bounded away from 0. This is not essential: every result extends to the case where the agent remembers a fixed number C of "anchored memories" in addition to those prescribed by our current memory process.

Next, we associate each action distribution with the distribution of beliefs that it induces. Let $F_\alpha^{\mu_0}$ be the distribution of beliefs induced by the distribution η_α of databases and prior μ_0, i.e., for all measurable $\mathcal{C} \subseteq \Delta(\Theta)$,

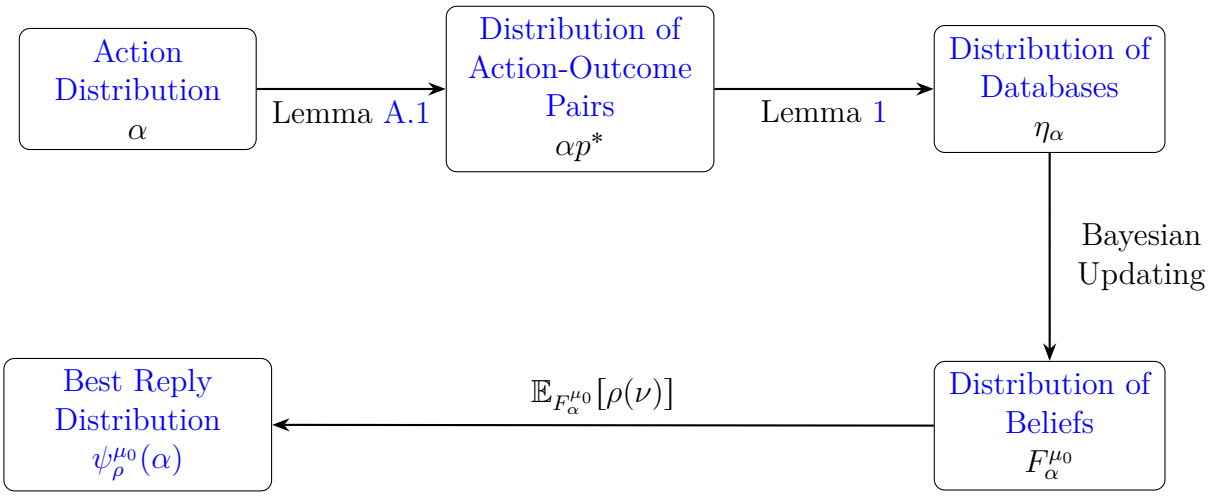
$$F_\alpha^{\mu_0}(\mathcal{C}) = \eta_\alpha\left(\{d : \mu(\cdot|d) \in \mathcal{C}\}\right). \tag{3}$$

Let $\mathcal{O}$ denote the set of measurable selections from the (mixed) best reply correspondence: i.e., $\rho : \Delta(\Theta) \to \Delta(A)$ is in $\mathcal{O}$ if and only if ρ is measurable and $\rho(\nu) \in \Delta(BR(\nu))$ for all $\nu \in \Delta(\Theta)$. For any $\rho \in \mathcal{O}$, let $\psi_\rho^{\mu_0}$ be the function that maps $\alpha \in \Delta(A)$ to the action distribution generated when the agent uses policy ρ and their beliefs are distributed according to $F_\alpha^{\mu_0}$. Formally,

$$\psi_\rho^{\mu_0}(\alpha) = \int_{\Delta(\Theta)} \rho(\nu) dF_\alpha^{\mu_0}(\nu). \tag{4}$$

The operations leading to $\psi_\rho^{\mu_0}(\alpha)$ are visualized in Figure 1, if action distribution α is played forever, it induces distribution αp^* over histories, where $\alpha p^*(a, y) = \alpha(a) p_a^*(y)$ (see Lemma A.1 in the Appendix). This distribution and the memory function m together induce a distribution of databases η_α, (Lemma 1) and Bayesian updating on each database generates distribution $F_\alpha^{\mu_0}$ over posterior beliefs. Assigning $\rho(\nu)$ to each posterior belief ν generates action distribution $\psi_\rho^{\mu_0}(\alpha)$. Stochastic memory equilibrium requires that the composition of these operations maps back to α.

Figure 1: Illustration of $\psi_\rho^{\mu_0}$.



Definition 1. A *stochastic memory equilibrium* is an $\alpha \in \Delta(A)$ for which there is $\rho \in \mathcal{O}$ such that $\alpha = \psi_\rho^{\mu_0}(\alpha)$.

Note that the set of stochastic memory equilibria depends on the prior μ_0 through its effect on the posterior beliefs. Below, we discuss this dependence further and show that it vanishes as k grows.

The notion of stochastic memory equilibrium and the ancillary functions used to define it are justified by the following result, which shows that whenever the action frequency converges the limit is a stochastic memory equilibrium.

Theorem 1. *If α is a limit frequency, then α is a stochastic memory equilibrium.*[10]

The first step of the proof is Lemma 1's characterization of the limit distribution of databases. The second step uses stochastic approximation to characterize the long-run outcome of the empirical action process when beliefs are generated by the constant action distribution α. To do this we apply the Benaim et al. [2005] extension of stochastic approximation to differential inclusions, which requires that the correspondence obtained by averaging best replies under $F_\alpha^{\mu_0}$ is well behaved, which Lemma A.3 verifies. Finally, if the differential inclusion enters a sufficiently small neighborhood of an α that is not a stochastic memory equilibrium, it leaves it within a bounded time, contradicting convergence to α.

The classic one-armed bandit problem provides an easy example in which the limit frequency need not be a point mass, so the "stochastic" part of stochastic memory equilibrium is needed. Suppose that the agent's prior belief is that the risky arm is better than the safe arm, but there is a sequence $(a_i, y_i)_{i=1}^t$ that has (objectively) positive probability and induces the agent to play the safe arm. The agent cannot converge to always play the risky arm, because then they would sometimes only recall $(a_i, y_i)_{i=1}^t$ and switch to the safe arm. However, the agent will play the risky action whenever they don't remember any past outcomes or only recall successes with the risky arm, which occurs with positive probability.

The following theorem shows that a stochastic memory equilibrium exists even in the absence of limit frequencies.[11]

Proposition 1. *A stochastic memory equilibrium exists.*

[10] Sections 4.3 and B.7 give examples with multiple stochastic memory equilibria; we do not know whether there can be multiple limit frequencies.

[11] An analogy to Nash equilibrium and learning in games may be useful here: Nash equilibria exist in all finite games, but unstable Nash equilibria cannot be limit points of stochastic fictitious play, and in some finite games the only ω-limit of stochastic fictitious play is a cycle [Benaım and Hirsch, 1999].

To prove this, we show that the correspondence that maps each α to the union over $\rho \in \mathcal{O}$ of $\psi_\rho^{\mu_0}(\alpha)$ satisfies the conditions of the Kakutani fixed-point theorem.

4 Applications

4.1 Random Utility and Stochastic Choice

This section relates limited memory to the random utility model, the most widely used model of stochastic choice in single-agent settings. In a random utility model, the agent's utility vector is independently drawn from a fixed distribution each period. In a stochastic memory equilibrium, by contrast, randomness in the perceived expected utilities of the actions comes from the agent's memories. This connection provides a learning foundation for random utility and some of its common specifications. Block and Marschak [1960] already emphasized that stochastic choice may reflect either random tastes or differences in information, and suggested that a learning model could help distinguish the two.[12]

Let $\mathcal{M}$ be the collection of non-empty subsets of A. A *stochastic choice function* is a map $c : \mathcal{M} \to \Delta(A)$ such that $\sum_{a \in M} c(a, M) = 1$ for all $M \in \mathcal{M}$. Let $\mathcal{P}$ be the linear orders on A. A stochastic choice function c has a *random utility representation* $\zeta \in \Delta(\mathcal{P})$ if for all $M \in \mathcal{M}$,[13] $c(a, M) = \zeta\left(\{P \in \mathcal{P} \mid \forall a' \in M \backslash \{a\}, aPa'\}\right) =: c_\zeta(a, M)$.

To relate our model of memory and learning to random utility and stochastic choice, suppose that in addition to the sequential learning process of the basic model, an experimenter elicits the agent's choice distribution on each menu (i.e., subset of A) at some single time t. During this elicitation process, no feedback is given to the subject, so that no learning happens throughout that stage. As usual, this requires many observations of choice from each menu. We say that the behavior is approaching a random utility model ζ if, when such elicitation is conducted sufficiently late, the resulting stochastic choice rule is arbitrarily close to c_ζ. We assume that when confronted with one of these menus, the agent breaks ties deterministically in a menu-independent way.[14]

[12] Block and Marschak [1960]: "All the various definitions of utility given in this paper will be related to the empirical entities [...] [that combine] information and desirability, in some unknown though presumably not-too-changeable fashion. [...] If discrimination is learned, through experience, in some more or less-determined manner, the model has to be modified, of course, but we have not attempted to bring in learning." Caplin [2025] discusses these issues.

[13] This is equivalent to a random utility representation that uses an additional probability space. See, e.g., Proposition 1.9 in Strzalecki [2025].

[14] Formally, we require that for all $M, M' \in \mathcal{M}$ if $a, a' \in BR(\nu|M) \cap M'$ and $\pi(\nu|M) = a$, then $\pi(\nu|M') \neq$

Definition 2. Behavior *approaches a random utility representation* ζ on a history sequence $(h_t)_{t\in\mathbb{N}}$ if menu choices conditional at measurement time t converge to those of ζ, i.e.,

$$\lim_{t\to\infty} \mathbb{P}_{\mu_0,\pi}[a_{t+1} = a|h_t, M] = c_\zeta(a, M) \qquad \forall M \in \mathcal{M}.$$

The next result shows that when the distribution of the agent's actions converges, the agent's menu choices converge to a limit that has a random utility representation. Moreover, the limit empirical action distribution coincides with the choice distribution induced by that random utility model on the complete action set. Together, Theorems 1 and 2 show how the information structure induced by η_α determines the choice probabilities. Conversely, they also show that any form of stochastic choice that can be rationalized by a random utility model can be approximated as the result of learning with limited memory for some true distribution p^* and outcome space Y.

Theorem 2 (RUM Learning Foundation)**.** *1. For every optimal Markovian policy π and $\alpha^* \in \Delta(A)$, and on $\mathbb{P}_{\mu_0,\pi}$-every sequence of histories such that $\lim_{t\to\infty} \alpha_t = \alpha^*$, the behavior converges to a RUM ζ. In particular, $\alpha^*(a) = c_\zeta(a, A)$ for all $a \in A$.*

2. For every RUM ζ and ϵ, there is a (Y, u, p^*, μ, k) such that under every optimal Markovian policy the behavior converges to a RUM ζ' such that $|c_\zeta(a, M) - c_{\zeta'}(a, M)| \leqslant \epsilon$ for all $M \in \mathcal{M}$ and $a \in M$.

The proof of the first part of Theorem 2 constructs the target random utility representation. To do this, we associate to every database d a strict preference relation in $\mathcal{P}$ where a is preferred to a' if and only if a is chosen by π from $\{a, a'\}$ conditional on $\mu(\cdot|d)$, and assign each ranking the limiting probability of the databases that induce it.[15] Lemma A.2 guarantees that the distribution over databases converges and the set of menus is finite, this pushforward measure also converges. The construction also yields an information representation in the sense of Lu [2016]: the agent maximizes a fixed utility function with respect to a distribution of posteriors over states. The proof of the second part constructs an environment in which outcomes are linear orders over actions and shows how to approximate the target RUM arbitrarily well by varying the probability distribution over $\mathcal{P}$ as the number of expected recalled experiences goes to infinity. (The proofs of this and the

a', where for any $M \subseteq A$, $BR(\nu|M) = \operatorname{argmax}_{a\in M} \int_\Theta \sum_{y\in Y} u(a, y)p_a(y)d\nu(p)$.

[15] Using information for binary comparisons alone will turn out to be sufficient because the agent uses a deterministic Markov policy to map beliefs to actions and a menu-independent tie-breaking rule.

next proposition are in Appendix A.3; the proofs of Propositions 3 and 4 are in the Online Appendix.)

Remark. The conclusion of Theorem 2 is false if we fix vector (Y, u, p^*, μ) and vary only the memory strength k: as the examples and applications throughout the paper illustrate, only a strict subset of $\Delta(A)$ can arise as stochastic memory equilibria. And if a choice frequency ζ cannot be matched in the complete menu A, it cannot be matched in the stronger sense of convergence to a RUM. See Example 1 in the Online Appendix for an illustration.

4.1.1 Monotonicity

For the next result, suppose that $Y \subseteq \mathbb{R}$, and thus the prior belief of the agent induces a (subjective) joint distribution $\tilde{\mu}_0^a$ over pairs $(\mathbb{E}_{p_a}[y], y)$.[16] The next result shows that when there are only two actions, the action with higher payoffs is played more frequently if the joint distributions are affiliated.[17] For simplicity, we also make the (generically satisfied) assumption that two actions are never indifferent after any observable database.

Proposition 2 (Monotonicity)**.** *Suppose that $A = \{a, a'\}$, $\mu_0 = \mu_0^a \times \mu_0^{a'}$, that for all $\hat{a} \in A$, the random variables $\mathbb{E}_{p_{\hat{a}}}[y]$ and y are affiliated under $\tilde{\mu}_0^{\hat{a}}$ and that $u(\hat{a}, y) = y$. If p_a^* increases in the sense of first-order stochastic dominance keeping $p_{a'}^*$ fixed, the frequency of a in the stochastic memory equilibria with the highest and lowest values of $\alpha(a)$ both increase.*[18]

Proposition 2 lets us connect the stochastic choice rule with the quality of the decisions, enabling predictions on how the agent's choices vary with the objective environment they face. Previous decision-theoretic models, such as Matějka and McKay [2015], have obtained the monotonicity in Proposition 2 from optimization of the signal structure by an agent who makes a one-time choice. In contrast, in our setting the agent repeatedly makes consumption decisions under a given information structure.

[16] Formally, for every measurable $C \subseteq \mathbb{R}$ and $c \in Y$, $\tilde{\mu}_0^a(C, c) = \int_{\{p : \mathbb{E}_{p_a}[y] \in C\}} p_a(c) d\mu(p)$.

[17] The expected and realized outcomes are affiliated with respect to the prior probability measure if for all $c, c' \in \mathbb{R}$ and $a \in A$ $\tilde{\mu}_0^a[\mathbb{E}_{p_a}[y] \geqslant c, y \geqslant c'] \geqslant \tilde{\mu}_0^a[\mathbb{E}_{p_a}[y] \geqslant c] \tilde{\mu}_0^a[y \geqslant c']$.

[18] Section B.7 in the Online Appendix shows there can be multiple equilibria here.

4.1.2 Specific Payoff Structures

For some specifications the fixed-point condition defining stochastic memory equilibrium admits an explicit solution, and the equilibria correspond to important stochastic choice models. This section gives two tractable examples, the normal and binomial cases. The section also connects stochastic memory equilibria with the Probit model (also in the normal environment) and the Elimination by Aspects model (Tversky [1972]).

Probit Probit (Thurstone [1927]) has two desirable features: payoff monotonicity and diminishing sensitivity.[19] With normally distributed payoffs our learning model delivers the related *mixed probit* (Hausman and Wise [1978], Greene [2000]) specification, which also has these two properties.[20] Moreover, the mixed probit we obtain also implies that the agent has more precise estimates of the values of actions they take more frequently, as found by Frydman and Jin [2022]. As Strzalecki [2025] (Example 1.15) points out, this frequency dependence is not accommodated by baseline probit.

In a *normal environment*, $u(a, y) = y$, and the payoffs of each action a are i.i.d. normally distributed with means $\bar{y}_a$ and variance σ^2. The agent knows σ^2, and their prior belief is that that $(\bar{y}_1, \ldots, \bar{y}_{|A|})$ are independently normally distributed with means $(y_1^0, \ldots, y_{|A|}^0)$ and common variance σ_0^2.[21]

Proposition 3. *In a normal environment, for every optimal Markovian policy π and every $\alpha^* \in \Delta(A)$, on $\mathbb{P}_{\mu_0,\pi}$-every sequence of histories such that $\lim_{t\to\infty} \alpha_t = \alpha^*$, the observed distribution of actions approaches a mixed probit distribution with mean parameter vector $\left(\frac{y_a^0/\sigma_0^2 + n_a \bar{y}_a/\sigma^2}{1/\sigma_0^2 + n_a/\sigma^2}\right)_{a\in A}$ and a diagonal covariance matrix with entries $\left(\frac{n_a/\sigma^2}{(1/\sigma_0^2 + n_a/\sigma^2)^2}\right)_{a\in A}$, where n_a has a Poisson distribution with parameter $\alpha^*(a)k$.*

In the mixed probit specification generated by our model, the variance of the payoff to a is stochastically decreasing (in the sense of first-order stochastic dominance) in $\alpha(a)$.

[19] Payoff monotonicity means that more preferable alternatives are more likely to be chosen. Probit satisfies payoff monotonicity when what is subject to the normal shock is the payoff of the alternatives; Thurstone [1927]'s original formulation was a bit different and doesn't necessarily satisfy payoff monotonicity. Diminishing sensitivity requires that, for a given difference between the alternatives, the better one is more likely to be chosen when the absolute desirability of both alternatives is lower. See Gescheider [2013] for empirical evidence about payoff monotonicity, diminishing sensitivity, and frequency dependence.

[20] Mixed probit is a hierarchical stochastic choice rule: First, the probit parameters (mean and covariance matrix) are drawn with some mixing probabilities, and the choice probabilities are determined as in probit.

[21] Here we allow infinitely many outcomes. Online Appendix B.8 extends our definitions and results to cover this.

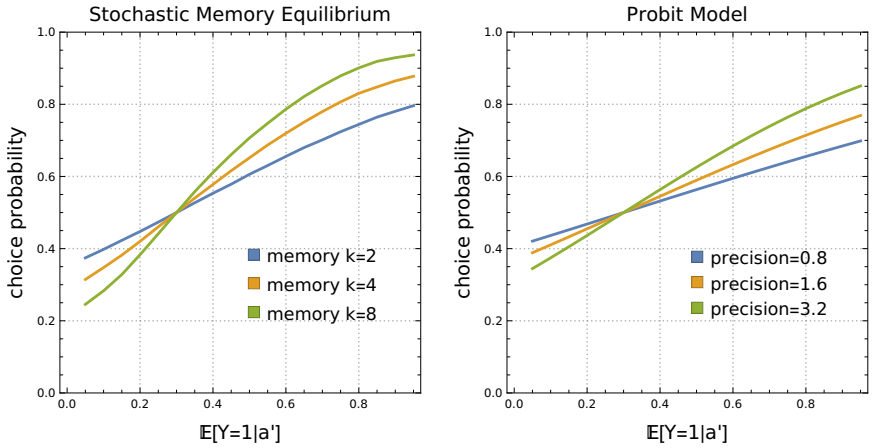


Figure 2: Binomial Beta vs. Probit. Left: Probability of choosing a' when $A = \{a', a''\}$, $Y = \{0, 1\}$, a symmetric beta prior with $\beta = \gamma = 1$, $k = 2$ (blue), $k = 4$ (orange), $k = 8$ (green), and a'' is known to give outcome 1 with probability .3. Right: Probit

Hence, the payoffs of more frequently chosen alternatives are more precisely estimated.[22]

Binomial Beta Model Another particularly tractable special case of our model is the combination of binary outcomes with a Beta prior. Suppose there are two outcomes $Y = \{0, 1\}$, that for each action a the the agent's prior belief about each success probability $p_a(1)$ is independently and identically beta distributed with parameters $\gamma, \beta \in \mathbb{R}_{++}$, so the posterior mean the agent assigns to action a given database d is

$$r_a(d) = \frac{\gamma + d(a, 1)}{\gamma + \beta + d(a, 1) + d(a, 0)}. \tag{5}$$

Assume also that $u(a, y) = y$. We assume that the agent uniformly randomizes over actions when they are indifferent. Then, given a database d the probability of choosing a is given as

$$\mathbb{P}[a \text{ is chosen} \mid d] = \frac{\mathbf{1}\{r_a(d) = \max_{a' \in A} r_{a'}(d)\}}{|\arg\max_{a' \in A} r_{a'}(d)|}.$$

As the distribution over databases given the long-run distribution over actions equals η_α a stochastic memory equilibrium is then a solution α to the fixed point equation

$$\alpha(a) = \mathbb{E}_{d \sim \eta_\alpha}\left[\frac{\mathbf{1}\{r_a(d) = \max_{a'} r_{a'}(d)\}}{|\arg\max_{a'} r_{a'}(d)|}\right].$$

[22] Danenberg and Spiegler [2026] makes a similar point about the relation between frequency and precision in a setting with exogenous normally distributed signals and no memory limitations.

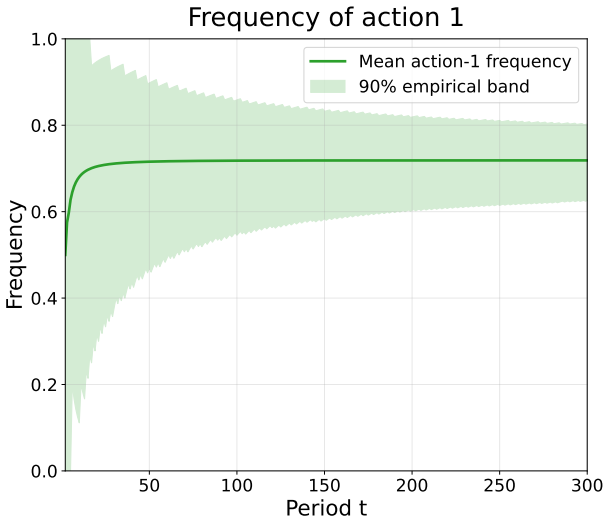


Figure 3: Belief dynamics in the binomial-beta model. Empirical frequency of action 1 across simulated histories. Parameters: $p_1^* = 0.6$, $p_2^* = 0.3$, $\gamma = \beta = 1/2$, and $k = 5$.

With endogenous data, the agent has more precise estimates of the payoff of actions that are more frequently chosen. This effect can be seen in Figure 2. Its left panel shows three values of k that mimic the probit model's behavior at three precision levels. The effect can also be directly observed from equation (5), which can be rearranged as

$$r_a = \frac{\gamma/n_a + d(a,1)/n_a}{(\gamma+\beta)/n_a + 1}.$$

Because $d(a,1)/n_a$ is the average payoff produced by action a, the agent's beliefs are less biased about actions they have taken (and thus recall) more often. Similarly, beliefs for actions the agent has taken more often will be more concentrated.

Despite the concentration, the distribution of posterior means remains diffuse, with substantial overlap in the 90% confidence bands for r_1 and r_2 even at long horizons. This highlights that behavior remains stochastic because the agent recalls only a finite number of experiences. To illustrate convergence to stochastic memory equilibrium, Figure 3 reports a Monte Carlo simulation of the binomial-beta environment with two actions, objective success probabilities $p_1^* = 0.6$ and $p_2^* = 0.3$, prior parameters $\gamma = \beta = 1/2$, and memory capacity $k = 5$. In each period, the agent recalls each past experience independently with probability $\min\{1, k/t\}$, computes the posterior means r_1 and r_2 from the recalled sample, and chooses the action with the higher posterior mean. Because action 1 is objectively

better, favorable recalled experiences with that action become more common, so, as the figure shows, the realized frequency of taking the optimal action $\alpha_t(1)$ tends to converge relatively quickly to the equilibrium.

Elimination by Aspects Tversky [1972] In Tversky [1972]'s *elimination by aspects* (EBA) model, the agent randomly orders attributes, successively eliminates alternatives that are not maximal on the current attribute, and continues until one alternative remains. EBA was designed to capture the following observed violation of IIA: Starting from a situation with two alternatives $\{a, a'\}$, the addition of a third alternative that is closer to a, without dominating or being dominated by a (i.e., not a "decoy" in the sense of Huber et al. [1982]) draws relatively more probability away from a than from a'. To fix ideas, consider an example from Tversky [1972]: suppose a manager needs to decide whether to hire a worker based on their intelligence and motivation scores. Adding a worker with (intelligence, motivation) scores $(78, 25)$ to a choice between $(75, 35)$ and $(60, 90)$ has been shown to remove significantly more choice probability from $(75, 35)$.

Although EBA lacks an axiomatic foundation, it can be derived from limited memory. To see how this works in the example above, suppose that outcomes are tasks in which either intelligence or motivation is the key feature driving the hired worker's performance. The manager's payoff depends on whether intelligence or motivation is relevant that period. The manager assigns a 50/50 prior that either motivation is relevant 90% of the time, or intelligence is. At the end of the period, the manager observes which factor was relevant for this period's task. Then, in every limit frequency, the probability that $(75, 35)$ is chosen over $(60, 90)$ equals the probability that the manager recalls more past periods in which intelligence was important. And, as predicted by EBA, the addition of $(78, 25)$ will reduce the probability of $(75, 35)$ but not that of $(60, 90)$.[23]

Our model predicts that more frequently relevant aspects are more likely to be used to determine choice, unlike Tversky [1972], where these probabilities are unrestricted. To see the general link between our model and EBA, define an *aspects environment* as one where $A \subseteq \{0, 1\}^{\mathcal{A}}$ for some finite set of aspects $\mathcal{A}$, $Y = \mathcal{A}$, $u(a, y) = a_y$, $p_a = p_{a'}$ for all $p \in \Theta$, $a, a' \in A$, and such that the prior is *responsive*: for every database $\sum_{a \in A} d(a, y) >$

[23]Let n_I and n_M be the number of recalled intelligence and motivation tasks, respectively, and observe that the posterior likelihood ratio between $(0.9, 0.1)$ and $(0.1, 0.9)$ is $9^{n_I - n_M}$. When $n_M > n_I$, the posterior probability of DGP $(0.9, 0.1)$ is no more than 0.1, and $(60, 90)$ is the unique best reply. And when $n_I > n_M$ the posterior probability of DGP $(0.9, 0.1)$ is at least 0.9, so $(78, 25)$ is the unique best reply.

$\sum_{a\in A} d(a, y') > \sum_{a\in A} d(a, y)$ and $a_y > a'_y$ implies that $\int_\Theta \mathbb{E}_p[u(a,\cdot)]d\mu(p|d) > \int_\Theta \mathbb{E}_p[u(a',\cdot)]d\mu(p|d)$.[24] The interpretation is that actions with more 1's have more desirable features, and the outcome is the period's most important feature.[25]

Corollary 1. *In an aspects environment, every limit frequency coincides with that of an EBA stochastic choice rule with aspects $\mathcal{A}$.*

4.2 The Effect of Skewness

Our next example illustrates how limited memory leads agents to overlook rare events, so that the probability of choosing an action is not determined by its expected payoff and can depend on higher-order moments of the utility function, such as variance and skewness. Suppose that the agent chooses between two actions. Action 0 always produces the outcome $c \in (0,1)$, e.g., $\mu(\{p : p_0(c) = 1\}) = 1$. The agent is uncertain whether action 1 follows process p (with prior probability q) or p'. Under p, action 1 yields $1/b$ with probability $b \in (0,1)$ and 0 otherwise; under p', it always yields 0.

The agent's payoff function is $u(a, y) = y$. Suppose $\frac{q(1-b)}{q(1-b)+(1-q)} < c < q$. The agent chooses action 1 if a success $(1/b)$ is recalled or if no outcomes are recalled (since the prior expected payoff is $q > c$), and chooses action 0 if only failures (0) are recalled. The first success reveals the true expected payoff of action 1 is $1 > c$, but limited memory makes such evidence rare.[26]

This setting generates the equilibrium probability of choosing action 1 displayed in Figure 4 as a function of the value of b under the actual data generating process $p^* = p$. Because rare events are unlikely to be recalled, they are absent from most databases. Consequently, actions with a given expected payoff tend to be chosen more often when they deliver a good payoff with a high probability than when they deliver a very good payoff more rarely.[27] This is consistent with the evidence in Hertwig et al. [2004], which also shows

[24] The latter condition is trivially satisfied when the prior is symmetric, and either there are two aspects as in the example above or each action has a unique strictly positive entry.

[25] Like Tversky [1972], we analyze a model with binary attributes but consider a multi-valued example. Our results extend to Gul et al. [2014]'s axiomatization of nonbinary EBA when (i) a single attribute is decision-relevant at a time and (ii) players observe noisy signals about attribute levels and their relevance.

[26] See Online Appendix B.5 for details.

[27] Ellison and Fudenberg [1993] makes this point in a model where each agent sees two signals. Conversely, when a rare event is recalled, it tends to be overrepresented in the database, which typically triggers an overreaction, but this cannot happen in our example due to its "perfect good news" structure. See Ba et al. [2024] for a recent theoretical and empirical analysis of which information structures tend to induce under-

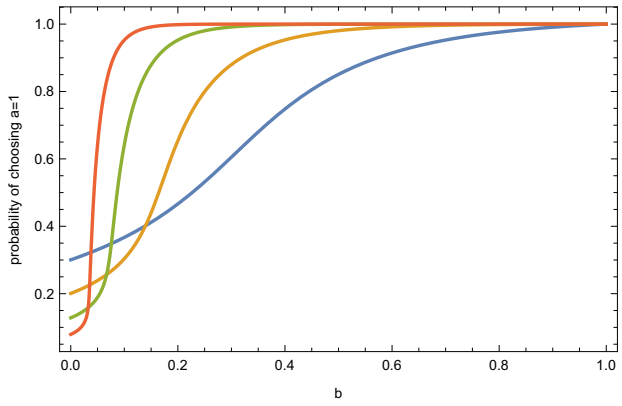


Figure 4: Effect of memory size depends on skewness. Probability of choosing $a = 1$ when $\mathbb{E}_{p^*}[u(1, y)] = 1$, and $k = 4$ (blue), $k = 8$ (orange), $k = 16$ (green), and $k = 32$ (red).

that, as our model predicts, this effect is not obtained if agents are told the probabilities instead of learning them from experience. Importantly, this occurs through the channel that unrepresentative samples are used to form the decision maker beliefs, as pointed out in Fox and Hadar [2006] and Wulff et al. [2018].

Figure 4 also illustrates that the effect decreases in the memory capacity k, and that in the limit $k \to \infty$, the agent chooses optimally.[28] In addition, the fact that *limited* memory tends to favor actions that are negatively skewed suggests an advantage for a form of *selective* memory that has been often observed: a higher probability of remembering more extreme observations can offset the excessive reliance on frequent and less extreme observations highlighted in Figure 4. Moreover, the benefit of selective memory depends critically on limited total memory; without this constraint, selective memory would only harm the agent (Fudenberg et al. [2024]).

4.3 The Effect of Reminders

Reminders are informational interventions that have a significant effect on behavior, for example, by helping people recall the positive impact of actions such as attending the gym or saving for the future.[29] Our model explains how reminders affect the agent's actions and why reminders about infrequent events have a greater effect than those about frequent ones. Thus, although our memory model predicts choices consistent with the random

or over-reaction.

[28] Theorem 3 in the next section gives a more general form of this observation.

[29] See, e.g., Karlan et al. [2016].

utility model in a fixed environment, its predictions about the effects of changes in the informational environment can differ substantially.

We interpret a reminder as leading the agent to remember a single past experience. Suppose the agent receives reminders about the activity as follows: First, the agent remembers past experiences according to our baseline model. Second, with probability $\phi \in (0,1)$, the agent is reminded and remembers one experience where they took the action $a = 1$, which is uniformly drawn at random (if such an experience exists).

For example, suppose that, in each period, the agent decides whether to go to the gym. If the agent does not engage in the activity ($a = 0$), they know they receive a payoff of 0. If they engage in the activity ($a = 1$), they pay an immediate cost $c \in (0,1)$ and, with some unknown probability $p_1(1)$, receive a benefit of 1 (e.g., feeling good about themselves after the workout). The agent's prior belief is that $p_1(1) = p > 0$ with probability $\pi \in (0,1)$ and $p_1(1) = 0$ with probability $1 - \pi$, and the true probability p^* is equal to p. We assume that $\pi p < c$ so that absent recall of a positive experience, the agent would not engage in the activity, and that $p > c$ so it is beneficial to engage in the activity. This simple example can be applied to any activity that yields uncertain benefits at certain costs. Another natural example is saving for future financial hardships, where the remembered benefit corresponds to past experiences in which savings helped the agent overcome an urgent financial need.

As we prove in Lemma B.1 in the Online Appendix, the stochastic memory equilibrium with maximal engagement in the activity is given by:[30]

$$\alpha_{\phi,p}(1) = \begin{cases} 0 & \text{if } \phi = 0 \text{ and } kp \leqslant 1 \\ 1 + \frac{1}{kp} W\left(-e^{-kp} kp(1 - p\phi)\right) & \text{otherwise} \end{cases}.$$

where $W : [-1/e, 0] \to (-1, 0]$ is the larger solution of $W(x)e^{W(x)} = x$. Intuitively, for $kp \leqslant 1$ if the agent takes action 1 with frequency $\alpha > 0$, then without reminders, they will remember the good outcome 1 with frequency less than α, so this cannot be an equilibrium. For $pk > 1$, the good outcome is sufficiently likely that if the agent takes the action $a = 1$ with some small probability α they remember the good outcome with probability greater α, so $a = 0$ cannot be a fixed point that is robust to trembles.

[30] There is always an equilibrium where the agent does not engage in the activity. We select the equilibrium where the agent engages in the activity if such an equilibrium exists because if the agent "trembles" and plays each action with at least a small probability ϵ, the stochastic memory equilibrium is unique and converges to the equilibrium where $\alpha(1) > 0$ as $\epsilon \searrow 0$. Thus, by Theorem 1, this equilibrium is the only possible limit frequency for $\epsilon \searrow 0$. See Lemma B.2 in the Online Appendix for details.

We have the following Proposition:

Proposition 4 (The effect of reminders)**.**

(i) The frequency $\alpha_{\phi,p}(1)$ *is increasing in the frequency of reminders* ϕ*.*

(ii) The effect of reminders $\alpha_{\phi,p}(1) - \alpha_{0,p}(1)$ *is decreasing in* $\alpha_{0,p}(1)$ *if* $\alpha_{0,p}(1) > 0$*.*

Figure 5 illustrates how the frequency of reminders increases the likelihood that the agent engages in the activity. A natural question is who is more affected by reminders:

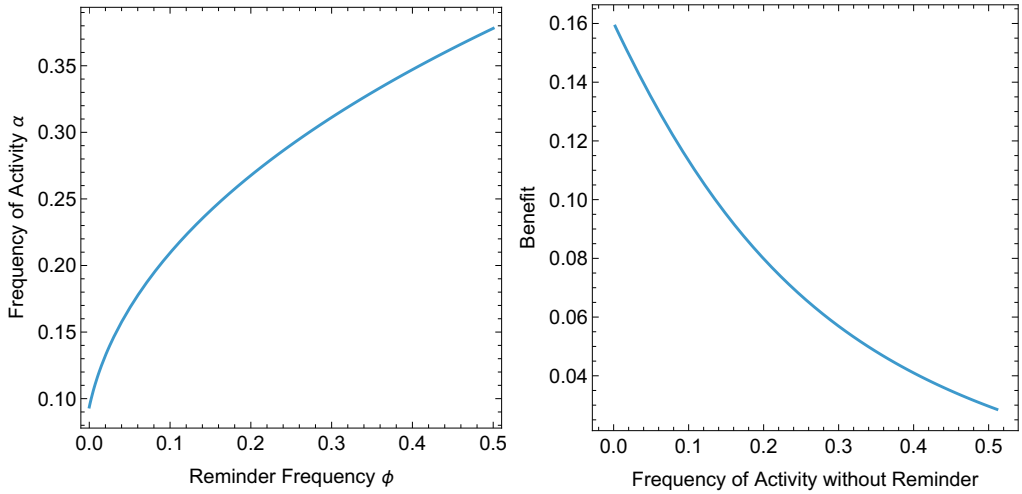


Figure 5: Reminders increase activity frequency most for agents with low baseline recall. Left: effect of reminders on activity frequency for $k = 7$, $p^* = 0.15$. Right: effect of reminders by baseline activity frequency.

people who already frequently engage in the activity without reminders or those who do not. Part (ii) of Proposition 4 speaks to that question by showing that an agent for whom $p = p^*$ is smaller benefits more from a higher frequency of reminders. Figure 5 illustrates that the more often a person engages in the activity without a reminder, the less their behavior is affected by a reminder about it. Intuitively, agents who frequently undertake the activity are likely to remember having benefited from it in the past, so when a single good outcome is enough to induce them to take the action, the reminder is likely to have no effect. This prediction aligns with the evidence presented in Calzolari and Nardotto [2017], which finds that reminders increase gym attendance among individuals with below-median attendance without reminders by 27%. In contrast, reminders have no measurable effect on those with above-median attendance without reminders. It is also consistent with the effect

of informational intervention documented in Augenblick et al. [2026], which shows that, for the same person, informational interventions about one's own spending (and thus the value of saving) are more effective when they include reminders of atypical expenditures. Finally, the effect of reminders in our model is driven by the finiteness of memory and vanishes as the agent's memory capacity k increases.

Remark. Instead of bringing to mind a random period when $a = 1$, reminders may bring to mind an experience in which action 1 was taken and the outcome was good. This would lead to the equilibrium $\alpha_{\phi,p}(1) = 1 + \frac{1}{kp} W\left(-e^{-kp} kp(1-\phi)\right)$. An alternative is that the agent observes the (potential) benefit of action 1 even if action 0 is taken. Our model allows us to describe the effect of reminders in either case.

4.4 Adding Alternatives and Choice Overload

This section shows how adding an objectively worse alternative leads an agent with limited memory to play the optimal action less frequently, even in settings where unlimited memory would lead to convergence on the best action, with or without the new alternative. Suppose $Y = \{0, 1\}$, $k = 1$, and $u(a, y) = y$. The agent believes that each action either always generates the good outcome 1 or always generates the bad outcome 0, so that using an action once perfectly reveals its data generating process. The agent's prior is uniform over $\Theta = \{(\delta_1, \delta_1), (\delta_1, \delta_0), (\delta_0, \delta_1), (\delta_0, \delta_0)\}$, so it is the product of independent distributions $(.5\delta_1, .5\delta_0)$. In reality, $p^* = (\delta_1, \delta_0)$, i.e., b is better than c. We suppose that the agent mixes uniformly when they do not recall any past experiences.

As we show in Online Appendix B.6, when $A = \{b, c\}$ the unique stochastic memory equilibrium where the agent chooses with equal probability b and c when they do not recall any experience has $\alpha(b) \sim 0.82$, i.e., the best action is chosen more frequently, but because of limited memory, with probability $\frac{1}{e}$ the agent does not recall any experience and then randomizes uniformly over both actions.

Suppose now that the environment is the same, with the only difference that $A = \{b, c, \hat{c}\}$, and the belief is again the product of independent distributions $(.5\delta_1, .5\delta_0)$, and both c and $\hat{c}$ are objectively bad.[31] In this case, in the unique stochastic memory equilibrium α^{new} where the agent chooses with equal probability if the alternatives are indifferent, we

[31] Therefore $\Theta = \{(\delta_1, \delta_1, \delta_1), (\delta_1, \delta_1, \delta_0), (\delta_1, \delta_0, \delta_1), (\delta_1, \delta_0, \delta_0), (\delta_0, \delta_1, \delta_1), (\delta_0, \delta_1, \delta_0), (\delta_0, \delta_0, \delta_1), (\delta_0, \delta_0, \delta_0)\}$, the prior is uniform over those models, and $p^* = (\delta_1, \delta_0, \delta_0)$.

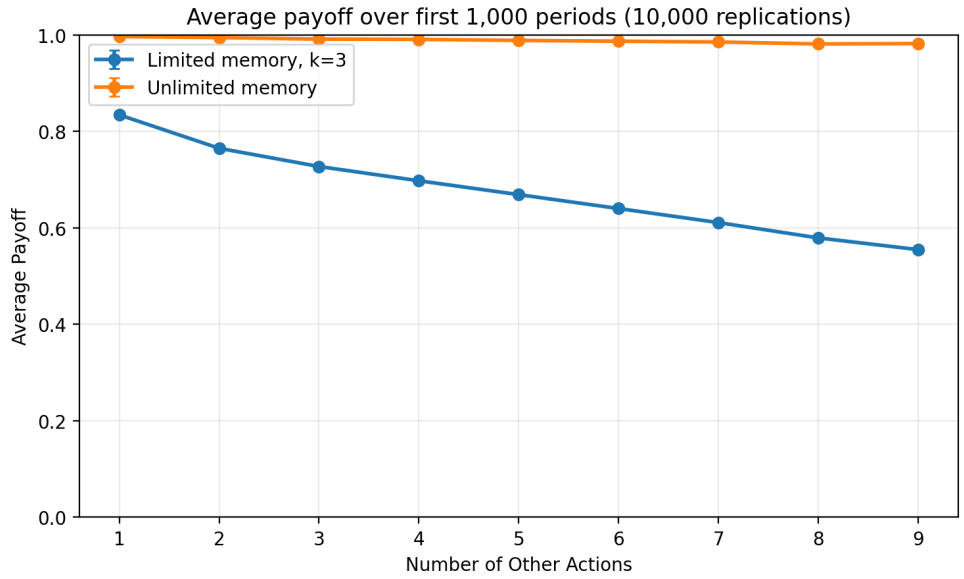


Figure 6: Limited vs unlimited memory: the prior mean return of each arm is normal with mean 0 and variance 1, and the realized mean return for the first arm is 1 and -1 for all others.

have $\alpha^{new}(b) \sim 0.694$, and $\alpha^{new}(c) = \alpha^{new}(\hat{c}) = 0.153$. That is, by adding an additional dominated option, the cognitive overload due to tracking the performance of all the options decreases the probability that the best option is taken, whereas the outcome under learning with perfect memory would be convergence to play b almost surely under both cases.

Assuming there is conclusive evidence allows a closed-form solution, but it is not needed. Figure 6 plots simulations of the probit of Section 4.1.2 as more bad actions are added. As in our analytical example, it has a good arm with an objective mean payoff of 1 and different numbers of bad alternatives, each with mean payoff of -1. Adding arms has basically no effect under unlimited memory, with the agent discovering and almost surely fully adopting the good arm. However, it significantly degrades long-run performance under limited memory, since the decision maker's limited cognitive resources must be devoted to tracking the performance of an increasing number of actions.

4.5 Underreaction to Evidence

Benjamin [2019] reports that underreaction of beliefs and insensitivity to sample size are among the most persistent departures from rationality in probabilistic reasoning. Underre-

action is often explained by the Phillips and Edwards [1966] underinference model. Limited memory also predicts underreaction to the data because the agent sometimes recalls only a small number of experiences. However, underinference predicts that a sufficiently long sequence of observations always leads beliefs to concentrate around the observed frequency. In contrast, our model predicts that the agent perceives uncertainty even in the limit, in line with Kahneman and Tversky [1972]'s "universal distribution" conditional on large samples. Also, our model suggests that underreaction will be more severe when people are shown data sequentially without written records of past outcomes, a prediction that the underinference model does not make.[32]

5 Almost Unlimited Memory

Since real-world memory has finite capacity, the perfect memory of standard models is, at best, an approximation. Here, we examine how well the perfect memory model captures the limits of stochastic memory equilibria as memory capacity k increases and the expected number of recalled experiences grows without bound. For every action distribution α, let

$$\Theta(\alpha) := \operatorname*{argmax}_{p \in \Theta} \left(\sum_{a \in A} \alpha(a) \sum_{y \in Y} p_a^*(y) \log p_a(y) \right)$$

be the set of models that maximize the log-likelihood of the true data generating process, where the weights depend on the frequency of each action. These are the models in the support of the agent's prior that exactly match the objective outcome distribution induced by α.[33] Consequently, when the outcome distribution identifies p^*, it is the unique element of $\Theta(\alpha)$.

Definition 3. A *(unitary-belief) self-confirming equilibrium* is an $\alpha \in \Delta(A)$ such that there is a $\nu^\alpha \in \Delta(\Theta(\alpha))$ such that $\alpha \in \Delta(\mathrm{BR}(\nu^\alpha))$.[34]

Unlike stochastic memory equilibrium, this concept only depends on the prior's support and not on the relative weights the prior assigns to various models.

[32] See Fudenberg and Peysakhovich [2016] and Esponda et al. [2024] for experimental evidence on the effect of providing agents with records and/or summary statistics.

[33] There is at least one such model because the agent's prior is correctly specified.

[34] This is called a unitary belief because a single belief is used to rationalize all of the actions in the support of α; the heterogeneous-belief version allows each action to be rationalized by different beliefs. See Section A.4 for more on heterogeneity.

Theorem 3. *Suppose* $(\alpha^k)_{k\in\mathbb{N}}$ *is a sequence of stochastic memory equilibria, each with memory capacity* k *and that* $\lim_{k\to\infty}\alpha^k = \hat{\alpha}$*. Then* $\hat{\alpha}$ *is a self-confirming equilibrium.*[35]

The first step of the proof is to show that when α^k converges, the distribution of databases converges as well. In particular, we show that the sizes of these databases have a vanishing probability of being smaller than any constant M, and that they are representative of the true frequencies. A standard argument then shows that the agent's beliefs concentrate on $\Theta(\hat{\alpha})$. The fact that $\hat{\alpha}$ is a self-confirming equilibrium then follows since each action $\tilde{a}$ for which $\hat{\alpha}(\tilde{a}) > 0$ is a best reply to some belief concentrated on $\Theta(\hat{\alpha})$. The last step is to show that because the agent is correctly specified, there is a single belief that makes every $a \in \text{supp}(\hat{\alpha})$ a best reply.

Theorem 3 exemplifies a broader principle: as memory capacity grows, agents gain a sharper long-run understanding of the consequences of their actions, which tends to improve their decisions. This is also illustrated by Figure 2 for the Binomial-Beta model, and formally established for the probit model in Proposition 3.

6 Rehearsal and Recency

We now extend the model to incorporate two additional psychological forces: *rehearsal* and *recency*. Here, rehearsal means that if an experience is recalled in one period, it is more likely to be recalled in subsequent periods, as in Kandel et al. [2000] and the references therein.[36] Recency is the idea that the agent gives more weight to more recent outcomes. There is an extensive psychology literature documenting various forms of the recency effect and proposing explanations for it; see e.g., Davelaar et al. [2005]. As in Mullainathan [2002], we study a very simple form of recency bias, where the previous period's experience is more likely to be remembered, while the periods before that do not receive an extra boost. To model rehearsal, we assume that the experiences recalled in the last period are also more likely to be recalled now.[37]

[35] If the agent is misspecified, they need not learn the path of play, so the limit outcome need not be self-confirming. Moreover, it need not be supportable by unitary beliefs; see Appendix A.4.

[36] Conlon [2026] finds substantial rehearsal effects, with an estimated baseline probability of recalling an instance of around 47% and an increase in the probability of being recalled due to rehearsal of 25%.

[37] We could extend the analysis to allow both recency and rehearsal to depend on a finite number of past periods, but allowing an unbounded number of past periods to matter would cause significant complications.

Formally, we now assume that the agent's memory at time $t+1$ is

$$m_{t+1}((a,y)|d_t,(a_t,y_t)) = \min\left\{1, \frac{k + r\mathbb{1}_{\{(a',y'):d_t(a',y')\geqslant 1\}\cup\{(a_t,y_t)\}}(a,y))}{t}\right\}, \tag{6}$$

where d_t is the database recalled in period t and $r \geqslant 0$ is the weight on rehearsal and recency; $r = 0$ reduces to the baseline model.[38] Thus the recalled periods at time $t+1$ given the previous period's database d_t and experience (a_t, y_t) are distributed as

$$\mathbb{P}\left[r_t = R \mid d_t,(a_t,y_t)\right] = \prod_{i\in R} m_{t+1}((a_i,y_i)|d_t,(a_t,y_t)) \prod_{i\in\{1,\ldots,t\}\setminus R} (1-m_{t+1}((a_i,y_i)|d_t,(a_t,y_t))), \quad \forall R \subseteq \{1,\ldots,t\}.$$

We continue to assume that, upon seeing a database, the agent updates their beliefs according to equation (2). Thus, when updating, they are also naive about the effect of rehearsal, which is consistent with the findings of Conlon [2026].

6.1 Ergodic Memory Equilibrium

Limit Distribution of Databases We next construct the relevant equilibrium notion in a way that parallels our approach in the baseline model. All proofs for this section are in Appendix A.5. When recent experiences or recently recalled memories are more memorable, the limiting distribution of databases becomes path-dependent: what is recalled today depends on what was recalled and happened yesterday. However, the Poisson approximation still applies, now as a stationary Markov chain. As before, we will show that after a long history with action frequency α, the limit distribution of databases is a product of Poisson distributions—but these distributions depend on the database recalled in the previous period, in addition to the action frequency and outcome probabilities. The key insight is that even with rehearsal and recency, the stationary distribution of the resulting Markov chain over databases is determined solely by α and p^*, which allows the same fixed-point characterization to extend to this setting.

Definition 4. For every $\alpha \in \Delta(A)$ the Markov chain η_α has state space $\mathcal{D}$ and Markov kernel $(\eta_{\alpha,d})_{d\in\mathcal{D}} = \left(\sum_{y'\in Y} p^*_{\pi(\mu(\cdot|d))}(y')\eta^{y'}_{\alpha,d}\right)_{d\in\mathcal{D}} \in \Delta(\mathcal{D})^{\mathcal{D}}$ where $\eta^{y'}_{\alpha,d}$ is the product of inde-

[38] Our model implies that more recent experience are more likely to be recalled, but are not more precisely recalled. This is consistent with the evidence surveyed in Section 4.1 of Gershman et al. [2025].

pendent Poisson distributions with parameter for $(a, y) \in A \times Y$ equal to[39]

$$\begin{array}{ll} \alpha(a)p_a^*(y)\,[k+r] & \text{if } d(a,y) \geqslant 1 \text{ or } (a,y) = (\pi(\mu(\cdot|d)), y') \\ \alpha(a)p_a^*(y)\,k & \text{otherwise.} \end{array}$$

Intuitively, conditional on the previous database d, the expected number of recalls of experience (a, y) is proportional to $\alpha(a)p_a^*(y)$. The proportionality factor is $k + r$ if the experience either occurred last period or appeared in d, and k otherwise. Lemma 2 shows that this Markov chain has a unique stationary distribution, and Corollary 2 in the Appendix shows it is the limiting time-average distribution of databases whenever action frequencies converge to α.

To see why uniqueness holds, note that any sub-database of the current recalled set can be recalled next with positive probability. In particular, the null database can be recalled in every period, so the chain is irreducible on the class of databases reachable from the empty database. A calculation also shows positive recurrence, giving the following lemma.

Lemma 2. *η_α admits a unique stationary distribution $\mathcal{H}_\alpha \in \Delta(\mathcal{D})$.*

Exactly as in the baseline model, the distribution over databases $\eta_{\alpha,d}$ induces, through Bayesian updating a distribution over beliefs $F^{\mu_0}_{\alpha,d}$; the only difference is that here both of them will depend on the database recalled in the previous period. For each $d \in \mathcal{D}$ and all measurable $\mathcal{C} \subseteq \Delta(\Theta)$,

$$F^{\mu_0}_{\alpha,d}(\mathcal{C}) = \eta_{\alpha,d}\left(\{d' : \mu(\cdot|d') \in \mathcal{C}\}\right). \tag{7}$$

Definition 5. An *ergodic memory equilibrium* is an $\alpha \in \Delta(A)$ such that there exists $\rho \in \mathcal{O}$ with $\alpha = \mathbb{E}_{\mathcal{H}_\alpha}[\mathbb{E}_{F^{\mu_0}_{\alpha,d}}[\rho(\nu)]]$.

Like stochastic memory equilibria, ergodic memory equilibria are fixed points, but here the relevant correspondence has an additional step that averages with respect to the stationary distribution of databases. For every database d, any mixed action α determines a probability distribution over what is recalled in the next period, and thus over the next period's beliefs. The agent's policy applied to those beliefs determines a mixed action $\alpha_d = \mathbb{E}_{F^{\mu_0}_{\alpha,d}}[\rho(\nu)]$; ergodic memory equilibrium requires that the expectation of α_d with respect to the induced stationary distribution over databases is α.

Proposition 5. *An ergodic memory equilibrium exists.*

[39] That is, the probability of a transition from d to d' is $\eta_{\alpha,d}(d')$.

The proof extends that of Proposition 1 by showing that the average of the best reply correspondence over the database with weights $\mathcal{H}_{(\cdot)}$ has the properties needed to appeal to a fixed-point theorem.

Theorem 4. *If α is a limit frequency, then α is an ergodic memory equilibrium.*

The proof of this theorem has three main steps. We first show that the transition matrices of the inhomogeneous Markov chain over databases converge as $t \to \infty$ and characterize their limit, and that the limiting chain is ergodic because the empty database is reachable in one step from any state. The second step generalizes Lemma A.2 on the convergence of beliefs conditional on the databases, with the key difference that the limit belief distribution is now database-dependent. The final step repeats the stochastic approximation argument from the proof of Theorem 1.

Effect of Rehearsal and Recency on Action Correlation Both rehearsal and recency bias can affect the correlation of actions. Consider first the effect of introducing a small amount of rehearsal in a setting with binary outcomes (e.g., success and failure). Rehearsal makes it such that if for a given action a at least one success is remembered today all past successes become more likely to be remembered tomorrow. If the likelihood of taking the action a increases in the number of remembered successes this implies that the action a becomes more likely both today and tomorrow when at least one success is remembered today which induces positive correlation of actions.

Under recency there are two competing effects. If an action a was taken today and led to a success the action becomes more likely tomorrow. If the action was taken today and let to a failure it becomes less likely tomorrow. As a consequence a small amount of recency bias leads to positive correlation if and only if the first effect dominates the latter. We illustrate the two effects in Figure 7 and prove the aforementioned result for the Binomial Beta model in the Online Appendix.

6.2 Applications of Memory Rehearsal to Finance

Our model of finite expected memory and rehearsal lets us generalize the findings of Mullainathan [2002] about income forecasts beyond the specific parametric structure it assumes. It also lets us provide a novel memory-based explanation of the equity-premium puzzle. In this subsection, we specialize to an exogenous forecasting problem: each period the agent

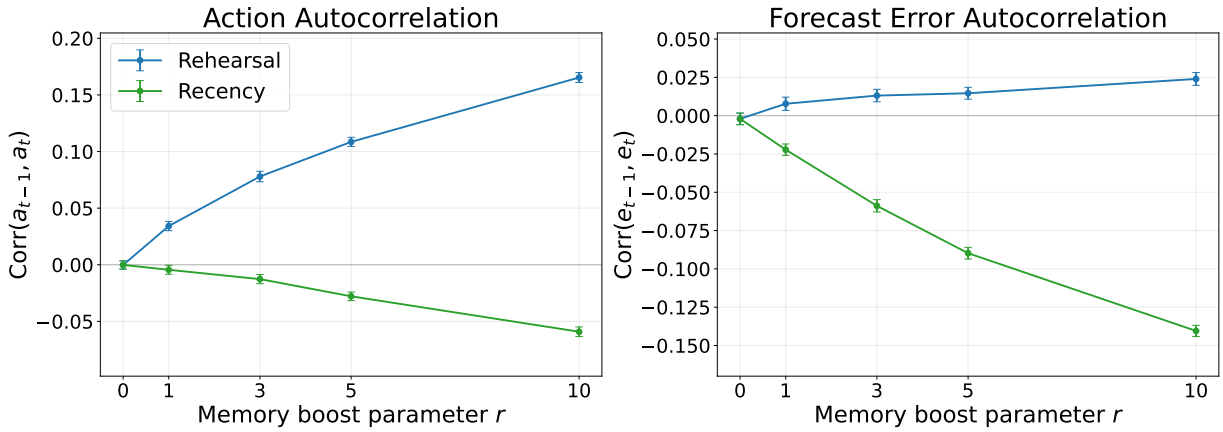


Figure 7: Autocorrelation of actions in the Binomial beta case with two actions, $k = 5$, $\gamma = \beta = 0.5$, and expected payoffs of $\mathbb{E}[y|a = 1] = 0.6$ and $\mathbb{E}[y|a = 2] = 0.3$.

observes a public outcome y_t, independent of their action, and $y_t = \theta + \epsilon_t$, where $\theta \in \mathbb{R}$ and the ϵ_t are mean-0 i.i.d. shocks.[40]

Correlated Prediction Errors The rehearsal memory function (6) generates the same predictions about one-period correlations as Mullainathan [2002], without assuming a specific functional form. First, under recency a high outcome in the last period makes that particular high outcome more likely to be recalled next period, so the forecasting error will be negatively correlated with the most recent information.[41] Second, when the baseline probability of remembering an event is low, and the rehearsal effect is strong, the forecast errors in successive periods are positively correlated for the same reason as in Mullainathan [2002]: memories that are remembered are more likely to be remembered again.

Asset Pricing Suppose that each of a continuum of risk-neutral agents indexed by $x \in [0, 1]$ has a constant per-period amount w to invest. Each period, each agent decides whether to buy, sell, or not trade a unit of a representative equity portfolio and to invest the wealth net of the expenditure/revenues from the risky asset in the risk-free asset.[42] The

[40]Both Mullainathan [2002] and Weitzman [2007] assume that outcomes follow an AR1 process. Our assumption of finite expected memory has the same implication even in an i.i.d. setting.

[41]Mullainathan [2002] supposes that y has a positive density on the real line so that some form of associativeness is needed for rehearsal to have any effect.

[42]Under our assumption of risk neutrality, the agent can do no better than invest all income in a single asset. We restrict to this case to directly apply our results, which assume finite actions.

safe asset has net return $i \in [-1, \infty)$ per period, while the risky asset provides per-period net return $i + \theta^T + \varepsilon_t$, where from the point of view of the agents θ^T is a random variable θ with unknown distribution and ε is a symmetric, zero-mean, and period-specific shock. The risky asset is in net-zero supply and prices p_0 and p_1 are determined by market clearing: $p_1 - p_0 = \text{Median}[\mathbb{E}_x[\theta]]$. The agents' actions impact their payoffs, but all agents observe the sequence of realized prices and returns.

In this setting, the equity premium puzzle is that the observed price of the risky asset is much lower than predicted by the above equation if the distribution of θ were known and equal to that observed in the data. Reconciling this gap requires implausibly large risk aversion. Weitzman [2007] explains it with the combination of an overly pessimistic prior and the assumption that the agent believes θ changes over time, so they discard old observations. Ergodic memory equilibrium predicts the same effect even with a perceived constant θ and with risk neutrality.[43]

The proof of Theorem 4 shows that if the action distribution converges to α the distribution of recalled experiences converges to $\mathcal{H}_\alpha$. In particular, even in the long run, the agents will rely on a limited number of observations. Any Dirichlet prior with predictive mean $\underline{\theta} < \theta^T$ can sustain the premium: the distribution of experiences $\mathcal{H}_\alpha$ centered around θ^T combined with the prior centered around $\underline{\theta}$ yields a median posterior expected value of θ strictly below θ^T. This explanation is consistent with empirical evidence that investors hold systematically overpessimistic beliefs about stock returns, rather than correct beliefs with extreme risk aversion (see, e.g., Dominitz and Manski [2007] for a systematic analysis). It is also consistent with the observation that investors tend not to use available historical return sequences but rather rely on their lived experience (see Malmendier and Nagel [2016] and the references therein).

7 Conclusion

This paper provides a simple and general model of limited memory that can be applied to many economic problems. In particular, the role of memory in belief formation is important for behavioral economics and macroeconomics, so our work will be useful there. The paper shows how limited memory can serve as a foundation for several well-known stochastic

[43] A memory function that is more likely to recall negative stock performance would create an additional force towards the equity premium puzzle.

choice models. It characterizes how asymptotic outcomes with limited memory relate to those with unlimited memory capacity.

Our model was designed to highlight the effects of limited and stochastic memory, so it left out many other important aspects of how memories are formed and used. However, our analysis can be extended in several interesting directions. Figure 1 provides a blueprint for how to do this, namely by modifying one or more of the maps in the composition of functions that define $\psi_\rho^{\mu_0}(\alpha)$. Proceeding counterclockwise from the final step, relaxing subjective expected utility amounts to replacing BR with another correspondence in the definition of the admissible policies ρ, and allowing non-Bayesian updating permits other maps from the distribution of databases to the distribution of beliefs.

One important class of generalizations is to other forms of imperfect memory, corresponding to different maps from histories to databases than those considered in Lemma 1. Section 6 demonstrates one way of doing this, namely, making the map into a Markov chain whose state variable is what was recalled in the previous period. Other ways to generalize the map from histories to databases include introducing selective memory, as in Fudenberg et al. [2024], and allowing the probability of recalling an experience to depend on the number of prior occurrences. More speculatively, if one modifies the objective data-generating process from a conditionally (on actions) i.i.d. process to a Markov process over outcomes, what ultimately changes is the distribution of databases, as characterized by Lemma A.1. Our framework enables these generalizations, provided that the correspondences involved in the various steps remain well-behaved.

It is easy to generalize the model to allow memory to be cued by an exogenous *signal* $s \in S$ that the agent observes before choosing an action, as in Fudenberg et al. [2024]. The data-generating process becomes $p^*_{s,a} \in \Delta(Y)$, an experience is a triple (s, a, y), and the constant recall probability $m_{t+1} = \min\{1, k/t\}$ is replaced by a signal-dependent memory function where $km_{s'}(s, a, y)/t$ is the probability that experience (s, a, y) is recalled when the $t + 1$ period signal is s'. Independence across periods is preserved, so the recalled database $d_{s'} \in \mathbb{N}^{S\times A\times Y}$ remains a sufficient statistic for beliefs and, by the same Poisson-limit argument as Lemma 1, has limit distribution with mean $\mathbb{E}[d_{s'}(s, a, y)] = k\,\zeta(s)\,\alpha(a \mid s)\,p^*_{s,a}(y)\,m_{s'}(s, a, y)$, where ζ is the signal distribution and $\alpha(\cdot \mid s)$ is the action frequency conditional on s. This formulation accommodates two natural interpretations of "context": the signal can encode the *set of available actions* , and it can represent the widely documented associativeness effect (see, e.g., Kahana, 2012) with exogenous *cues* such as

macroeconomic regime or location that makes some past experiences more or less likely to be recalled. Although associativeness does not change the set of limit points with unbounded memory (Fudenberg et al. [2024]), we conjecture that under limited memory it might lead the agent to behave as if they were using a kernel estimator.[44]

A Appendix

A.1 Preliminaries

Let $(\upsilon_t)_{t\in\mathbb{N}} \in \Delta(A\times Y)^{\mathbb{N}}$ be a sequence of empirical joint distributions over actions and outcomes. The next lemma shows that almost surely, if the action frequency converges to some α^*, then the joint empirical distribution of actions and outcomes converges to the distribution where each pair (a,y) has frequency $\alpha^*(a)p_a^*(y)$. Let d_t denote the random variable describing the database recalled at time t, and μ_t denote the random (beginning of) period-t belief induced by d_t.

Lemma A.1. *We have*

$$\mathbb{P}_{\mu_0,\pi}\left[\max_{(a,y)\in A\times Y}\limsup_{t\to\infty}|\upsilon_t(a,y)-\alpha_t(a)p_a^*(y)|\neq 0\right]=0$$

and in particular for any $\alpha^*\in\Delta(A)$,

$$\mathbb{P}_{\mu_0,\pi}\left[\lim_{t\to\infty}\alpha_t=\alpha^* \text{ and } \max_{(a,y)\in A\times Y}\limsup_{t\to\infty}|\upsilon_t(a,y)-\alpha^*(a)p_a^*(y)|\neq 0\right]=0.$$

Lemma A.2. *For $\mathbb{P}_{\mu_0,\pi}$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$ if $\lim_{t\to\infty}\alpha_t=\alpha^*$, the distribution of μ_t given h_{t-1} weakly converges to $F^{\mu_0}_{\alpha^*}$, and $F^{\mu_0}_{(\cdot)}$ is continuous in α.*

The proofs of Proposition 1 and Theorem 1 use the correspondence $\Psi^{\mu_0}:\Delta(A)\rightrightarrows\Delta(A)$ defined by $\Psi^{\mu_0}(\alpha)=\{\psi^{\mu_0}_\rho(\alpha):\rho\in\mathcal{O}\}$.

Lemma A.3.

1. *Ψ^{μ_0} is non-empty valued;*
2. *Ψ^{μ_0} is closed valued;*
3. *Ψ^{μ_0} is upper hemicontinuous;*
4. *Ψ^{μ_0} is convex valued;*
5. *$\alpha'\in\Delta(A)$ is a stochastic memory equilibrium if and only if $\alpha'\in\Psi^{\mu_0}(\alpha')$.*

[44] Relatedly, Ilut and Valchev [2023] studies the learning traps that can arise under associative reasoning.

A.2 Proofs for Section 3

Proof of Lemma 1. By Lemma A.1, $\mathbb{P}_{\mu_0,\pi}$ assigns probability 0 to the sequences of histories $(h_t)_{t\in\mathbb{N}}$ in which $\limsup_{t\to\infty} |\upsilon_t(a,y) - \alpha_t(a)p_a^*(y)| \neq 0$ for at least one $(a,y) \in A \times Y$. We prove that the stated convergence holds on every sequence of histories $(h_t)_{t\in\mathbb{N}}$ that is outside of that null set.

The database at time $t \geqslant k$ is distributed as a product of binomial distributions. If $d(a,y) > \upsilon_t(a,y)t$ for some $(a,y) \in A\times Y$, then $\mathbb{P}_{\mu_0,\pi}[d_t = d] = 0$. Otherwise, $\mathbb{P}_{\mu_0,\pi}[d_t = d] = \prod_{(a,y)\in A\times Y} \binom{\upsilon_t(a,y)t}{d(a,y)} (k/t)^{d(a,y)}((t-k)/t)^{\upsilon_t(a,y)t-d(a,y)}$, where for every $x \geqslant y \geqslant 0$, $\binom{x}{y}$ denotes the binomial coefficient between x and y. Set $E := A \times Y$, and, for $z = (a,y) \in E$, define $N_t(z) := t\upsilon_t(a,y)$, and $q_t(z) := \alpha_t(a)p_a^*(y)$. Conditional on h_t, the coordinates of d_{t+1} are independent and

$$d_{t+1}(z) \sim \text{Bin}\left(N_t(z), \frac{k}{t}\right).$$

Let

$$\widetilde{\eta}_t := \bigotimes_{z\in E} \text{Pois}\big(k\upsilon_t(z)\big).$$

Le Cam's inequality and the product inequality for total variation give

$$\|\mathcal{L}(d_{t+1} \mid h_t) - \widetilde{\eta}_t\|_{TV} \leqslant \sum_{z\in E} \left\|\text{Bin}\left(N_t(z), \frac{k}{t}\right) - \text{Pois}\left(\frac{kN_t(z)}{t}\right)\right\|_{TV} \leqslant 2\sum_{z\in E} N_t(z)\left(\frac{k}{t}\right)^2 = \frac{2k^2}{t},$$

where the last equality uses $\sum_{z\in E} N_t(z) = t$.

By definition,

$$\eta_{\alpha_t} = \bigotimes_{z=(a,y)\in E} \text{Pois}\big(kq_t(z)\big).$$

For Poisson random variables,

$$\|\text{Pois}(\lambda) - \text{Pois}(\lambda')\|_{TV} \leqslant 1 - e^{-|\lambda-\lambda'|} \leqslant |\lambda - \lambda'|.$$

Applying this inequality coordinatewise and again using the product inequality yields

$$\|\widetilde{\eta}_t - \eta_{\alpha_t}\|_{TV} \leqslant \sum_{z\in E} |k\upsilon_t(z) - kq_t(z)| = k \sum_{(a,y)\in A\times Y} |\upsilon_t(a,y) - \alpha_t(a)p_a^*(y)|.$$

Consequently,

$$\|\mathcal{L}(d_{t+1} \mid h_t) - \eta_{\alpha_t}\|_{TV} \leqslant \frac{2k^2}{t} + k \sum_{(a,y)\in A\times Y} |\upsilon_t(a,y) - \alpha_t(a)p_a^*(y)|. \tag{8}$$

On every history outside the null set identified at the beginning of the proof, the right-hand

side converges to zero by Lemma A.1. Hence

$$\|\mathcal{L}(d_{t+1} \mid h_t) - \eta_{\alpha_t}\|_{TV} \longrightarrow 0.$$

In particular, for every bounded $f : \mathcal{D} \to \mathbb{R}$,

$$\left|\int_{\mathcal{D}} f \, d\mathcal{L}(d_{t+1} \mid h_t) - \int_{\mathcal{D}} f \, d\eta_{\alpha_t}\right| \leqslant 2\|f\|_\infty \, \|\mathcal{L}(d_{t+1} \mid h_t) - \eta_{\alpha_t}\|_{TV} \longrightarrow 0.$$

This proves the lemma. □

Lemma A.4. *For $\mathbb{P}_{\mu_0,\pi}$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$, the distribution of actions given h_t converges to $\psi_\pi^{\mu_0}(\alpha_t)$.*

The proofs of Theorems 1 and 4 use a continuous-time approximation of the process of empirical frequencies. Set $\alpha_0 \colon = \alpha_1$, $\tau_0 \colon = 0$, and $\tau_t \colon = \sum_{i=1}^{t} \frac{1}{i}$ for all $t \in \mathbb{N}$. Following Benaim et al. [2005], define the continuous-time interpolation of $(\alpha_t)_{t\in\mathbb{N}}$ to be the function $w : \mathbb{R}_+ \to \Delta(A)$

$$w(\tau_t + c) = \alpha_t + c\frac{\alpha_{t+1} - \alpha_t}{\tau_{t+1} - \tau_t}, \quad \forall t \in \mathbb{N}, \forall c \in \left[0, \frac{1}{t+1}\right]. \tag{9}$$

Proof of Theorem 1. We want to show that if α is not a stochastic memory equilibrium, it is not a limit frequency. To do this, we extend Esponda et al. [2021a]'s application of Benaim et al. [2005]'s stochastic approximation techniques for differential inclusion to settings where beliefs remain stochastic in the limit. In particular, we will show that (9) converges to a solution of

$$\dot{\boldsymbol{\alpha}}(t) \in \Psi^{\mu_0}(\boldsymbol{\alpha}(t)) - \boldsymbol{\alpha}(t). \tag{10}$$

A solution to (10) with initial point $\alpha^* \in \Delta(A)$ is a mapping $x : \mathbb{R}_+ \to \Delta(A)$ that is absolutely continuous over compact intervals, with $x(0) = \alpha^*$, and (10) satisfied for almost every t. For every $T \in \mathbb{N}$ and $\alpha_0 \in \Delta(A)$, let $X^T_{\alpha^*}$ be the set of solutions to (10) over $[0, T]$ with initial conditions α^*, and let $X^T = \bigcup_{\alpha_0\in\Delta(A)} X^T_{\alpha^*}$.

Now we show that the continuous-time interpolation of α defined in (9) can, in the long run, be approximated arbitrarily well by a solution to (10). Recall that π is the optimal Markovian policy employed by the agent, and that for any measurable selection ρ from the best response correspondence, $\psi_\rho^{\mu_0}(\alpha) = \int_{\Delta(\Theta)} \rho(\nu) dF_\alpha^{\mu_0}(\nu)$. Define the stochastic process $\tilde{U}_t = t(\alpha_t - \frac{t-1}{t}\alpha_{t-1}) - \psi_\pi^{\mu_0}(\alpha_{t-1}) = \delta_{a_t} - \psi_\pi^{\mu_0}(\alpha_{t-1})$ and let $b_t = \mathbb{E}[\tilde{U}_t | h_{t-1}]$, $U_t = \tilde{U}_t - b_t$.

Then

$$\alpha_{t+1} - \alpha_t - \frac{U_{t+1} + b_{t+1}}{t+1} = \frac{\psi_\pi^{\mu_0}(\alpha_t) - \alpha_t}{t+1}. \tag{11}$$

Since both $\psi_\pi^{\mu_0}(\alpha_t)$ and α_{t+1} are uniformly bounded, U_t is a uniformly bounded martingale difference process. Moreover, by Lemma A.4, b_{t+1} converges to 0 on $\mathbb{P}_{\mu_0,\pi}$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$ where $\lim_{t\to\infty}\alpha_t = \alpha$, so condition (i) of Proposition 1.3 in Benaim et al. [2005] is satisfied (setting $\gamma_n = 1/n$ in their formula) by Remark 4.5 in Benaim and Hirsch [1999]. Condition (ii) is also satisfied because $||\alpha_{t+1} - \alpha_t||_\infty < 1/(t+1)$, w is Lipschitz continuous of order 1, and α_t is uniformly bounded because it takes values in $\Delta(A)$. Therefore, since w is the interpolated process for the stochastic process in equation (11), it is a perturbed solution of (10). Thus, by Theorem 4.2 in Benaim et al. [2005], on $\mathbb{P}_{\mu_0,\pi}$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$ where $\lim_{t\to\infty}\alpha_t = \alpha$[45]

$$\lim_{t\to\infty}\inf_{\tilde{\boldsymbol{\alpha}}\in X^T}\sup_{s\in[0,T]}||w(t+s)-\tilde{\boldsymbol{\alpha}}(s)|| = 0 \qquad \forall T\in\mathbb{N}. \tag{12}$$

If α is not a stochastic memory equilibrium, then parts 1,2, and 4 of Lemma A.3 and the separating hyperplane theorem (see, e.g., Section 14.5 of Royden and Fitzpatrick [2010] for the version used here) guarantee that there exists $f\in\mathbb{R}^A$ with $\alpha\cdot f > \sup_{\bar\alpha\in\Psi^{\mu_0}(\alpha)}\bar\alpha\cdot f$. By part 4 of Lemma A.3, $\alpha\cdot f > \max_{\bar\alpha\in\Psi^{\mu_0}(\alpha)}\bar\alpha\cdot f$. Let $K = \alpha\cdot f - \max_{\bar\alpha\in\Psi^{\mu_0}(\alpha)}\bar\alpha\cdot f$. By part 3 of Lemma A.3, there exists $\varepsilon\in\mathbb{R}_{++}$ such that for all $\alpha'\in B_\varepsilon(\alpha)$, $\max_{\bar\alpha\in\Psi^{\mu_0}(\alpha')}\bar\alpha\cdot f < \max_{\bar\alpha\in\Psi^{\mu_0}(\alpha)}\bar\alpha\cdot f + K/4$ and $\alpha'\cdot f > \alpha\cdot f - K/4$. Therefore, for every initial condition $\alpha_0\in B_\varepsilon(\alpha)$ and every solution in $X^T_{\alpha*}$, $\boldsymbol{\alpha}(t)\cdot f$ decreases at rate at least $K/2$ until the solution leaves $B_\varepsilon(\alpha)$. So there exists $T\in\mathbb{N}$ such that for every initial condition $\alpha_0\in B_\varepsilon(\alpha)$ every solution $\tilde{\boldsymbol{\alpha}}$ in $X^T_{\alpha*}$ leaves $B_\varepsilon(\alpha)$ by time T, that is,[46]

$$\sup_{\tilde{\boldsymbol{\alpha}}\in X^T_{\alpha*}}\inf\{t:\tilde{\boldsymbol{\alpha}}(t)\notin B_\varepsilon(\alpha)\}\leqslant T \qquad \forall\alpha_0\in B_\varepsilon(\alpha). \tag{13}$$

To conclude the proof, we will show that α_t does not asymptotically lie in an $\varepsilon/2$ ball around α on any path $(h_t)_{t\in\mathbb{N}}$ where (12) applies. Since the set of samples where (12) does not apply has 0 probability under $\mathbb{P}_{\mu_0,\pi}$, this implies that α is not a limit frequency. If there is no $\hat T\in\mathbb{N}$ such that $w(c)\in B_{\varepsilon/2}(\alpha)$ for all $c>\hat T$, $(\alpha_t)_{t\in\mathbb{N}}$ does not converge to α. So let $\hat T\in\mathbb{N}$ be such that on the chosen path $(h_t)_{t\in\mathbb{N}}$, $w(c)\in B_{\varepsilon/2}(\alpha)$ for all $c>\hat T$ and

[45]The proof of Theorem 4.2 in Benaim et al. [2005] invokes an implication of their Theorem 4.1 that is not correct. However, the weaker statement we are invoking is correct, as shown by equation (3) in Esponda et al. [2021b].

[46]To see that T can be taken to be the same for every $\alpha_0\in B_\varepsilon$ let $C = \max_{\alpha'\in B_\varepsilon(\alpha)}\alpha' f - \min_{\alpha'\in B_\varepsilon(\alpha)}\alpha' f$ and take $T = 2C/K+1$.

$\inf_{\tilde{\boldsymbol{\alpha}} \in X^T} \sup_{0 \leqslant s \leqslant T} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}(s)|| < \varepsilon/4$, and take $\tilde{\boldsymbol{\alpha}} \in X^T$ with

$$\sup_{0 \leqslant s \leqslant T} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}(s)|| < \varepsilon/4. \tag{14}$$

Then (13) implies that every solution to differential inclusion leaves $B_\varepsilon(\alpha)$ at least once between $\hat{T}$ and $\hat{T}+T$, and by (14), α_t must leave $B_{\varepsilon/2}(\alpha)$ at least once between $\hat{T}$ and $\hat{T}+T$. This proves Theorem 1. □

Proof of Proposition 1. By point 5 of Lemma A.3, every fixed point of Ψ^{μ_0} is a stochastic memory equilibrium. By points 2 and 3 of Lemma A.3 and the closed-graph theorem (e.g., Theorem 17.11 in Aliprantis and Border [2013]), Ψ^{μ_0} has a closed graph. The Lemma also shows it is non-empty valued (Point 1) and convex valued (Point 4), so, since $\Delta(A)$ is a non-empty, closed, convex and bounded subset of $\mathbb{R}^{|A|}$, it admits a fixed point by the Kakutani fixed point theorem (see, e.g., page 283 in Franklin [2002]). □

A.3 Proofs for Section 4.1

Proof of Theorem 2. 1) Suppose that α_t converges to some α^*. We first construct the target random utility representation ζ and then prove that α_t approaches it. Consider the map $G : \mathcal{D} \to \mathcal{P}$ such that $aG(d)a'$ if and only if $a = \pi(\mu(\cdot|d) \mid \{a, a'\})$. The map is well-defined because the binary relation is total by definition, and it is transitive because of the assumption that ties are broken in a menu-independent way.[47]

Define ζ by $\zeta(P) = \eta_{\alpha^*}(G^{-1}(P))$, and let c_η be the associated stochastic choice. Because the sets of menus and actions are finite, the proposition can be established by showing that the probabilities that any given action a is chosen from any given menu M converge to those of c_ζ. Fix $M \in \mathcal{M}$ and $a \in M$. Lemma 1 shows that for $\mathbb{P}_{\mu_0,\pi}$ almost any sequence of histories such that $\mathbb{P}_{\mu_0,\pi}[\lim_{t\to\infty} \alpha_t = \alpha^*] > 0$, the probability that the number of recalled (a, y) experiences is equal to c for every $c \in \mathbb{N}$ converges to its probability under a Poisson random variable with parameter $\alpha^*(a)p^*_a(y)k$, and these distributions are independent across (a, y) pairs. Therefore, as $\tau \to \infty$, the probability distribution η^{h_τ} over databases recalled at period $\tau+1$ given h_τ converges to η_{α^*}.

To link this with the convergence of the stochastic choice rule, let $D_M(a) = \{d \in \mathcal{D} : \forall a' \in M, aG(d)a'\}$. Because the agent's tiebreaking is menu-independent, the difference

[47] If for some $d \in \mathcal{D}$ we have $a, a', a'' \in A$ such that $aG(d)a'$, $a'G(d)a''$, and $a''G(d)a$, and (without loss of generality) $a \in \pi(\mu(\cdot|d) \mid \{a, a', a''\})$. Then $a'', a \in BR(\mu(\cdot|d)|\{a, a', a''\}) \cap \{a', a''\}$, by optimality of π, $a \in \pi(\mu(\cdot|d) \mid \{a, a', a''\})$ by $a'G(d)a''$, and $a \in \pi(\mu(\cdot|d) \mid \{a, a''\})$, a contradiction.

between $c_\zeta(a,M)$ and $\mathbb{P}_{\mu_0,\pi}[a_{\tau+1}=a|h_\tau,M]$ is bounded by $|\eta_{\alpha^*}(D_M(a))-\eta^{h_\tau}(D_M(a))|$. For every $l\in\mathbb{N}$, this is less than

$$\left|\eta_{\alpha^*}\left(d\in D_M(a):\sum_{(a,y)\in A\times Y}d(a,y)>l\right)-\eta^{h_\tau}\left(d\in D_M(a):\sum_{(a,y)\in A\times Y}d(a,y)>l\right)\right|$$
$$+\left|\eta_{\alpha^*}\left(d\in D_M(a):\sum_{(a,y)\in A\times Y}d(a,y)\leqslant l\right)-\eta^{h_\tau}\left(d\in D_M(a):\sum_{(a,y)\in A\times Y}d(a,y)\leqslant l\right)\right|.$$

Both addends can be made arbitrarily small by taking l and τ large.

This establishes the random utility representation.

To see that this stochastic choice rule admits an information representation in the sense of Lu [2016] with posterior distribution $F^{\mu_0}_{\alpha^*}$, identify the state space with Θ, acts with A, and outcomes with $Z=\{u(a,y):(a,y)\in A\times Y\}$; map each action a to the Anscombe-Aumann act assigning state p the lottery with mass $p_a(\{y:u(a,y)=\mathfrak{u}\})$ on each $\mathfrak{u}\in Z$. Finally, $\alpha^*(a)=c_\zeta(a,A)$ follows by the martingale SLLN (see, e.g., Theorem 2.7 of Hall and Heyde [2014]).

2) Consider an arbitrary RUM ζ on the set of alternatives A. Since we only need to approximate it to an ε term, we can assume it assigns strictly positive probability to every linear order in $\mathcal{P}$. We start by defining the learning environment. Let outcomes and utility be given by $Y=\{y_\succ\}_{\succ\in\mathcal{P}}$ and $u(a,y_\succ)=|\{a'\in A:a\succ a'\}|$. Moreover, for every $\succ\in\mathcal{P}$, and $\varepsilon\in(0,1)$ define the (action-independent) DGP

$$p^\succ_a(y)=\begin{cases}1-\varepsilon & y=y_\succ\\ \frac{\varepsilon}{|\mathcal{P}|-1} & \text{otherwise}\end{cases}$$

and let the prior be uniform over those models:

$$\mu_0(\{p^\succ\})=\frac{1}{|\mathcal{P}|}\qquad\forall\succ\in\mathcal{P}$$

so that $\Theta=\{p^\succ\}_{\succ\in\mathcal{P}}$. Observe that given these utility functions and DGPs, there exists $\underline{\varepsilon}\in\left(0,\frac{1}{2}\right)$ such that for all $\nu\in\Delta(\Theta)$ and for all $\succ\in\mathcal{P}$

$$\frac{\nu(p^\succ)}{1-\nu(p^\succ)}>\frac{1-\underline{\varepsilon}}{\underline{\varepsilon}|\mathcal{P}|}\text{ and }a\succ a'\implies\mathbb{E}_\nu[\mathbb{E}_p[u(a,y)]]>\mathbb{E}_\nu[\mathbb{E}_p[u(a',y)]].\tag{15}$$

Consider a database $d\in\mathbb{N}^{A\times Y}$ such that for $\succ\in\mathcal{P}$

$$\sum_{a\in A}|d(a,y_\succ)|>\sum_{a\in A}|d(a,y)|\qquad\forall y\in Y\backslash\{y_\succ\}.\tag{16}$$

We have

$$\begin{aligned}\frac{\mu\left(p^{\succ}|d\right)}{1-\mu\left(p^{\succ}|d\right)} &= \frac{\frac{1}{|\mathcal{P}|}\left(1-\varepsilon\right)^{\sum_{a\in A}|d(a,y_{\succ})|}\varepsilon^{\sum_{y\in Y\setminus\{y_{\succ}\}}\sum_{a\in A}|d(a,y)|}}{\sum_{\succ'\in\mathcal{P}\setminus\{\succ\}}\frac{1}{|\mathcal{P}|}\left(1-\varepsilon\right)^{\sum_{a\in A}|d(a,y_{\succ'})|}\varepsilon^{\sum_{y\in Y\setminus\left\{y_{\succ'}\right\}}\sum_{a\in A}|d(a,y)|}} \\ &\geqslant \frac{\frac{1}{|\mathcal{P}|}\left(1-\varepsilon\right)^{\sum_{a\in A}|d(a,y_{\succ})|}\varepsilon^{\sum_{y\in Y\setminus\{y_{\succ}\}}\sum_{a\in A}|d(a,y)|}}{\sum_{\succ'\in\mathcal{P}\setminus\{\succ\}}\frac{1}{|\mathcal{P}|}\left(1-\varepsilon\right)^{\sum_{a\in A}|d(a,y_{\succ})|}\varepsilon^{\sum_{y\in Y\setminus\{y_{\succ}\}}\sum_{a\in A}|d(a,y)|}\frac{\varepsilon}{(1-\varepsilon)}} \geqslant \frac{1-\varepsilon}{\varepsilon|\mathcal{P}|}.\end{aligned}$$

Therefore, Equation (16) implies that

$$a\succ a' \implies \mathbb{E}_{\mu(\cdot|d)}\left[\mathbb{E}_{p}\left[u\left(a,y\right)\right]\right]>\mathbb{E}_{\mu(\cdot|d)}\left[\mathbb{E}_{p}\left[u\left(a',y\right)\right]\right],$$

i.e., if the agent recalls database d they choose accordingly to $\succ$ from every menu. Call $\mathcal{D}'$ the set of databases for which (16) holds for some $\succ$.

Given this (Y,u,μ_0) we now shows that there is a sequence of DGPs such that for every $\succ\in\mathcal{P}$ the probability of recalling a database that satisfies Equation (16) converges to $\zeta\left(\succ\right)$. We can do this by directly taking a sequence of Poisson parameters $\left(\left(\lambda_{\succ,n}\right)_{\succ\in\mathcal{P}}\right)_{n\in\mathbb{N}}$ where $\lambda_{\succ,n}$ is the Poisson parameter describing the distribution of $\sum_{a\in A}d\left(a,y_{\succ}\right)$. By Lemma A.4, taking

$$k_n=\sum_{\succ\in\mathcal{P}}\lambda_{\succ,n}\text{ and }p^{*}_{a,n}\left(y_{\succ}\right)=\frac{\lambda_{\succ,n}}{k_n}\quad\forall a\in A,\succ\in\mathcal{P}$$

these will be the relevant Poisson parameters for η^n_α for every $\alpha\in\Delta\left(A\right)$.

Let $\lambda_{\succ,n}=n+c_{\succ}\sqrt{n}$ for all $\succ\in\mathcal{P}$ and $n\in\mathbb{N}$, where $\left(c_{\succ}\right)_{\succ\in\mathcal{P}}\in\mathbb{R}^{\mathcal{P}}$ is an arbitrary vector that we will later determine. Since $\sum_{a\in A}d(a,y_{\succ})\sim\text{Poisson}(\lambda_{\succ,n})$ and $\lambda_{\succ,n}/n=1+c_{\succ}/\sqrt{n}\to1$, the Poisson CLT and Slutsky's theorem yield

$$\frac{\sum_{a\in A}d\left(a,y_{\succ}\right)-n}{\sqrt{n}}=\left(\frac{\sum_{a\in A}d\left(a,y_{\succ}\right)-\lambda_{\succ,n}}{\sqrt{\lambda_{\succ,n}}}\right)\sqrt{\frac{\lambda_{\succ,n}}{n}}+c_{\succ}\xrightarrow{d}N\left(c_{\succ},1\right).\tag{17}$$

Therefore,

$$\lim_{n\to\infty}\mathbb{P}_{\eta^n_\alpha}\left[\sum_{a\in A}d\left(a,y_{\succ}\right)>\sum_{a\in A}d\left(a,y_{\succ'}\right),\forall\succ'\in\mathcal{P}\right]=\mathbb{P}\left[X_{\succ}+c_{\succ}>X_{\succ'}+c_{\succ'},\forall\succ'\in\mathcal{P}\setminus\{\succ\}\right]$$

where $\left(X_{\succ}\right)_{\succ\in\mathcal{P}}$ are i.i.d. $N\left(0,1\right)$. The remaining task is to prove the existence of a vector $\left(c_{\succ}\right)_{\succ\in\mathcal{P}}$ such that

$$\mathbb{P}\left[X_{\succ}+c_{\succ}>X_{\succ'}+c_{\succ'},\forall\succ'\in\mathcal{P}\right]=\zeta\left(\succ\right)\qquad\forall\succ\in\mathcal{P}.\tag{18}$$

To do so, define the function $f : \mathbb{R}^{\mathcal{P}} \to \mathbb{R}$ as

$$f\left((c_{\succ})_{\succ \in \mathcal{P}}\right) = \mathbb{E}\left[\max_{\succ \in \mathcal{P}} X_{\succ} + c_{\succ}\right] - \sum_{\succ \in \mathcal{P}} \zeta\left(\succ\right) c_{\succ}.$$

Differentiating with respect to $c_{\succ'}$ and applying the dominated convergence theorem we get

$$\frac{\partial f\left((c_{\succ})_{\succ \in \mathcal{P}}\right)}{\partial c_{\succ'}} = \mathbb{P}\left[X_{\succ'} + c_{\succ'} > X_{\succ''} + c_{\succ''}, \forall \succ'' \in \mathcal{P} \backslash \{\succ'\}\right] - \zeta\left(\succ'\right).$$

Therefore, points $(c_{\succ})_{\succ \in \mathcal{P}}$ where the gradient of $f\left((c_{\succ})_{\succ \in \mathcal{P}}\right)$ is 0 satisfies equation (18). Restrict f to the hyperplane

$$H = \left\{(c_{\succ})_{\succ \in \mathcal{P}} : \sum_{\succ \in \mathcal{P}} c_{\succ} = 0\right\}.$$

Observe first that

$$\begin{aligned} f\left((c_{\succ})_{\succ \in \mathcal{P}}\right) &= \mathbb{E}\left[\max_{\succ \in \mathcal{P}} X_{\succ} + c_{\succ}\right] - \frac{\sum_{\succ' \in \mathcal{P}} c_{\succ'}}{|\mathcal{P}|} - \sum_{\succ \in \mathcal{P}} \zeta\left(\succ\right) c_{\succ} + \frac{\sum_{\succ' \in \mathcal{P}} c_{\succ'}}{|\mathcal{P}|} \\ &= \mathbb{E}\left[\max_{\succ \in \mathcal{P}} X_{\succ} + c_{\succ} - \frac{\sum_{\succ' \in \mathcal{P}} c_{\succ'}}{|\mathcal{P}|}\right] - \sum_{\succ \in \mathcal{P}} \zeta\left(\succ\right)\left(c_{\succ} - \frac{\sum_{\succ' \in \mathcal{P}} c_{\succ'}}{|\mathcal{P}|}\right) \\ &= \mathbb{E}\left[\max_{\succ \in \mathcal{P}} X_{\succ} + \tilde{c}_{\succ}\right] - \sum_{\succ \in \mathcal{P}} \zeta\left(\succ\right) \tilde{c}_{\succ} = f\left((\tilde{c}_{\succ})_{\succ \in \mathcal{P}}\right) \end{aligned}$$

where

$$\tilde{c}_{\succ} = c_{\succ} - \frac{\sum_{\succ' \in \mathcal{P}} c_{\succ'}}{|\mathcal{P}|} \quad \forall \succ \in \mathcal{P}.$$

Therefore, a minimum on H is a global minimum of the function, and thus needs to have a gradient equal to 0.

Moreover, f is coercive on H. To see this, let $M((c_{\succ})_{\succ \in \mathcal{P}}) = \max_{\succ \in \mathcal{P}} c_{\succ}$, and let $\succ^* = \arg\max_{\succ \in \mathcal{P}} c_{\succ}$. By definition, $\max_{\succ \in \mathcal{P}}(X_{\succ} + c_{\succ}) \geqslant X_{\succ^*} + c_{\succ^*} = X_{\succ^*} + M((c_{\succ})_{\succ \in \mathcal{P}})$.

Taking the expectation on both sides, and noting that $\mathbb{E}[X_{\succ^*}] = 0$:

$$\mathbb{E}\left[\max_{\succ \in \mathcal{P}}(X_{\succ} + c_{\succ})\right] \geqslant M((c_{\succ})_{\succ \in \mathcal{P}}). \tag{19}$$

On the other hand, let $\zeta_{\min} = \min_{\succ \in \mathcal{P}} \zeta\left(\succ\right) > 0$. By equation (19) we have

$$f((c_{\succ})_{\succ \in \mathcal{P}}) \geqslant \sum_{\succ' \in \mathcal{P}} \zeta(\succ')(M((c_{\succ})_{\succ \in \mathcal{P}}) - c_{\succ'})$$

By definition of M, every term in the summation is non-negative. Replacing each $\zeta(\succ')$

with $\zeta_{\min}$ yields:

$$f((c_\succ)_{\succ\in\mathcal{P}}) \geqslant \zeta_{\min}\left(|\mathcal{P}|M((c_\succ)_{\succ\in\mathcal{P}}) - \sum_{\succ'\in\mathcal{P}} c_{\succ'}\right)$$

Because $c \in H$, the $\sum_{\succ'\in\mathcal{P}} c_{\succ'} = 0$, so $f((c_\succ)_{\succ\in\mathcal{P}}) \geqslant \zeta_{\min}|\mathcal{P}|M((c_\succ)_{\succ\in\mathcal{P}})$. Since $\sum_{\succ'\in\mathcal{P}} c_{\succ'}$, $M((c_\succ)_{\succ\in\mathcal{P}}) \geqslant 0$. For any coordinate $\succ'\in\mathcal{P}$, we have: $-c_{\succ'} = \sum_{\succ''\in\mathcal{P}\setminus\{\succ'\}} c_{\succ''} \leqslant (|\mathcal{P}| - 1)M((c_\succ)_{\succ\in\mathcal{P}})$.

This bounds the absolute value of every component:

$$|c_{\succ'}| \leqslant (|\mathcal{P}| - 1)M((c_\succ)_{\succ\in\mathcal{P}}) \implies \|(c_\succ)_{\succ\in\mathcal{P}}\|_\infty \leqslant (|\mathcal{P}| - 1)M((c_\succ)_{\succ\in\mathcal{P}})$$

Using the equivalence of norms in $\mathbb{R}^n$, we know $\|(c_\succ)_{\succ\in\mathcal{P}}\|_2 \leqslant \sqrt{|\mathcal{P}|}\|(c_\succ)_{\succ\in\mathcal{P}}\|_\infty$. Combining these inequalities yields:

$$\|(c_\succ)_{\succ\in\mathcal{P}}\|_2 \leqslant \sqrt{|\mathcal{P}|}(|\mathcal{P}| - 1)M((c_\succ)_{\succ\in\mathcal{P}}) \implies M((c_\succ)_{\succ\in\mathcal{P}}) \geqslant \frac{1}{\sqrt{|\mathcal{P}|}(|\mathcal{P}| - 1)}\|(c_\succ)_{\succ\in\mathcal{P}}\|_2$$

Substituting this back into our lower bound for $f((c_\succ)_{\succ\in\mathcal{P}})$:

$$f((c_\succ)_{\succ\in\mathcal{P}}) \geqslant \frac{\zeta_{\min}|\mathcal{P}|}{\sqrt{|\mathcal{P}|}(|\mathcal{P}| - 1)}\|(c_\succ)_{\succ\in\mathcal{P}}\|_2$$

As $\|(c_\succ)_{\succ\in\mathcal{P}}\|_2 \to \infty$ within H, the right-hand side goes to infinity, meaning $f((c_\succ)_{\succ\in\mathcal{P}}) \to \infty$. Thus, f is coercive on H.

Claim 1. *f is strictly convex on H.*

Thus because f is coercive it admits a minimum. This establishes that for every $\epsilon > 0$ there is a $n \in \mathbb{N}$ such that with (Y, u, p_n^*, μ, k_n) and under every optimal Markovian policy the behavior converges to a RUM ζ' such that $|c_\zeta(a, M) - c_{\zeta'}(a, M)| \leqslant \epsilon/2$ for all $M \in \mathcal{M}$ and $a \in M$.

To conclude, observe that we can make (Y, u, p_n^*, μ, k_n) correctly specified (and so satisfy Assumption 1) by modifying μ to μ' by assigning probability $\varepsilon'' > 0$ to p_n^*. Indeed, for every $K > 0$ there is an $\varepsilon''(K) > 0$ sufficiently small that only databases $d \in \mathcal{D}'$ with at least K entries can generate different behaviors under $\mu(\cdot|d)$ and $\mu'(\cdot|d)$. However, given the memory capacity k_n, by taking a sufficiently large K the probability of recalling at least K experiences can be made arbitrarily small. Therefore, with a sufficiently small ε'' probability assigned to p_n^*, under every optimal Markovian policy the behavior converges to a RUM ζ'' such that $|c_{\zeta''}(a, M) - c_{\zeta'}(a, M)| \leqslant \epsilon/2$ for all $M \in \mathcal{M}$ and $a \in M$. By the triangle inequality this ζ'' such that $|c_{\zeta''}(a, M) - c_\zeta(a, M)| \leqslant \epsilon$ for all $M \in \mathcal{M}$ and $a \in M$. □

Proof of Proposition 2. With only two actions, Ψ^{μ_0} can be seen as a correspondence from $[0,1]$ to $[0,1]$ where these real numbers represent the probability assigned to the action a whose distribution has been FOSD increased. Moreover, the assumption that two actions are never indifferent after any observable database implies that Ψ^{μ_0} is a function, and it is continuous by Lemma A.3. By Theorem 5 in Milgrom and Weber [1982], this function is pointwise larger when p^*_a increases in the FOSD order. The result then follows by Theorem 1 of Villas-Boas [1997]. □

A.4 Proofs for Section 5

As a first step towards proving Theorem 3, we establish a concentration inequality for ratios of random variables whose distributions converge to Poisson distributions. In this proof we let k grow, so we explicitly index the distributions η_{α^k} by k, i.e., as $\eta^k_{\alpha^k}$.

Lemma A.5. *Suppose that* $\lim_{k\to\infty}\alpha^k = \hat{\alpha}$. *For every* $\varepsilon > 0$ *and* $(a,a',y,y') \in A^2 \times Y^2$ *with* $\hat{\alpha}(a') > 0$ *and* $p^*_{a'}(y') > 0$, $\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}\left[\left|\frac{d(a,y)}{d(a',y')} - \frac{\hat{\alpha}(a)p^*_a(y)}{\hat{\alpha}(a')p^*_{a'}(y')}\right| > \varepsilon\right] = 0$.[48] *Moreover,* $\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}\left[d(a',y') - \frac{\hat{\alpha}(a')p^*_{a'}(y')k}{2} < \varepsilon\right] = 0$.

Definition 6. A *unitary-data self-confirming equilibrium* is an $\alpha \in \Delta(A)$ such that for all $a \in \text{supp}(\alpha)$ there is $\nu^a \in \Delta(\Theta(\alpha))$ such that $a \in BR(\nu^a)$.

Unitary-data self-confirming equilibrium allows each action in the support of α to be rationalized by a different belief but requires that all beliefs are supported over the likelihood maximizers given the same data. This is more restrictive than heterogeneous-belief self-confirming equilibrium (Fudenberg and Levine [1993]), which only requires that each action a in the support of α is a best response to a belief over the maximizers corresponding to data about the consequences of the particular pure action a. This difference is due to the different origins of heterogeneity in the two equilibrium concepts.[49]

The next lemma shows that this distinction is irrelevant if the agent is correctly specified.

Lemma A.6. *When* $p^* \in \Theta$ *(as we have assumed here), every unitary-data self-confirming equilibrium is a unitary-belief self-confirming equilibrium.*[50]

[48] We adopt the convention that $0/0 = \infty$.

[49] Heterogeneous-belief self-confirming equilibria are steady states of models with many agents in each player role, and heterogeneity comes from the fact that different agents in the same role may behave differently and thus find that different models best fit their data.

[50] Our working paper gives an example where $p^* \notin \Theta$ and the limit as $k \to \infty$ of the stochastic memory equilibria is a unitary-data self-confirming equilibrium that unitary beliefs cannot support.

Proof of Theorem 3. Let $f(d) \in \Delta(A \times Y)$ denote the empirical joint distribution over action-outcome pairs corresponding to $d \in \mathcal{D}$. By Lemma A.5, for every $M \in \mathbb{N}$ and $\varepsilon > 0$, $\lim_{k\to\infty} \mathbb{P}_{\eta^k_{\alpha^k}} \left[d : \sum_{a\in A, y\in Y} d(a,y) > M, \max_{a\in A, y\in Y} |f(d)(a,y) - \hat{\alpha}(a)p^*_a(y)| < \varepsilon \right]$ is equal to 1. That is, with probability approaching 1, the database is large, and the recalled frequency of pair (a,y) is approximately proportional to $\hat{\alpha}(a)p^*_a(y)$.

Let $\varepsilon > 0$. By Assumption 1, $p \in \Theta(\hat{\alpha})$ if and only if $p_a = p^*_a$ for all $a \in \operatorname{supp}\hat{\alpha}$. Therefore, there exist $\varepsilon' < \varepsilon$ and $K > 0$ such that $\left(\sum_{a\in A} \hat{\alpha}(a) \sum_{y\in Y} p^*_a(y) \log p_a(y)\right) - \left(\sum_{a\in A} \hat{\alpha}(a) \sum_{y\in Y} p^*_a(y) \log p'_a(y)\right) > K$ for all $p \in B_{\varepsilon'}(\Theta(\hat{\alpha}))$, $p' \notin B_{\varepsilon}(\Theta(\hat{\alpha}))$. Thus, there is a set of databases d that has $\mathbb{P}_{\eta^k_{\hat{\alpha}^k}}$ probability going to 1, whose length $\sum_{a\in A, y\in Y} d(a,y)$ is growing to ∞ and such that

$$\frac{\mu(B_\varepsilon(\Theta(\hat{\alpha}))|d)}{1-\mu(B_\varepsilon(\Theta(\hat{\alpha}))|d)} \geqslant \frac{\int_{B_{\varepsilon'}(\Theta(\hat{\alpha}))} \prod_{(a,y)\in A\times Y} (p_a(y))^{d(a,y)} \mathrm{d}\mu_0(p)}{\int_{\Theta\setminus B_\varepsilon(\Theta(\hat{\alpha}))} \prod_{(a,y)\in A\times Y} (p_a(y))^{d(a,y)} \mathrm{d}\mu_0(p)} \geqslant \mu_0(B_{\varepsilon'}(\Theta(\hat{\alpha}))) \exp\left(K/2 \sum_{a\in A, y\in Y} d(a,y)\right).$$

As k grows the RHS goes to ∞, so since ε can be arbitrarily small, the agent's beliefs concentrate on $\Theta(\hat{\alpha})$. Now fix any action $a \in \operatorname{supp}(\hat{\alpha})$. Since $\lim_{k\to\infty} \alpha^k(a) = \hat{\alpha}(a) > 0$, the posterior probability of drawing a belief under which a is optimal is at least $\alpha^k(a)$. Because beliefs asymptotically concentrate on $\Delta(\Theta(\hat{\alpha}))$, there must exist a sequence of beliefs ν_k and a limit $\nu^* \in \Delta(\Theta(\hat{\alpha}))$ such that $\nu_k \to \nu^*$ and $a \in BR(\nu_k)$ for all large k. Because best reply correspondence has closed graph, $a \in BR(\nu^*)$. Since this holds for every $a \in \operatorname{supp}(\hat{\alpha})$, $\hat{\alpha}$ satisfies the definition of a unitary-data self-confirming equilibrium. Finally, by Lemma A.6 and Assumption 1, every unitary-data self-confirming equilibrium is a self-confirming equilibrium. □

A.5 Proofs for Section 6

Proof of Lemma 2. Let $D' \subseteq \mathcal{D}$ denote the set of databases that under η_α have a positive probability of being reached with a finite number of transitions from the empty database. At any $d \in D'$, the probability of a transition to the empty history is bounded below by $Q := \exp(-[k+r]|A \times Y|) > 0$, so D' is a closed irreducible class.

Moreover, for any $d' \in D'$, there is a simple path of length $\tau \in \mathbb{N}$ connecting the empty database and d', i.e., a finite sequence of distinct databases $(\tilde{d}_0, ..., \tilde{d}_\tau)$ with $\tilde{d}_0$ the empty database, $\tilde{d}_\tau = d'$, and $\eta_{\alpha,\tilde{d}_i}(\tilde{d}_{i+1}) > 0$ for all $i \in \{0, ..., \tau-1\}$. Let $M = \prod_{i=0}^{\tau-1} \eta_{\alpha,\tilde{d}_i}(\tilde{d}_{i+1})$.

Thus the expected return time to d' is bounded from above by

$$\sum_{i=1}^{\infty}(1 - P(\text{return time} \leqslant i)) \leqslant \tau + \sum_{i=1}^{\infty}(1 - P(\text{return time} \leqslant \tau + i))$$
$$\leqslant \tau + \sum_{i=1}^{\infty}\prod_{j=1}^{i}(1 - P(d_{j+l} = \tilde{d}_l, \forall l \in \{0, ..., \tau\})) \leqslant \tau + \sum_{i=1}^{\infty}(1 - QM)^i < \infty,$$

so d' is positive recurrent. It remains to rule out invariant mass outside D'. For every $d \in \mathcal{D}$, the probability of transitioning to the empty database is bounded below by $Q > 0$. Since the empty database belongs to D', and D' is closed, the hitting time $\tau_{D'} := \inf\{t \geqslant 0 : d_t \in D'\}$ satisfies $\mathbb{P}_d(\tau_{D'} > n) \leqslant (1 - Q)^n$ for all $d \in \mathcal{D}$, $\forall n \in \mathbb{N}$. Thus no invariant distribution can assign positive mass to $\mathcal{D} \backslash D'$. Since there is zero probability of leaving D' and all the states in D' are positive recurrent, η_α has a unique invariant distribution (see Theorem 6.5.3 in Durrett [2019]). □

To state the next lemma, we introduce the following notation. Let $b(d) := \pi(\mu(\cdot \mid d))$. Also, let $\mathcal{G}_t$ be the beginning-of-period filtration after d_t has been drawn but before y_t is realized, so that $a_t = b(d_t)$ and α_t are $\mathcal{G}_t$-measurable. but y_t has not yet been realized. For each $d \in \mathcal{D}$, let $Q_t(d, \cdot)$ be the transition law generated from the history through period $t - 1$ if the current database is d, and put

$$\alpha_t^d := \frac{t-1}{t}\alpha_{t-1} + \frac{1}{t}\delta_{b(d)}.$$

Thus $Q_t(d_t, \cdot) = \mathcal{L}(d_{t+1} \mid \mathcal{G}_t)$ and $\alpha_t^{d_t} = \alpha_t$. Write $P_\alpha(d, \cdot) := \eta_{\alpha,d}$ and set

$$\varphi(d) := \delta_{b(d)},$$
$$G(\alpha, d) := P_\alpha\varphi(d) = \int_{\Delta(\Theta)} \delta_{\pi(\nu)}\, dF_{\alpha,d}^{\mu_0}(\nu),$$
$$\overline{G}(\alpha) := \int_{\mathcal{D}} G(\alpha, d)\, d\mathcal{H}_\alpha(d).$$

Also write $q_\alpha(a, y) := \alpha(a)p_a^*(y)$.

Lemma A.7 (Controlled Markov averaging)**.** *Then the following statements hold.*

1. *On every path for which $\|\upsilon_{t-1} - q_{\alpha_{t-1}}\|_1 \to 0$,*

$$\sup_{d\in\mathcal{D}} \left\| Q_t(d, \cdot) - P_{\alpha_t^d}(d, \cdot) \right\|_{TV} \longrightarrow 0.$$

Moreover, the conditional belief law converges uniformly to $F_{\alpha_t^d,d}^{\mu_0}$ on the same set of histories. In particular, the conclusion holds $\mathbb{P}_{\mu_0,\pi}$-a.s.

2. *The family* $(P_\alpha)_{\alpha\in\Delta(A)}$ *is uniformly geometrically ergodic and Lipschitz in* α *in the uniform total-variation norm. The maps* $\alpha \mapsto \mathcal{H}_\alpha$ *and* $\alpha \mapsto F^{\mu_0}_{\alpha,d}$ *are Lipschitz in total variation, the latter uniformly in* d.

The proofs of the lemmas and corollaries of this section are in Online Appendix B.2.

Corollary 2. *Consider any sequence of histories for which*

$$\|\upsilon_t - q_{\alpha_t}\|_1 \to 0 \qquad \text{and} \qquad \alpha_t \to \alpha.$$

Let Q_t *be the actual transition kernel of the recalled database at time* $t+1$ *given the current database at time* t, *and let* λ_t *be the law of the recalled database* d_t *generated by the sequence of kernels* $Q_1, Q_2, \ldots, Q_{t-1}$. *Then*

$$\|\lambda_t - \mathcal{H}_\alpha\|_{TV} \to 0.$$

Let $\Psi_{d'}(\alpha)$ denote the distributions over actions induced by an optimal Markovian mixed policy ρ and random beliefs ν: $\Psi^{\mu_0}(\alpha, d') = \{\int_{\Delta(\Theta)} \rho(\nu) dF^{\mu_0}_{\alpha,d'}(\nu) : \rho \in \mathcal{O}\}$.

Lemma A.8.

1. $\int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ *is non-empty valued.*
2. $\int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ *is closed valued;*
3. $\int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ *is upper hemicontinuous;*
4. $\int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ *is convex valued;*
5. α' *is an ergodic memory equilibrium if and only if* $\alpha' \in \int_{\mathcal{D}} \Psi^{\mu_0}(\alpha', d') d\mathcal{H}_{\alpha'}(d')$.

Proof of Proposition 5. Lemma A.8 shows that every fixed point of $\bar{\Psi} := \int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ is an ergodic memory equilibrium, and that $\bar{\Psi}$ is non-empty valued and convex valued. Together with the closed-graph theorem, the lemma also shows $\bar{\Psi}$ has a closed graph, so it has a fixed point by the Kakutani fixed point theorem. □

Proof of Theorem 4. The proof has four steps. Lemma A.7 gives the limit of the transition matrices over databases and its uniform continuity in α. The second step of the proof establishes that the accumulated deviation of the recalled database from its theoretical average vanishes: for all $T > 0$,

$$\lim_{n\to\infty} \sup_{m\geqslant n:\ \sum_{t=n}^{m} 1/(t+1)\leqslant T} \left\| \sum_{t=n}^{m} \big(G(\alpha_t, d_t) - \overline{G}(\alpha_t)\big)/(t+1) \right\| = 0. \tag{20}$$

The third step of the proof uses stochastic approximation to show that the long-run behavior of (9) can be approximated by $\dot{\boldsymbol{\alpha}}(t) \in \mathbb{E}_{\mathcal{H}_{\boldsymbol{\alpha}_t}}[\Psi^{\mu_0}(\boldsymbol{\alpha}(t), d)] - \boldsymbol{\alpha}(t)$. The last step parallels the last step of the proof of Theorem 1 with $\int_{\mathcal{D}} \Psi^{\mu_0}(\alpha, d')d\mathcal{H}_{\boldsymbol{\alpha}}$ in place of $\Psi^{\mu_0}(\alpha)$ after observing that $\int_{\mathcal{D}} \Psi^{\mu_0}(\alpha, d')d\mathcal{H}_{\boldsymbol{\alpha}}$ inherits the key properties of Ψ^{μ_0}, as shown by Lemma A.8. The details of the second and third step are given in Online Appendix B.3; we omit the remaining details. □

References

Aliprantis, C. and K. Border (2013). *Infinite dimensional analysis: A hitchhiker's guide*. Berlin. Springer-Verlag.

Artstein, Z. and R. J. Wets (1988). "Approximating the integral of a multifunction." *Journal of Multivariate Analysis*, 24, 285–308.

Augenblick, N., B. K. Jack, S. Kaur, F. Masiye, and N. Swanson (2026). "Retrieval failures and consumption smoothing: A field experiment on seasonal poverty." *Quarterly Journal of Economics*, Forthcoming.

Aumann, R. J. (1965). "Integrals of set-valued functions." *Journal of mathematical analysis and applications*, 12, 1–12.

Ba, C., J. A. Bohren, and A. Imas (2024). "Over-and underreaction to information." *Working paper*.

Banerjee, A. and D. Fudenberg (2004). "Word-of-mouth learning." *Games and economic behavior*, 46, 1–22.

Benaim, M., J. Hofbauer, and S. Sorin (2005). "Stochastic approximations and differential inclusions." *SIAM Journal on Control and Optimization*, 44, 328–348.

Benaım, M. and M. W. Hirsch (1999). "Mixed equilibria and dynamical systems arising from fictitious play in perturbed games." *Games and Economic Behavior*, 29, 36–72.

Benaïm, M. (2006). "Dynamics of stochastic approximation algorithms." *Seminaire de probabilites XXXIII*. Springer, 1–68.

Benjamin, D. J. (2019). "Errors in probabilistic reasoning and judgment biases." *Handbook of Behavioral Economics: Applications and Foundations*, 2, 69–186.

Block, H. D. and J. Marschak (1960). *Random orderings and stochastic theories of response*. Springer.

Bordalo, P., J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer (2023). "Memory and probability." *The Quarterly Journal of Economics*, 138, 265–311.

Bordalo, P., N. Gennaioli, G. Lanzani, and A. Shleifer (2025). "A cognitive theory of reasoning and choice."

Calzolari, G. and M. Nardotto (2017). "Effective reminders." *Management Science*, 63, 2915–2932.

Caplin, A. (2025). *An introduction to cognitive economics: The science of mistakes*. Springer Nature.

Che, Y.-K. and K. Mierendorff (2019). "Optimal dynamic allocation of attention." *American Economic Review*, 109, 2993–3029.

Conlon, J. J. (2026). *Memory Rehearsal and Belief Biases*. Tech. rep. Working paper.

d'Acremont, M., W. Schultz, and P. Bossaerts (2013). "The human brain encodes event frequencies while forming subjective beliefs." *Journal of Neuroscience*, 33, 10887–10897.

Danenberg, T. and R. Spiegler (2026). "Representative sampling equilibrium." *Journal of Political Economy: Microeconomics, Forthcoming*.

Davelaar, E. J., Y. Goshen-Gottstein, A. Ashkenazi, H. J. Haarmann, and M. Usher (2005). "The Demise of Short-Term Memory Revisited: Empirical and Computational Investigations of Recency Effects." *Psychological Review*, 112, 3–42.

Dominitz, J. and C. F. Manski (2007). "Expected equity returns and portfolio choice: Evidence from the Health and Retirement Study." *Journal of the European Economic Association*, 5, 369–379.

Duncan, K. and D. Shohamy (2020). "Memory, Reward, and Decision-Making." *The Cognitive Neurosciences*. MIT Press, 617–630.

Durrett, R. (2019). *Probability: theory and examples*. Vol. 49. Cambridge university press.

Ellison, G. and D. Fudenberg (1993). "Rules of thumb for social learning." *Journal of political Economy*, 101, 612–643.

Erev, I. and E. Haruvy (2016). "Learning and the economics of small decisions." *Handbook of Experimental Economics, Vol. 2*. Princeton University Press, 638–716.

Esponda, I., D. Pouzo, and Y. Yamamoto (2021a). "Asymptotic behavior of Bayesian learners with misspecified models." *Journal of Economic Theory*, 195, 105260.

— (2021b). "Corrigendum: Asymptotic behavior of Bayesian learners with misspecified models." *Journal of Economic Theory*, 195, 105260.

Esponda, I., E. Vespa, and S. Yuksel (2024). "Mental models and learning: The case of base-rate neglect." *American Economic Review*, 114, 752–782.

Fox, C. R. and L. Hadar (2006). ""Decisions from experience"= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004)." *Judgment and Decision making*, 1, 159–161.

Franklin, J. N. (2002). *Methods of mathematical economics: linear and nonlinear programming, fixed-point theorems*. SIAM.

Frydman, C. and L. J. Jin (2022). "Efficient Coding and Risky Choice." *Quarterly Journal of Economics*.

Fudenberg, D., G. Lanzani, and P. Strack (2024). "Selective Memory Equilibrium." *Journal of Political Economy*, 132, 3978–4020.

Fudenberg, D. and D. K. Levine (1993). "Self-confirming equilibrium." *Econometrica*, 61, 523–545.

Fudenberg, D. and A. Peysakhovich (2016). "Recency, records, and recaps: Learning and nonequilibrium behavior in a simple decision problem." *ACM Transactions on Economics and Computation (TEAC)*, 4, 1–18.

Fudenberg, D., P. Strack, and T. Strzalecki (2018). "Speed, accuracy, and the optimal timing of choices." *American Economic Review*, 108, 3651–3684.

Gershman, S. J., I. Fiete, and K. Irie (2025). "Key-value memory in the brain." *Neuron*, 113, 1694–1707.

Gescheider, G. (2013). *Psychophysics: The fundamentals*.

Gonçalves, D. (2023). "Sequential Sampling Equilibrium." *Working paper*.

Gottlieb, D. (2014). "Imperfect memory and choice under risk." *Games and Economic Behavior*, 85, 127–158.

Greene, W. H. (2000). *Econometric analysis*. 4th ed. International edition, New Jersey: Prentice Hall, 201–215.

Gul, F., P. Natenzon, and W. Pesendorfer (2014). "Random choice as behavioral optimization." *Econometrica*, 82, 1873–1912.

Hall, P. and C. C. Heyde (2014). *Martingale limit theory and its application*. Academic press.

Hausman, J. A. and D. A. Wise (1978). "A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences." *Econometrica*, 46, 403–426.

Hébert, B. and M. Woodford (2023). "Rational inattention when decisions take time." *Journal of Economic Theory*, 208, 105612.

Heidhues, P., B. Kőszegi, and P. Strack (2023). "Misinterpreting yourself." *Available at SSRN 4325160*.

Hertwig, R., G. Barron, E. U. Weber, and I. Erev (2004). "Decisions from experience and the effect of rare events in risky choice." *Psychological science*, 15, 534–539.

Huber, J., J. W. Payne, and C. Puto (1982). "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis." *Journal of consumer research*, 9, 90–98.

Ilut, C. and R. Valchev (2023). "Economic agents as imperfect problem solvers." *The Quarterly Journal of Economics*, 138, 313–362.

Jehiel, P. and J. Steiner (2020). "Selective Sampling with Information-Storage Constraints." *The Economic Journal*, 130, 1753–1781.

Kaanders, P., P. Sepulveda, T. Folke, P. Ortoleva, and B. De Martino (2022). "Humans actively sample evidence to support prior beliefs." *Elife*, 11, e71768.

Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press.

Kahneman, D. and A. Tversky (1972). "Subjective probability: A judgment of representativeness." *Cognitive Psychology*, 3, 430–454.

Kallenberg, O. (2002). *Foundations of modern probability*. Springer.

— (2017). *Random Measures, Theory and Applications*. Springer.

Kandel, E. R. et al. (2000). *Principles of neural science*. Vol. 4. McGraw-Hill.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016). "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science*, 62, :3393–3411.

Ke, T. T. and J. M. Villas-Boas (2019). "Optimal learning before choice." *Journal of Economic Theory*, 180, 383–437.

Koszegi, B., G. Loewenstein, and T. Murooka (2021). "Fragile self-esteem." *Review of Economics Studies*.

Loève, M. (1977). *Probability Theory I*. Springer.

Lu, J. (2016). "Random choice and private information." *Econometrica*, 84, 1983–2027.

Malmendier, U. and S. Nagel (2016). "Learning from inflation experiences." *The Quarterly Journal of Economics*, 131, 53–87.

Matějka, F. and A. McKay (2015). "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105, 272–298.

Milgrom, P. R. and R. J. Weber (1982). "A theory of auctions and competitive bidding." *Econometrica*, 50, 1089–1122.

Molchanov, I. (2017). *Theory of random sets*. Springer.

Mullainathan, S. (2002). "A memory-based model of bounded rationality." *The Quarterly Journal of Economics*, 117, 735–774.

Natenzon, P. (2019). "Random choice and learning." *Journal of Political Economy*, 127, 419–457.

Osborne, M. J. and A. Rubinstein (1998). "Games with procedurally rational players." *American Economic Review*, 88, 834–847.

Parthasarathy, K. R. (2005). *Probability measures on metric spaces*. American Mathematical Soc.

Phillips, L. D. and W. Edwards (1966). "Conservatism in a simple probability inference task." *Journal of Experimental Psychology*, 72, 346.

Reder, L. M. (2014). *Implicit memory and metacognition*. Psychology Press.

Royden, H. and P. M. Fitzpatrick (2010). *Real analysis*. China Machine Press.

Salant, Y. and J. Cherry (2020). "Statistical inference in games." *Econometrica*, 88, 1725–1752.

Shadlen, M. N. and D. Shohamy (2016). "Decision making and sequential sampling from memory." *Neuron*, 90, 927–939.

Sial, A. Y., J. R. Sydnor, and D. Taubinsky (2025). "Biased Memory and Perceptions of Self-Control." *Working Paper*.

Steele, J. M. (1994). "Le Cam's inequality and Poisson approximations." *The American Mathematical Monthly*, 101, 48–54.

Stroock, D. W. (2010). *Probability theory: an analytic view*. Cambridge university press.

Strzalecki, T. (2025). *Stochastic Choice Theory*. Cambridge University Press.

Thurstone, L. L. (1927). "A law of comparative judgment." *Psychological Review*, 34, 273–278.

Tversky, A. (1972). "Elimination by aspects: A theory of choice." *Psychological Review*, 79, 281.

Villas-Boas, J. M. (1997). "Comparative statics of fixed points." *Journal of Economic Theory*, 73, 183–198.

Weitzman, M. L. (2007). "Subjective expectations and asset-return puzzles." *American Economic Review*, 97, 1102–1130.

Wilson, A. (2014). "Bounded Memory and Biases in Information Processing." *Econometrica*, 82, 2257–2294.

Wulff, D. U., M. Mergenthaler-Canseco, and R. Hertwig (2018). "A meta-analytic review of two modes of learning and the description-experience gap." *Psychological bulletin*, 144, 140.

B Online Appendix

B.1 Proofs of Propositions Stated in the Text.

Proof of Proposition 3. Because the agent is risk neutral, it is optimal for them to choose the action with the highest posterior mean. Let $\hat{y}_a$ be the average outcome over the n_a recalled experiences where action a was used. When the agent recalls n_a experiences for action a, each is normally distributed with mean $\bar{y}_a$ and variance σ^2, so $\hat{y}_a$ is normally distributed with mean $\bar{y}_a$ and variance σ^2/n_a. Thus the posterior mean of y_a is $\frac{y_a^0/\sigma_0^2+\hat{y}_a n_a/\sigma^2}{1/\sigma_0^2+n_a/\sigma^2}$, and the induced choice probabilities are those of a Probit model with means $\left(\frac{y_a^0/\sigma_0^2+n_a\bar{y}_a/\sigma^2}{1/\sigma_0^2+n_a/\sigma^2}\right)_{a\in A}$ and a diagonal covariance matrix with entries $\frac{n_a/\sigma^2}{(1/\sigma_0^2+n_a/\sigma^2)^2}$.

To derive the distribution over the number of recalled experiences, let $t \geqslant k$ and $a \in A$. We first consider infinite sequences of experiences $(a_t, y_t)_{t\in\infty}$ such that $\lim_{t\to\infty} \sum_{\tau=1}^{t} \mathbb{1}_{\{a\}}(a_\tau) = \infty$. By equation (1), the probability that $d_{t+1}(a, Y) = n_a \in \mathbb{N}$ conditional on the history (a^t, y^t) is $\binom{t\alpha_t(a)}{n_a}(k/t)^{n_a}((t-k)/t)^{t\alpha_t(a)-n_a}$. Because a occurs infinitely often, $\lim_{t\to\infty} t\alpha_t(a)\left(\frac{k}{t}\right) = \alpha(a)k$, so the Poisson limit theorem (e.g., page 15 of Loève [1977]) implies the probability that $d_{t+1}(a, Y) = n_a \in \mathbb{N}$ converges to $e^{-\alpha(a)k}\frac{(\alpha(a)k)^{n_a}}{n_a!}$.

Now suppose that for the pair $(a, (a_t, y_t)_{t\in\infty})$, $\lim_{t\to\infty} \sum_{\tau=1}^{t} \mathbb{1}_{\{a\}}(a_\tau) \neq \infty$. Then the distribution over recalled experiences given (a^t, y^t) is eventually first-order stochastically dominated by the one given (a'^t, y^t), where each $(a'^t)_{t\in\mathbb{N}}$ is an arbitrary action sequence with the properties that $\lim_{t\to\infty} \mathbb{1}_{\{a\}}(a'_t) = \infty$ and $\lim_{t\to\infty} \mathbb{1}_{\{a\}}(a'_t)/t = 0$. Since this dominating distribution converges to a Dirac at 0 by the previous part of the proof, so does the distribution associated with (a^t, y^t). Thus the number of recalled experiences of action n_a is Poisson distributed with parameter $\alpha(a)k$. □

The next proof requires the following lemma:

Lemma B.1. *The choice probabilities of the agent are*

$$\alpha_{\phi,p}(1) = \begin{cases} 0 & \text{if } \phi = 0 \text{ and } kp \leqslant 1 \\ 1 + \frac{1}{kp} W\left(-e^{-kp} kp(1-p\phi)\right) & \text{otherwise} \end{cases}.$$

Furthermore, $\alpha_{0,p}(1)$ is increasing in p.

Proof of Proposition 4. Part (i) holds because W is an increasing function so $\alpha_{\phi,p}$ is increasing in ϕ. By Lemma B.1, there is an equilibrium where the agent engages in the activity with strictly positive probability absent reminders if and only if $pk > 1$, and thus

we can assume $pk > 1$ when establishing part (ii). The marginal effect of reminders is $\frac{\partial \alpha_{\phi,p}(1)}{\partial \phi} = \frac{1}{k(1-p\phi)} \times \frac{-W\left(-e^{-kp}kp(1-p\phi)\right)}{1+W(-e^{-kp}kp(1-p\phi))}$. To simplify notation define $\psi(p) = -e^{-kp}kp(1-p\phi)$. Taking logarithms and differentiating with respect to p yields

$$\frac{\partial}{\partial p}\log\left(\frac{\partial \alpha_{\phi,p}(1)}{\partial \phi}\right) = \frac{\phi}{1-p\phi} + \frac{1}{(1+W(\psi(p)))W(\psi(p))}W'(\psi(p))\psi'(p).$$

As $\psi'(p) = e^{-kp}\left[(kp-1)k(1-p\phi) + kp\phi\right]$, we get that

$$\frac{\partial}{\partial p}\log\left(\frac{\partial \alpha_{\phi,p}(1)}{\partial \phi}\right) = \frac{\phi}{1-p\phi} - \frac{1}{(1+W(\psi(p)))^2}\frac{(kp-1)k(1-p\phi)+kp\phi}{kp(1-p\phi)}.$$

As $kp > 1$ and $W \in [-1, 0]$ we obtain that $\frac{\partial}{\partial p}\log\left(\frac{\partial \alpha_{\phi,p}(1)}{\partial \phi}\right) \leqslant -\frac{kp-1}{p} < 0$. Thus, $\alpha_{\phi,p}(1)$ is sub-modular in (ϕ, p). This and Lemma B.1 imply (ii). □

B.2 Proofs of Ancillary Lemmas and Claims

Proof of Lemma A.1. Let $h_t = (a_1, y_1, \ldots, a_t, y_t)$ denote the period-t history and let $\{\mathcal{F}_t\}_{t\geqslant 0}$ be the filtration $\mathcal{F}_t = \sigma(h_t)$. Consider the $\{\mathcal{F}_t\}$–adapted stochastic processes $(\mathbf{X}_t^{(\hat{a},\hat{y})})_{(\hat{a},\hat{y})\in A\times Y, t\in\mathbb{N}}$ defined by $\mathbf{X}_t^{(\hat{a},\hat{y})} = (\mathbb{1}_{\{\hat{y}\}}(y_t) - p^*_{\hat{a}}(\hat{y}))\mathbb{1}_{\{\hat{a}\}}(a_t)$ for all $(\hat{a},\hat{y}) \in A \times Y$, and for all $t \in \mathbb{N}$.

For each $(\hat{a},\hat{y})$, $\mathbf{X}_t^{(\hat{a},\hat{y})}$ is 0 if action $\hat{a}$ is not played in period t. If $\hat{a}$ is played, $\mathbf{X}_t^{(\hat{a},\hat{y})}$ reports the difference between the indicator for $y_t = \hat{y}$ minus its expected value given $\hat{a}$, $p^*_{\hat{a}}(\hat{y})$. They are not i.i.d., but for each $(a, y) \in A \times Y$,

$$\begin{aligned}\mathbb{E}[\mathbf{X}_t^{(a,y)} \mid \mathcal{F}_{t-1}] &= \mathbb{P}_{\mu_0,\pi}[a_t = a \mid \mathcal{F}_{t-1}]\mathbb{E}[\mathbf{X}_t^{(a,y)} \mid \mathcal{F}_{t-1}, a_t = a] + \mathbb{P}_{\mu_0,\pi}[a_t \neq a \mid \mathcal{F}_{t-1}]\mathbb{E}[\mathbf{X}_t^{(a,y)} \mid \mathcal{F}_{t-1}, a_t \neq a]\\ &= \mathbb{P}_{\mu_0,\pi}[a_t = a \mid \mathcal{F}_{t-1}]0 + \mathbb{P}_{\mu_0,\pi}[a_t \neq a \mid \mathcal{F}_{t-1}]0 = 0.\end{aligned}$$

Thus $(\mathbf{X}_t^{(a,y)})_{t\in\mathbb{N}}$ is a martingale difference sequence, so from the strong law of large numbers (see Theorem 2.7 in Hall and Heyde [2014]), $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \mathbf{X}_t^{(a,y)} = 0$, $\mathbb{P}_{\mu_0,\pi}$-a.s.

Note that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \mathbf{X}_t^{(a,y)} = \lim_{n\to\infty}\sum_{t=1}^n \frac{\mathbb{1}_{\{a,y\}}(a_t,y_t) - p^*_a(y)\mathbb{1}_{\{a\}}(a_t)}{n} = \lim_{n\to\infty} \upsilon_n(a,y) - \alpha_n(a)p^*_a(y)$$

whenever the limits exist. Taking the maximum over the finite set $A\times Y$ yields the uniform statement. The "in particular" clause follows immediately when $\alpha_t \to \alpha^*$. □

Proof of Lemma A.2. Let

$$T_{\mu_0} : \mathcal{D} \to \Delta(\Theta), \qquad T_{\mu_0}(d) := \mu(\cdot \mid d),$$

denote the map from a database to the corresponding posterior. By definition,

$$F_\alpha^{\mu_0} = (T_{\mu_0})_\# \eta_\alpha,$$

where $(T_{\mu_0})_\# \eta_\alpha$ denotes the pushforward of η_α under T_{μ_0}.

We first establish continuity in α. Recall that

$$\eta_\alpha = \bigotimes_{(a,y)\in A\times Y} \mathrm{Pois}\big(k\alpha(a)p_a^*(y)\big).$$

For any $\lambda, \lambda' \geqslant 0$, a coupling of Poisson random variables gives

$$\|\mathrm{Pois}(\lambda) - \mathrm{Pois}(\lambda')\|_{TV} \leqslant 1 - e^{-|\lambda-\lambda'|} \leqslant |\lambda - \lambda'|.$$

Consequently, the product inequality for total variation implies that, for every $\alpha, \beta \in \Delta(A)$,

$$\begin{aligned}
\|\eta_\alpha - \eta_\beta\|_{TV} &\leqslant \sum_{(a,y)\in A\times Y} |k\alpha(a)p_a^*(y) - k\beta(a)p_a^*(y)| \\
&= k \sum_{a\in A} |\alpha(a) - \beta(a)| \sum_{y\in Y} p_a^*(y) \\
&= k\|\alpha - \beta\|_1. \qquad (21)
\end{aligned}$$

Total variation cannot increase under a measurable pushforward. Therefore,

$$\big\|F_\alpha^{\mu_0} - F_\beta^{\mu_0}\big\|_{TV} \leqslant \|\eta_\alpha - \eta_\beta\|_{TV} \leqslant k\|\alpha - \beta\|_1. \qquad (22)$$

Thus $\alpha \mapsto F_\alpha^{\mu_0}$ is Lipschitz in total variation and, in particular, weakly continuous. Indeed, if $\alpha_n \to \alpha^*$, then for every bounded measurable $f : \Delta(\Theta) \to \mathbb{R}$,

$$\left|\int f\, dF_{\alpha_n}^{\mu_0} - \int f\, dF_{\alpha^*}^{\mu_0}\right| \leqslant 2\|f\|_\infty \big\|F_{\alpha_n}^{\mu_0} - F_{\alpha^*}^{\mu_0}\big\|_{TV} \leqslant 2k\|f\|_\infty \|\alpha_n - \alpha^*\|_1 \longrightarrow 0.$$

It remains to prove the asserted convergence of conditional beliefs. Since

$$\mathcal{L}(\mu_t \mid h_{t-1}) = (T_{\mu_0})_\# \mathcal{L}(d_t \mid h_{t-1}),$$

the contraction of total variation under pushforwards, Lemma 1, and (21) give

$$\begin{aligned}
\|\mathcal{L}(\mu_t \mid h_{t-1}) - F_{\alpha^*}^{\mu_0}\|_{TV} &\leqslant \|\mathcal{L}(d_t \mid h_{t-1}) - \eta_{\alpha^*}\|_{TV} \\
&\leqslant \big\|\mathcal{L}(d_t \mid h_{t-1}) - \eta_{\alpha_{t-1}}\big\|_{TV} + \big\|\eta_{\alpha_{t-1}} - \eta_{\alpha^*}\big\|_{TV} \\
&\leqslant \big\|\mathcal{L}(d_t \mid h_{t-1}) - \eta_{\alpha_{t-1}}\big\|_{TV} + k\|\alpha_{t-1} - \alpha^*\|_1 \longrightarrow 0
\end{aligned}$$

on every history on which $\alpha_t \to \alpha^*$ and the conclusion of Lemma 1 holds. This occurs $\mathbb{P}_{\mu_0,\pi}$-almost surely. Total-variation convergence implies the claimed weak convergence. □

Proof of Lemma A.3.

1. Since the set of actions is finite, there is at least one measurable selection from the best reply correspondence.
2. $\Delta(A)$ is finite-dimensional and bounded, $\cup_{\rho\in\mathcal{O}}\rho(\nu)$ is closed for every $\nu\in\Delta(\Theta)$, and $\Psi^{\mu_0}(\alpha)$ is the Aumann integral Aumann [1965] of the (mixed) best reply correspondence with respect to the distribution of beliefs $F_\alpha^{\mu_0}$. Therefore, it satisfies the assumptions of case (i) of Theorem 2.1.37 of Molchanov [2017], so it is closed.
3. By Lemma A.2, $F_{(\cdot)}^{\mu_0}$ is continuous in α, and so by Artstein and Wets [1988], Theorem 4.2, Ψ^{μ_0} is upper hemicontinuous.
4. This follows immediately from the definition of Ψ^{μ_0}.
5. This follows immediately from the definition of stochastic memory equilibrium.

□

Proof of Lemma A.4. Since A is finite, it is enough that for every $a\in A$ the probability that a is used at period $t+1$ given h_t converges to $\psi_\pi^{\mu_0}(\alpha_t)$. By definition of $\psi_\pi^{\mu_0}$, a sufficient condition for this is that the distribution of databases given h_t converges to η_{α_t}. The result thus follows by Lemma 1. □

Proof of Lemma B.1. Suppose the agent takes action 1 with frequency α. If there is a reminder, the agent takes action 0 only if they do not get a reminder and they do not recall a good outcome, which has probability $\exp^{-\alpha kp}(1-\phi)$, or they do not recall a good outcome and are given a reminder, but the reminder is about a bad outcome, which has probability $\exp^{-(\alpha kp)}\phi(1-p)$. Thus, the equilibrium probability that the agent takes action 1 is given by the fixed-point equation $\alpha_{\phi,p}(1)=1-e^{-kp\alpha_{\phi,p}(1)}(1-\phi+\phi(1-p))$. First, observe that for $\phi>0$ any solution to the equation

$$\alpha_{\phi,p}(1)=1-e^{-k\alpha_{\phi,p}(1)p}(1-\phi+\phi(1-p)) \tag{23}$$

must satisfy $\alpha_{\phi,p}(1)>0$ as the right-hand side is strictly positive. For $\phi=0$ the function $\hat\alpha\mapsto\hat\alpha-\left(1-e^{-kp\,\hat\alpha}\right)$ has derivative $1-kp\,e^{-kp\,\hat\alpha}$ and is convex. As for $kp\leqslant 1$, the function crosses 0 from below at $\hat\alpha=0$; this is the unique solution to (23).

For $kp>1$ or $\phi>0$ a positive solution exists, namely $\alpha_{\phi,p}(1)=1+\frac{1}{kp}W\left(-e^{-kp}kp(1-p\phi)\right)$, which we select by assumption. The function $\hat\alpha\mapsto\hat\alpha-\left(1-e^{-kp\,\hat\alpha}\right)$ has positive derivative at $\alpha_{0,p}(1)$, so $0\leqslant 1-kp\,e^{-kp\,\alpha_{0,p}(1)}=1-kp(1-\alpha_{0,p}(1))$. To show that $\alpha_{0,p}(1)$ is increasing in p, observe that by equation (23) it is the solution to $\alpha_{0,p}(1)=1-e^{-kp\,\alpha_{0,p}(1)}$. By the

implicit function theorem

$$\frac{\partial \alpha_{0,p}(1)}{\partial p} = \left[k\alpha_{0,p}(1) + kp\frac{\partial \alpha_{0,p}(1)}{\partial p}\right] e^{-kp\alpha_{0,p}(1)}$$
$$\Rightarrow \frac{\partial \alpha_{0,p}(1)}{\partial p}\left[1 - kp(1-\alpha_{0,p}(1))\right] = k\alpha_{0,p}(1)(1-\alpha_{0,p}(1))\,.$$

Since $1 - kp(1-\alpha_{0,p}(1)) \geqslant 0$, $\frac{\partial \alpha_{0,p}(1)}{\partial p} > 0$ for $kp > 0$. □

Lemma B.2. *Suppose $\phi = 0$ and there are multiple equilibria when $\epsilon = 0$. Then for any $\epsilon \in (0,1)$, there is a unique equilibrium, and it has $\alpha(1) > 0$. Moreover, this equilibrium converges to the equilibrium with $\alpha(1) > 0$ as $\epsilon \searrow 0$.*

Proof. There are multiple equilibria for $\epsilon = 0$ if and only if $pk > 1$. To see that the equilibrium is unique if with probability $\epsilon > 0$ the agent takes the action $a = 1$, observe that the frequency with which the agent takes the action $a = 1$ must solve $\alpha_p(1) = \epsilon + (1-\epsilon)[1 - e^{-k\alpha_p(1)p}]$. Any solution of this equation must satisfy $\alpha > \epsilon$. The function $\alpha \mapsto \epsilon + (1-\epsilon)[1 - e^{-k\alpha p}] - \alpha$ is concave, so it has at most 2 roots. At $\alpha = 0$ it is positive and has derivative $(1-\epsilon)kp - 1$, which as $kp > 1$, is also positive for ϵ small enough. Thus, this equation can only have the single positive solution $\alpha(1) = 1 + \frac{1}{kp}W\left(-e^{-kp}kp(1-\epsilon)\right)$, which converges to the equilibrium where $\alpha(1) > 0$ as $\epsilon \to 0$ because W is continuous. □

Proof of Lemma A.5. Let $\mu_k := k\alpha^k(a)p^*_a(y)$ and $\nu_k := k\alpha^k(a')p^*_{a'}(y')$. Since $\hat{\alpha}(a') > 0$ and $p^*_{a'}(y') > 0$, we have $\lim_{k\to\infty}\nu_k = \infty$. If $p^*_a(y) = 0$, then $\mu_k = 0$ for all $k \in \mathbb{N}$ and $d(a,y) = 0$ almost surely under η^k_α. Hence $\frac{d(a,y)}{d(a',y')} = 0$ whenever $d(a',y') > 0$, and since $\lim_{k\to\infty}\nu_k = \infty$, $\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}[d(a',y') = 0] = 0$. Thus the first claim holds because both the ratio and its limit are equal to 0 with probability converging to 1. The second claim concerns only $d(a',y')$ and is unaffected by $p^*_a(y)$, so it follows from the argument below. Therefore, in what follows, suppose $p^*_a(y) > 0$. Similarly, if $(a,y) = (a',y')$ the first claim trivially holds, and since the second claim concerns only $d(a',y')$, it follows from the argument below. Therefore, in what follows, suppose $(a,y) \neq (a',y')$.

Case 1: $\hat{\alpha}(a) > 0$. Then also $\lim_{k\to\infty}\mu_k = \infty$. For any $\beta \in (0,1)$, Chernoff's theorem (see, e.g., Theorem 1.3.12 in Stroock [2010]) applied to the Poisson moment generating function $M(t) = e^{\mu_k(e^t-1)}$ yields $\mathbb{P}_{\eta^k_{\alpha^k}}[d(a,y) \geqslant (1+\beta)\mu_k] \leqslant \exp\big(-\mu_k\big[(1+\beta)\log(1+\beta) - \beta\big]\big)$ and $\mathbb{P}_{\eta^k_{\alpha^k}}[d(a,y) \leqslant (1-\beta)\mu_k] \leqslant \exp\big(-\mu_k\big[\beta + (1-\beta)\log(1-\beta)\big]\big)$ with analogous bounds for $d(a',y')$. For any $\beta \in (0,1)$, the Chernoff bounds above imply that

$$\mathbb{P}_{\eta^k_{\alpha^k}}[|d(a,y) - \mu_k| > \beta\mu_k] \leqslant 2\exp\big(-c(\beta)\mu_k\big) \quad \text{and} \quad \mathbb{P}_{\eta^k_{\alpha^k}}[|d(a',y') - \nu_k| > \beta\nu_k] \leqslant 2\exp\big(-c(\beta)\nu_k\big),$$

where $c(\beta) := \min\{(1+\beta)\log(1+\beta) - \beta,\ \beta + (1-\beta)\log(1-\beta)\} > 0$.

Using a union bound (and independence under $\eta^k_{\alpha^k}$), we obtain

$$\mathbb{P}_{\eta^k_{\alpha^k}}[|d(a,y) - \mu_k| > \beta\mu_k \text{ or } |d(a',y') - \nu_k| > \beta\nu_k] \leqslant 2e^{-c(\beta)\mu_k} + 2e^{-c(\beta)\nu_k} \leqslant 4e^{-c(\beta)\min\{\mu_k,\nu_k\}}.$$

Since $\lim_{k\to\infty} \min\{\mu_k, \nu_k\} = \infty$, the right-hand side converges to 0. Now set $r_k := \frac{\mu_k}{\nu_k} = \frac{\alpha^k(a)p^*_a(y)}{\alpha^k(a')p^*_{a'}(y')}$ and $r := \frac{\hat{\alpha}(a)p^*_a(y)}{\hat{\alpha}(a')p^*_{a'}(y')}$.

Since $\lim_{k\to\infty} r_k = r \in (0,\infty)$, there exists $M > 0$ and $K \in \mathbb{N}$ such that $r_k \leqslant M$ for all $k \geqslant K$. Choose $\beta \in (0,1)$ small enough that $M\dfrac{2\beta}{1-\beta} < \varepsilon$. On the event $E_k := \Big\{|d(a,y) - \mu_k| \leqslant \beta\mu_k \text{ and } |d(a',y') - \nu_k| \leqslant \beta\nu_k\Big\}$, we have

$$\frac{1-\beta}{1+\beta} r_k \leqslant \frac{d(a,y)}{d(a',y')} \leqslant \frac{1+\beta}{1-\beta} r_k,$$

so for all $k \geqslant K$

$$\left|\frac{d(a,y)}{d(a',y')} - r_k\right| \leqslant r_k \frac{2\beta}{1-\beta} \leqslant M\frac{2\beta}{1-\beta} < \varepsilon.$$

Therefore,

$$\lim_{k\to\infty} \mathbb{P}_{\eta^k_{\alpha^k}}\left[\left|\frac{d(a,y)}{d(a',y')} - \frac{\alpha^k(a)p^*_a(y)}{\alpha^k(a')p^*_{a'}(y')}\right| > \varepsilon\right] \leqslant \lim_{k\to\infty} \mathbb{P}_{\eta^k_{\alpha^k}}[E^c_k] = 0.$$

Case 2: $\hat{\alpha}(a) = 0$**.** Then $\lim_{k\to\infty} r_k = 0$. Fix $\varepsilon > 0$. For all sufficiently large k,

$$\left|\frac{d(a,y)}{d(a',y')} - r_k\right| > \varepsilon \quad \Rightarrow \quad \frac{d(a,y)}{d(a',y')} > \varepsilon/2.$$

Hence, for all sufficiently large k,

$$\mathbb{P}_{\eta^k_{\alpha^k}}\left[\left|\frac{d(a,y)}{d(a',y')} - r_k\right| > \varepsilon\right] \leqslant \mathbb{P}_{\eta^k_{\alpha^k}}\left[\frac{d(a,y)}{d(a',y')} > \varepsilon/2\right].$$

Now

$$\mathbb{P}_{\eta^k_{\alpha^k}}\left[\frac{d(a,y)}{d(a',y')} > \varepsilon/2\right] \leqslant \mathbb{P}_{\eta^k_{\alpha^k}}\left[\frac{d(a,y)}{d(a',y')} > \varepsilon/2 \middle| d(a',y') \geqslant \frac{1}{2}\nu_k\right] + \mathbb{P}_{\eta^k_{\alpha^k}}\left[d(a',y') < \frac{1}{2}\nu_k\right] \tag{24}$$

$$\leqslant \mathbb{P}_{\eta^k_{\alpha^k}}\left[d(a,y) > \frac{\varepsilon}{4}\nu_k\right] + \mathbb{P}_{\eta^k_{\alpha^k}}\left[d(a',y') < \frac{1}{2}\nu_k\right]. \tag{25}$$

The first term is bounded by Markov's inequality: $\mathbb{P}_{\eta^k_{\alpha^k}}\big[d(a,y) > \frac{\varepsilon}{4}\nu_k\big] \leqslant \frac{4\mu_k}{\varepsilon\nu_k} \to 0$ since $\lim_{k\to\infty} r_k = \lim_{k\to\infty} \mu_k/\nu_k = 0$. The second term goes to 0 by the lower-tail Chernoff bound above (with $\beta = 1/2$), because $\lim_{k\to\infty} \nu_k = \infty$. This establishes the second claim of

the lemma and also implies that

$$\lim_{k\to\infty} \mathbb{P}_{\eta^k_{\alpha^k}}\left[\left|\frac{d(a,y)}{d(a',y')} - \frac{\alpha^k(a)p^*_a(y)}{\alpha^k(a')p^*_{a'}(y')}\right| > \varepsilon\right] = 0.$$

□

Proof of Lemma A.6. If $\hat{\alpha}$ is a unitary-data self-confirming equilibrium then for all $a' \in \operatorname{supp}(\hat{\alpha})$ there is $\nu^{a'} \in \Delta(\Theta(\hat{\alpha}))$ such that $a' \in BR(\nu^{a'})$. Moreover, because the agent is correctly specified for all $a', a'' \in \operatorname{supp}(\alpha)$ and for every $p \in \operatorname{supp}\nu^{a'}$, $p_{a''} = p^*_{a''}$, so the agent has a single and correct belief $\hat{u}_{a'}$ about the expected payoff of every $a' \in \operatorname{supp}(\alpha)$. And since $\hat{\alpha}$ is a unitary-data self-confirming equilibrium, $\hat{u}_{a''} = \hat{u}_{a'} \geqslant \int_\Theta \sum_{y\in Y} p_a(y)u(a,y)d\nu^{a'}(p)$ for all $a'', a' \in \operatorname{supp}(\alpha)$ and $a \in A$. Thus, $\hat{\alpha}$ is a unitary-belief self-confirming equilibrium. □

Proof of Lemma A.7. Put $E := A \times Y$, let $N_{t-1}(a,y)$ be the number of occurrences of (a,y) through period $t-1$, and, for $z = (a,y) \in E$, set

$$c_z(d,y') := k + r\mathbb{1}_{\{d(z)>0 \text{ or } z=(b(d),y')\}}.$$

For fixed d and y', abbreviate

$$\begin{aligned}
n_{t,z}(d,y') &:= N_{t-1}(z) + \mathbb{1}_{\{z=(b(d),y')\}},\\
\lambda_{t,z}(d,y') &:= \frac{c_z(d,y')}{t} n_{t,z}(d,y'),\\
\overline{\lambda}_{t,z}(d,y') &:= c_z(d,y') q_{\alpha^d_t}(z).
\end{aligned}$$

For all sufficiently large t, conditional on $\mathcal{G}_t$, $d_t = d$, and $y_t = y'$, the coordinates of d_{t+1} are independent and their joint law is

$$B^{d,y'}_t := \bigotimes_{z\in E} \operatorname{Bin}\left(n_{t,z}(d,y'), \frac{c_z(d,y')}{t}\right).$$

Let

$$\widetilde{P}^{d,y'}_t := \bigotimes_{z\in E} \operatorname{Pois}\big(\lambda_{t,z}(d,y')\big),$$

and write

$$P^{d,y'}_{\alpha^d_t} := \bigotimes_{z\in E} \operatorname{Pois}\big(\overline{\lambda}_{t,z}(d,y')\big).$$

Thus

$$Q_t(d,\cdot) = \sum_{y'\in Y} p^*_{b(d)}(y')B^{d,y'}_t, \qquad P_{\alpha^d_t}(d,\cdot) = \sum_{y'\in Y} p^*_{b(d)}(y')P^{d,y'}_{\alpha^d_t}.$$

Convexity of total variation and the triangle inequality give

$$\begin{aligned}&\left\|Q_t(d,\cdot) - P_{\alpha_t^d}(d,\cdot)\right\|_{TV} \\ &\quad\leqslant \sum_{y'\in Y} p^*_{b(d)}(y') \left(\left\|B_t^{d,y'} - \widetilde{P}_t^{d,y'}\right\|_{TV} + \left\|\widetilde{P}_t^{d,y'} - P_{\alpha_t^d}^{d,y'}\right\|_{TV}\right).\end{aligned} \tag{26}$$

We bound the two terms separately. Le Cam's inequality and the product inequality for total variation imply

$$\begin{aligned}\left\|B_t^{d,y'} - \widetilde{P}_t^{d,y'}\right\|_{TV} &\leqslant 2\sum_{z\in E} n_{t,z}(d,y') \left(\frac{c_z(d,y')}{t}\right)^2 \\ &\leqslant \frac{2(k+r)^2}{t^2}\sum_{z\in E} n_{t,z}(d,y') = \frac{2(k+r)^2}{t},\end{aligned} \tag{27}$$

where the last equality uses $\sum_z n_{t,z}(d,y') = t$. Next, Poisson random variables with means λ and λ' can be coupled so that their mismatch probability is at most $1-e^{-|\lambda-\lambda'|} \leqslant |\lambda-\lambda'|$. Coupling the coordinates independently and using the product inequality therefore gives

$$\begin{aligned}\left\|\widetilde{P}_t^{d,y'} - P_{\alpha_t^d}^{d,y'}\right\|_{TV} &\leqslant \sum_{z\in E} \left|\lambda_{t,z}(d,y') - \overline{\lambda}_{t,z}(d,y')\right| \\ &\leqslant (k+r)\sum_{z\in E}\left|\frac{n_{t,z}(d,y')}{t} - q_{\alpha_t^d}(z)\right|.\end{aligned} \tag{28}$$

For $z=(a,y)$, the difference in the last line is exactly

$$\begin{aligned}\frac{n_{t,(a,y)}(d,y')}{t} - q_{\alpha_t^d}(a,y) &= \frac{t-1}{t}\left(\upsilon_{t-1}(a,y) - \alpha_{t-1}(a)p^*_a(y)\right) \\ &\quad + \frac{1}{t}\mathbb{1}_{\{a=b(d)\}}\left(\mathbb{1}_{\{y=y'\}} - p^*_{b(d)}(y)\right).\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{z\in E}\left|\frac{n_{t,z}(d,y')}{t} - q_{\alpha_t^d}(z)\right| &\leqslant \frac{t-1}{t}\left\|\upsilon_{t-1} - q_{\alpha_{t-1}}\right\|_1 + \frac{1}{t}\sum_{y\in Y}\left|\mathbb{1}_{\{y=y'\}} - p^*_{b(d)}(y)\right| \\ &\leqslant \frac{t-1}{t}\left\|\upsilon_{t-1} - q_{\alpha_{t-1}}\right\|_1 + \frac{2}{t},\end{aligned} \tag{29}$$

uniformly in d and y'. Substituting (27)–(29) into (26) yields

$$\begin{aligned}\varepsilon_t &:= \sup_{d\in\mathcal{D}}\left\|Q_t(d,\cdot) - P_{\alpha_t^d}(d,\cdot)\right\|_{TV} \\ &\leqslant \frac{2(k+r)^2 + 2(k+r)}{t} + (k+r)\frac{t-1}{t}\left\|\upsilon_{t-1} - q_{\alpha_{t-1}}\right\|_1.\end{aligned} \tag{30}$$

Along the realized state, $d = d_t$ and $\alpha_t^d = \alpha_t$. If one does not combine the current-

observation term with the definition of α_t^d, the same calculation gives directly

$$\begin{aligned}
&\|Q_t(d_t,\cdot) - P_{\alpha_t}(d_t,\cdot)\|_{TV} \\
&\quad\leqslant \frac{2(k+r)^2}{t} + (k+r)\sum_{z\in E}\left|\frac{N_{t-1}(z) + \mathbb{1}_{\{z=(b(d_t),y')\}}}{t} - q_{\alpha_t}(z)\right| \\
&\quad\leqslant \frac{2(k+r)^2 + (k+r)}{t} + (k+r)\sum_{(a,y)\in E}\left|\frac{N_{t-1}(a,y)}{t} - \alpha_t(a)p_a^*(y)\right|,
\end{aligned}$$

uniformly in the current-outcome component y'. Averaging over y' preserves this bound. This is the realized-path form of (30). In particular this has the stated form $C/t + C\|\upsilon_{t-1} - q_{\alpha_{t-1}}\|_1$, with a constant independent of the current database. Consequently, Lemma A.1 implies

$$\varepsilon_t \longrightarrow 0 \qquad \mathbb{P}_{\mu_0,\pi}\text{-a.s.} \tag{31}$$

This proves part 1. The belief-law conclusion follows because measurable pushforwards contract total variation.

For part 2, let $\mathbf{0}$ be the empty database. For every α and d,

$$\begin{aligned}
P_\alpha(d,\{\mathbf{0}\}) &= \sum_{y'\in Y} p^*_{b(d)}(y')\exp\left\{-\sum_{(a,y)\in E} c_{(a,y)}(d,y')\alpha(a)p_a^*(y)\right\} \\
&\geqslant \exp\{-(k+r)\} =: \epsilon_0 > 0.
\end{aligned}$$

Thus

$$P_\alpha(d,\cdot) = \epsilon_0\delta_{\mathbf{0}} + (1-\epsilon_0)R_\alpha(d,\cdot)$$

for a Markov kernel R_α. Its Dobrushin coefficient is at most $1-\epsilon_0$, and hence

$$\sup_{d\in\mathcal{D}}\|P_\alpha^j(d,\cdot) - \mathcal{H}_\alpha\|_{TV} \leqslant 2(1-\epsilon_0)^j \qquad \forall j \geqslant 0, \quad \forall\alpha\in\Delta(A). \tag{32}$$

A coupling argument gives the required uniform Lipschitz bound. For fixed $d\in\mathcal{D}$ and $y'\in Y$, let

$$P_\alpha^{d,y'} := \bigotimes_{z=(a,y)\in E} \mathrm{Pois}\left(\lambda_{\alpha,z}^{d,y'}\right), \qquad \lambda_{\alpha,z}^{d,y'} := c_z(d,y')\alpha(a)p_a^*(y),$$

so that

$$P_\alpha(d,\cdot) = \sum_{y'\in Y} p^*_{b(d)}(y')P_\alpha^{d,y'}.$$

The mixture weights $p^*_{b(d)}(y')$ do not depend on α.

For each coordinate z, Poisson random variables with means $\lambda_{\alpha,z}^{d,y'}$ and $\lambda_{\beta,z}^{d,y'}$ can be

coupled as follows. Let

$$Z_z \sim \mathrm{Pois}\Big(\min\{\lambda^{d,y'}_{\alpha,z}, \lambda^{d,y'}_{\beta,z}\}\Big),$$

and independently let

$$R^{\alpha}_z \sim \mathrm{Pois}\Big((\lambda^{d,y'}_{\alpha,z} - \lambda^{d,y'}_{\beta,z})_+\Big),$$
$$R^{\beta}_z \sim \mathrm{Pois}\Big((\lambda^{d,y'}_{\beta,z} - \lambda^{d,y'}_{\alpha,z})_+\Big).$$

Then $X_z := Z_z + R^{\alpha}_z$ and $Y_z := Z_z + R^{\beta}_z$ have the desired Poisson marginal distributions. Taking these couplings independently across $z \in E$, the coupling inequality gives

$$\begin{aligned}\left\|P^{d,y'}_{\alpha} - P^{d,y'}_{\beta}\right\|_{TV} &\leqslant \mathbb{P}[(X_z)_{z\in E} \neq (Y_z)_{z\in E}] \\ &\leqslant 1 - \exp\left(-\sum_{z\in E}\left|\lambda^{d,y'}_{\alpha,z} - \lambda^{d,y'}_{\beta,z}\right|\right) \\ &\leqslant \sum_{z\in E}\left|\lambda^{d,y'}_{\alpha,z} - \lambda^{d,y'}_{\beta,z}\right|.\end{aligned}$$

Because $c_z(d, y') \leqslant k + r$, the last expression satisfies

$$\begin{aligned}\sum_{z=(a,y)\in E}\left|\lambda^{d,y'}_{\alpha,z} - \lambda^{d,y'}_{\beta,z}\right| &= \sum_{a\in A}\sum_{y\in Y} c_{(a,y)}(d, y')p^*_a(y)|\alpha(a) - \beta(a)| \\ &\leqslant (k + r)\sum_{a\in A}|\alpha(a) - \beta(a)|\sum_{y\in Y} p^*_a(y) \\ &= (k + r)\|\alpha - \beta\|_1.\end{aligned}$$

Finally, convexity of total variation under mixtures with common weights implies

$$\begin{aligned}\|P_\alpha(d, \cdot) - P_\beta(d, \cdot)\|_{TV} &\leqslant \sum_{y'\in Y} p^*_{b(d)}(y')\left\|P^{d,y'}_{\alpha} - P^{d,y'}_{\beta}\right\|_{TV} \\ &\leqslant (k + r)\|\alpha - \beta\|_1.\end{aligned}$$

The bound is independent of d, and therefore

$$\sup_{d\in\mathcal{D}}\|P_\alpha(d, \cdot) - P_\beta(d, \cdot)\|_{TV} \leqslant (k + r)\|\alpha - \beta\|_1. \tag{33}$$

Using invariance and the preceding contraction,

$$\|\mathcal{H}_\alpha - \mathcal{H}_\beta\|_{TV} \leqslant \frac{k + r}{\epsilon_0}\|\alpha - \beta\|_1. \tag{34}$$

Finally, $F^{\mu_0}_{\alpha,d}$ is the pushforward of $P_\alpha(d, \cdot)$ under $d' \mapsto \mu(\cdot \mid d')$, so

$$\sup_{d\in\mathcal{D}}\|F^{\mu_0}_{\alpha,d} - F^{\mu_0}_{\beta,d}\|_{TV} \leqslant (k + r)\|\alpha - \beta\|_1.$$

This proves part 2. □

Proof of Corollary 2. By Lemma A.7, along the specified history sequence,

$$\sup_{d\in\mathcal{D}} \|Q_t(d,\cdot) - P_{\alpha_t^d}(d,\cdot)\|_{TV} \to 0,$$

where

$$\alpha_t^d = \frac{t-1}{t}\alpha_{t-1} + \frac{1}{t}\delta_{\pi(\mu(\cdot|d))}.$$

Since $\alpha_t \to \alpha$, we also have

$$\sup_{d\in\mathcal{D}} \|\alpha_t^d - \alpha\|_1 \to 0.$$

Lemma A.7 also gives Lipschitz continuity of $\alpha \mapsto P_\alpha$ in the uniform total-variation norm. Hence

$$\varepsilon_t := \sup_{d\in\mathcal{D}} \|Q_t(d,\cdot) - P_\alpha(d,\cdot)\|_{TV} \to 0.$$

The proof of Lemma A.7 shows that P_α satisfies the uniform minorization

$$P_\alpha(d,\cdot) = \epsilon_0\delta_{\mathbf{0}} + (1-\epsilon_0)R_\alpha(d,\cdot), \qquad \epsilon_0 := e^{-(k+r)} > 0,$$

for some Markov kernel R_α. Therefore, for any two probability measures λ, λ' on $\mathcal{D}$,

$$\|\lambda P_\alpha - \lambda' P_\alpha\|_{TV} \leqslant (1-\epsilon_0)\|\lambda - \lambda'\|_{TV}.$$

Using the invariance $\mathcal{H}_\alpha P_\alpha = \mathcal{H}_\alpha$, we get

$$\begin{aligned}
\|\lambda_{t+1} - \mathcal{H}_\alpha\|_{TV} &= \|\lambda_t Q_t - \mathcal{H}_\alpha P_\alpha\|_{TV} \\
&\leqslant \|\lambda_t Q_t - \lambda_t P_\alpha\|_{TV} + \|\lambda_t P_\alpha - \mathcal{H}_\alpha P_\alpha\|_{TV} \\
&\leqslant \varepsilon_t + (1-\epsilon_0)\|\lambda_t - \mathcal{H}_\alpha\|_{TV}.
\end{aligned}$$

Since $\varepsilon_t \to 0$ and $1-\epsilon_0 < 1$, this recursion implies

$$\|\lambda_t - \mathcal{H}_\alpha\|_{TV} \to 0.$$

□

Proof of Lemma A.8. Observe that for every $\alpha \in \Delta(A)$, $\int_{\mathcal{D}} \Psi^{\mu_0}(\alpha, d')d\mathcal{H}_\alpha(d')$ is the Aumann [1965] integral of the mixed best reply correspondence with respect to the measure $\int_{\mathcal{D}} F^{\mu_0}_{\alpha,d'} d\mathcal{H}_{\boldsymbol{\alpha}}(d')$. Then 1) follows from the finiteness of A and 2) follows from the finite dimensionality of $\Delta(A)$ and Theorem 2.1.37, case (i) of Molchanov [2017]. From Lemma A.7, $F^{\mu_0}_{(\cdot)}$ is continuous in α. Moreover, part 2 of Lemma A.7 shows that $\alpha \mapsto \mathcal{H}_\alpha$ is Lipschitz in total variation and that $\alpha \mapsto F^{\mu_0}_{\alpha,d}$ is Lipschitz in total variation uniformly in d. Therefore,

by Artstein and Wets [1988], Theorem 4.2, Ψ^{μ_0} is upper hemicontinuous, which proves 3). Finally, 4) is immediate from the definition of $\int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d')d\mathcal{H}_{(\cdot)}(d')$, and 5) is immediate from the definition of ergodic memory equilibrium and Ψ^{μ_0}. □

Proof of Claim 1. For every $\succ' \neq \succ''$

$$\frac{\partial f\left((c_\succ)_{\succ\in\mathcal{P}}\right)}{\partial c_{\succ'}\partial c_{\succ''}} = -\int_{-\infty}^{+\infty} \phi\left(x - c_{\succ'}\right)\phi\left(x - c_{\succ''}\right) \prod_{\succ\in\mathcal{P}\setminus\{\succ',\succ''\}} \Phi\left(x - c_\succ\right) dx := -q_{\succ'\succ''}\left((c_\succ)_{\succ\in\mathcal{P}}\right) < 0.$$

and thus, since $\sum_{\succ*\in\mathcal{P}} \frac{\mathbb{P}[X_{\succ'}+c_{\succ'}>X_{\succ''}+c_{\succ''},\forall\succ''\in\mathcal{P}\setminus\{\succ'\}]}{\partial c_{\succ*}} = 0$, we have

$$\frac{\partial f\left((c_\succ)_{\succ\in\mathcal{P}}\right)}{\partial^2 c_{\succ'}} = \sum_{\succ''\in\mathcal{P}} q_{\succ'\succ''}\left((c_\succ)_{\succ\in\mathcal{P}}\right).$$

Now, evaluate the quadratic form of the Hessian for an arbitrary non-zero vector $v \in \mathbb{R}^{\mathcal{P}}$: $v^T H_f((c_\succ)_{\succ\in\mathcal{P}})v = \sum_{\succ'\in\mathcal{P}} v_\succ^2 \left(\sum_{\succ''\neq\succ'} q_{\succ'\succ''}((c_\succ)_{\succ\in\mathcal{P}})\right) - \sum_{\succ'\in\mathcal{P}}\sum_{\succ''\neq\succ'} q_{\succ'\succ''}((c_\succ)_{\succ\in\mathcal{P}})v_{\succ'}v_{\succ''}$.

Since $q_{\succ'\succ''}((c_\succ)_{\succ\in\mathcal{P}}) = q_{\succ''\succ'}((c_\succ)_{\succ\in\mathcal{P}})$, this simplifies to:

$$v^T H_f(c)v = \frac{1}{2}\sum_{\succ'\in\mathcal{P}}\sum_{\succ''\neq\succ'} q_{\succ'\succ''}((c_\succ)_{\succ\in\mathcal{P}})(v_{\succ'} - v_{\succ''})^2$$

Because $q_{\succ'\succ''}((c_\succ)_{\succ\in\mathcal{P}}) > 0$ for all $\succ''\neq\succ'$, the quadratic form equals zero ($v^T H_f(c)v = 0$) if and only if v is constant. This means the null space of the Hessian is exactly the one-dimensional subspace spanned by the all-ones vector: $\ker(H_f(c)) = \text{span}(1, 1, \ldots, 1)^T$.

By definition, the hyperplane H is the orthogonal complement to this null space ($H = \text{span}(\mathbf{1})^\perp$). Consequently, for any non-zero tangent vector $v \in H$, it cannot be a uniform vector, so $v^T H_f(c)v > 0$ for all $v \in H\setminus\{\mathbf{0}\}$.

This establishes that the Hessian restricted to H is strictly positive definite, proving f is strictly convex on H. □

B.3 Missing Steps of the Proofs of Theorem 4

We maintain the notation of Lemma A.7. For a bounded function $f : \mathcal{D} \to \mathbb{R}^A$, let

$$\text{osc}(f) := \sup_{d,d'\in\mathcal{D}} \|f(d) - f(d')\|,$$

and let $\mathbb{B}_{\text{osc}}$ denote the space of bounded functions modulo constant functions, equipped with the oscillation norm. Write $[f]$ for the equivalence class of f. Since Markov operators preserve constants, P_α induces an operator $\widehat{P}_\alpha$ on $\mathbb{B}_{\text{osc}}$.

Recall from the proof of Lemma A.7 that

$$P_\alpha(d,\cdot) = \epsilon_0 \delta_{\mathbf{0}} + (1-\epsilon_0) R_\alpha(d,\cdot), \qquad \epsilon_0 = e^{-(k+r)},$$

for a Markov kernel R_α. Consequently, for every bounded f,

$$\operatorname{osc}(P_\alpha f) \leqslant (1-\epsilon_0)\operatorname{osc}(f).$$

Indeed, the contribution of the common atom is independent of d, whereas the difference between two averages of f under probability measures is bounded by $\operatorname{osc}(f)$. Hence

$$\|\widehat{P}_\alpha\|_{\mathrm{osc}} \leqslant 1-\epsilon_0$$

uniformly in α. It follows that $I - \widehat{P}_\alpha$ is invertible and that

$$\mathcal{R}_\alpha := (I-\widehat{P}_\alpha)^{-1} = \sum_{j=0}^{\infty} \widehat{P}_\alpha^{\,j}, \qquad \|\mathcal{R}_\alpha\|_{\mathrm{osc}} \leqslant \frac{1}{\epsilon_0}. \tag{35}$$

Set

$$h_\alpha(d) := G(\alpha,d) - \overline{G}(\alpha).$$

By the definition of $\overline{G}(\alpha)$,

$$\int_{\mathcal{D}} h_\alpha(d)\, d\mathcal{H}_\alpha(d) = 0. \tag{36}$$

Moreover, the boundedness of G, together with (33) and (34), gives constants $M_h, L_h < \infty$ such that

$$\sup_{\alpha,d} \|h_\alpha(d)\| \leqslant M_h, \qquad \operatorname{osc}(h_\alpha - h_\beta) \leqslant L_h \|\alpha-\beta\|_1. \tag{37}$$

For example, the second assertion follows by writing

$$\overline{G}(\alpha) - \overline{G}(\beta) = \int_{\mathcal{D}} \big(G(\alpha,d) - G(\beta,d)\big)\, d\mathcal{H}_\alpha(d) + \int_{\mathcal{D}} G(\beta,d)\, d(\mathcal{H}_\alpha - \mathcal{H}_\beta)(d)$$

and applying the uniform Lipschitz bound for G and (34).

Define the equivalence class of the corrector by

$$[U_\alpha] := \mathcal{R}_\alpha[h_\alpha],$$

and select its unique representative satisfying $U_\alpha(\mathbf{0}) = 0$. Equivalently,

$$U_\alpha(d) = \sum_{j=0}^{\infty} \left(P_\alpha^j h_\alpha(d) - P_\alpha^j h_\alpha(\mathbf{0})\right). \tag{38}$$

The series converges uniformly because

$$\left\|P_\alpha^j h_\alpha(d) - P_\alpha^j h_\alpha(\mathbf{0})\right\| \leqslant (1-\epsilon_0)^j \operatorname{osc}(h_\alpha).$$

The quotient-space identity

$$(I - \widehat{P}_\alpha)[U_\alpha] = [h_\alpha]$$

initially implies that, for some constant vector c_α,

$$U_\alpha - P_\alpha U_\alpha = h_\alpha + c_\alpha.$$

Integrating this equality against the invariant distribution $\mathcal{H}_\alpha$, using invariance and (36), gives

$$0 = \int_{\mathcal{D}} (U_\alpha - P_\alpha U_\alpha)\, d\mathcal{H}_\alpha = c_\alpha.$$

Thus $c_\alpha = 0$, and U_α solves

$$U_\alpha - P_\alpha U_\alpha = h_\alpha, \qquad U_\alpha(\mathbf{0}) = 0. \tag{39}$$

We next establish the uniform and Lipschitz bounds. Since $U_\alpha(\mathbf{0}) = 0$,

$$\sup_d \|U_\alpha(d)\| \leqslant \operatorname{osc}(U_\alpha) \leqslant \frac{\operatorname{osc}(h_\alpha)}{\epsilon_0} \leqslant \frac{2M_h}{\epsilon_0}.$$

Thus one may take $C_U = 2M_h/\epsilon_0$.

To obtain the Lipschitz bound, first observe that (33) implies, for some $C_P < \infty$,

$$\|(\widehat{P}_\alpha - \widehat{P}_\beta)[f]\|_{\operatorname{osc}} \leqslant C_P \|\alpha - \beta\|_1 \|[f]\|_{\operatorname{osc}}. \tag{40}$$

For instance, under the convention $\|\mu - \nu\|_{TV} = \sup_B |\mu(B) - \nu(B)|$, one may take $C_P = 4(k+r)$: subtract a constant from f, apply the total-variation integration inequality, and then take the oscillation over the initial database.

The resolvent identity gives

$$\mathcal{R}_\alpha - \mathcal{R}_\beta = \mathcal{R}_\alpha(\widehat{P}_\alpha - \widehat{P}_\beta)\mathcal{R}_\beta.$$

Therefore,

$$\begin{aligned}
\|[U_\alpha] - [U_\beta]\|_{\operatorname{osc}} &\leqslant \|\mathcal{R}_\alpha\|_{\operatorname{osc}} \|[h_\alpha] - [h_\beta]\|_{\operatorname{osc}} \\
&\quad + \|\mathcal{R}_\alpha\|_{\operatorname{osc}} \|\widehat{P}_\alpha - \widehat{P}_\beta\|_{\operatorname{osc}} \|\mathcal{R}_\beta\|_{\operatorname{osc}} \|[h_\beta]\|_{\operatorname{osc}} \\
&\leqslant \left(\frac{L_h}{\epsilon_0} + \frac{2C_P M_h}{\epsilon_0^2}\right) \|\alpha - \beta\|_1.
\end{aligned}$$

Finally, $U_\alpha(\mathbf{0}) = U_\beta(\mathbf{0}) = 0$, so

$$\sup_d \|U_\alpha(d) - U_\beta(d)\| \leqslant \operatorname{osc}(U_\alpha - U_\beta).$$

Consequently, with

$$L_U := \frac{L_h}{\epsilon_0} + \frac{2C_P M_h}{\epsilon_0^2},$$

we obtain

$$\sup_{\alpha,d} \|U_\alpha(d)\| \leqslant C_U, \qquad \sup_d \|U_\alpha(d) - U_\beta(d)\| \leqslant L_U \|\alpha - \beta\|_1. \tag{41}$$

For a bounded f, write $Q_t f(d_t) := \int f(d') Q_t(d_t, dd')$, and put $U_t := U_{\alpha_t}$. Part 1 of Lemma A.7 gives

$$\varepsilon_t := \|Q_t(d_t, \cdot) - P_{\alpha_t}(d_t, \cdot)\|_{TV} \longrightarrow 0 \qquad \mathbb{P}_{\mu_0,\pi}\text{-a.s.}$$

The Poisson equation gives the pathwise decomposition

$$\begin{aligned} h_{\alpha_t}(d_t) &= [U_t(d_t) - U_t(d_{t+1})] + M_{t+1} + \rho_{t+1}, \\ M_{t+1} &:= U_t(d_{t+1}) - Q_t U_t(d_t), \\ \rho_{t+1} &:= (Q_t - P_{\alpha_t}) U_t(d_t). \end{aligned} \tag{42}$$

The sequence (M_{t+1}) is a uniformly bounded martingale-difference sequence. Since $\sum_t \gamma_{t+1}^2 < \infty$, where $\gamma_{t+1} = 1/(t+1)$, the martingale series $\sum_t \gamma_{t+1} M_{t+1}$ converges almost surely. Moreover,

$$\|\rho_{t+1}\| \leqslant C\varepsilon_t \longrightarrow 0 \qquad \mathbb{P}_{\mu_0,\pi}\text{-a.s.}$$

Writing $\lambda_t := \gamma_{t+1}$, summation by parts and (41) give, uniformly over $m \geqslant n$,

$$\begin{aligned} &\left\| \sum_{t=n}^{m} \lambda_t [U_t(d_t) - U_t(d_{t+1})] \right\| \\ &\quad \leqslant C_U(\lambda_n + \lambda_m) + C_U \sum_{t=n+1}^{m} |\lambda_t - \lambda_{t-1}| + L_U \sum_{t=n+1}^{m} \lambda_{t-1} \|\alpha_t - \alpha_{t-1}\|_1 \\ &\quad \leqslant \frac{C}{n} + C \sum_{t=n}^{\infty} \frac{1}{t^2} \longrightarrow 0, \end{aligned}$$

where $\|\alpha_t - \alpha_{t-1}\|_1 \leqslant 2/t$. On a window for which $\sum_{t=n}^m \lambda_t \leqslant T$,

$$\left\| \sum_{t=n}^{m} \lambda_t \rho_{t+1} \right\| \leqslant CT \sup_{t \geqslant n} \varepsilon_t \longrightarrow 0.$$

Together with convergence of the martingale series, these estimates prove that, equation

(20) holds almost surely for every $T > 0$. It remains to insert this estimate into the empirical-frequency recursion. Because $a_{t+1} = b(d_{t+1})$, $\alpha_{t+1} - \alpha_t = \gamma_{t+1}\big(\varphi(d_{t+1}) - \alpha_t\big)$. Define $\xi_{t+1} := \varphi(d_{t+1}) - Q_t\varphi(d_t)$ and $r_{t+1} := (Q_t - P_{\alpha_t})\varphi(d_t)$. Then (ξ_{t+1}) is a bounded martingale-difference sequence and $\|r_{t+1}\| \leqslant C\varepsilon_t \to 0$ almost surely. Hence

$$\alpha_{t+1} - \alpha_t = \gamma_{t+1}\left[\overline{G}(\alpha_t) - \alpha_t + \xi_{t+1} + r_{t+1} + G(\alpha_t, d_t) - \overline{G}(\alpha_t)\right]. \tag{43}$$

The accumulated contribution of the last three terms vanishes uniformly over every bounded stochastic-approximation time window: for ξ_{t+1} by martingale convergence, for r_{t+1} by $\varepsilon_t \to 0$, and for the controlled Markov term by (20).

Lemma A.8 shows that $\overline{\Psi}$ is nonempty, compact- and convex-valued, and upper hemi-continuous. Because π is an optimal Markovian policy, $G(\alpha, d) \in \Psi^{\mu_0}(\alpha, d)$ for every (α, d). Hence

$$\overline{G}(\alpha) = \int_{\mathcal{D}} G(\alpha, d)\, d\mathcal{H}_\alpha(d) \in \overline{\Psi}(\alpha).$$

The usual perturbed-solution criterion for stochastic approximation with differential inclusions therefore applies to (43). Thus the interpolation in (9) is a perturbed solution of

$$\dot{\boldsymbol{\alpha}}(s) \in \overline{\Psi}(\boldsymbol{\alpha}(s)) - \boldsymbol{\alpha}(s). \tag{44}$$

By Theorem 4.2 of Benaim et al. [2005], in the same tracking form used in the proof of Theorem 1, its sufficiently late shifts are uniformly approximated on every bounded interval by solutions of (44).

The rest is the separating-hyperplane and exit-time argument in the proof of Theorem 1, with $\overline{\Psi}$ in place of Ψ^{μ_0}. If $\alpha \notin \overline{\Psi}(\alpha)$, a separating linear functional decreases at a uniform positive rate along every solution of (44) while the solution remains in a sufficiently small neighborhood of α. Every such solution therefore leaves that neighborhood within a common finite time. The tracking property then implies that the interpolated empirical frequencies cannot remain forever in a smaller neighborhood of α. Hence α is not a limit frequency. The contrapositive, together with part 5 of Lemma A.8, proves that every limit frequency is an ergodic memory equilibrium.

B.4 Additional Examples

Example 1. *Let* $A = \{b, c\}, Y = \{y, b, c\}$, *and* $u(a, y) = 1$ *if* $a = y$ *and* 0 *otherwise. Suppose* $p^* = \delta_y, p' = 0.9\delta_b + 0.1\delta_{y,}$, $p'' = 0.9\delta_c + 0.1\delta_y$ *and* $\mu = \frac{1}{2}\delta_{p'} + \frac{1}{4}\delta_{p''} + \frac{1}{4}\delta_{p^*}$. *For*

any k, the unique stochastic memory equilibrium is $\alpha^* = \delta_b$ because the only outcome that happens with positive probability is y, and conditional on remembering any database d with only outcomes y, $\mu(p'|d) > \mu(p''|d)$ and so the best reply is b.

B.5 Computations for Section 4.2

If the agent recalls $n_0 \geqslant 1$ outcomes equal to 0 from action 1 and no occurrences of $1/b$, the posterior probability of DGP p is

$$\Pr(p \mid n_0 \text{ failures}) = \frac{q(1-b)^{n_0}}{q(1-b)^{n_0} + (1-q)}.$$

Since under p action 1 has expected payoff 1 and under p' it has expected payoff 0, the posterior expected payoff of action 1 is $\dfrac{q(1-b)^{n_0}}{q(1-b)^{n_0} + (1-q)}$. Because $0 < 1-b < 1$ and the map $x \mapsto \frac{qx}{qx+(1-q)}$ is increasing, this is at most $\dfrac{q(1-b)}{q(1-b)+(1-q)}$. Thus, if $d\frac{q(1-b)}{q(1-b)+(1-q)} < c$, then after recalling at least one failure and no successes, the posterior expected payoff of action 1 is strictly below the sure payoff c from action 0, so the agent optimally chooses action 0.

B.6 Computations for Section 4.4

The equilibrium with two actions is computed by solving the fixed point

$$\alpha^{new}(b) = \left(1 - e^{-\alpha^{new}(b)}\right) + \left(1 - e^{-(1-\alpha^{new}(b))}\right) e^{-\alpha^{new}(b)} + \frac{1}{2e}.$$

The equilibrium with three actions is computed by solving the fixed point

$$\alpha^{new}(b) = \left(1 - e^{-\alpha^{new}(b)}\right) + \left(1 - e^{\frac{-(1-\alpha^{new}(b))}{2}}\right)^2 e^{-\alpha^{new}(b)} + e^{-\alpha^{new}(b)} \left(1 - e^{\frac{-(1-\alpha^{new}(b))}{2}}\right) e^{\frac{-(1-\alpha^{new}(b))}{2}} + \frac{1}{3e}.$$

B.7 Multiplicity and Convergence in the Normal Environment

The normal-environment application in Proposition 3 does not imply uniqueness of stochastic memory equilibrium. The next proposition shows that multiplicity can arise.

Proposition 6. *There exist parameter values in the normal environment for which there are at least 3 stochastic memory equilibria.*

Proof. Restrict attention to two actions, $A = \{1, 2\}$. Let the payoff variance be $\sigma^2 = 1$, the prior variance be $\sigma_0^2 = \frac{1}{2}$, and the memory parameter be $k = 3$. For every candidate frequency of action 1 $\alpha \in [0, 1]$, Lemma 1 implies that the recalled sample sizes satisfy $N_1 \sim \text{Poisson}(3\alpha)$ and $N_2 \sim \text{Poisson}(3(1-\alpha))$, independently, where N_1 (resp. N_2) is the recalled number of periods where 1 (resp. 2) was chosen. Conditional on $(N_1, N_2) = (n_1, n_2)$, the posterior mean of action a is normally distributed with mean

$$m_a(n_a) := \frac{y_a^0/\sigma_0^2 + n_a\bar{y}_a/\sigma^2}{1/\sigma_0^2 + n_a/\sigma^2} = \frac{2y_a^0 + n_a\bar{y}_a}{2 + n_a}$$

and variance

$$v_a(n_a) := \frac{n_a/\sigma^2}{(1/\sigma_0^2 + n_a/\sigma^2)^2} = \frac{n_a}{(2 + n_a)^2}.$$

Hence, conditional on $(n_1, n_2) \neq (0, 0)$, the probability that action 1 is chosen is

$$Q(n_1, n_2; \bar{y}_1, \bar{y}_2, y_1^0, y_2^0) = \Phi\left(\frac{m_1(n_1) - m_2(n_2)}{\sqrt{v_1(n_1) + v_2(n_2)}}\right).$$

At $(n_1, n_2) = (0, 0)$, we adopt the convention

$$Q(0, 0; \bar{y}_1, \bar{y}_2, y_1^0, y_2^0) = \begin{cases} 1 & \text{if } y_1^0 > y_2^0, \\ 0 & \text{if } y_1^0 < y_2^0, \\ \frac{1}{2} & \text{if } y_1^0 = y_2^0. \end{cases}$$

Define

$$\Psi(\alpha; \bar{y}_1, \bar{y}_2, y_1^0, y_2^0) = \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} e^{-3\alpha}\frac{(3\alpha)^{n_1}}{n_1!}e^{-3(1-\alpha)}\frac{(3(1-\alpha))^{n_2}}{n_2!}Q(n_1, n_2; \bar{y}_1, \bar{y}_2, y_1^0, y_2^0).$$

Now fix the parameter vector $(\bar{y}_1, \bar{y}_2, y_1^0, y_2^0) = (1.75, 1.75, 0, 0)$.

For this parameter vector, the map $\alpha \mapsto \Psi(\alpha; 1.75, 1.75, 0, 0)$ is continuous on $[0, 1]$. Indeed, for each (n_1, n_2) the corresponding summand is continuous in α, and since

$$0 \leqslant Q(n_1, n_2; 1.75, 1.75, 0, 0) \leqslant 1,$$

the absolute value of the summand is bounded by

$$e^{-3\alpha}\frac{(3\alpha)^{n_1}}{n_1!}e^{-3(1-\alpha)}\frac{(3(1-\alpha))^{n_2}}{n_2!} \leqslant \frac{3^{n_1}}{n_1!}\frac{3^{n_2}}{n_2!},$$

and the latter is summable over $(n_1, n_2) \in \mathbb{N}^2$. Therefore the dominated convergence theorem implies continuity of $\Psi(\cdot; 1.75, 1.75, 0, 0)$.

Direct computation of the above series gives

$$\Psi(0.2; 1.75, 1.75, 0, 0) - 0.2 \approx 0.001657 > 0, \qquad \Psi(0.25; 1.75, 1.75, 0, 0) - 0.25 \approx -0.001098 < 0,$$
$$\Psi(0.45; 1.75, 1.75, 0, 0) - 0.45 \approx -0.001283 < 0, \qquad \Psi(0.55; 1.75, 1.75, 0, 0) - 0.55 \approx 0.001283 > 0,$$
$$\Psi(0.75; 1.75, 1.75, 0, 0) - 0.75 \approx 0.001098 > 0, \qquad \Psi(0.8; 1.75, 1.75, 0, 0) - 0.8 \approx -0.001657 < 0.$$

Hence the continuous function $\alpha \mapsto \Psi(\alpha; 1.75, 1.75, 0, 0) - \alpha$ changes sign on each of the intervals $(0.2, 0.25)$, $(0.45, 0.55)$, and $(0.75, 0.8)$.

By the intermediate value theorem, it has at least one zero in each of these three disjoint intervals. Therefore the fixed-point equation defining stochastic memory equilibrium has at least three solutions for these parameter values. □

Proposition 7. *Consider a normal environment with $A = \{1, 2\}$ and $y_1^0 = y_2^0 = 0$. Then the empirical frequency $\alpha_t(1)$ converges $\mathbb{P}_{\mu_0,\pi}$-almost surely to a random variable $X_\infty \in [0, 1]$, so a limit frequency exists.*

Proof. Write $x_t := \alpha_t(1) \in [0, 1]$, so that $\alpha_t(2) = 1 - x_t$. We first define the one-dimensional map that gives the probability of choosing action 1 when the current frequency is x.

Fix $x \in [0, 1]$. By Lemma 1, if the action frequency is x, then the numbers of recalled observations of the two actions are independent Poisson random variables $N_1 \sim \text{Poisson}(kx)$ and $N_2 \sim \text{Poisson}(k(1-x))$. Conditional on $(N_1, N_2) = (n_1, n_2)$, the posterior means of the two actions are independent normal random variables. By the normal-normal conjugacy calculation in Proposition 3, the posterior mean of action $a \in \{1, 2\}$ is normal with mean

$$m_a(n_a) := \bar{y}_a \frac{n_a/\sigma^2}{1/\sigma_0^2 + n_a/\sigma^2}$$

and variance

$$v_a(n_a) := \frac{n_a/\sigma^2}{(1/\sigma_0^2 + n_a/\sigma^2)^2}.$$

Hence, conditional on (n_1, n_2), the probability that action 1 is chosen is

$$Q(n_1, n_2) := \Phi\left(\frac{m_1(n_1) - m_2(n_2)}{\sqrt{v_1(n_1) + v_2(n_2)}}\right),$$

with the obvious interpretation that $v_1(n_1) + v_2(n_2) = 0$, which only occurs at $(n_1, n_2) = (0, 0)$, is resolved by the deterministic tie-breaking rule.

Define

$$q(x) := \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} e^{-kx} \frac{(kx)^{n_1}}{n_1!} e^{-k(1-x)} \frac{(k(1-x))^{n_2}}{n_2!} Q(n_1, n_2).$$

Thus $q(x)$ is the probability that action 1 is chosen when the current empirical frequency is x. Since the summand is bounded by

$$e^{-kx}\frac{(kx)^{n_1}}{n_1!}\,e^{-k(1-x)}\frac{(k(1-x))^{n_2}}{n_2!},$$

and Poisson probabilities are continuous in their means, the dominated convergence theorem implies that $q : [0,1] \to [0,1]$ is continuous and admits a fixed point by Brouwer's fixed-point theorem.

We next write the empirical frequency process as a stochastic approximation. Let $I_{t+1} := \mathbb{1}_{\{1\}}(a_{t+1})$. Then

$$x_{t+1} = \frac{t}{t+1}x_t + \frac{1}{t+1}I_{t+1} = x_t + \frac{1}{t+1}(I_{t+1} - x_t).$$

By Lemma A.4, in the normal environment the conditional distribution of the next action given h_t converges to the distribution induced by the recalled Poisson sample associated with x_t. Therefore, $\mathbb{P}_{\mu_0,\pi}[a_{t+1} = 1 \mid h_t] = q(x_t) + \beta_t$, where $\beta_t \to 0$ $\mathbb{P}_{\mu_0,\pi}$-almost surely. Hence

$$x_{t+1} = x_t + \gamma_{t+1}\big(q(x_t) - x_t + \xi_{t+1} + \beta_t\big), \qquad \gamma_{t+1} := \frac{1}{t+1},$$

where $\xi_{t+1} := I_{t+1} - \mathbb{E}[I_{t+1} \mid h_t]$ is a bounded martingale-difference sequence.

Let $\bar{x}$ denote the affine interpolation of (x_t), namely

$$\bar{x}(\tau_t + s) := x_t + \frac{s}{\gamma_{t+1}}(x_{t+1} - x_t), \qquad 0 \leqslant s \leqslant \gamma_{t+1}, \qquad \tau_t := \sum_{j=1}^{t}\gamma_j.$$

Since the noise is a bounded martingale-difference sequence and $\gamma_t = 1/t$, we have, for every $q > 2$, $\sup_t \mathbb{E}[\|\xi_{t+1}\|^q] < \infty$, and $\sum_{t=1}^{\infty}\gamma_t^{1+q/2} = \sum_{t=1}^{\infty} t^{-(1+q/2)} < \infty$. Hence, by Benaïm [2006] Propositions 4.1 and 4.2, the interpolated process $\bar{x}$ is almost surely an asymptotic pseudo-trajectory of the flow generated by $\dot{x} = q(x) - x$. Define $V(x) := \int_0^x (u - q(u))\,du$. Then V is continuously differentiable and $V'(x) = x - q(x)$ so along every solution $x(\cdot)$ of the ODE,

$$\frac{d}{dt}V(x(t)) = V'(x(t))\dot{x}(t) = (x(t) - q(x(t)))(q(x(t)) - x(t)) = -(q(x(t)) - x(t))^2 \leqslant 0,$$

with equality if and only if $q(x(t)) = x(t)$, so V is a strict Lyapunov function for the ODE. Since $[0,1]$ is compact, if the set $E := \{x \in [0,1] : q(x) = x\}$ is infinite it admits an accumulation point. Therefore, since $x \mapsto q(x) - x$ is analytic, we would have $E = [0,1]$ a contradiction with $q(0) > 0$. Therefore Corollary 6.6 of Benaïm [2006] implies that $\bar{x}(t) \to X_\infty$, $\mathbb{P}_{\mu_0,\pi}$-a.s., for some random variable X_∞ taking values in E. Since $x_t = \bar{x}(\tau_t)$,

it follows that $x_t \to X_\infty$ $\mathbb{P}_{\mu_0,\pi}$-a.s.

It remains to be verified that the statement about limit frequencies holds. Let x belong to the support of the law of X_∞. Then for every $\varepsilon > 0$, $\mathbb{P}_{\mu_0,\pi}[|X_\infty - x| \leqslant \varepsilon] > 0$. Since $x_t \to X_\infty$ almost surely, $\mathbb{P}_{\mu_0,\pi}[\limsup_{t\to\infty} |x_t - x| \leqslant \varepsilon] \geqslant \mathbb{P}_{\mu_0,\pi}[|X_\infty - x| \leqslant \varepsilon] > 0$. Thus x is a limit frequency in the sense of the definition in Section 3. □

B.8 A Continuum of Outcomes

The extensions of Theorem 1 and Proposition 1 to a continuum of outcomes require three main changes. First, databases will be modeled as locally finite integer-valued Borel measures on $E = A \times Y$. Second, the coordinatewise law of large numbers and Poisson limit arguments must be replaced by weak convergence of empirical action-outcome measures and Poisson point process approximations. Third, one must show that Bayesian updating remains continuous with respect to the induced law of random databases, so that the fixed-point correspondence Ψ^{μ_0} retains the compactness and upper-hemicontinuity properties needed for Kakutani's theorem and for the limit-frequency argument. This last step is the most delicate, because once databases become random point measures, the induced posterior is a nonlinear function of an infinite-dimensional random object.

We assume Y is a locally compact Polish space with Borel sigma-algebra $\mathcal{B}(Y)$, that for each $a \in A$ and $p \in \Theta$, the outcome distribution p_a is dominated by a common σ-finite Borel measure λ on Y, with density

$$f_a(y \mid p) := \frac{dp_a}{d\lambda}(y)$$

that is jointly continuous and strictly positive on $Y \times \Theta$, and that Θ is compact or that the densities $f_a(y|p)$ are uniformly bounded.[51] We also assume that for each $a \in A$, the expected utility

$$m_a(p) := \int_Y u(a, y)\, p_a(dy)$$

is well defined for every $p \in \Theta$, and the map $m_a : \Theta \to \mathbb{R}$ is bounded and continuous.

The argument has six steps: (i) empirical action–outcome measures converge to q_α (Lemma B.3); (ii) the recalled database is close in total variation to a Poisson point process, hence converges to η_α (Lemma B.4); (iii) the posterior map, and so the induced belief law, is continuous, giving convergence of the conditional belief law along histories

[51] The normal example satisfies this uniform boundedness condition

(Lemmas B.5–B.6 and Corollary 3); (iv) Ψ^{μ_0} is nonempty, convex-, and compact-valued and upper hemicontinuous, which yields existence via Kakutani (Lemmas B.7–B.8, Theorem 5); (v) any subsequential limit of the conditional action distribution β_{t_n} lies in $\Psi^{\mu_0}(\alpha)$ when $\alpha_{t_n} \to \alpha$ (Lemma B.9); and (vi) a separating-hyperplane and drift argument, as in the proof of Theorem 1, shows every limit frequency is a stochastic memory equilibrium (Theorem 6).

Let $E := A \times Y$. For $\alpha \in \Delta(A)$ define $q_\alpha \in \Delta(E)$ by

$$q_\alpha(\{a\} \times B) = \alpha(a) p_a^*(B) \qquad \forall a \in A,\ \forall B \in \mathcal{B}(Y).$$

Let $\mathcal{N}(E)$ denote the space of locally finite integer-valued Borel measures on E, endowed with the vague topology, and let $\mathcal{D} := \{d \in \mathcal{N}(E) : d(E) < \infty\}$ be the subset of such measures that are finite. Equivalently, each $d \in \mathcal{D}$ can be written in the form $d = \sum_{i=1}^n \delta_{(a_i,y_i)}$ for some finite collection of points $(a_i, y_i) \in E$. We identify databases with elements of $\mathcal{D}$. For notational convenience, for $(a, y) \in E$, we write $d(a, y) := d(\{(a, y)\})$. When Y is finite, this is equivalent to representing databases as vectors in $\mathbb{N}^{A\times Y}$ as in Section 3. Since each intensity measure considered below has finite total mass, the relevant point processes are almost surely in $\mathcal{D}$.

For each database $d \in \mathcal{D}$, write $d = \sum_{i=1}^n \delta_{(a_i,y_i)}$ for some finite collection of points $(a_i, y_i) \in E$, and define the posterior

$$\mu\big(C \mid d\big) = \frac{\int_C \prod_{(a,y)\in E} (f_a(y|p))^{d(a,y)} d\mu_0(p)}{\int_\Theta \prod_{(a,y)\in E} (f_a(y|p))^{d(a,y)} d\mu_0(p)}. \qquad \forall C \subseteq \Theta \text{ Borel}.$$

Observe that $\mu(\cdot \mid d)$ depends only on the counting measure d, not on the order in which the points are listed.

For any $\nu \in \Delta(E)$, let η_ν denote the law of the Poisson point process on E with intensity measure $k\nu$, and define $F_\nu^{\mu_0} \in \Delta(\Delta(\Theta))$ by $F_\nu^{\mu_0}(\mathcal{C}) := \eta_\nu\big(\{d \in \mathcal{D} : \mu(\cdot \mid d) \in \mathcal{C}\}\big)$ for all $\mathcal{C} \subseteq \Delta(\Theta)$ Borel.

Since the intensity measure $k\nu$ has finite total mass, $\eta_\nu(\mathcal{D}) = 1$, this is exactly the distribution of posterior beliefs induced by the Poisson point process; if Y is finite, it reduces to the product Poisson distribution defined in Section 3. Because a Poisson point process with intensity $k\nu$ can be generated by first drawing $N \sim \text{Poisson}(k)$ and then drawing N i.i.d. points with law ν, this is exactly the distribution of posterior beliefs induced by η_ν.

For $\alpha \in \Delta(A)$, write $\eta_\alpha := \eta_{q_\alpha}$ and $F_\alpha^{\mu_0} := F_{q_\alpha}^{\mu_0}$. Finally, for $m \in \Delta(\Delta(\Theta))$ define

$$\Gamma(m) := \left\{ \int_{\Delta(\Theta)} \rho(\nu)\, dm(\nu) : \rho \in \mathcal{O} \right\} \subseteq \Delta(A),$$

so that

$$\Psi^{\mu_0}(\alpha) = \Gamma(F_\alpha^{\mu_0}) \qquad \forall \alpha \in \Delta(A).$$

Lemma B.3 (Empirical convergence of action-outcome measures)**.** *For each $t \geqslant 1$, let*

$$\upsilon_t := \frac{1}{t} \sum_{\tau=1}^{t} \delta_{(a_\tau, y_\tau)} \in \Delta(E)$$

denote the empirical measure of action-outcome pairs. Then for every bounded continuous function $g : E \to \mathbb{R}$,

$$\int_E g\, d\upsilon_t - \int_E g\, dq_{\alpha_t} \to 0 \qquad \mathbb{P}_{\mu_0,\pi} \text{ almost surely.}$$

Equivalently, $\frac{1}{t}\sum_{\tau=1}^{t} g(a_\tau, y_\tau) - \sum_{a \in A} \alpha_t(a) \int_Y g(a,y)\, dp_a^*(y) \to 0$, $\mathbb{P}_{\mu_0,\pi}$ almost surely, so

$$\lim_{t\to\infty} \alpha_t = \alpha \implies \upsilon_t \text{ weakly converges to } q_\alpha \qquad \mathbb{P}_{\mu_0,\pi} \text{ almost surely.}$$

Moreover, there exists an event Ω_0 with $\mathbb{P}_{\mu_0,\pi}(\Omega_0) = 1$ such that for every sample path in Ω_0 and every subsequence t_n with $\alpha_{t_n} \to \alpha$, υ_{t_n} weakly converges to q_α.

Proof. Fix a bounded continuous function $g : E \to \mathbb{R}$. For each $a \in A$, define

$$\bar{g}(a) := \int_Y g(a,y)\, dp_a^*(y).$$

Since A is finite and g is bounded, $\bar{g} : A \to \mathbb{R}$ is bounded.

Let $\{\mathcal{F}_t\}_{t\geqslant 0}$ be the filtration generated by histories, $\mathcal{F}_t := \sigma(a_1, y_1, \ldots, a_t, y_t)$, and set $X_t := g(a_t, y_t) - \bar{g}(a_t)$. Then $(X_t)_{t\geqslant 1}$ is adapted to $\{\mathcal{F}_t\}$, and

$$\begin{aligned}\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] &= \mathbb{E}[\mathbb{E}[X_t \mid \mathcal{F}_{t-1}, a_t] \,|\, \mathcal{F}_{t-1}] \\ &= \mathbb{E}[\mathbb{E}[g(a_t, y_t) - \bar{g}(a_t) \mid \mathcal{F}_{t-1}, a_t] \,|\, \mathcal{F}_{t-1}]\,.\end{aligned}$$

Conditional on $(\mathcal{F}_{t-1}, a_t)$, the outcome y_t is drawn according to $p_{a_t}^*$, so

$$\mathbb{E}[g(a_t, y_t) \mid \mathcal{F}_{t-1}, a_t] = \int_Y g(a_t, y)\, dp_{a_t}^*(y) = \bar{g}(a_t).$$

Hence $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = 0$, so $(X_t)_{t\geqslant 1}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}$.

Since g is bounded, there exists $M < \infty$ such that $|g(a,y)| \leqslant M$ for all $(a,y) \in A \times Y$, and therefore $|X_t| \leqslant 2M \qquad \forall t \geqslant 1$.

Hence

$$\sum_{t=1}^{\infty} \mathbb{E}\left[\left(\frac{X_t}{t} \right)^2 \middle| \mathcal{F}_{t-1} \right] \leqslant 4M^2 \sum_{t=1}^{\infty} \frac{1}{t^2} < \infty \qquad \mathbb{P}_{\mu_0,\pi} \text{ almost surely.}$$

Thus $\sum_{t=1}^{n} X_t/t$ converges almost surely (see, e.g., Theorem 2.15 in Hall and Heyde [2014]), and Kronecker's lemma implies $\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} X_\tau = 0$, $\mathbb{P}_{\mu_0,\pi}$ almost surely, so that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} g(a_\tau, y_\tau) - \frac{1}{t} \sum_{\tau=1}^{t} \bar{g}(a_\tau) = 0 \qquad \mathbb{P}_{\mu_0,\pi} \text{ almost surely.}$$

Now

$$\frac{1}{t} \sum_{\tau=1}^{t} \bar{g}(a_\tau) = \sum_{a\in A} \left(\frac{1}{t} \sum_{\tau=1}^{t} \mathbf{1}_{\{a\}}(a_\tau) \right) \bar{g}(a) = \sum_{a\in A} \alpha_t(a) \int_Y g(a,y)\, dp_a^*(y) = \int_E g\, dq_{\alpha_t},$$

while

$$\frac{1}{t} \sum_{\tau=1}^{t} g(a_\tau, y_\tau) = \int_E g\, d\upsilon_t.$$

This proves the first claim.

If $\lim_{t\to\infty} \alpha_t = \alpha$, then

$$\lim_{t\to\infty} \int_E g\, dq_{\alpha_t} = \lim_{t\to\infty} \sum_{a\in A} \alpha_t(a) \int_Y g(a,y)\, dp_a^*(y) = \sum_{a\in A} \alpha(a) \int_Y g(a,y)\, dp_a^*(y) = \int_E g\, dq_\alpha.$$

Hence $\lim_{t\to\infty} \int g\, d\upsilon_t = \int g\, dq_\alpha$ for every bounded continuous g, so υ_t weakly converges to q_α.

For the final claim, let $\{g_m\}_{m\in\mathbb{N}}$ be a sequence of bounded and continuous functions on E which form a convergence-determining class for weak convergence, which exists by the separability of E (see, e.g., Theorem 6.2 in Parthasarathy [2005]). For each m there is an event Ω_m with $\mathbb{P}_{\mu_0,\pi}(\Omega_m) = 1$ on which $\lim_{t\to\infty} \int_E g_m\, d\upsilon_t - \int_E g_m\, dq_{\alpha_t} = 0$. Set $\Omega_0 := \bigcap_{m\in\mathbb{N}} \Omega_m$, so $\mathbb{P}_{\mu_0,\pi}(\Omega_0) = 1$. Fix a sample path in Ω_0 and a subsequence t_n such that $\lim_{n\to\infty} \alpha_{t_n} = \alpha$. Then for every m,

$$\lim_{n\to\infty} \int_E g_m\, d\upsilon_{t_n} - \int_E g_m\, dq_{\alpha_{t_n}} = 0,$$

and since $\lim_{n\to\infty} \alpha_{t_n} = \alpha$,

$$\lim_{n\to\infty} \int_E g_m\, dq_{\alpha_{t_n}} = \int_E g_m\, dq_\alpha.$$

Hence

$$\lim_{n\to\infty} \int_E g_m\, d\upsilon_{t_n} = \int_E g_m\, dq_\alpha \qquad \forall m \in \mathbb{N},$$

which implies $\lim_{n\to\infty} \upsilon_{t_n} = q_\alpha$. □

Lemma B.4 (Poisson approximation around the empirical measure)**.** *Fix a history $h_t = ((a_1, y_1), \ldots, (a_t, y_t))$ with $t \geqslant k$, and let $\upsilon_t := \frac{1}{t}\sum_{i=1}^{t} \delta_{(a_i,y_i)}$. Let $d_{t+1} \in \mathcal{D}$ be the recalled database obtained by retaining each past experience independently with probability k/t. Then*

$$\|\mathcal{L}(d_{t+1} \mid h_t) - \eta_{\upsilon_t}\|_{TV} \leqslant \frac{k^2}{t}.$$

In particular, if $\lim_{t\to\infty} \upsilon_t = q_\alpha$ along a sequence $(h_t)_{t\in\mathbb{N}}$, then η_{υ_t} weakly converges to η_α and $\mathcal{L}(d_{t+1} \mid h_t)$ weakly converges to η_α.[52]

Proof. Let $x^1, \ldots, x^m$ be the distinct points in the support of υ_t, and let n_j be the multiplicity of x^j among the first t experiences, so $\sum_{j=1}^{m} n_j = t$. Conditional on h_t, the number of recalled copies of x^j is distributed as Binomial$(n_j, k/t)$, and these counts are independent across j. Under η_{υ_t}, the point process has intensity measure $k\upsilon_t$, so the counts at the atoms $x^1, \ldots, x^m$ are independent and the count at x^j is distributed as Poisson(kn_j/t).

For each j, let P_j denote the law of Binomial$(n_j, k/t)$ and let Q_j denote the law of Poisson(kn_j/t). A standard form of Le Cam's inequality (see, e.g., Steele [1994]) states that if $S = \sum_{\ell=1}^{N} \xi_\ell$ with independent Bernoulli(p_ℓ) summands and $\lambda = \sum_{\ell=1}^{N} p_\ell$, then $\|\mathcal{L}(S) - \text{Poisson}(\lambda)\|_{TV} \leqslant \sum_{\ell=1}^{N} p_\ell^2$. Applying this with $N = n_j$ and $p_\ell = k/t$ for all ℓ yields

$$\|P_j - Q_j\|_{TV} \leqslant n_j \left(\frac{k}{t}\right)^2.$$

Now let $P := \bigotimes_{j=1}^{m} P_j$ and $Q := \bigotimes_{j=1}^{m} Q_j$. These are the laws of the count vectors under $\mathcal{L}(d_{t+1} \mid h_t)$ and η_{υ_t}, respectively. Using the telescoping identity for product measures,

$$P - Q = \sum_{j=1}^{m} \left(\bigotimes_{\ell<j} P_\ell\right) \otimes (P_j - Q_j) \otimes \left(\bigotimes_{\ell>j} Q_\ell\right).$$

Each term in the telescoping sum differs only in one coordinate. The other coordinates are probability measures, so they do not affect the total variation norm of that term. Hence, by the triangle inequality,

$$\|P - Q\|_{TV} \leqslant \sum_{j=1}^{m} \|P_j - Q_j\|_{TV} \leqslant \sum_{j=1}^{m} n_j \left(\frac{k}{t}\right)^2 = \frac{k^2}{t}.$$

Define $\Phi : \mathbb{N}^m \to \mathcal{D}$ by $\Phi(c_1, \ldots, c_m) = \sum_{j=1}^{m} c_j \delta_{x_j}$. Then the laws of the recalled database and the Poisson approximation respectively are $\mathcal{L}(d_{t+1} \mid h_t) = P \circ \Phi^{-1}$ and

[52] We use the convention $\|P - Q\|_{TV} := \sup_B |P(B) - Q(B)|$.

$\eta_{\upsilon_t} = Q \circ \Phi^{-1}$. Since taking pushforwards cannot increase total variation distance, the bound on $\|P - Q\|_{TV}$ implies

$$\|\mathcal{L}(d_{t+1} \mid h_t) - \eta_{\upsilon_t}\|_{TV} \leqslant \|P - Q\|_{TV} \leqslant \frac{k^2}{t}.$$

This proves the first claim.

Now suppose $\lim_{t\to\infty} \upsilon_t = q_\alpha$. For every nonnegative and continuous function g on E with compact support, by the formula of the Laplace functional of Poisson random variables (see, e.g., Kallenberg, 2002, Lemma 12.2), we have

$$\int_{\mathcal{N}(E)} e^{-\sum_{(a,y)\in\operatorname{supp} d} d(a,y)g(a,y)}\, d\eta_{\upsilon_t}(d) = \exp\left(-k \int_E (1 - e^{-g(x)})\, d\upsilon_t(x)\right)$$

Moreover,

$$\lim_{t\to\infty} \exp\left(-k \int_E (1 - e^{-g(x)})\, d\upsilon_t(x)\right) = \exp\left(-k \int_E (1 - e^{-g(x)})\, dq_\alpha(x)\right),$$

which by Kallenberg [2002], Lemma 12.2 is the Laplace functional of η_α. By the characterization of vague convergence in distribution of random measures through convergence of Laplace functionals (Kallenberg [2017], Theorem 4.11), η_{υ_t} weakly converges to η_α. Since $\lim_{t\to\infty} k^2/t = 0$, the total variation bound implies that for every bounded continuous $\varphi : \mathcal{N}(E) \to \mathbb{R}$,

$$\lim_{t\to\infty} \left| \int \varphi\, d\mathcal{L}(d_{t+1} \mid h_t) - \int \varphi\, d\eta_{\upsilon_t} \right| \leqslant \lim_{t\to\infty} 2\|\varphi\|_\infty \|\mathcal{L}(d_{t+1} \mid h_t) - \eta_{\upsilon_t}\|_{TV} = 0.$$

Thus $\eta_{\upsilon_t} \Rightarrow \eta_\alpha$ implies $\mathcal{L}(d_{t+1} \mid h_t) \Rightarrow \eta_\alpha$.

□

Lemma B.5 (Continuity of the posterior map on ordered samples)**.** *For each $n \in \mathbb{N}$, the map $\mathbf{z} \mapsto \mu(\cdot \mid \mathbf{z})$ from E^n to $\Delta(\Theta)$ is continuous for the weak topology.*

Proof. Fix $n \in \mathbb{N}$. If $n = 0$, the map is constant, hence continuous. So suppose $n \geqslant 1$. Since E^n and $\Delta(\Theta)$ are metrizable, it is enough to prove sequential continuity.

Let $\mathbf{z}^m = ((a_1^m, y_1^m), \ldots, (a_n^m, y_n^m)) \to \mathbf{z} = ((a_1, y_1), \ldots, (a_n, y_n))$ in E^n, and fix $h \in C_b(\Theta)$. Write

$$L_m(p) := \prod_{i=1}^n f_{a_i^m}(y_i^m \mid p), \qquad L(p) := \prod_{i=1}^n f_{a_i}(y_i \mid p).$$

Since A is finite and each f_a is jointly continuous in (y, p), we have $L_m(p) \to L(p)$ for every $p \in \Theta$.

It remains to justify dominated convergence. Let $K := \{\mathbf{z}\} \cup \{\mathbf{z}^m : m \in \mathbb{N}\} \subseteq E^n$. Since $\mathbf{z}^m \to \mathbf{z}$, the set K is compact. If Θ is compact, then continuity of $(\mathbf{w}, p) \mapsto \prod_{i=1}^n f_{\tilde{a}_i}(\tilde{y}_i \mid p)$ on $K \times \Theta$ implies a uniform bound. If instead the densities are uniformly bounded, then $\prod_{i=1}^n f_{a_i^m}(y_i^m \mid p)$ is uniformly bounded by M^n, where $M := \max_{a \in A} \sup_{(y,p) \in Y \times \Theta} f_a(y \mid p) < \infty$. Thus in either case there is $C < \infty$ such that $L_m(p) \leqslant C$ for all m and p.

Therefore $h(p)L_m(p) \to h(p)L(p)$ pointwise and $|h(p)L_m(p)| \leqslant \|h\|_\infty C$, so dominated convergence gives

$$\int_\Theta h(p)L_m(p)\, d\mu_0(p) \to \int_\Theta h(p)L(p)\, d\mu_0(p) \quad \text{and} \quad \int_\Theta L_m(p)\, d\mu_0(p) \to \int_\Theta L(p)\, d\mu_0(p).$$

Because the densities are strictly positive, the denominators are strictly positive, so

$$\int_\Theta h(p)\, d\mu(p \mid \mathbf{z}^m) \to \int_\Theta h(p)\, d\mu(p \mid \mathbf{z}).$$

Since this holds for every $h \in C_b(\Theta)$, we conclude that $\mu(\cdot \mid \mathbf{z}^m) \Rightarrow \mu(\cdot \mid \mathbf{z})$. □

Lemma B.6 (Continuity of the induced belief law)**.** *If $\nu_m \Rightarrow \nu$ in $\Delta(E)$, then $F_{\nu_m}^{\mu_0} \Rightarrow F_\nu^{\mu_0}$, so the map $\alpha \mapsto F_\alpha^{\mu_0} = F_{q_\alpha}^{\mu_0}$ from $\Delta(A)$ to $\Delta(\Delta(\Theta))$ is weakly continuous.*

Proof. Fix a bounded continuous function $\Phi : \Delta(\Theta) \to \mathbb{R}$. For each $n \in \mathbb{N}$, define $\psi_n : E^n \to \mathbb{R}$ and $\psi_n(\mathbf{z}) := \Phi(\mu(\cdot \mid \mathbf{z}))$. By Lemma B.5, ψ_n is continuous, and it is bounded by $\|\Phi\|_\infty$. By definition of $F_\nu^{\mu_0}$,

$$\int_{\Delta(\Theta)} \Phi(\beta)\, dF_\nu^{\mu_0}(\beta) = \sum_{n=0}^\infty \iota_n \int_{E^n} \psi_n(\mathbf{z})\, d\nu^{\otimes n}(\mathbf{z}).$$

If $\nu_m \Rightarrow \nu$, then for each fixed n, $\nu_m^{\otimes n} \Rightarrow \nu^{\otimes n} \in \Delta(E^n)$, so $\int_{E^n} \psi_n\, d\nu_m^{\otimes n} \to \int_{E^n} \psi_n\, d\nu^{\otimes n}$. Since $|\psi_n| \leqslant \|\Phi\|_\infty$ and $\sum_{n=0}^\infty \iota_n = 1$, dominated convergence yields

$$\int_{\Delta(\Theta)} \Phi\, dF_{\nu_m}^{\mu_0} \to \int_{\Delta(\Theta)} \Phi\, dF_\nu^{\mu_0}.$$

Thus $F_{\nu_m}^{\mu_0} \Rightarrow F_\nu^{\mu_0}$. The continuity of $\alpha \mapsto F_\alpha^{\mu_0}$ follows because $\alpha \mapsto q_\alpha$ is continuous. □

Corollary 3 (Conditional belief law along convergent subsequences)**.** *Fix a deterministic history sequence $h_t = ((a_1, y_1), \ldots, (a_t, y_t))$, and let μ_{t+1} denote the posterior belief at the beginning of period $t+1$. Then for each $t \geqslant k$,*

$$\left\| \mathcal{L}(\mu_{t+1} \mid h_t) - F_{\upsilon_t}^{\mu_0} \right\|_{TV} \leqslant \frac{k^2}{t}.$$

Consequently, if $t_n \to \infty$ is a subsequence such that $\upsilon_{t_n} \Rightarrow q_\alpha$, then $\mathcal{L}(\mu_{t_n+1} \mid h_{t_n}) \Rightarrow F_\alpha^{\mu_0}$. In particular, by Lemma B.3, on every history sequence in the event Ω_0 of that lemma and

along every subsequence with $\alpha_{t_n} \to \alpha$, *we have* $\mathcal{L}(\mu_{t_n+1} \mid h_{t_n}) \Rightarrow F_\alpha^{\mu_0}$.

Proof. Fix $t \geqslant k$ and the history h_t. Let $x^1, \dots, x^m$ be the distinct points in the support of υ_t. Every database recalled at time $t+1$ can be identified with a count vector $c = (c_1, \dots, c_m) \in \mathbb{N}^m$, where c_j is the number of recalled copies of x^j. Let $R_t : \mathbb{N}^m \to \Delta(\Theta)$ be the posterior map that assigns to each count vector the posterior belief generated by the corresponding database.

Under the actual recall process, let P_t denote the law of the count vector. Under the Poisson point process with intensity $k\upsilon_t$, let Q_t denote the law of the count vector. By the first part of Lemma B.4, $\|P_t - Q_t\|_{TV} \leqslant \frac{k^2}{t}$. The conditional law of μ_{t+1} given h_t is the pushforward $P_t \circ R_t^{-1}$, while $F_{\upsilon_t}^{\mu_0}$ is the pushforward $Q_t \circ R_t^{-1}$. Since pushforwards do not increase total variation, $\left\|\mathcal{L}(\mu_{t+1} \mid h_t) - F_{\upsilon_t}^{\mu_0}\right\|_{TV} \leqslant \frac{k^2}{t}$. This proves the first claim.

Now suppose $\upsilon_{t_n} \Rightarrow q_\alpha$. By Lemma B.6, $F_{\upsilon_{t_n}}^{\mu_0} \Rightarrow F_{q_\alpha}^{\mu_0} = F_\alpha^{\mu_0}$. Combining this with the total-variation bound gives $\mathcal{L}(\mu_{t_n+1} \mid h_{t_n}) \Rightarrow F_\alpha^{\mu_0}$. The last claim follows from the last part of Lemma B.3. □

For $f \in \mathbb{R}^A$ and $\nu \in \Delta(\Theta)$ define $m_f(\nu) := \max_{a \in BR(\nu)} f_a$.

Lemma B.7 (Support function of $\Gamma(m)$)**.** *For every* $f \in \mathbb{R}^A$, *the function* m_f *is bounded and upper semicontinuous on* $\Delta(\Theta)$, *and for every* $m \in \Delta(\Delta(\Theta))$,

$$\sup_{\beta \in \Gamma(m)} \beta \cdot f = \int_{\Delta(\Theta)} m_f(\nu)\, dm(\nu).$$

Proof. Because each $m_a : \Theta \to \mathbb{R}$ is bounded and continuous, the expected payoff of action a under belief ν is $\int_\Theta m_a(p)\, d\nu(p)$, which is continuous in ν. Since A is finite, the best-reply correspondence BR is nonempty-valued and upper hemicontinuous.

To prove that m_f is upper semicontinuous, let $\nu_n \to \nu$. For each n, choose $a_n \in BR(\nu_n)$ such that $f_{a_n} = m_f(\nu_n)$. By passing to a subsequence, we may assume $a_n = a$ for all n. Upper hemicontinuity of BR implies $a \in BR(\nu)$, so $\limsup_{n\to\infty} m_f(\nu_n) = \limsup_{n\to\infty} f_{a_n} = f_a \leqslant m_f(\nu)$. Thus m_f is upper semicontinuous. It is bounded because A is finite.

Now fix $m \in \Delta(\Delta(\Theta))$. For any $\rho \in \mathcal{O}$,

$$\left(\int \rho\, dm\right) \cdot f = \int \sum_{a \in A} \rho_a(\nu) f_a\, dm(\nu) \leqslant \int m_f(\nu)\, dm(\nu),$$

so

$$\sup_{\beta \in \Gamma(m)} \beta \cdot f \leqslant \int m_f\, dm.$$

Conversely, define $a_f(\nu)$ to be the smallest-index action in $BR(\nu)$ that attains the maximum in $m_f(\nu)$, and let $\rho_f(\nu) := \delta_{a_f(\nu)}$. Then $\rho_f \in \mathcal{O}$ and $\left(\int \rho_f \, dm\right) \cdot f = \int m_f(\nu) \, dm(\nu)$. □

Lemma B.8 (Properties of Γ and Ψ^{μ_0})**.** *The correspondence $\Gamma : \Delta(\Delta(\Theta)) \rightrightarrows \Delta(A)$ is nonempty, convex-valued, compact-valued, and upper hemicontinuous. Consequently, $\Psi^{\mu_0}(\alpha) = \Gamma(F_\alpha^{\mu_0})$ is nonempty, convex-valued, compact-valued, and upper hemicontinuous on $\Delta(A)$.*

Proof. Fix $m \in \Delta(\Delta(\Theta))$. We first prove that $\Gamma(m)$ is compact. For each $a \in A$, let $B_a := \{\nu \in \Delta(\Theta) : a \notin BR(\nu)\}$. These sets are Borel measurable because the expected payoff of every action is continuous in ν and A is finite.

Consider the following subset of $L^2(m; \mathbb{R}^A)$:

$$\mathcal{K}_m := \left\{ \rho : \begin{array}{ll} \rho_a \geqslant 0 & m\text{-a.e. for every } a, \\ \sum_{a \in A} \rho_a = 1 & m\text{-a.e.}, \\ \rho_a = 0 & m\text{-a.e. on } B_a \quad \text{for every } a \end{array} \right\}.$$

Thus $\mathcal{K}_m$ is the set, modulo m-null sets, of measurable mixed best replies. It is nonempty because BR admits a measurable selection. It is also convex and norm closed in $L^2(m; \mathbb{R}^A)$. Moreover, every $\rho \in \mathcal{K}_m$ satisfies $0 \leqslant \rho_a \leqslant 1$ almost everywhere, and hence $\mathcal{K}_m$ is bounded.

Since $L^2(m; \mathbb{R}^A)$ is reflexive, every closed, bounded, convex subset is weakly compact. It follows that $\mathcal{K}_m$ is weakly compact. The linear map $T_m : L^2(m; \mathbb{R}^A) \to \mathbb{R}^A, T_m(\rho) := \int_{\Delta(\Theta)} \rho(\nu) \, dm(\nu)$, is continuous from the weak topology of L^2 to the ordinary topology of $\mathbb{R}^A$, since each coordinate of T_m is a continuous linear functional.

We claim that

$$\Gamma(m) = T_m(\mathcal{K}_m). \tag{45}$$

One inclusion is immediate: every $\rho \in \mathcal{O}$ determines an element of $\mathcal{K}_m$. Conversely, let $\rho \in \mathcal{K}_m$, and choose a measurable representative of its L^2-equivalence class. The conditions defining $\mathcal{K}_m$ may fail only on an m-null set N. Let $\rho^0 \in \mathcal{O}$ be any fixed measurable best-reply selection and define

$$\widetilde{\rho}(\nu) := \begin{cases} \rho(\nu), & \nu \notin N, \\ \rho^0(\nu), & \nu \in N. \end{cases}$$

Then $\widetilde{\rho} \in \mathcal{O}$ and $\int \widetilde{\rho} \, dm = \int \rho \, dm$. This proves (45).

Because T_m is weakly continuous and $\mathcal{K}_m$ is weakly compact, $\Gamma(m) = T_m(\mathcal{K}_m)$ is compact. Its convexity follows either from (45) or directly by taking pointwise mixtures of

policies. Nonemptiness was established above.

We may now use the support-function result without a closure gap. A nonempty closed convex subset $C \subseteq \mathbb{R}^A$ satisfies $C = \left\{\beta \in \mathbb{R}^A : \beta \cdot f \leqslant \sup_{\gamma \in C} \gamma \cdot f \quad \forall f \in \mathbb{R}^A\right\}$; the reverse inclusion follows from the separating-hyperplane theorem. Applying this characterization to the compact convex set $\Gamma(m)$, and then applying Lemma B.7, gives

$$\Gamma(m) = \left\{\beta \in \Delta(A) : \beta \cdot f \leqslant \int_{\Delta(\Theta)} m_f(\nu)\, dm(\nu) \quad \forall f \in \mathbb{R}^A\right\}. \tag{46}$$

We next prove upper hemicontinuity. Let $m_n \Rightarrow m$, and suppose that $\beta_n \in \Gamma(m_n)$, $\beta_n \to \beta$. For every $f \in \mathbb{R}^A$, Lemma B.7 gives $\beta_n \cdot f \leqslant \sup_{\gamma \in \Gamma(m_n)} \gamma \cdot f = \int_{\Delta(\Theta)} m_f(\nu)\, dm_n(\nu)$. Since m_f is bounded and upper semicontinuous, the portmanteau theorem implies

$$\limsup_{n\to\infty} \int_{\Delta(\Theta)} m_f(\nu)\, dm_n(\nu) \leqslant \int_{\Delta(\Theta)} m_f(\nu)\, dm(\nu).$$

Therefore, $\beta \cdot f \leqslant \int_{\Delta(\Theta)} m_f(\nu)\, dm(\nu)$ for all $f \in \mathbb{R}^A$. The characterization (46) implies that $\beta \in \Gamma(m)$. Hence Γ has a closed graph.

Because the values of Γ are compact subsets of the compact metric space $\Delta(A)$, the closed-graph property implies that Γ is upper hemicontinuous. Finally, Lemma B.6 shows that $\alpha \mapsto F_\alpha^{\mu_0}$ is continuous, so $\Psi^{\mu_0}(\alpha) = \Gamma(F_\alpha^{\mu_0})$ is nonempty, convex-valued, compact-valued, and upper hemicontinuous. □

Theorem 5 (Continuum-outcome extension of Proposition 1)**.** *Under the assumptions of this section, a stochastic memory equilibrium exists.*

Proof. By Lemma B.8, the correspondence $\Psi^{\mu_0} : \Delta(A) \rightrightarrows \Delta(A)$ is nonempty, convex-valued, compact-valued, and upper hemicontinuous. Since $\Delta(A)$ is a nonempty compact convex subset of a finite-dimensional Euclidean space, Kakutani's fixed point theorem implies that Ψ^{μ_0} has a fixed point. By definition, every fixed point of Ψ^{μ_0} is a stochastic memory equilibrium. □

Fix an optimal Markovian policy π. For each t, let $\beta_t := \mathbb{P}_{\mu_0,\pi}[a_{t+1} \in \cdot \mid h_t] \in \Delta(A)$ denote the conditional distribution of the next action given the history through period t.

Lemma B.9 (Subsequence bound for conditional action distributions)**.** *Let (h_t) be a history sequence belonging to the almost-sure event Ω_0 of Lemma B.3. If $t_n \to \infty$ is a subsequence such that $\alpha_{t_n} \to \alpha$, then for every $f \in \mathbb{R}^A$, $\limsup_{n\to\infty} \beta_{t_n} \cdot f \leqslant \sup_{\beta \in \Psi^{\mu_0}(\alpha)} \beta \cdot f$.*

Proof. For each $t \geqslant k$, define $\widehat{\beta}_t := \int_{\Delta(\Theta)} \delta_{\pi(\nu)}\, dF^{\mu_0}_{\upsilon_t}(\nu) \in \Gamma(F^{\mu_0}_{\upsilon_t})$. By Corollary 3,

$$\left\| \mathcal{L}(\mu_{t+1} \mid h_t) - F^{\mu_0}_{\upsilon_t} \right\|_{TV} \leqslant \frac{k^2}{t}.$$

Pushing forward by the measurable map $\nu \mapsto \pi(\nu)$ yields $\|\beta_t - \widehat{\beta}_t\|_{TV} \leqslant \frac{k^2}{t}$, so $\beta_t \cdot f - \widehat{\beta}_t \cdot f \to 0$. Since $\widehat{\beta}_t \in \Gamma(F^{\mu_0}_{\upsilon_t})$, Lemma B.7 gives $\widehat{\beta}_t \cdot f \leqslant \sup_{\beta \in \Gamma(F^{\mu_0}_{\upsilon_t})} \beta \cdot f = \int m_f\, dF^{\mu_0}_{\upsilon_t}$. By Lemma B.3, $\upsilon_{t_n} \Rightarrow q_\alpha$, so Lemma B.6 implies that $F^{\mu_0}_{\upsilon_{t_n}} \Rightarrow F^{\mu_0}_{\alpha}$. Since m_f is bounded upper semicontinuous,

$$\limsup_{n\to\infty} \int m_f\, dF^{\mu_0}_{\upsilon_{t_n}} \leqslant \int m_f\, dF^{\mu_0}_{\alpha} = \sup_{\beta \in \Psi^{\mu_0}(\alpha)} \beta \cdot f,$$

where the last equality follows again from Lemma B.7. Combining the last three displays proves the claim. □

Theorem 6 (Continuum-outcome extension of Theorem 1)**.** *Under the assumptions of this section, every limit frequency is a stochastic memory equilibrium.*

Proof. The proof parallels that of Theorem 1. Suppose $\alpha \in \Delta(A)$ is not a stochastic memory equilibrium, i.e. $\alpha \notin \Psi^{\mu_0}(\alpha)$. Since $\Psi^{\mu_0}(\alpha)$ is nonempty, compact, and convex, the separating hyperplane theorem yields $f \in \mathbb{R}^A$ and $c > 0$ such that $\alpha \cdot f \geqslant \sup_{\beta \in \Psi^{\mu_0}(\alpha)} \beta \cdot f + 3c$. By upper hemicontinuity of Ψ^{μ_0}, after shrinking to a neighborhood U of α if necessary, we may assume that for every $\alpha' \in U$,

$$\alpha' \cdot f \geqslant \sup_{\beta \in \Psi^{\mu_0}(\alpha')} \beta \cdot f + 2c. \tag{47}$$

Fix an optimal Markovian policy π, and let Ω_0 be the almost-sure event of Lemma B.3. We claim that on every sample path in Ω_0 there exists $T < \infty$ such that whenever $t \geqslant T$ and $\alpha_t \in U$,

$$\beta_t \cdot f \leqslant \alpha_t \cdot f - c. \tag{48}$$

Suppose not. Then on some path in Ω_0 there exists a subsequence $t_n \to \infty$ such that $\alpha_{t_n} \in U$ for all n and $\beta_{t_n} \cdot f > \alpha_{t_n} \cdot f - c$. By compactness of $\Delta(A)$, after passing to a further subsequence we may assume $\alpha_{t_n} \to \alpha' \in U$. Lemma B.9 then implies $\limsup_{n\to\infty} \beta_{t_n} \cdot f \leqslant \sup_{\beta \in \Psi^{\mu_0}(\alpha')} \beta \cdot f \leqslant \alpha' \cdot f - 2c$, where the last inequality is (47). But $\alpha_{t_n} \cdot f \to \alpha' \cdot f$, so the strict inequality $\beta_{t_n} \cdot f > \alpha_{t_n} \cdot f - c$ yields $\liminf_{n\to\infty} \beta_{t_n} \cdot f \geqslant \alpha' \cdot f - c$, a contradiction. This proves (48).

Now define

$$M_t := \sum_{s=1}^{t-1} \frac{f(a_{s+1}) - \beta_s \cdot f}{s+1}, \qquad t \geqslant 1.$$

Since $f(a_{s+1}) - \beta_s \cdot f$ is a bounded martingale difference sequence and $\sum_s (s+1)^{-2} < \infty$, there exists an event

$$\Omega_M := \{\omega : (M_t(\omega))_{t \geqslant 1} \text{ converges}\}$$

such that

$$\mathbb{P}_{\mu_0,\pi}(\Omega_M) = 1.$$

Define

$$\Omega := \Omega_0 \cap \Omega_M.$$

Since both Ω_0 and Ω_M have probability one,

$$\mathbb{P}_{\mu_0,\pi}(\Omega) = 1. \tag{49}$$

Let

$$E_U := \{\omega : \alpha_t(\omega) \in U \text{ for all sufficiently large } t\}.$$

We claim that

$$\Omega \cap E_U = \varnothing.$$

Suppose, to the contrary, that $\omega \in \Omega \cap E_U$. Because $\omega \in \Omega_0$, the pathwise drift estimate (48) provides a time $T_0 = T_0(\omega)$ such that

$$\beta_t(\omega) \cdot f \leqslant \alpha_t(\omega) \cdot f - c$$

whenever $t \geqslant T_0$ and $\alpha_t(\omega) \in U$. Because $\omega \in E_U$, there also exists $T_U = T_U(\omega)$ such that

$$\alpha_t(\omega) \in U \qquad \forall t \geqslant T_U.$$

Set $T := \max\{T_0, T_U\}$. Then, for every $t \geqslant T$,

$$\begin{aligned}
\alpha_{t+1}(\omega) \cdot f &= \alpha_T(\omega) \cdot f + \sum_{s=T}^{t} \frac{\beta_s(\omega) \cdot f - \alpha_s(\omega) \cdot f}{s+1} + M_{t+1}(\omega) - M_T(\omega) \\
&\leqslant \alpha_T(\omega) \cdot f - c \sum_{s=T}^{t} \frac{1}{s+1} + M_{t+1}(\omega) - M_T(\omega).
\end{aligned}$$

Since $\omega \in \Omega_M$, the sequence $M_t(\omega)$ converges, whereas $\sum_{s=T}^{t} \frac{1}{s+1} \longrightarrow +\infty$. The right-hand side therefore tends to $-\infty$. This is impossible because $\alpha_{t+1}(\omega) \in \Delta(A)$ and the continu-

ous function $\gamma \mapsto \gamma \cdot f$ is bounded below on the compact set $\Delta(A)$. Hence $\Omega \cap E_U = \varnothing$. Using (49), we conclude that $\mathbb{P}_{\mu_0,\pi}(E_U) = \mathbb{P}_{\mu_0,\pi}(E_U \cap \Omega) = 0$. Equivalently, $\mathbb{P}_{\mu_0,\pi}[\alpha_t \in U \text{ eventually}] = 0$. Since U is an open neighborhood of α in $\Delta(A)$, there exists $r > 0$ such that $B_r(\alpha) := \{\alpha' \in \Delta(A) : \|\alpha' - \alpha\|_\infty < r\} \subseteq U$. Set $\varepsilon := r/2$. If $\limsup_{t\to\infty} \|\alpha_t - \alpha\|_\infty \leqslant \varepsilon$, then, since $\varepsilon < r$, there exists $T < \infty$ such that $\|\alpha_t - \alpha\|_\infty < r$, for all $t \geqslant T$. Thus $\alpha_t \in B_r(\alpha) \subseteq U$ eventually. Consequently, $\{\limsup_{t\to\infty} \|\alpha_t - \alpha\|_\infty \leqslant \varepsilon\} \subseteq \{\alpha_t \in U \text{ eventually}\}$. The event on the right has probability zero, and therefore $\mathbb{P}_{\mu_0,\pi}\left[\limsup_{t\to\infty} \|\alpha_t - \alpha\|_\infty \leqslant \varepsilon\right] = 0$. Since the definition of a limit frequency requires this probability to be strictly positive for every $\varepsilon > 0$, α is not a limit frequency. endproof

B.9 The Binomial-Beta Example Extended to Rehearsal and Recency

Suppose action $a = 0$ has known deterministic payoff $q \in \mathbb{R}$, while action $a = 1$ results in a random payoff in $Y = \{0, 1\}$ with unknown success probability $\theta := p_1^*(1)$. The agent has a Beta(γ, β) prior over θ. Let $p = p_1^*(1)$ be the true success probability, and let α be the frequency with which action $a = 1$ is chosen in an interior ergodic memory equilibrium for $r = 0$. For a database $d = (d(0), d(1))$, define $R(d) = \frac{\gamma + d(1)}{\gamma + \beta + d(0) + d(1)}$ as the posterior mean of action $a = 1$, where $d(0)$ and $d(1)$ are the recalled numbers of failures and successes of action $a = 1$. Assume that the known payoff q is such that the agent is never indifferent.[53]

Consider the two polar cases in which only the rehearsal term or only the recency term is present. Let $a^*(d) = \mathbf{1}_{(q,\infty)}(R(d))$. For any $\lambda = (\lambda(0), \lambda(1)) \in \mathbb{R}^2_+$, let $\mathbb{E}_\lambda$ denote expectation when $D(0)$ and $D(1)$ are independent Poisson random variables with means $\lambda(0)$ and $\lambda(1)$. Define $m_F(\lambda) = \frac{\lambda(0)}{k}\mathbb{E}_\lambda[a^*(D + e_0) - a^*(D)], m_S(\lambda) = \frac{\lambda(1)}{k}\mathbb{E}_\lambda[a^*(D + e_1) - a^*(D)]$. Here e_y is the unit database with one additional recalled outcome y. The term $\mathbb{E}_\lambda[a^*(D + e_y) - a^*(D)]$ is the average effect of one additional remembered outcome y on the probability of choosing action $a = 1$. The factor $\lambda(y)/k$ is the marginal increase in the expected number of remembered y outcomes caused by recency at $r = 0$. Define the recency imbalance $\Delta_R(\lambda) = (1 - p)(-m_F(\lambda)) - p\, m_S(\lambda)$. This is the failure-probability-weighted marginal effect of a recent failure minus the success-probability-weighted marginal effect of a recent success. The next result formalizes the local effect on $\text{corr}(a_{t-1}, a_t)$.

Proposition 8. *Let $\alpha(r)$ be a selection of ergodic memory equilibria defined in a neigh-*

[53] That is, there is no finite database $d = (d(0), d(1))$ such that $q = R(d)$. This removes payoff ties, so the chosen action is locally well defined for every recalled database.

borhood of $r = 0$ and continuous at $r = 0$. Suppose $\alpha(0) \in (0,1)$, and write $\lambda_0 = (k\alpha(0)(1-p), k\alpha(0)p)$. Then for all sufficiently small $r > 0$:

(i) Under pure rehearsal, $\operatorname{corr}(a_{t-1}, a_t) > 0$

(ii) Under pure recency, if $\Delta_R(\lambda_0) \neq 0$, then $\Delta_R(\lambda_0)\operatorname{corr}(a_{t-1}, a_t) < 0$.

Proof of Proposition 8. The choice rule $a^*(d)$ is increasing in $d(1)$ and decreasing in $d(0)$. For a candidate action frequency α, write $\lambda_\alpha = (k\alpha(1-p), k\alpha p)$ and let η_α be the corresponding product-Poisson distribution over databases. Since recalled successes make the uncertain action more attractive and recalled failures make it less attractive, $m_S(\lambda_\alpha) \geqslant 0$ and $m_F(\lambda_\alpha) \leqslant 0$.

Fix one of the two polar mechanisms and suppress its label. In the notation of the main text, let $\eta^r_{\alpha,d}$ be the transition law of the next database conditional on the current database d, and let $\mathcal{H}^r_\alpha$ be its stationary distribution where the superscript tracks dependence on r. Under the stationary chain, set $a_t = a^*(D_t)$. At $r = 0$ the transition law is independent of the current database, so $\eta^0_{\alpha,d} = \eta_\alpha$ for every d, $\mathcal{H}^0_\alpha = \eta_\alpha$, and $\operatorname{corr}_{\alpha,0}(a_t, a_{t+1}) = 0$, where the subscript on corr identifies the candidate frequency and recency/rehearsal parameter.

We first record how the change in the stationary distribution enters the derivative. By stationarity, $\mathbb{E}_{\mathcal{H}^r_\alpha}[a^*(D)] = \mathbb{E}_{\mathcal{H}^r_\alpha}\left[\mathbb{E}_{\eta^r_{\alpha,D}}[a^*(D')]\right]$. At $r = 0$, the inner expectation is constant in D and equal to $\mathbb{E}_{\eta_\alpha}[a^*(D)]$. Thus, when the last display is differentiated, the change in the stationary distribution has no effect on the expectation of this constant, and $\frac{\partial}{\partial r}\mathbb{E}_{\mathcal{H}^r_\alpha}[a^*(D)]\big|_{r=0} = \mathbb{E}_{\eta_\alpha}\left[\frac{\partial}{\partial r}\mathbb{E}_{\eta^r_{\alpha,D}}[a^*(D')]\Big|_{r=0}\right]$. Similarly, $\mathbb{E}[a_t a_{t+1}] = \mathbb{E}_{\mathcal{H}^r_\alpha}\left[a^*(D)\mathbb{E}_{\eta^r_{\alpha,D}}[a^*(D')]\right]$. Differentiating this expression and subtracting the derivative of the square of the stationary mean shows that the terms generated by the changing stationary marginal cancel, leaving $\frac{\partial}{\partial r}\operatorname{Cov}(a_t, a_{t+1})\big|_{r=0} = \operatorname{Cov}_{\eta_\alpha}\left(a^*(D), \frac{\partial}{\partial r}\mathbb{E}_{\eta^r_{\alpha,D}}[a^*(D')]\Big|_{r=0}\right)$. Since the covariance is zero at $r = 0$, differentiation of the correlation gives

$$\frac{\partial}{\partial r}\operatorname{corr}_{\alpha,r}(a_t, a_{t+1})\Big|_{r=0} = \frac{1}{\operatorname{Var}_{\eta_\alpha}(a^*(D))}\operatorname{Cov}_{\eta_\alpha}\left(a^*(D), \frac{\partial}{\partial r}\mathbb{E}_{\eta^r_{\alpha,D}}[a^*(D')]\Big|_{r=0}\right) \tag{50}$$

whenever $\operatorname{Var}_{\eta_\alpha}(a^*(D)) > 0$.

The conditional Poisson probabilities are continuously differentiable in (α, r). Moreover, in a neighborhood of $(\alpha_0, 0)$, the probability of an empty database is bounded away from zero in every period. The stationary process therefore regenerates uniformly, so its stationary distribution and stationary correlation inherit the required local differentiability, uniformly

in α. Hence, whenever $\widetilde{\alpha}(r) \to \alpha_0$,

$$\frac{\operatorname{corr}_{\widetilde{\alpha}(r),r}(a_t, a_{t+1})}{r} \longrightarrow \frac{\partial}{\partial r} \operatorname{corr}_{\alpha_0,r}(a_t, a_{t+1})\Big|_{r=0}. \tag{51}$$

□

Rehearsal Under pure rehearsal, conditional on $D_t = d$, the coordinates of D_{t+1} are independent Poisson random variables with means $\lambda_\alpha(y) + r\frac{\lambda_\alpha(y)}{k}\mathbf{1}\{d(y) > 0\}$, for all $y \in \{0, 1\}$. Differentiating the conditional probability of choosing the uncertain action tomorrow gives $\frac{\partial}{\partial r}\mathbb{E}_{\eta^r_{\alpha,d}}[a^*(D')]\Big|_{r=0} = \mathbf{1}\{d(0) > 0\}m_F(\lambda_\alpha)+\mathbf{1}\{d(1) > 0\}m_S(\lambda_\alpha)$. By (50), the derivative of the stationary correlation is

$$\frac{1}{\operatorname{Var}_{\eta_\alpha}(a^*(D))}\big[\operatorname{Cov}_{\eta_\alpha}(a^*(D), \mathbf{1}\{D(0) > 0\})m_F(\lambda_\alpha) + \operatorname{Cov}_{\eta_\alpha}(a^*(D), \mathbf{1}\{D(1) > 0\})m_S(\lambda_\alpha)\big].$$

The first covariance is nonpositive: $a^*(D)$ decreases in failures, while $\mathbf{1}\{D(0) > 0\}$ increases in failures. The second covariance is nonnegative because both terms are increasing in successes. Combining these covariance signs with $m_F(\lambda_\alpha) \leqslant 0$ and $m_S(\lambda_\alpha) \geqslant 0$, the correlation derivative is nonnegative. It is in fact strictly positive: at $\alpha = \alpha_0$, the equilibrium condition gives $\alpha_0 = \mathbb{E}_{\eta_{\alpha_0}}[a^*(D)]$. Since $\alpha_0 \in (0, 1)$, $a^*(D)$ is not almost surely constant, so at least one coordinate has a strict marginal effect and the corresponding covariance above is strict. Since $\operatorname{Var}_{\eta_{\alpha_0}}(a^*(D)) = \alpha_0(1 - \alpha_0) > 0$, equation (50) implies $\frac{\partial}{\partial r}\operatorname{corr}_{\alpha_0,r}(a_t, a_{t+1})\big|_{r=0} > 0$. Taking $\widetilde{\alpha}(r) = \alpha(r)$ in (51) proves part (i).

Recency Under pure recency, the boost matters only after the uncertain action was chosen. Conditional on that action, the current outcome is a success with probability p and a failure with probability $1 - p$. Hence the derivative of tomorrow's probability of choosing the uncertain action after choosing it today is $p\, m_S(\lambda_\alpha) + (1 - p)m_F(\lambda_\alpha)$. If the known action was chosen today, the recency boost is payoff-irrelevant. Therefore $\frac{\partial}{\partial r}\mathbb{E}_{\eta^r_{\alpha,d}}[a^*(D')]\Big|_{r=0} = a^*(d)\big(p\, m_S(\lambda_\alpha) + (1 - p)m_F(\lambda_\alpha)\big)$.

Equation (50) now gives $\frac{\partial}{\partial r}\operatorname{corr}_{\alpha,r}(a_t, a_{t+1})\big|_{r=0} = p\, m_S(\lambda_\alpha)+(1-p)m_F(\lambda_\alpha) = -\Delta_R(\lambda_\alpha)$. Taking $\widetilde{\alpha}(r) = \alpha(r)$ in equation (51) shows that $\frac{\operatorname{corr}_{\alpha(r),r}(a_t,a_{t+1})}{r} \longrightarrow -\Delta_R(\lambda_0)$. Since $\Delta_R(\lambda_0) \neq 0$, the product in part (ii) is therefore strictly negative for all sufficiently small $r > 0$. □