Learning, Limited Memory, and Stochastic Choice*

Drew Fudenberg† Giacomo Lanzani‡
Philipp Strack§

July 12, 2024

Abstract

We study learning by agents whose information depends on their actions and whose decisions are based on a random finite subset of their past experience. We show that if the empirical distribution of actions converges, it must be a *limited memory equilibrium*, and that limited memory equilibrium generates the stochastic choices of random utility models. We relate limited memory equilibrium to the selective memory equilibrium concept that applies when the characterizes the number of recalled experiences goes to infinity as the agent's sample size increases. We then extend the model to allow experiences to be more likely to be remembered if they were remembered in the previous period,

JEL codes: D83, D90

*We thank Simone Cerreia-Vioglio, Roberto Corrao, Amanda Freidenberg, Nicola Gennaioli, Fabio Maccheroni, David Miller, and Tomasz Strzalecki for their helpful comments.

†Department of Economics, MIT

‡Department of Economics, Harvard University

§Department of Economics, Yale University

1 Introduction

People typically only remember a relatively limited number of their past experiences, and what they remember is stochastic– they might remember some things one week and different things the next. People are also typically unaware of the full extent of their limited memory, so they update their beliefs as if the experiences they remember are the only ones that occurred.

We analyze the long-run implications of these memory biases for myopic agents whose information can depend on their actions. Because the agent only recalls a small subset of their experiences, their beliefs and behavior remain stochastic as their number of experiences goes to infinity. We say that a distribution of actions is a *limited-memory equilibrium* if the action distribution is generated by a best response to the distribution of memories it induces. We show that such equilibria exist and that whenever the empirical frequency of actions converges, it converges to a limited-memory equilibrium.

We show that limited memory generates the stochastic choices of a random utility model and specifically that of Lu [2016]'s information representation of stochastic choice rules. When the data generating process consists of a vector of signals about the quality of the possible alternatives and memory is limited but unselective, the stochastic choice rule is monotone in the sense that actions that have higher utility are chosen more often. Thus, although limit behavior is stochastic, the environment disciplines the errors, making more costly mistakes less likely. When the outcomes and prior are normally distributed, we obtain the particular case of the mixed probit random utility model. Here, the variance-covariance matrix of the resulting probit accommodates both payoff monotonicity and diminishing sensitivity, as in baseline probit [Thurstone, 1927], and it also captures frequency dependence because less frequently chosen actions have noisier perceived values.[1] If instead actions are described as vectors of desirable features, and outcomes correspond to situations in which those features proved useful, we show that the limit frequency corresponds to

[1]Kaanders, Sepulveda, Folke, Ortoleva, and De Martino [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.

a different random utility model, the Elimination by Aspect (EBA) model of [Tversky, 1972], where the distribution of the random lexicographic preferences is given by the probability of recalling instances where a particular aspect was useful. We then show that phenomena such as sample size insensitivity [Kahneman and Tversky, 1972] and underreaction to signals, usually attributed to "underinference," [Phillips and Edwards, 1966] can alternatively be explained through limited memory.

We also relate our findings to the limit outcomes when the number of recalled experiences goes to infinity as the agent's sample size increases. We show that every accumulation point of a sequence of limited-memory equilibria with an increasing expected number of recalled events must be a *heterogeneous-belief selective memory equilibrium*, a mixed-strategy generalization of the selective-memory equilibrium that Fudenberg, Lanzani, and Strack [2024] uses to characterizes limit outcomes when the agent's memory is selective but unlimited.

Finally, we expand the model to allow experiences that were remembered in one period to be more likely to be recalled in the next one. This addition makes the model more realistic and lets it fit evidence about the importance of rehearsal and recency. We show that, as with limited memory, a fixed point condition characterizes the limit action distribution. But now, limit beliefs are autocorrelated instead of i.i.d., and the limit action distribution must be consistent with the stationary distribution of the Markov chain of beliefs it induces. This property, which we call *ergodic memory equilibrium*, lets us extend Mullainathan [2002]'s analysis of the effect of rehearsal on income forecasts from short-run predictions to the long run and to more general functional forms. It also lets us provide an explanation of the equity premium puzzle that is similar to that in Weitzman [2007] but does not require misspecified beliefs about the evolution of the state.

Related work Memory has been informally described as stochastic since the early stages of the psychology literature, and the finiteness of memory has been documented at least since Miller [1956]. Shadlen and Shohamy [2016] provides more recent evidence of stochastic memory. d'Acremont, Schultz, and Bossaerts [2013] and Sial, Sydnor, and Taubinsky [2023] provide fMRI and choice evidence, respectively, that agents access their accumulated evidence each period when updating beliefs. Reder [2014], Zimmermann [2020], and Gödker, Jiao, and Smeets [2022] provide evidence of partial or complete unawareness of memory biases.

In limited-memory equilibrium, agents' actions are stochastic because they remember a random sample of their (endogenous) experiences. Several different classes of models derive random choice from randomness in exogenous or endogenous signals. Perhaps the oldest example of this is the Wald optimal stopping problem, where the agent wants to match a binary action with a binary state and pays a flow cost to observe a Brownian signal; once the agent is sufficiently certain of the state, they stop. Fudenberg, Strack, and Strzalecki [2018] extends this to settings where the agent is uncertain of the payoff difference between the actions, and Che and Mierendorff [2019] further extends to more general signal structures. Lu [2016] and Natenzon [2019] axiomatize stochastic choice due to Bayesian updating, where the distribution and number of signals are exogenous.

Wilson [2014] and Jehiel and Steiner [2020] study the optimal use of a finite memory by an agent who receives a stream of exogenous signals until they stop at an exogenous time 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, again with a fixed sample size; Danenberg and Spiegler [2023] extends this to the case where agents receive signals of their payoff to each action whose variance is inversely proportional to the probability the action is played. Gonçalves [2023] defines an equilibrium concept for games based on Bayesian-optimal sequential sampling from the equilibrium distribution.

A large literature in psychology documents the recency effect; see, e.g., the summaries in Lee [1971] and Erev and Haruvy [2016]. There is also extensive evidence of the importance of rehearsal; see, e.g., the Kandel et al. [2000] textbook. Schacter [2008] discusses evidence that some experiences are recalled more often than others. Mullainathan [2002] analyzes the short-run implications of rehearsal in a specific parametric context but did not study its long-run effects. Bordalo, Coffman, Gennaioli, Schwerter, and Shleifer [2021] shows how memory depends on the relative frequency of various characteristics and can be manipulated by making some observations stand out.

2 The Model

We study a sequence of choices made by a single agent. In every period $t \in \mathbb{N}_+$, the agent chooses an action a from the finite set A. In the periods action a is chosen, it induces the objective probability distribution $p^*_a \in \Delta(Y)$ over the finite set of possible outcomes Y.[2] The agent's flow payoff is given by the utility function $u : A \times Y \to \mathbb{R}$.

The agent knows that the map from actions to probability distributions over outcomes is fixed and depends only on their current action but is uncertain about the outcome distributions each action induces. We suppose 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. The support of μ_0 is Θ; its elements are the p the agent initially thinks are possible. We maintain the following assumption throughout:

Assumption 1. For all $p \in \Theta$, $y \in Y$, and $a \in A$, $p^*_a(y) > 0$ if and only if $p_a(y) > 0$.

This assumption guarantees that no data generating process is ruled out in finite time and that posteriors are well defined.[3]

Objective Histories and Recalled Periods We call action-outcome pairs $(a, y) \in A \times Y$ *experiences*. Period $t \in \mathbb{N}$ *histories* are sequences $h_t \in H_t = (A \times Y)^t$, and $H = \bigcup_t H_t$ is the set of all histories. We assume that the agent's memory of past experiences at the beginning of period $t+1$ is distorted by a *memory function* m_{t+1}, where

$$m_{t+1}(a, y) = \min\{1, k/t\}\, m(a, y), \tag{1}$$

[2] We denote objective distributions with a superscript *.

[3] The assumption can be considerably relaxed, as in Fudenberg, Lanzani, and Strack [2021], but the relaxation would create additional technical subtleties unrelated to this paper's focus.

for some fixed function $m : A \times Y \to (0,1]$, and some fixed *memory capacity* $k \in \mathbb{N}$. The way m varies with (a, y) captures the relative memorability of the various experiences.

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 for now that each past experience has an independent probability of being recalled, so[4]

$$\mathbb{P}\left[r_t = R \mid h_t\right] = \prod_{i \in R} m_{t+1}(a_i, y_i) \prod_{i \in \{1,\ldots,t\} \setminus R} (1 - m_{t+1}(a_i, y_i)). \tag{2}$$

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 they realized.

Beliefs We assume the agent recomputes their beliefs each period based on all of their remembered 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 selective memory and naïvely updates their beliefs as if the experiences they remember are the only ones that occurred. Allowing m to vary with a and y lets the model capture various memory biases that have been documented in the literature, such as positive memory bias, cognitive dissonance-reducing memory, associative memory, and confirmatory memory bias (see Fudenberg, Lanzani, and Strack, 2024).

We let μ_{t+1} denote the random (beginning of) period-$t+1$ belief induced by the recalled history, so that the posterior probability of every (measurable) $C \subseteq \Theta$ after recalled history $(a_\tau, y_\tau)_{\tau \in r_t}$ is

$$\mu\big(C \mid (a_\tau, y_\tau)_{\tau \in r_t}\big) = \frac{\int_C \prod_{\tau \in r_t} p_{a_\tau}(y_\tau) d\mu_0(p)}{\int_\Theta \prod_{\tau \in r_t} p_{a_\tau}(y_\tau) d\mu_0(p)}. \tag{3}$$

An implication of equation (3) is that the agent's beliefs depend on the recalled

[4] Section 6 allows experiences that were recalled at $t-1$ to be more likely to be recalled at t.

[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.

history only through how many times each (a, y) pair occurs and not on their order.

Optimal Policies Denote by $BR(\nu)$ the actions that maximize expected utility when the agent's belief is $\nu \in \Delta(\Theta)$:[6]

$$BR(\nu) = \underset{a \in A}{\operatorname{argmax}} \int_{\Theta} \sum_{y \in Y} u(a, y) p_a(y) d\nu(p).$$

A *Markovian policy* $\pi : \Delta(\Theta) \to A$ specifies a pure action for every belief. We assume that the agent is myopic and uses an *optimal Markovian policy* π, i.e., for every $\nu \in \Delta(\Theta)$, $\pi(\nu) \in BR(\nu)$. Together, a true data generating process p^*, a memory function $(m_t)_{t \in \mathbb{N}}$, and an optimal Markovian policy function uniquely induce a probability measure over histories, denoted as $\mathbb{P}_\pi$.

Because the agent's beliefs only depend on the number of times each (a, y) pair is recalled, they can be written as functions of the agent's *database* d of recalled experiences. We let $\mathcal{D} = \mathbb{N}^{A \times Y}$ denote the set of databases, and denote by $\mu(\cdot|d)$ the posterior belief obtained by applying the formula in (3) to an arbitrary history whose database is equal to d.

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

$$\mathbb{P}_\pi \left[\lim_{t \to \infty} \alpha_t = \alpha \right] > 0\,.$$

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

3 Limited-Memory Equilibrium

This section defines limited-memory equilibrium and shows that it characterizes the action distributions that can arise as the limit of the empirical action frequencies.

Limit Distribution of Databases The first step is to derive the distribution over databases that is induced by a fixed action distribution α. For every action distribution α, define $\eta_\alpha^m \in \Delta(\mathcal{D})$ by

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

This is a product distribution where the marginal distribution for each action-outcome pair (a, y) is Poisson with mean $\alpha(a)p_a^*(y)\,k\,m(a, y)$. We will show that η_α^m is the limit distribution of databases if the action frequencies converge to α. Intuitively, 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 how memorable that experience is, i.e., $m(a, y)$.

Lemma 1. *Let $\alpha \in \Delta(A)$. For $\mathbb{P}_\pi$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$, if $\lim_{t\to\infty} \alpha_t = \alpha$ then the distribution of databases given h_{t-1} converges to η_α^m.*

To prove the lemma, we first use a law of large numbers for martingale differences to show that if the empirical action distribution converges to α, then the joint frequency of each action and outcome pair (a, y) converges to $\alpha(a)p_a^*(y)$; Lemma 1 then follows from the Poisson limit theorem on the sum of binomials. The proofs of this and all other results stated in this section are in Appendix A.2.

Limit Distribution of Beliefs The second step is to associate the candidate action distribution with the distribution of beliefs that it induces. We do so by taking the image measure of the databases with respect to the Bayesian updating operator. Let F_α^{m,μ_0} be the distribution of beliefs induced by the distribution η_α^m of databases and prior μ_0, i.e., for all measurable $\mathcal{C} \subseteq \Delta(\Theta)$,

$$F_\alpha^{m,\mu_0}(\mathcal{C}) = \eta_\alpha^m \left(\{d : \mu(\cdot|d) \in \mathcal{C}\}\right). \tag{4}$$

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 $\psi_\rho : \Delta(A) \to \Delta(A)$ be the function that maps α to the action distribution generated when the agent uses policy ρ and their beliefs are distributed according to F_α^{m,μ_0}:

$$\psi_\rho^{m,\mu_0}(\alpha) = \int_{\Delta(\Theta)} \rho(\nu) dF_\alpha^{m,\mu_0}(\nu). \tag{5}$$

The function ψ_ρ^{m,μ_0} is illustrated in the following figure. In words, if action distribution α is played forever, it induces distribution d_α over histories. This distribution and the memory function m together induce a distribution of databases η_α^m, and Bayesian updating on each database generates distribution F_α^{m,μ_0} over posterior beliefs. Finally, assigning $\rho(\nu)$ to each posterior belief ν generates action distribution $\psi_\rho^{m,\mu_0}(\alpha)$.

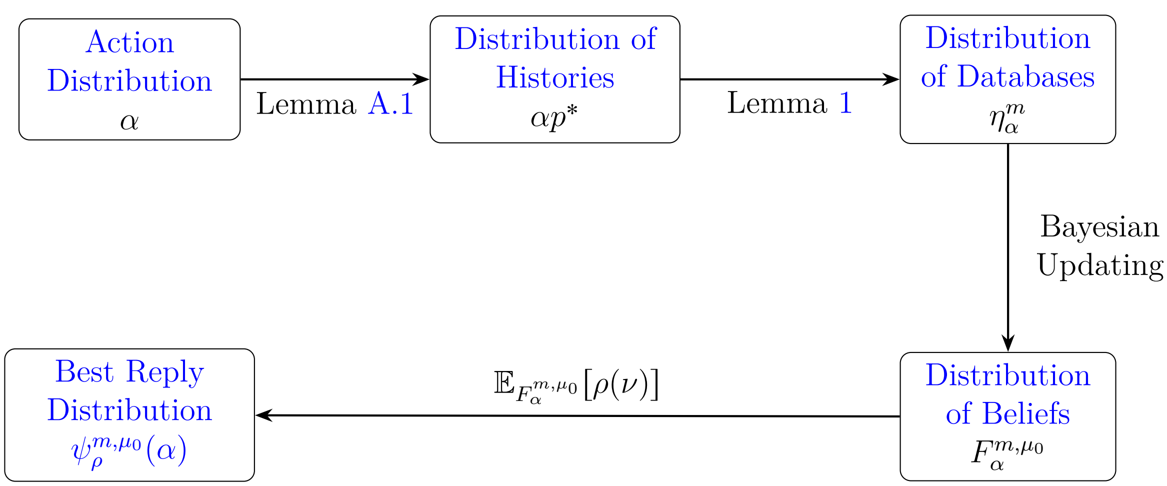


Illustration of ψ_ρ^{m,μ_0}.

Limited-memory equilibrium requires that the agent's behavior best replies to the distribution of memories it induces:

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

Note that the set of limited-memory equilibria depends on the prior μ_0 through its effect on the posterior beliefs. Below, we say more about this dependence and show that it vanishes as k grows.

Theorem 1. *A limited-memory equilibrium exists.*

To prove this, we show that the correspondence that maps each α to the union over $\rho \in \mathcal{O}$ of $\psi_\rho^{m,\mu_0}(\alpha)$ satisfies the conditions of the Kakutani fixed point theorem. The definition of limited-memory equilibrium is justified by the following result, which shows that whenever the behavior converges to an action distribution, that distribution is a limited-memory equilibrium.

Theorem 2. *If α is a limit frequency, then α is a limited-memory equilibrium.*

The first step of the proof is the characterization of the limit beliefs in Lemma 1. The second step of the proof uses the Benaim, Hofbauer, and Sorin [2005] extension of stochastic approximation to differential inclusions to show that the asymptotic behavior of the empirical distribution can be characterized by looking at the limit points of the solution to an associated differential inclusion. In particular, the correspondence defining the inclusion is shown to be a well-behaved integral of the best reply correspondence with respect to F_α^{m,μ_0} (Lemma A.3 in the Appendix). We conclude the proof by showing that if the differential inclusion enters a sufficiently small neighborhood of α, it leaves it after a bounded time interval, which contradicts convergence to α.[7]

[7]There can be multiple limited-memory equilibria, see Example 2 in the Online Appendix. Whether multiple limited-memory equilibria can have a positive probability of arising as the limit behavior remains an open question.

4 Applications

4.1 Random Utility and Stochastic Choice

As Theorem 2 makes clear, a decision-maker with limited memory will behave stochastically even in the long run. This section relates random behavior in limited-memory equilibrium to the most widely used model of non-deterministic behavior in single-agent problems, the random utility model of stochastic choice. In a random utility model, the agent's utility function for the various actions is independently drawn from a fixed distribution in every period; in a limited-memory equilibrium, the agent's beliefs about the expected utility of each action are determined by their random memories. By connecting the two concepts, we will provide a long-run learning foundation for the random utility model and some of its specifications.

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 $\sum_{x \in M} c(x, M) = 1$ for all $M \in \mathcal{M}$. Let $\mathcal{P}$ be the strict orders on A. A stochastic choice function c has a *random utility representation* if there is $\nu \in \Delta(\mathcal{P})$ such that for all $M \in \mathcal{M}$,[8]

$$c_\nu(x, M) = \nu\left(\{P \in \mathcal{P} \mid \forall y \in M, xPy\}\right).$$

To relate our model of memory and learning to random utility, we suppose that the analyst, as in the standard decision theory exercise, can elicit the agent's choice from restricted sets of actions. We assume that the decision maker uses an optimal Markovian policy and chooses from the restricted sets in an optimal Markovian way, and breaks ties in a menu-independent way (i.e., it satisfies *uniform tiebreaking* in the sense of Chapter 1.6 of Strzalecki [2023]).[9]

Definition 2. Behavior converges to a random utility representation ν on a history sequence $(h_t)_{t \in \mathbb{N}}$ if for every $\varepsilon > 0$ there is $t \in \mathbb{N}$ such that $|\mathbb{P}_\pi[a_{\tau+1} = a | h_\tau, M] -$

[8]This is equivalent to the form of the random utility representation that uses an additional probability space. See, e.g., Proposition 1.9 in Strzalecki [2023].

[9]To state this condition formally, let $BR(\nu|M)$ denote the set of best replies to the posterior restricted to menu M for any $M \subseteq A$. Uniform tie breaking requires that for all $M, M' \in \mathcal{M}$ if $a, a' \in BR(\nu|M) \cap M'$ and $\pi(\nu|M) = a$, then $\pi(\nu|M') \neq a'$.

$c_\nu(a, M)| \leqslant \varepsilon$ for all $\tau \geqslant t$.

The next result shows that when the decision maker's behavior stabilizes, it converges to a random utility representation. Moreover, the behavior will be consistent with Lu [2016]'s information representation, where there is a state space (for us, the set Θ), a set of outcomes (for us, the utility levels $u(a, y)$), and the decision maker is choosing between acts (for us, the actions in A) to maximize expected utility with respect to a distribution over posteriors over states. (For us this distribution is $F_{\alpha^*}^{m,\mu_0}$.) Intuitively, Lu models random choice as the result of a random posterior belief; in our model, random posterior beliefs come from finite samples of past events.

Proposition 1. *Suppose the agent uses an optimal Markovian policy π, then for every $\alpha \in \Delta(A)$ and on any sequence of histories such that $\mathbb{P}_\pi[\lim_{t\to\infty} \alpha_t = \alpha^*] > 0$, the agent converges to a random utility representation. In particular, it has an information representation.*

The proofs for this section are in Appendix A.3. The proof of the first constructs the target random utility representation. To do this, we associate to every database d a 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 $\mu(\cdot|d)$.[10] The random utility representation is then defined by assigning each ranking the limit probability of the recalled histories that induce it. Since Lemma A.2 guarantees that the distribution over recalled histories converges and the set of menus is finite, this pushforward measure also converges. To see that the stochastic choice rule admits an information representation, we map our objects to those in Lu [2016]'s representation of stochastic choice as utility maximization with a fixed utility index and a random posterior.

4.1.1 Monotonicity

To further develop the connections between limited memory and stochastic choice, we now assume that memory is imperfect but not selective, i.e., $m(a, y) = 1$,[11]

[10] Here, we use the assumption that when two actions have the same expected utility, the agent uses a deterministic and history-independent tie-breaking rule.

[11] With non-selective memory restricting to $c = 1$ is without loss of generality, as any combination of c and k for which ck is the same induces the same memory function in equation (1).

and what the agent observes is independent of their chosen action, so that outcome $y = (y_a)_{a \in A}$ is the $|A|$ dimensional vector corresponding to the payoff that would have been obtained taking the different actions: $u(a, y) = y_a$. Moreover, we assume that the mean payoffs $(\mathbb{E}_p[y_a])_{a \in A}$ are sufficient statistics for the agent's beliefs about the outcome distributions. For the results about basic stochastic choice properties, no requirements are imposed on these conditional distributions.

The prior belief of the agent induces a joint distribution over the pair $(\mathbb{E}_p[y_a], y_a)$ through the combination of the prior over the means and the conjectured conditional distribution given the mean. Our first result shows that under a prior for which each of these joint distributions is affiliated, the limit choice probability is well-behaved in the real quality of the different actions: actions with higher payoffs are played more frequently.[12]

Proposition 2 (Monotonicity)**.** *Suppose that for all $a \in A$, the agent prior is correct about the distribution of y_a conditional on $\mathbb{E}_p[y_a]$ and that for all $a \in A$ $\mathbb{E}_p[y_a]$ and y_{ta} are affiliated under the prior. Then, if α, is a limit frequency with $\mathbb{E}_{p*}[y_a]$ and we increase $\mathbb{E}_{p*}[y_a]$, keeping fixed $\mathbb{E}_{p*}[y'_a]$, $a' \neq a$, $\alpha(a)$ increases.*

The classical explanations of stochastic choice are purely subjective and unrelated to objective quality measures. In contrast, 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.

The objective environment does have a role in models where the decision maker with perfect memory stops and makes a choice after acquiring information from costly sequential sampling, as in Fudenberg, Strack, and Strzalecki [2018], Che and Mierendorff [2019], Ke and Villas-Boas [2019] and Hébert and Woodford [2023]. The key difference is that agents repeatedly make decisions in our setting, while in the optimal stopping papers, it is made once and for all. Callaway, Rangel, and Griffiths [2020] measures the repeated allocation of attention and shows it is directed towards alternatives that are perceived to have higher values and more uncertainty. The first

[12] The expected and realized outcomes are affiliated with respect to the prior probability measure if for all $c, c' \in \mathbb{R}$, $\mathbb{P}_\mu[\mathbb{E}_p[y_a] \geqslant c, y_a \geqslant c'] \geqslant \mathbb{P}_\mu[\mathbb{E}_p[y_a] \geqslant c]\mathbb{P}_\mu[y_a \geqslant c']$.

property is consistent with our result; the second requires an experimentation motive that is ruled out by our assumption that agents are myopic.

4.1.2 Specific Signal Structures

The fixed-point condition defining limited-memory equilibrium can be quite tractable with specific signal structures. This section gives two examples, one with normally distributed signals and one where signals have a beta distribution; Appendix B.2 presents the case of exponentially distributed signals.

Probit *Probit* [Thurstone, 1927] is a stochastic choice model with two desirable features: payoff monotonicity and diminishing sensitivity. Our learning model delivers the related *mixed probit* [Hausman and Wise, 1978; Greene, 2000] specification, which also has these two properties.[13] Moreover, the form of mixed probit that we obtain is also consistent with frequency-dependence (i.e., agents have more precise estimates of the values of actions they take more frequently), which, as Strzalecki [2023] points out, is not accommodated by baseline probit.

To relate mixed probit to our model, suppose the payoff y_a of each action $a \in A$ is i.i.d. normally distributed $y_a \sim \mathcal{N}(\bar{y}_a, \sigma^2)$, and that the agent's prior belief is that $(\bar{y}_1, \ldots, \bar{y}_{|A|})$ are independently normally distributed with mean 0 and variance σ_0^2. We call this a *normal environment*. (Here, we allow infinitely many outcomes, so our general results do not directly apply, but the definitions extend in the obvious way.)

Proposition 3. *Let α be a limit frequency in a normal environment. Its induced choice probabilities correspond to a Mixed Probit Model where the mean parameter is constant, and the variance parameter is a diagonal matrix where the entry in position (a, a) is σ^2/n and n has a Poisson distribution with parameter $\alpha(a)k$.*

In this mixed probit specification, the variance σ^2/n of the payoff to a is stochastically decreasing (i.e., in the sense of first-order stochastic dominance) in $\alpha(a)$, so

[13] A mixed probit is a hierarchical stochastic choice rule. First, the probit parameters (mean and covariance matrix) are drawn with some mixing probabilities; given the realized probit parameters, the usual stochastic probit rule determines the choice.

the payoffs of more frequently chosen alternatives are more precisely estimated.[14]

Elimination by Aspects Elimination by aspects [Tversky, 1972] postulates that agents use "random lexicographic order" when choosing between alternatives with multiple attributes: they randomly choose an attribute to focus on and restrict their choice to the alternatives with the largest values of that attribute. If there are multiple such alternatives, a second attribute is randomly chosen, and only maximal alternatives (within the restricted set) in that second attribute are considered. The procedure continues in this way until a single alternative is left.

EBA was designed to capture the following observed violations of the 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 from a (i.e., not a "decoy" in the sense of Huber, Payne, and Puto, 1982) draws relatively more probability away from a than from a'. To fix idea ideas using an example from Tversky [1972], suppose a manager needs to decide whether to hire a worker based on their intelligence and motivation score. Adding a worker with (intelligence, motivation) scores $(78, 25)$ to a choice between $(75, 35)$ and $(60, 90)$ has been shown to draw away significantly more choice probability from $(78, 25)$.

The EBA model is a RUM. It does not have an axiomatic foundation [Strzalecki, 2023]; our limited memory model gives it a foundation based on learning. To see this 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 receives payoff equal to the worker's skill in the dimension that is relevant in this period (intelligence or motivation). Further, the manager is uncertain about the probabilities with which each is relevant. In particular, the prior is 50-50 on two DGPs: either there is probability 0.9 that motivation is the relevant factor or probability .9 that success only depends on the worker's intelligence, and the memory function is constant. 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

[14] Danenberg and Spiegler [2023] makes a similar point about the relation between frequency and precision in a setting with exogenous normally distributed signals and no memory limitations.

$(75, 35)$ is chosen over $(60, 90)$ equals the probability that the manager recalls more past periods in which intelligence was important. This is equal to the probability that a draw from a Poisson distribution with parameter proportional to the probability of an intelligence task is larger than a draw from a Poisson with parameter proportional to the probability of a motivation task. And, as predicted by the EBA prediction, the addition of $(78, 25)$ will reduce the probability of $(75, 35)$ but not that of $(60, 90)$.[15] Also, observe that the probability that a particular aspect will be used for valuations, which is not specified in Tversky [1972], is partially disciplined by the true data generating process: The EBA criterion obtained from our limited-memory foundation (with a uniform prior) predicts that more frequently relevant aspects have a higher probability of being used to determine choice.

This relation extends beyond this example. Define an *aspects environment* as one where $A \subseteq \{0,1\}^{\mathcal{A}}$ for some finite $\mathcal{A}$, $Y = \mathcal{A}$, $u(a, y) = a_y$, for every $p \in \Theta$, $a, a' \in A$ $p_a = p'_a$ and that the prior is *responsive*: for every database $d(A, y) > d(A, y')$ and $a_y > a'_y$ implies that $\int u(a, \cdot) d\mu(\cdot|d) > \int u(a', \cdot) d\mu(\cdot|d)$.[16] The interpretation is that actions with more 1's have more desirable features, and the problem is nontrivial only if no alternative is pointwise weakly larger than the others.[17]

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

---

[15] 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 $n_M > n_I$ the posterior probability of DGP $(0.9, 0.1)$ is at least 0.9, so $(78, 25)$ the unique best reply.

[16] The latter condition is trivially satisfied when there are two aspects like in the intelligence/motivation example above or when each $a \in A$ has a unique strictly positive entry.

[17] We follow Tversky [1972] in considering a formal framework with binary discrete attributes but an example where attributes can take on more values. The latter can be reduced to a special case of the former by transforming each option into a vector of 0 and 1, where 1 means that they are the alternative with the highest value in that entry. Gul, Natenzon, and Pesendorfer [2014] axiomatizes an extension of EBA where attributes are nonbinary (with a different tie-breaking rule). Our corollary extends to that case if the outcomes are enriched to encode noisy signals about each attribute's level and whether it is currently relevant.

Binomial Beta Model Suppose there are only two outcomes $Y = \{0, 1\}$ and that for each action $a \in A$, the prior is independently and identically beta distributed with parameters γ, β, and that the agent only observes the outcome of the action they chose. The posterior mean the agent's assigns to action a then is given as

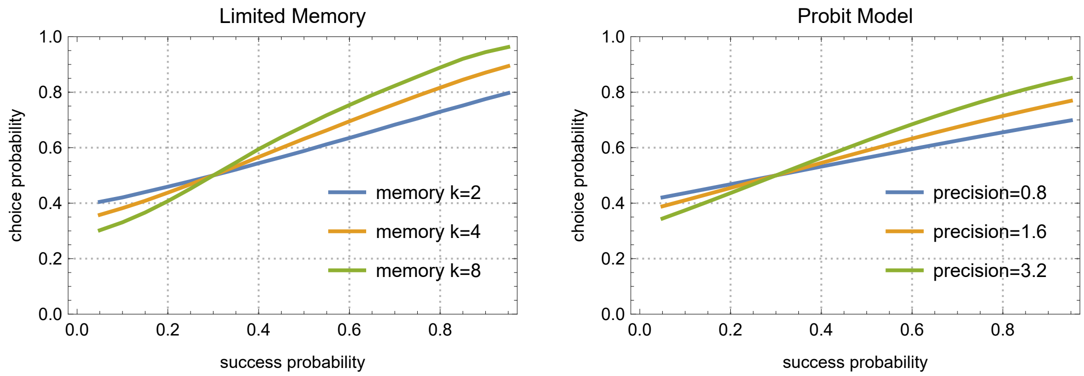
$$r_a = \frac{\gamma + n_{1a}}{\gamma + \beta + n_{1a} + n_{0a}},$$

where n_{1a}, n_{0a} is the number of events the agent remembers where the action was a, and the outcome was a success ($y = 1$) or a failure ($y = 0$). As we show in Lemma A.2, n_{ya} is Poisson distributed with parameter $\alpha(a)p_a^*(y)k\,m(a, y)$, so $n_a = n_{1a} + n_{0a}$ is Poisson distributed with rate $\alpha(a)\sum_y p_a^*(y)k\,m(a, y)$ because n_{1a} and n_{0a} are independent. We can thus explicitly describe a limited-memory equilibrium as a solution α to the equation

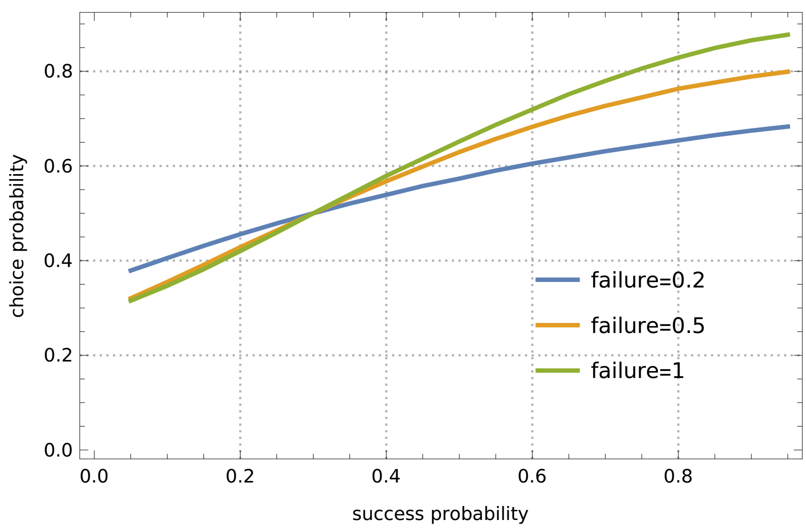
$$\begin{aligned}
\alpha(a) &= \mathbb{E}\left[\frac{\mathbf{1}\{r_a = \max_{a'} r_{a'}\}}{|\arg\max_{a'} r_{a'}|}\right] \\
n_a &\sim \text{Poisson}\left(k\,\alpha(a)\sum_y p_a^*(y)m(a, y)\right) \\
n_{1a} &\sim \text{Poisson}\left(k\,\alpha(a)p_a^*(1)m(a, 1)\right),
\end{aligned}$$

where we assume that the agent uniformly randomizes over actions when they are indifferent. The above is a fixed point equation as r_a is a function of n_a, n_{1a} whose distribution is a function of α.

With endogenous data, the signal's precision about the action's quality depends on how frequently that action is chosen. This effect can be seen in the previous figure. There, the left panel contains three different expected numbers of recalled experiences that mimic the behavior under three different precisions in the probit model when the success probability of the uncertain action is high, and thus, it is played frequently. However, differences in memory capacity played a more significant role at lower success probabilities and, thus, lower equilibrium use of the uncertain action. Indeed, when the success probability is low, the endogenous low number of



Binomial beta: Probability with which the first action is chosen when there are two actions $a = \{0, 1\}$ with binary outcomes, a symmetric beta prior with $\beta = \gamma = 1$, no memory bias $(m = 1)$, $k = 2$ (blue), $k = 4$ (orange), $k = 8$ (green), the second action leads to the good outcome with probability 30%. The probability the first action leads to the good outcome is on the x-axis, and the probability with which this action is chosen is on the y-axis.



Probability the first action is chosen when there are two actions $a = \{0, 1\}$ with binary outcomes, and a symmetric beta prior with $\beta = \gamma = 1$, $k = 2$. The three lines correspond to different relative memorability of a failure, $m(0) \in \{0.2, 0.5, 1\}$ and $m(1) = 1$. The second action leads to the good outcome with probability 30%. The probability the first action leads to the good outcome is on the x-axis, and the probability with which this action is chosen is on the y-axis.

recalled experiences makes an additional observation more valuable.

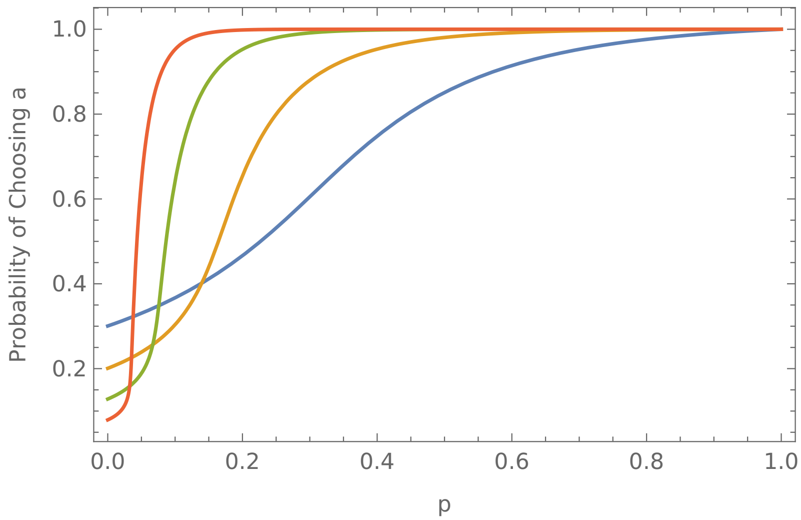
4.2 The Effect of Skewness

Our next example shows that the probability that an action is chosen need not be determined by its expected payoff but can also depend on higher-level moments, such as variance and skewness. Suppose that the agent chooses between two actions $A = \{0, 1\}$ with outcomes $Y = \{0, c\} \cup [1, \infty)$. The action 0 always produces the outcome $c \in (0, 1)$, e.g., $\mu(\{p : p_0(c) = 1\}) = 1$. The action 1 generates outcome $y = 0$ with probability $1 - q$. With probability q, action 1 produces the outcome $y = 1/b$ with probability b and the outcome 0 with probability $1 - b$. The agent does not know b, which is uniformly distributed on $[0, 1]$.

The agent's payoff function is $u(a, y) = y$. As the first success reveals that the action $a = 1$ has an expected payoff of 1, the agent will always choose that action whenever they remember a success. The prior expected value of $a = 1$ is q, so for $q > c$, the agent will take the action $a = 1$ if they don't remember any outcomes of the risky arm. After remembering the outcome $y = 0$ once after taking the action $a = 1$, the posterior expected value associated with the action $a = 1$ is given $0.5q/(0.5q + (1 - q))$. If $0.5q/(0.5q + (1 - q)) < c$, it is optimal for the agent to take the safe action if they remember the outcome $y = 0$ at least once after taking the action $a = 1$. Thus, for a prior as described above with $0.5q/(0.5q + (1 - q)) < c < q$, the agent takes the action $a = 1$ if they don't remember an outcome with $a = 1$, or if they remember at least one occurrence of $y \geqslant 1$. This generates the equilibrium probability of choosing action 1 displayed in Figure 4.

As rare events are likely not to be recalled, they will not be present in most databases and so be ignored. As a consequence, actions that are objectively equally desirable as the one in Figure 4 tend to be chosen more often when they deliver a good payoff with high probability than when they deliver a very good payoff more rarely.[18] This is consistent with the evidence in Hertwig, Barron, Weber, and Erev [2004],

[18] Ellison and Fudenberg [1993] makes this point in a model where each agent sees two signals. Conversely, when a rare event is recalled, it will tend to be over-represented in the database and trigger over-reaction.



Probability of $a = 1$ when $\mathbb{E}_p[u(1, y)] = 1$, $m = 1$, and $k = 4$ (blue), $k = 8$ (orange), and $k = 16$ (green), $k = 32$ (red).

which also shows how this effect is not obtained if probabilities are given rather than learned. The figure also illustrates that the effect decreases in the memory capacity k, and in the limit $k \to \infty$, the agent chooses optimally.[19]

4.3 Underreaction to Evidence

In a limited-memory equilibrium, the agent sometimes relies on a small dataset to make decisions. This can induce long-run underreaction of beliefs and insensitivity to sample size, which Benjamin [2019] reports are some of the most persistent departures from rationality in probabilistic reasoning. The main model that has been used to explain this underreaction, Phillips and Edwards [1966])'s underinference model, modifies Bayes rule to

$$\tilde{mu}(C|(a_i, y_i)_{i=1}^t) = \frac{\int_C \prod_{i=1}^t (p_{a_i}(y_i))^c d\mu(p)}{\int_\Theta \prod_{i=1}^t (p'_{a_i}(y_i))^c d\mu(p')}, \text{ for some } c \in (0, 1).$$

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

with $c \in (0, 1)$. Although underinference and limited memory predict underreaction to the data, they have distinctive features that can be tested in the lab and the field. First, 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 being provided written records of past outcomes, while underinference does not.[20]

We can calibrate the memory capacity parameter k to match various aspects of the evidence. For example, the limit of the probability that the agent doesn't remember any relevant experiences is equal to[21]

$$\prod_{(a,y)\in A\times Y} \exp(-\alpha(a)p_a^*(y)\, k\, m(a,y)), \tag{6}$$

There is a similar formula for the probability of making a decision based on at most n experiences.[22]

5 Unlimited Memory

This section compares the equilibria and long-run outcomes under the memory function $m_{t+1}(\cdot) = \min\{1, k/t\}m(\cdot)$ to those with a time-independent version of the same memory function, i.e., $m_{t+1}(\cdot) = m(\cdot)$. When m is independent of time, the expected

[20]The persistent uncertainty comes from the finite expected memory; the effect of reminders and written records also arises with the unlimited-memory model of Fudenberg, Lanzani, and Strack [2024], as that paper noted. See Fudenberg and Peysakhovich [2016] and Esponda, Vespa, and Yuksel [2023] for experimental evidence on the effect of providing agents with records and/or summary statistics.

[21]Asymptotically, the probability the agent remembers nothing converges to the product of the probabilities that the agent doesn't remember each (a, y) pair, and each of these is Poisson with parameter $\alpha(a)p_a^*(y)km(a, y)$.

[22]Miller [1956] claims that people can store at most 7 "chunks" in working memory, while Cowan [2001] says they can only access 4, but it is not clear how to map these chunks to our experiences.

number of recalled experiences converges to infinity with t, as in Fudenberg, Lanzani, and Strack [2024].

To facilitate this, we extend that paper's definition of selective-memory equilibrium to mixed actions, which requires an extension of the paper's definition of the memory-weighted likelihood maximizers from pure actions to action distributions. This is the set

$$\Theta^m(\alpha) := \underset{p\in\Theta}{\operatorname{argmax}} \left(\sum_{a\in A} \alpha(a) \sum_{y\in Y} m(a,y) p^*_a(y) \log p_a(y) \right). \tag{7}$$

For every action distribution α, $\Theta(\alpha)$ consists of the models that maximize a weighted version of the log-likelihood of the true data generating process, where the weights depend both on the memory function and the frequency of each action.

Definition 3. (i) A *pure selective memory equilibrium* is an action a such that there is $\nu \in \Delta(\Theta^m(\delta_a))$ such that $a \in BR(\nu)$.

(ii) A *unitary-beliefs selective memory equilibrium* is an $\alpha \in \Delta(A)$ such that there is $\nu \in \Delta(\Theta^m(\alpha))$ with $a \in BR(\nu)$ for all $a \in \operatorname{supp}(\alpha)$.

(iii) A *heterogeneous-beliefs selective memory equilibrium* is an $\alpha \in \Delta(A)$ such that for all $a \in \operatorname{supp}(\alpha)$ there is $\nu^a \in \Delta(\Theta^m(\alpha))$ such that $a \in BR(\nu^a)$.

Note that, unlike limited-memory equilibrium, these equilibrium concepts only depend on the prior's support and not on the relative weights the prior assigns to various models. Note also that heterogeneous-belief selective memory equilibrium requires that every action in $\operatorname{supp}(\alpha)$ is justified by a (possibly different) belief over the *same* set of likelihood maximizers $\Theta^m(\alpha)$, corresponding to the mixed action α. This differs from 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 $\Theta^{\mathbf{1}}(a)$ corresponding to data about the consequences of a. This difference is a consequence of the different origins of the heterogeneity for the two equilibrium concepts. 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. In contrast, in heterogeneous-belief selective memory equilibrium is a single-person equilibrium concept, where all the beliefs are based on the consequences of the same mixed action α. Here, as we show in Example 1 below, the heterogeneity arises because, with limited memory, the agent's database and beliefs are stochastic even when k is arbitrarily large, as we show in Example 1 below.

Theorem 3. *Suppose $(\alpha^k)_{k\in\mathbb{N}}$ is a sequence of limited-memory equilibria each with memory capacity k and that $\lim_{k\to\infty}\alpha^k = \hat{\alpha}$. Then $\hat{\alpha}$ is a heterogeneous-beliefs selective memory equilibrium.*

The first step of the proof is to show that when $\alpha^k \to \alpha$, the distributions of databases also converge, so the agent's beliefs concentrate on $\Theta^m(\hat{\alpha})$. The fact that $\hat{\alpha}$ is a selective memory equilibrium then follows from the fact that each action $\tilde{a}$ for which $\hat{\alpha}(a) > 0$ is a best reply to some belief concentrated on $\Theta^m(\hat{\alpha})$. Importantly, there may not be a single belief that makes all of these actions best replies, so the limit need not be a unitary-beliefs selective memory equilibrium.

Combining Theorems 1 and 3 shows that a heterogeneous-beliefs selective memory equilibrium exists, but considering a game between the agent and Nature, where Nature chooses the agent's beliefs to maximize the memory-weighted log-likelihood of the agent's data, shows that a stronger result is true:

Proposition 4. *A unitary-beliefs selective memory equilibrium exists.*

Definition 4. We say Θ has a *product structure* if there exists $(\Theta_a)_{a\in A} \subseteq \Delta(Y)^A$ such that $\Theta = \times_{a\in A}\Theta_a$, so that for any $q', q'' \in \Delta(Y)$ and any $a', a'' \in A$, if there are $p', p'' \in \Theta$ with $p'_{a'} = q'$ and $p''_{a''} = q''$ then there is $\tilde{p} \in \Theta$ with $\tilde{p}_{a'} = q'$ and $\tilde{p}_{a''} = q''$.

Note that the assumption that Θ has a product structure does not imply that the agent's prior is a product of independent marginals.

Proposition 5. *When Θ has a product structure and is convex, any heterogeneous-belief selective memory equilibrium can be supported with unitary beliefs.*

Combining Theorem 3 and Proposition 5 shows that when Θ is convex and has a product structure if $(\alpha^k)_{k\in\mathbb{N}}$ is a sequence of limited-memory equilibria converging to $\hat{\alpha}$, $\hat{\alpha}$ is a unitary-beliefs selective memory equilibrium. The next example shows that without the assumption of Proposition 5, there are heterogeneous-beliefs selective memory equilibria that are not unitary-beliefs selective memory equilibria. It also shows that there exist pure selective memory equilibria that are not the limit of a sequence of limited-memory equilibria.

Example 1. *Let $Y = \{-1, 1\}$, $A = \{-1, 0, 1\}$, $u(a, y) = ay - 0.1\mathbb{I}_{\{-1,1\}}(a)$. The agent believes that p, the probability of $y = 1$, is independent of a and is either equal to 0.9 or 0.1, so the myopic best response is 1 if $\mu(.9)>9/16$, -1 if $\mu(.9) < 7/16$ and 0 if $7/16 < \mu(.9) < 9/16$. The true probability of $y = 1$ is indeed independent of a, and equal to 0.5. Here both $p = .1$ and $p = .9$ are likelihood maximizers, and $a = 0$ is a unitary-belief selective memory equilibrium supported by the belief that both models are equally likely. There aren't any unitary beliefs equilibria with a positive probability of both $a = -1$ and $a = 1$ because the unique belief that makes actions -1 and 1 indifferent assigns equal probability to 0.9 and 0.1, and at that belief action, 0 is preferred to both. However, if the prior is uniform over $.9$ and $.1$, then after observing j occurrences of $y = 1$ and k occurrences of -1, the posterior probability that $p = .9$ is 9^{j-k}, so the agent strictly prefers either action 1 or -1 unless $j = k$. As k grows to infinity, the probability that $j = k$ goes to 0, so the limited-memory equilibria converge to the heterogeneous-belief equilibrium $(1/2, 0, 1/2)$.*

Corollary 2. *Fix $\Theta \subseteq \Delta(Y)^A$ and $m : A \times Y \to [0, 1]$. Suppose that all $p \in \Theta$, p^*, and m do not depend on actions. Then:*

1. *When the agent has infinite expected memory, if there is a unique selective memory equilibrium $\hat{a}$, then $\mathbb{P}_\pi[\sup\{t\colon a_t \neq \hat{a}\} < \infty] = 1$ for every optimal policy π.*
2. *For every action a that is a strict best reply to some beliefs on Θ, and every memory capacity $k \in \mathbb{N}$, there exist a prior belief with support Θ and an optimal policy π such that if $\mathbb{P}_\pi[\lim_{t\to\infty} \alpha_t = \alpha] > 0$ then $\alpha(a) > 0$.*

When the data generating process is exogenous and memory is unbounded, the empirical distribution of recalled outcomes converges almost surely, and the agent ends up playing the best reply to this distribution. Since there is a unique selective memory equilibrium, this best response is unique, and the agent eventually converges to it. With finite memory, there is a positive fraction of periods in which the agent recalls so little that they play a best reply to their prior, although the probability that this occurs becomes smaller and smaller as k goes to infinity. More generally, when the action does influence the distribution of outcomes, the prior may affect the probability of converging to a specific selective memory equilibrium. Still, the set of selective memory equilibria is the same for priors that share a common support. This is not the case with limited memory.

6 Rehearsal and Recency

This section extends the model to incorporate the effects of *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, and recency is the idea that the agent gives more weight to more recent outcomes. To model these phenomena, we assume that the agent's memory at time $t+1$ is distorted through a *rehearsal memory function* that can depend on the last period's experience and the experiences that were recalled then. Formally, let d_t denote the database recalled at period t, and define

$$m_{t+1}((a,y)|d_t,(a_t,y_t)) = \min\{1, k/t\}\left(m(a,y) + r\mathbb{1}_{\{(a',y'):d_t(a',y')\geqslant 1\}\cup\{(a_t,y_t)\}}(a,y)\right), \tag{8}$$

where the experiences that were recalled or experienced last period have an additional probability of $r \in [0, 1-\max_{a\in A, y\in Y} m(a,y)]$ of being recalled, and $r=0$ reduces to equation (1).[23] Throughout this section, we fix an optimal Markovian policy π. The

[23] 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 in (2), with m_{t+1} replaced by $m_{t+1}(\cdot|E_t,(a_t,y_t))$.

Both rehearsal and recency typically depend on more than just the last period's outcome and recollections. We think we could extend the analysis to allow these effects to depend on a finite number of past periods, but allowing an unbounded number of past periods to matter would cause proofs for this section are in Appendix A.5.

6.1 Ergodic Memory Equilibrium

Limit Distribution of Databases As in the baseline model, the stationary distribution of databases is a product of Poisson distributions, but now they depend on what was recalled in the previous period, in addition to the frequency of each action a, the probability of the outcomes given a, and how memorable that experience is. We now define a Markov chain over databases for each action distribution $\alpha \in \Delta(A)$.

Definition 5. The Markov chain η_α^m has state space $\mathcal{D}$ and Markov kernel $\eta_{\alpha,d}^m(d')$,[24] where for every $\alpha \in \Delta(A)$, $t \in \mathbb{N}$, and $d \in \mathcal{D}$, let $\eta_{\alpha,d}^m \in \Delta(\mathcal{D})$ be a product of independent Poisson distributions with parameter each $(a, y) \in A \times Y$ equal to

$$\begin{cases} \alpha(a)p_a^*(y)\, k\, [m(a,y) + r] & \text{if } d(a,y) \geqslant 1 \\ \alpha(a)p_a^*(y)\, k\, [m(a,y) + rp_a^*(y))] & \text{if } a = \pi(\mu(\cdot|d)) \text{ and } d(a,y) = 0 \\ \alpha(a)p_a^*(y)\, k\, m(a,y) & \text{otherwise.} \end{cases}$$

Intuitively, the expected number of times experience (a, y) is recalled given previous database d is proportional to the frequency of a, the probability of y given a, how memorable that experience is, and whether it either occurred last period or was recalled in d. We will show that this Markov chain has a unique stationary distribution (Lemma 2) and that this distribution is the limit time-average distribution over databases when the distribution over actions is α (Claim 1 in the Appendix).

The first step is to note that at any time, every subdatabase of what is currently recalled has a positive probability of being the subsequent database. In particular, every period, the null database has a positive probability of being recalled, so the chain is irreducible on the subsets of databases that can be reached with a positive probability starting from the empty database. A calculation in the appendix shows the Markov chain is also positive recurrent, which yields the following lemma.

significant complications. Mullainathan [2002] also assumes a one-period recency window.

[24] That is, the distribution of d given d' is $\eta_{\alpha,d'}^m$.

Lemma 2. *η^m_α admits a unique stationary distribution $\mathcal{H}_\alpha \in \Delta(\mathbb{N}^{A\times Y})$.*

Let $F^{m,\mu_0}_{\alpha,d'}$ be the database-dependent distributions of beliefs induced by $\eta^m_{\alpha,d'}$: For each $d \in \mathcal{D}$ and all measurable $\mathcal{C} \subseteq \Delta(\Theta)$,

$$F^{m,\mu_0}_{\alpha,d'}(\mathcal{C}) = \eta^m_{\alpha,d'}\left(\{d : \mu_0(\cdot|d) \in \mathcal{C}\}\right). \tag{9}$$

Definition 6. An *ergodic memory equilibrium* is an $\alpha \in \Delta(A)$ such that there exists $(\alpha_d)_{d\in\mathcal{D}} \in \Delta(A)^{\mathcal{D}}$ and $\rho \in \mathcal{O}$ such that

1. Each α_d equals the action frequencies induced by ρ, i.e.,

$$\alpha_d = \mathbb{E}[\rho(\nu)|\nu \sim F^{m,\mu_0}_{\alpha,d}],$$

2. $\alpha = \mathbb{E}_{\mathcal{H}_\alpha}[\alpha_d]$.

Ergodic memory equilibria are fixed points: for every database d, a mixed action α determines a probability distribution over what is recalled next period, and thus over the next period's beliefs. The agent's policy applied to those beliefs determines a mixed action α_d; ergodic memory equilibrium requires that the expectation of α_d with respect to the induced stationary distribution over databases is α.

Theorem 4. *An ergodic memory equilibrium exists.*

The proof extends that of Theorem 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 5. *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 inhomogeneous Markov chain over databases has the "Doeblin property" that there is a state that has positive probability of being reached in one period from every state, which guarantees convergence to the ergodic distribution. (In our case, the special state is the empty database.) The second step generalizes Lemma A.2 on the convergence of beliefs conditional on the databases, with the key difference that now the limit belief distribution is database-dependent. The final part of the proof repeats arguments from the proof of Theorem 2.

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 and equity-volatility puzzles. For this subsection we suppose that the outcome y_t is i.i.d. $y_t = \theta + \epsilon_t$, independent of the action of the agent, where $\theta \in \mathbb{R}$ and the ϵ_t are mean- 0 shocks.[25]

Correlated Prediction Errors The rehearsal memory function of equation (8) generates the same predictions about one-period correlations as Mullainathan [2002], without assuming associativeness. First, a high outcome last period triggers memories of equally high past realizations, so the forecasting error will be negatively correlated with the most recent information.[26] Second, when the baseline probability of remembering an event is low, and the rehearsal effect is strong, the forecast errors in successive periods will be positively correlated for the same reason as in as 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, and each period decides whether to buy, sell, or not trade a unit of a representative equity portfolio in net zero supply and invest the wealth net of the expenditure/revenues from the risky

[25]Both Mullainathan [2002] and Weitzman [2007] assumes that the outcomes follow an AR1 process so that a standard Bayesian would always place non-vanishing weight on the most recent outcome; our assumption of limited expected memory has the same implication even in an i.i.d. setting.

[26]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.

asset in the risk-free asset.[27] The safe asset has net return $i \in [-1, \infty)$ per period, while the risky asset provides per-period net return $i + \theta + \varepsilon_t$, where from the point of view of the agents θ is a random variable with unknown distribution and ε is symmetric, zero mean period specific shock. The risky asset is in net-zero supply and prices p_0 and p_1 are determined by market clearing:

$$p_1 - p_0 = \text{Median}[\mathbb{E}_x[\theta]].$$

In this setting, the equity premium puzzle is that if the distribution of θ were known and equal to that observed in the data, a very large amount of risk aversion would be needed to justify the observed difference in asset prices.

Weitzman [2007] explains this 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 an unbiased but selective constant memory function $m(a, y) = c$, $c \in (0, 1)$.[28] Here, the agents' actions impact their payoffs, but all agents observe the sequence of realized prices and returns. Therefore, Theorem 5, paired with an exact law of large numbers, guarantees that if the action distribution converges to α in the long run, the distribution of recalled experiences equal $\mathcal{H}_\alpha$. Therefore, even in the long run, the agents will rely on a limited number of observations. If the prior is symmetric and centered around a value $\underline{\theta} < \theta$, the pessimistic prior can sustain the premium. This is because the combination of the distribution of experiences $\mathcal{H}_\alpha$ centered around θ and the prior centered around $\underline{\theta}$ makes the median expected value of $\theta < \underline{\theta}$ under the posterior strictly smaller than θ^T.[29]

[27] Under our assumption of risk neutrality, a bang-bang solution in which all the income is invested in the same asset is without loss of optimality. We restrict to this case to be able to directly apply our results, which assume finite actions.

[28] It is easy to see that a memory function that is more likely to recall negative stock performance would create an additional force towards the equity premium puzzle.

[29] Of course, as the size of the average number of recalled events grows, the premium shrinks, just as the premium in Weitzman [2007] shrinks as the fundamental's rate of change goes to 0.

7 Conclusion

This paper has maintained the assumption that agents are completely myopic. Fudenberg, Lanzani, and Strack [2024] showed that the qualitative predictions of selective memory equilibrium extend to agents who are partially naive about their memory and believe they have a memory function $\hat{m}$ instead of their actual memory function m. We expect that limited-memory equilibrium has the same sort of qualitative robustness, but we have not checked the details. Under this extension, something that would be interesting to explore is the use of reminders to remember better, and especially the combination of this possibility with rehearsal.

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 $\alpha^*(a)p_a^*(y)$.

Lemma A.1. *For any optimal policy π and any $\alpha^* \in \Delta(A)$,*

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

Proof. Consider the stochastic processes $(\mathbf{X}_t^{(\hat{a},\hat{y})})_{(\hat{a},\hat{y})\in A\times Y, t\in\mathbb{N}}$ defined by

$$\mathbf{X}_t^{(\hat{a},\hat{y})} = (\mathbb{1}_{\{\hat{y}\}}(y_t) - p_{\hat{a}}^*(\hat{y}))\mathbb{1}_{\{\hat{a}\}}(a_t) \qquad \forall(\hat{a}, \hat{y}) \in A \times Y, \forall t \in \mathbb{N}.$$

These stochastic processes correspond to the deviations of the number of times each y has appeared from their expected frequencies given the actions chosen. The processes are measurable with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ generated by the stochastic process of histories $(h_t)_{t\in\mathbb{N}}$. They are not i.i.d., as previous outcome realizations affect current period choices, but for each $(a, y) \in A \times Y$, $\mathbb{E}[\mathbf{X}_t^{(a,y)} \mid \mathcal{F}_t] = 0$. Consequently, $(\mathbf{X}_t^{(a,y)})_{t\in\mathbb{N}}$ is a martingale difference sequence, and from the strong law of large numbers (see Theorem 2.7 in Hall and Heyde [2014] for the version that applies here) $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t^{(a,y)} = 0$, $\mathbb{P}_\pi$-a.s. And since

$$0 = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t^{(a,y)} = \lim_{n\to\infty} \sum_{t=1}^{n} \frac{\mathbb{1}_{\{a,y\}}(a_t, y_t) - p_a^*(y)\mathbb{1}_{\{a\}}(a_t)}{n} = \sum_{t=1}^{n} v_t(a, y) - \alpha_t(a)p_a^*(y),$$

we get $\lim_{t\to\infty} v_t(a, y) - \alpha^*(a)p_a^*(y) = 0$, $\mathbb{P}_\pi$-a.s. conditional on $\lim_{t\to\infty} \alpha_t = \alpha^*$. ☐

Lemma A.2. *Let $\alpha \in \Delta(A)$. For $\mathbb{P}_\pi$ 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_α^{m,μ_0}, and $F_{(\cdot)}^{m,\mu_0}$ is continuous in α.*

Proof of Lemma A.2. The convergence of the distribution of beliefs follows from Lemma 1 and the fact that the agent's posterior belief is the same for all recalled histories with the same database. Let $(\alpha_n)_{n\in\mathbb{N}} \in \Delta(A)^{\mathbb{N}}$ be a sequence converging to α^*, and fix some $\varepsilon > 0$. Let $(N_{a,y}^{\alpha^*})_{(a,y)\in A\times Y}$ be independent random variables with distributions corresponding to the marginal of $\eta_{\alpha^*}^m$ on (a, y). Since all the $N_{a,y}^{\alpha^*}$ have Poisson distributions, there is a $K \in \mathbb{N}$ such that $\mathbb{P}[\max_{(a,y)\in A\times Y} N_{a,y}^{\alpha^*} > K] < \varepsilon$. Let $M \in \mathbb{N}$ be such that $\mathbb{P}[\max_{(a,y)\in A\times Y} N_{a,y}^{\alpha_n} > K] < \varepsilon$ and $|\mathbb{P}[N_{a,y}^{\alpha_n} = c] - \mathbb{P}[N_{a,y}^{\alpha^*} = c]| < \varepsilon$ for all $(a, y) \in A\times Y$, for all $c \leqslant K$ and $n > M$. Then for any continuous and bounded $f : \Delta(\Theta) \to \mathbb{R}$, for all $n > M$ we have

$$\left| \int_{\Delta(\Theta)} f(\nu) dF_{\alpha_n}^{m,\mu_0} - \int_{\Delta(\Theta)} f(\nu) dF_\alpha^{m,\mu_0} \right| < 2 \max_{\nu\in\Delta(\Theta)} |f(\nu)|((K + 1)|A \times Y|)\varepsilon, \quad (10)$$

so $F_{\alpha_n}^{m,\mu_0}$ weakly converges to F_α^{m,μ_0}. Since the sequence was arbitrarily chosen, $F_{(\cdot)}^{m,\mu_0}$ is continuous in α. ☐

The proof of Theorem 3 applies a fixed-point theorem to the correspondence $\Psi^{m,\mu_0} : \Delta(A) \rightrightarrows \Delta(A)$ defined by $\Psi^{m,\mu_0}(\alpha) = \bigcup_{\rho\in\mathcal{O}} \psi_\rho^{m,\mu_0}(\alpha)$.

Lemma A.3.

1. *Ψ^{m,μ_0} is non-empty valued.*

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

The proof of this result is in Online Appendix B.1.

A.2 Proofs for Section 3

Proof of Lemma 1. By Lemma A.1, there is $\mathbb{P}_\pi$-probability 0 of sequences of histories $(h_t)_{t\in\mathbb{N}}$, in which $\lim_{t\to\infty} \alpha_t = \alpha$ but where we do not have $\lim_{t\to\infty} \upsilon_t(a, y) = \alpha(a)p^*_a(y)$ for all $(a, y) \in A \times Y$. Take any sequence of histories $(h_t)_{t\in\mathbb{N}}$, in which $\lim_{t\to\infty} \alpha_t = \alpha$ that is outside of that null set. We prove the stated convergence on that sequence of histories.

The database at time $t \geqslant k$ is distributed as a product of multinomial distributions: for all $d \in \mathcal{D}$

$$\mathbb{P}_\pi\left[d_t = d\right] = \prod_{(a,y)\in A\times Y} \binom{\upsilon_t(a,y)t}{d(a,y)} \left(\frac{k}{t}m(a,y)\right)^{d(a,y)} \left(\frac{t-km(a,y)}{t}\right)^{\upsilon_t(a,y)t-d(a,y)}.$$

By the Poisson limit theorem (e.g., page 15 of Loève [1977]), the probability that (a, y) is recalled $C \in \mathbb{N}$ times converges to $e^{-\lambda_{a,y}}\frac{\lambda_{a,y}^C}{C!}$, where

$$\lambda_{a,y} = \lim_{t\to\infty} \upsilon_t(a,y)\, t \left(\frac{k}{t}\left(m(a,y)\right)\right). \tag{11}$$

Thus, on this sequence of histories, the number of times (a, y) is recalled converges to a random variable Poisson distributed with parameter $\lambda_{a,y}$. □

Proof of Theorem 1. By point 5 of Lemma A.3, every fixed point of Ψ^{m,μ_0} is a limited-memory equilibrium. By points 1 and 2 of Lemma A.3 and the closed-graph theorem, Ψ^{m,μ_0} has a closed graph. The Lemma also shows it is non-empty valued and convex-valued, so it admits a fixed point by the Kakutani fixed point theorem. □

The proofs of Theorems 2 and 5 use the continuous-time approximation of the process of empirical frequencies. Set $\alpha_0 := \alpha_1$, $\tau_0 := 0$, and $\tau_t := \sum_{i=1}^{t} \frac{1}{i}$ for all $t \in \mathbb{N}$. Following Benaim, Hofbauer, and Sorin [2005], we define the continuous-time interpolation of $(\alpha_t)_{t\in\mathbb{N}}$ to be the function $w : \mathbb{R}_+ \to \Delta(A)$

$$w(\tau_t + c) = \alpha_t + c\frac{\alpha_{t+1} - \alpha_t}{\tau_{t+1} - \tau_t}, \quad \forall t \in \mathbb{N}, \forall c \in \left[0, \frac{1}{t+1}\right]. \tag{12}$$

Proof of Theorem 2. We extend the Esponda, Pouzo, and Yamamoto [2021a] application of Benaim, Hofbauer, and Sorin [2005]'s stochastic approximation techniques for differential inclusion to settings where beliefs remain stochastic in the limit. In particular, we will use the results of Benaim, Hofbauer, and Sorin [2005] to show that (12) can be approximated by a solution to

$$\dot{\boldsymbol{\alpha}}(t) \in \Psi^{m,\mu_0}(\boldsymbol{\alpha}(t)) - \boldsymbol{\alpha}(t). \tag{13}$$

A solution to (13) with initial point $x^* \in \Delta(A)$ is a mapping $x : \mathbb{R}_+ \to \Delta(A)$ that is absolutely continuous over compact intervals, with $x(0) = x^*$, and (13) satisfied for almost every t. By part 3 of Lemma A.3, a solution exists by Theorem 2.1.4 in Aubin and Cellina [2012]. For every $T \in \mathbb{N}$ and $x^* \in \Delta(A)$, let $X^T_{x^*}$ be the set of solutions to (13) over $[0, T]$ with initial conditions x^*, and let $X^T = \bigcup_{x^*\in\Delta(A)} X^T_{x^*}$.

Now we show that the continuous-time interpolation of α defined in (12) can, in the long run, be approximated arbitrarily well by a solution to (13). Define the random variable $U_t = \alpha_{t+1} - \tilde{U}_t$, where $\tilde{U}_t$ is an arbitrary element of $\operatorname{argmin}_{\alpha'\in\Psi(\alpha_t)} ||\alpha_{t+1} - \alpha'||$. Since both $\Psi^{m,\mu_0}(\alpha_t)$ and α_{t+1} are uniformly bounded, U_t is uniformly bounded. Moreover, by Lemma A.2 and the definition of $\Psi^{m,\mu_0}(\alpha_t)$, $\tilde{U}_t$ converges almost surely to 0, so condition (i) of Proposition 1.3 in Benaim, Hofbauer, and Sorin [2005] is satisfied. 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 $\Delta(A)$, so w is a perturbed solution of (13). Thus, by Theorem 4.2 in Benaim, Hofbauer, and Sorin [2005],[30]

$$\lim_{t\to\infty} \inf_{\tilde{\boldsymbol{\alpha}}\in X^T} \sup_{0\leqslant s\leqslant T} ||w(t+s) - \tilde{\boldsymbol{\alpha}}(s)|| = 0 \quad \mathbb{P}_\pi \text{ almost surely for all } T \in \mathbb{N}. \tag{14}$$

Suppose by contradiction that α is not a limited-memory equilibrium. We will show it is not a limit frequency. By parts 1,2, and 4 of Lemma A.3, the separating hyperplane theorem guarantees that there exists $f \in \mathbb{R}^A$ with $\alpha \cdot f > \max_{\bar{\alpha}\in\Psi(\alpha)} \bar{\alpha} \cdot f$. Let $K = \alpha \cdot f - \max_{\bar{\alpha}\in\Psi(\alpha)} \bar{\alpha} \cdot f$. By part 3 of Lemma A.3, there exists $\varepsilon \in \mathbb{R}_{++}$ such that for all $\alpha' \in B_\varepsilon(\alpha)$, $\max_{\bar{\alpha}\in\Psi(\alpha')} \bar{\alpha}\cdot f < \max_{\bar{\alpha}\in\Psi(\alpha)} \bar{\alpha}\cdot f + K/4$ and $\alpha' \cdot f > \alpha \cdot f - K/4$. Therefore, for every initial condition $\alpha^* \in B_\varepsilon(\alpha)$ 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 $\alpha^* \in B_\varepsilon(\alpha)$ and every solution in $X^T_{\alpha^*}$, the differential inclusion leaves $B_\varepsilon(\alpha)$ by time T, that is,[31]

$$\sup_{\tilde{\boldsymbol{\alpha}}\in X^T_{\alpha^*}} \inf\{t : \tilde{\boldsymbol{\alpha}}(t) \notin B_\varepsilon(\alpha)\} \leqslant T \qquad \forall \alpha^* \in B_\varepsilon(\alpha). \tag{15}$$

To conclude the proof, we will show that α_t does not converge to α on any path $(h_t)_{t\in\mathbb{N}}$ where (14) applies. Since the set of such sample paths has $\mathbb{P}_\pi$ probability 1, 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 $c > \tau_{\hat{T}}$, $(\alpha_t)_{t\in\mathbb{N}}$ 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 $c > \tau_{\hat{T}}$ and $\inf_{\tilde{\boldsymbol{\alpha}}\in X^T} \sup_{0\leqslant s\leqslant T} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}(s)|| < \varepsilon/4$.

Take a $\tilde{\boldsymbol{\alpha}} \in X^T$ with

$$\sup_{0\leqslant s\leqslant T} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}(s)|| < \varepsilon/4. \tag{16}$$

But then (15) implies that the differential inclusion leaves $B_\varepsilon(\alpha)$ at least once between

[30] The proof of Theorem 4.2 in Benaim, Hofbauer, and Sorin [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, Pouzo, and Yamamoto [2021b].

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

$\hat{T}$ and $\hat{T}+T$, and by (16), α_t must leave $B_{\varepsilon/2}(\alpha)$ at least once between $\hat{T}$ and $\hat{T}+T$. This proves Theorem 2. □

A.3 Proofs for Section 4

Proof of Proposition 1. We first construct the target random utility representation and then prove convergence to it. Consider the map $G : (A \times Y)^{\mathbb{N}} \to \mathcal{P}$ such that $aG(d)a'$ if and only if $a = \pi(\mu(\cdot|d) \mid \{a, a'\})$. Because the agent uses a Markov policy, the map G only depends on the agent's beliefs and not other aspects of the history, so the random utility representation ν given by $\nu(P) = \eta^m_{\alpha*}(G^{-1}(P))$ is well-defined.

Because the set of menus is finite, the claim can be established by proving pointwise convergence on each menu. Lemma A.2 shows that for every $c \in \mathbb{N}$ the probability that the number of recalled (a, y) experiences is equal to c converges to its probability under a Poisson random variable with parameter $\alpha^*(a)p^*_a(y)km(a, y)$. By Lebesgue's dominated convergence theorem with respect to the counting measures, the next period probability distribution over recalled histories converges to $\eta^m_{\alpha*}$ in L_1. This proves the distribution over databases converges, establishing the random utility representation given the assumed Markovian policy and uniform tiebreaking.

To see that the stochastic choice rule admits an Anscombe-Aumann/information representation, we map our objects to those in Lu [2016]. We identify the states S with Θ, the acts with actions, so $H = A$, and the outcomes with utility realizations $Z = u(A, Y)$ (where u is our u). We let the utility function u of Lu [2016] be the identity, with each action delivering the lottery over utility levels/outcome $\mathfrak{u}$ in state $p \in \Theta = S$ that is equal to $p_a(\{y : u(a, y) = \mathfrak{u}\})$. Given this mapping, the stochastic choice rule admits a representation with distribution over posteriors $\nu = F^{m,\mu_0}_{\alpha}$ to which convergence has been established in the first part of the proof. □

Proof of Proposition 2. Let α be a limit frequency under the original performance of action a. From the affiliation assumption, Theorem 5 in Milgrom and Weber [1982] implies that an increase in $\mathbb{E}_{p*}[y_a]$ increases all $(y_{\tau a})_\tau$ in first-order stochastic dominance. Therefore, by Lemma A.1, the limit distribution of action a payoff increases, too. Similarly, for any history $h_t = (a^t, y^t)$, the predicted expected value of action a, $\mathbb{E}_{\mu(p|h_t)}[\mathbb{E}_p[y_a]]$, increases in each entry y_{sa}. Thus, the predicted mean the agent assigns to action a increases in first-order stochastic dominance with respect to $\mathbb{E}_{p*}[y_a]$. The result follows since the data-generating process is exogenous, and an action is chosen only if it has the highest predicted mean. □

Proof of Proposition 3. Suppose first that the agent at time t remembers n_a experiences for each action. As the agent's prior is symmetric across the actions and Normal, it is optimal for the agent to choose the action with the highest *average* payoff among the experiences they remember. Since memory is unselective, when the agent remembers n_a experiences for action a, each of them has the entry corresponding to action a that is normally distributed with mean $\bar{y}_a$ and variance σ^2. Therefore, the average payoff of action a is Normally distributed with mean $\bar{y}_a$ and variance σ^2/n_a. Since the prior over alternatives is symmetric, by Lemma 1 in Natenzon [2019], the induced choice probabilities are equal to those in a Probit model.

Let $t \geqslant k$. The probability that the number of recalled instances equals some $n_a \in \mathbb{N}$ conditional on the history (a^t, y^t) is

$$\binom{t}{n_a}\left(\frac{k\alpha(a)}{t}\right)^n\left(\frac{t-k\alpha(a)}{t}\right)^{t-n}.$$

By the Poisson limit theorem (e.g., page 15 of Loève [1977]), the probability that n_a experiences are recalled converges to $e^{-\alpha(a)k}\frac{(\alpha(a)k)^n}{n!}$, as

$$\lim_{t\to\infty} t\left(\frac{\alpha(a)k}{t}\right) = \alpha(a)k. \tag{17}$$

This concludes the proof as it guarantees that the number of recalled experiences of action n_a is exponentially distributed with parameter $\alpha(a)k$. □

A.4 Proofs for Section 5

As a first step towards proving Theorem 3, we establish a deviation bound for ratios of random variables whose distributions converge to Poisson distributions. Since in this proof we let k grow, we explicitly index the distribution $\eta^m_{\alpha^k}$ by k, i.e., as $\eta^{m,k}_{\alpha^k}$.

Lemma A.4. *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* $p^*_{a'}(y') > 0$

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

Moreover,

$$\lim_{k\to\infty} \mathbb{P}_{\eta^{m,k}_{\alpha^k}} \left[n_{a',y'} - \frac{\hat{\alpha}(a')\, m(a',y')\, p^*_{a'}(y')\, k}{2} < \varepsilon \right] = 0.$$

Proof. Let $\beta \in (0,1)$, and $a \in \operatorname{supp} \hat{\alpha}$. By Chernoff's Theorem (e.g., Theorem 9.3 in Billingsley [2017])

$$\begin{aligned} & \mathbb{P}_{\eta^{m,k}_{\alpha^k}} \left[|n_{a,y} - k\alpha^k(a)\, m(a,y)\, p^*_a(y)| > \beta k \alpha^k(a)\, m(a,y)\, p^*_a(y) \right] \\ \leqslant\ & \inf_{c\in\mathbb{R}} M_{n_{a,y}}(c)\, e^{-c\beta k\alpha^k(a) m(a,y) p^*_a(y)} \end{aligned}$$

where $M_{n_{a,y}}$ is the moment generating function of the distribution associated to $n_{a,y}$. Because $n_{a,y}$ has a Poisson distribution with expected value $k\alpha^k(a)\, m(a,y)\, p^*_a(y)$,

$$\begin{aligned} & \mathbb{P}_{\eta^{m,k}_{\alpha^k}} \left[|n_{a,y} - k\alpha^k(a)\, m(a,y)\, p^*_a(y)| > \beta k\alpha^k(a)\, m(a,y)\, p^*_a(y) \right] \\ \leqslant\ & \inf_{c\in\mathbb{R}} \exp\left(k\alpha^k(a)\, m(a,y)\, p^*_a(y)\, (e^c - 1) \right) \exp\left(-c\beta k\alpha^k(a)\, m(a,y)\, p^*_a(y) \right) \\ =\ & \inf_{c\in\mathbb{R}} \exp\left((e^c - 1 + \beta c)\, k\alpha^k(a)\, m(a,y)\, p^*_a(y) \right) \\ \leqslant\ & \exp\left(c^*_\beta k\alpha^k(a)\, m(a,y)\, p^*_a(y) \right), \end{aligned}$$

where $c^*_\beta \in \mathbb{R}_{--}$ is any strictly negative real number that depend on β (but not on k) such that $\inf_{c\in\mathbb{R}} (e^c - 1 + \beta c) < c^*_\beta < 0$. (Such a number exists because for every $\beta \in (0,1)$, $\inf_{c\in\mathbb{R}} (e^c - 1 + \beta c) < (1/e - 1 - \beta) < 0$.) Taking the limit $k \to \infty$ gives

$$\lim_{k\to\infty} \mathbb{P}_{\eta^{m,k}_{\alpha^k}} \left[|n_{a,y} - k\alpha^k(a)\, m(a,y)\, p^*_a(y)| > \beta k\alpha^k(a)\, m(a,y)\, p^*_a(y) \right] = 0 \qquad \forall \beta \in (0,1). \tag{18}$$

Analogous steps show that for every $\bar{a} \notin \operatorname{supp} \hat{\alpha}(a)$ and any $\lambda > 0$ the limit as $k \to \infty$ of $\mathbb{P}_{\eta^{m,k}_{\alpha^k}}[n_{\bar{a},y} > k\lambda] = 0$.

Let $\beta \in (0,1)$ and $\beta' \in (0,1)$ be such that for all $\gamma \in [0,\beta]$ and $\gamma' \in [0,\beta']$

$$\frac{\alpha^k(a)\,m(a,y)\,p_a^*(y)\,(1+\gamma)}{\alpha^k(a')\,m(a',y')\,p_{a'}^*(y')\,(1-\gamma)} - \frac{\alpha^k(a)\,m(a,y)\,p_a^*(y)}{\alpha^k(a')\,m(a',y')\,p_{a'}^*(y')} < \varepsilon$$

and

$$\frac{\alpha^k(a)\,m(a,y)\,p_a^*(y)}{\alpha^k(a')\,m(a',y')\,p_{a'}^*(y')} - \frac{\alpha^k(a)\,m(a,y)\,p_a^*(y)\,(1-\gamma')}{\alpha^k(a')\,m(a',y')\,p_{a'}^*(y')\,(1+\gamma')} < \varepsilon.$$

Thus we have

$$\begin{aligned}
&\lim_{k\to\infty} \mathbb{P}_{\eta^{m,k}_{\alpha^k}}\left[\left|\frac{n^k_{a,y}}{n^k_{a',y'}} - \frac{\alpha^k(a)\,m(a,y)\,p_a^*(y)}{\alpha^k(a')\,m(a',y')\,p_{a'}^*(y')}\right| > \varepsilon\right] \\
\leqslant\ &\lim_{k\to\infty} \mathbb{P}_{\eta^{m,k}_{\alpha^k}}\left[\left|n^k_{a,y} - \alpha^k(a)\,m(a,y)\,p_a^*(y)\,k\right| > k\alpha^k(a)\,m(a,y)\,p_a^*(y)\min\{\beta,\beta'\}\right] \\
&+\mathbb{P}_{\eta^{m,k}_{\alpha^k}}\left[\left|n^k_{a',y'} - \alpha^k(a')\,m(a',y')\,p_{a'}^*(y')\,k\right| > k\alpha^k(a')\,m(a',y')\,p_{a'}^*(y')\min\{\beta,\beta'\}\right].
\end{aligned}$$

Equation (18) implies that the RHS goes to 0, and since α^k is converging to $\hat{\alpha}$, this proves the first part of the lemma.

The second part of the statement immediately follows by equation (18). □

Proof of Theorem 3. Let $f((n_{a,y})_{a\in A,y\in Y}) \in \Delta(A\times Y)$ denote the empirical joint distribution over action-outcome pairs corresponding to $(n_{a,y})_{a\in A,y\in Y} \in \mathbb{N}^{A\times Y}$. By Lemma A.4, for every $M \in \mathbb{N}$ and $\varepsilon > 0$,

$$\lim_{k\to\infty} \mathbb{P}_{\eta^{m,k}_{\alpha^k}}\left[(n_{a,y})_{a\in A,y\in Y} : \sum_{a\in A,y\in Y} n_{a,y} > M,\ \max_{a\in A,y\in Y} |f((n_{a,y})_{a\in A,y\in Y})(a,y) - \hat{\alpha}(a)m(a,y)p_a^*(y)| < \varepsilon\right]$$

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 $\hat{\alpha}(a)m(a,y)p_a^*(y)$. Let $\varepsilon > 0$. By Assumption 1 there exist $\varepsilon' < \varepsilon$ and $K > 0$ such that

$$\left(\sum_{a\in A}\hat{\alpha}(a)\sum_{y\in Y} m(a,y)p_a^*(y)\log p_a(y)\right) - \left(\sum_{a\in A}\hat{\alpha}(a)\sum_{y\in Y} m(a,y)p_a^*(y)\log p'_a(y)\right) > K$$

for all $p \in B_{\varepsilon'}(\Theta^m(\hat{\alpha}))$, $p' \notin B_\varepsilon(\Theta^m(\hat{\alpha}))$. Thus, there is a set of histories $(n_{a,y})_{a\in A,y\in Y})$

that has $\mathbb{P}_{\eta_{\hat{\alpha}^k}^{m,k}}$ probability going to 1, whose length $\sum_{a\in A,y\in Y} n_{a,y}$ is growing to ∞ and such that

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

Since the RHS is growing to ∞ as k grows and ε can be arbitrarily small, the agent's beliefs concentrate on $\Theta^m(\hat{\alpha})$. The fact that $\hat{\alpha}$ is a selective memory equilibrium then follows from the fact that every action $\tilde{a}$ for which $\hat{\alpha}(a) > 0$ is a best reply to a belief that assigns an arbitrary high probability to any ε ball around $\Theta^m(\hat{\alpha})$ and the upper hemicontinuity of the best reply correspondence. □

Proof of Proposition 5. Let α be a heterogeneous-beliefs selective memory equilibrium. When Θ has a product structure, the maximization in equation (7) can be done separately for each addend of the sum over $a \in A$, and when Θ is convex, there will be a unique maximizer for each $a \in \text{supp}(\alpha)$. Let $\bar{p}_a$ be the likelihood maximizer for a and let $\underline{p}_a = \text{argmin}_{q\in\Theta_a}\left(\sum_{y\in Y} u(a,y)q(y)\right)$. Define $\hat{p}_a = \bar{p}_a$ for $a \in \text{supp}(\alpha)$, $\hat{p}_a = \underline{p}_a$ for $a \notin \text{supp}(\alpha)$. By the product structure assumption, $\hat{p} \in \Theta$, and α is a unitary selective memory equilibrium with belief $\delta_{\hat{p}}$. □

Proof of Proposition 4. This proof builds on the proof of Proposition 3 in Lanzani [2024]. Consider the following two-player game. The action sets are $A_1 = \Delta(A)$, $A_2 = \Delta(\Theta)$ with arbitrary elements denoted as α, ν. The utility functions are

$$\begin{aligned} U_1(\alpha,\nu) &= \sum_{a\in A}\alpha(a)\int_\Theta \mathbb{E}_p[u(a,y)]\,\mathrm{d}\nu(p) \\ U_2(\alpha,\nu) &= \int_\Theta\sum_{a\in A}\alpha(a)\sum_{y\in Y} m(a,y)\,p_a^*(y)\log p_a(y)\,\mathrm{d}\nu(p), \end{aligned}$$

Since the compactness of Q implies that $\Delta(Q)$ is compact the action sets are compact. Moreover, they are clearly convex. The utility function U_1 is continuous in its second argument, and U_2 is continuous in its first argument, so the game is better-reply secure, and U_1 and U_2 are linear in A_1 and A_2 respectively. Therefore, by Reny [1999] Theorem 3.1, this game admits a pure-strategy equilibrium (α^*, ν^*). But $\alpha^* \in \operatorname{argmax}_{\alpha \in \Delta(A)} U_1(\alpha, \nu^*)$ implies that $\alpha^* \in \Delta(BR(\nu^*))$, and

$$\nu \in \underset{\nu \in \Delta(Q)}{\operatorname{argmax}} \int_\Theta \sum_{a \in A} \alpha(a) \sum_{y \in Y} m(a, y)\, p_a^*(y) \log p_a(y)\, \mathrm{d}\nu(p) \implies \nu^* \in \Delta(\Theta^m(\alpha)).$$

Therefore, α^* is a unitary-beliefs selective memory equilibrium. □

Proof of Corollary 2. Part 1 follows from the proof of Theorem 2 in Fudenberg, Lanzani, and Strack [2024]. For part 2, note that by Theorem 2, if $\mathbb{P}_\pi[\lim_{t\to\infty} \alpha_t = \alpha] > 0$, α is a limited-memory equilibrium, so by equation (6) the probability that the agent doesn't remember any relevant data is $\prod_{(a,y)\in A\times Y} \exp(-\alpha(a)p_a^*(y)\, k\, m(a,y)) > 0$. The result follows by choosing a prior that makes a the unique best reply. □

A.5 Proofs for Section 6

This section proves our results for the model that allows for rehearsal.

Proof of Lemma 2. Let $D' \subseteq \mathbb{N}^{A\times Y}$ denote the set of databases that can be reached with positive probability starting from the empty database under the Markov chain η_α^m. At any database $d \in D'$, the probability of a transition to the empty history is bounded below by $Q := \prod_{(a,y)\in A\times Y} \exp(-\alpha(a)p_a^*(y)\, k\, [m(a,y) + r]) > 0$, so D' is a closed communicating class. Moreover, for any $d' \in D'$, there is $\tau \in \mathbb{N}$ such that, given that the state at period $t \in \mathbb{N}$ is the empty database, the probability that the state at period $t + \tau$ is d' is some $M > 0$. Thus the expected time of return to d' is bounded from above by $\tau + \sum_{i=1}^\infty (1 - P(\text{return time} \leqslant \tau + i)) \leqslant \tau + \sum_{i=1}^\infty (1 - QM)^i \leqslant \infty$, so d' is positive recurrent. Since there is zero probability of leaving D', the Markov chain is irreducible on D', and all the states in D' are positive recurrent, η_α^m has a unique invariant distribution (see Theorem 5.5.9 in Durrett [2008]). □

Let $\Psi_{d'}(\alpha)$ denote the distributions over actions induced by an optimal Markovian mixed policy ρ and random beliefs μ:

$$\Psi^{m,\mu_0}(\alpha, d') = \bigcup_{\rho \in \mathcal{O}} \int_{\Delta(\Theta)} \rho(\nu) dF^{m,\mu_0}_{\alpha,d'}(\nu)$$

Lemma A.5.

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

The proof of this result is in Online Appendix B.1.

Proof of Theorem 4. By point 5 of Lemma A.5, every fixed point of $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ is a limited-memory equilibrium. By points 1 and 2 of Lemma A.5, and the closed-graph theorem, $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ has a closed graph. By Lemma A.5, it is also non-empty valued and convex-valued, so it has a fixed point by the Kakutani fixed point theorem. □

Proof of Theorem 5. The proof has four steps. First, Lemma A.6 characterizes the Markov chain on databases when the empirical distribution of actions and outcomes converges. Claim 1 then shows that this chain is ergodic. The third step uses stochastic approximation to show that play can only converge to a fixed point of the associated differential inclusion, as in the proof of Theorem 2. Finally, we show that if α is not an ergodic memory equilibrium, it cannot be a fixed point.

Lemma A.6. *For any* $\alpha \in \Delta(A)$ *and any sequence of histories* $(h_t)_{t\in\mathbb{N}}$ *such that* $\lim_{t\to\infty} \upsilon_t(a, y) = \alpha(a)p^*_a(y)$ *for all* $(a, y) \in A \times Y$, *the distribution of* d_{t+1} *when* $d_t = d'$ *converges to the product of independent Poisson random variables, with parameters*

$$\lambda(d')_{a,y} = \alpha(a)p^*_a(y)\, k\, (m(a, y) + r(\mathbb{1}_{\mathbb{N}\setminus 0}(d'(a, y)) + \mathbb{1}_{\{0\}}(d'(a, y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a)p^*_a(y))).$$

Moreover, the distribution of μ_t conditional on a database at time $t-1$ equal to d' weakly converges to $F^{m,\mu_0}_{\alpha,d'} \in \Delta(\Delta(\Theta))$, and $F^{m,\mu_0}_{\cdot,d'}$ is continuous in α.

The proof of this result is in Online Appendix B.1

Claim 1. *The distribution of databases converges to $\mathcal{H}_\alpha$.*

Proof. Lemma A.6 shows that the transition matrices over databases converge. Moreover, for every $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that the probability of a transition to the set of databases with K or more experiences is smaller than ε. Create a coarse state space where every database with K or more experiences and the same set of experiences with positive frequency in the database is pooled together. Transition to the null database always has positive probability in the Markov chain for the restricted process, so, in the language of Seneta [2006], the limit matrix is regular, and by Theorem 4.14 in Seneta [2006], the Markov chain converges to a stationary distribution that coincides with $\mathcal{H}_\alpha$ on the coarser history.[32] But since ε can be chosen arbitrarily small, the claim follows.

The third step of the proof uses stochastic approximation to show that the long-run behavior of (12) can be approximated by

$$\dot{\boldsymbol{\alpha}}(t) \in \mathbb{E}_{\mathcal{H}_{\boldsymbol{\alpha}_t}}[\Psi(\boldsymbol{\alpha}(t), d)] - \boldsymbol{\alpha}(t). \tag{19}$$

The last step parallels the last step of the proof of Theorem 2 with $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\alpha, d')d\mathcal{H}_{\boldsymbol{\alpha}}$ in place of $\Psi^{m,\mu_0}(\alpha)$; after observing that $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\alpha, d')d\mathcal{H}_{\boldsymbol{\alpha}}$ inherits the key properties of Ψ^{m,μ_0}, as shown by the next claim; we omit the remaining details. □

References

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

[32] Cohn [1981] and Cerreia-Vioglio, Corrao, and Lanzani [2024] prove related convergence results for finite-state inhomogeneous Markov chains.

Aubin, J.-P. and A. Cellina (2012). *Differential inclusions: set-valued maps and viability theory*. Vol. 264. Springer Science & Business Media.

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

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

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

Billingsley, P. (2017). *Probability and measure*. John Wiley & Sons.

Bordalo, P., K. Coffman, N. Gennaioli, F. Schwerter, and A. Shleifer (2021). “Memory and representativeness.” *Psychological Review*, 128, 71.

Callaway, F., A. Rangel, and T. L. Griffiths (2020). “Fixation patterns in simple choice are consistent with optimal use of cognitive resources.” *https://doi.org/10.31234/osf. io/57v6k*.

Cerreia-Vioglio, S., R. Corrao, and G. Lanzani (2024). “Dynamic opinion aggregation: Long-run stability and disagreement.” *Review of Economic Studies, forthcoming*.

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

Cohn, H. (1981). “On a paper by Doeblin on non-homogeneous Markov chains.” *Advances in Applied Probability*, 13, 388–401.

Cowan, N. (2001). “The magical number 4 in short-term memory: A reconsideration of mental storage capacity.” *Behavioral and brain sciences*, 24, 87–114.

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 (2023). “Representative sampling equilibrium.”

Durrett, R. (2008). *Probability models for DNA sequence evolution*. Springer Science & Business Media.

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 (2023). “Mental models and learning: The case of base-rate neglect.”

Fudenberg, D., G. Lanzani, and P. Strack (2021). “Limit points of endogenous misspecified learning.” *Econometrica*, 89, 1065–1098.

— (2024). “Selective Memory Equilibrium.” *Journal of Political Economy, forthcoming.*

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.

Gödker, K., P. Jiao, and P. Smeets (2022). “Investor memory.”

Gonçalves, D. (2023). “Sequential Sampling Equilibrium.” arXiv: 2212.07725.

Greene, W. H. (2000). “Econometric analysis 4th edition.” *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*, 403–426.

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

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.

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.

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

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

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

Lanzani, G. (2024). "Dynamic Concern for Misspecification."

Lee, W. (1971). *Decision theory and human behavior.* John Wiley & Sons.

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

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

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

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review*, 63, 81.

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*, 834–847.

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.

Reny, P. J. (1999). "On the existence of pure and mixed strategy Nash equilibria in discontinuous games." *Econometrica*, 67, 1029–1056.

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

Schacter, D. L. (2008). *Searching for memory: The brain, the mind, and the past.* Basic Books.

Schervish, M. J. (2012). *Theory of statistics.* Springer Science & Business Media.

Seneta, E. (2006). *Non-negative matrices and Markov chains.* Springer Science & Business Media.

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 (2023). "Biased Memory and Perceptions of Self-Control."

Strzalecki, T. (2023). "Stochastic Choice Theory." *preprint.*

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.

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.

Zimmermann, F. (2020). "The dynamics of motivated beliefs." *American Economic Review*, 110, 337–61.

B Online Appendix

B.1 Omitted Proofs

Lemma A.3. 1. Ψ^{m,μ_0} is non-empty valued.

2. Ψ^{m,μ_0} is closed valued;

3. Ψ^{m,μ_0} is upper hemicontinuous;

4. Ψ^{m,μ_0} is convex valued;

5. $\alpha' \in \Delta(A)$ is a limited-memory equilibrium if and only if $\alpha' \in \Psi(\alpha')$.

Proof.

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, $\cup_{\rho \in O} \rho(\nu)$ is closed for every $\nu \in \Delta(\Theta)$, and $\Psi^{m,\mu_0}(\alpha)$ is the Aumann integral [Aumann, 1965] of the (mixed) best reply correspondence with respect to the distribution of beliefs F_α^{m,μ_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, $F_{(\cdot)}^{m,\mu_0}$ is continuous in α, and so by Artstein and Wets [1988], Theorem 4.2, Ψ^{m,μ_0} is upper hemicontinuous.

4. This follows immediately from the definition of Ψ^{m,μ_0}.

5. This follows immediately from the definition of limited-memory equilibrium.

□

Lemma A.6. For any $\alpha \in \Delta(A)$ and any sequence of histories $(h_t)_{t \in \mathbb{N}}$ such that $\lim_{t \to \infty} v_t(a, y) = \alpha(a) p_a^*(y)$ for all $(a, y) \in A \times Y$, the distribution of d_{t+1} when $d_t = d'$ converges to the product of independent Poisson random variables, with parameters

$$\lambda(d')_{a,y} = \alpha(a) p_a^*(y)\, k\left(m(a, y) + r(\mathbb{1}_{\mathbb{N} \setminus 0}(d'(a, y)) + \mathbb{1}_{\{0\}}(d'(a, y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a) p_a^*(y))\right).$$

Moreover, the distribution of μ_t conditional on a database at time $t-1$ equal to d' weakly converges to $F^{m,\mu_0}_{\alpha,d'} \in \Delta(\Delta(\Theta))$, and $F^{m,\mu_0}_{\cdot,d'}$ is continuous in α.

Proof. Given a database d' recalled in period $t-1$, the database at time $t \geqslant k$ is distributed as a product of multinomial distributions:

$$\mathbb{P}_t\left[d|d'\right] = \prod_{(a,y)\in A\times Y} \binom{\upsilon_t(a,y)t}{d(a,y)}$$
$$\times \left(\frac{k}{t}(m(a,y) + r(\mathbb{1}_{\mathbb{N}\setminus 0}(d'(a,y)) + \mathbb{1}_{\{0\}}(d'(a,y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a)p^*_a(y)))\right)^{d(a,y)}$$
$$\times \left(1 - \frac{k}{t}(m(a,y) + r(\mathbb{1}_{\mathbb{N}\setminus 0}(d'(a,y)) + \mathbb{1}_{\{0\}}(d'(a,y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a)p^*_a(y)))\right)^{\upsilon_t(a,y)t-d(a,y)}.$$

By the Poisson limit theorem (e.g., page 15 of Loève [1977]), the probability that (a,y) is recalled $C \in \mathbb{N}$ times when the previous database was d' converges to $e^{-\lambda(d')_{a,y}}\frac{\lambda(d')^C_{a,y}}{C!}$, where

$$\begin{aligned}\lambda(d')_{a,y} &= \lim_{t\to\infty} \upsilon_t(a,y)t\left(\frac{k}{t}(m(a,y) + r(\mathbb{1}_{\mathbb{N}\setminus 0}(d'(a,y)) + \mathbb{1}_{\{0\}}(d'(a,y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a)p^*_a(y))))\right)\\ &= \alpha(a)p^*_a(y)\,k\,(m(a,y) + r(\mathbb{1}_{\mathbb{N}\setminus 0}(d'(a,y)) + \mathbb{1}_{\{0\}}(d'(a,y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a)p^*_a(y)))\,.\end{aligned} \tag{20}$$

Thus the random number of times (a,y) is recalled conditional on υ_t and the previous database being d' converges to a random variable $N^{\alpha}_{a,y}(d')$ that is Poisson distributed with parameter $\lambda(d')_{a,y}$. Moreover, let $(\alpha_n)_{n\in\mathbb{N}} \in \Delta(A)$ be a sequence converging to α^*, and fix some $\varepsilon > 0$. Since all the $N^{\alpha^*}_{a,y}(d')$ have Poisson distributions, there is a $K \in \mathbb{N}$ such that

$$\mathbb{P}[\max_{(a,y)\in A\times Y} N^{\alpha^*}_{a,y}(d') > K] < \varepsilon.$$

Let $M \in \mathbb{N}$ be such that $\mathbb{P}[\max_{(a,y)\in A\times Y} N^{\alpha_n}_{a,y}(d') > K] < \varepsilon$ and $|\mathbb{P}[N^{\alpha_n}_{a,y}(d') = c] - \mathbb{P}[N^{\alpha^*}_{a,y}(d') = c]| < \varepsilon$ for all $(a,y) \in A\times Y$, for all $c \leqslant K$ and $n > M$. Then for any continuous and bounded $f : \Delta(\Theta) \to \mathbb{R}$, for all $n > M$ we have

$$\left| \int_{\Delta(\Theta)} f(\nu) dF^{m,\mu_0}_{\alpha_n,d'} - \int_{\Delta(\Theta)} f(\nu) dF^{m,\mu_0}_{\alpha,d'} \right| < \max_{\nu \in \Delta(\Theta)} |f(\nu)|((K+1)|A \times Y|)\varepsilon, \qquad (21)$$

so $F^{m,\mu_0}_{\alpha_n,d'}$ weakly converges to $F^{m,\mu_0}_{\alpha,d'}$. □

Lemma A.5.

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

Proof. First, observe that $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ is the integral of the mixed best reply correspondence with respect to the measure $\int_{\mathcal{D}} F^{m,\mu_0}_{\alpha,d'} d\mathcal{H}_{\boldsymbol{\alpha}'}(d')$.

1. Follows from the finiteness of A.

2. Follows from the finite dimensionality of $\Delta(A)$ and Theorem 2.1.37, case (i) of Molchanov [2017].

3. By Lemma A.6, $F^{m,\mu_0}_{(\cdot)}$ is continuous in α. Moreover, since the stationary distribution is continuous in the entries of the corresponding Markov chains on the set of matrices that admit a unique station distribution, $\mathcal{H}_{(\cdot)}(d')$ is continuous in and so is $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$. Therefore, by Artstein and Wets [1988], Theorem 4.2, Ψ^{m,μ_0} is upper hemicontinuous.

4. Immediate from the definition of $\int_{\mathcal{D}} \Psi^{m,\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$.

5. Immediate from the definition of ergodic memory equilibrium and Ψ^{m,μ_0}.

□

B.2 Exponential Signals

Suppose that the payoffs y_a of the $a \in A$ are i.i.d. exponentially distributed with mean $\bar{y}_a$, and that the agent's prior is that $(\bar{y}_1, \ldots, \bar{y}_{|A|})$ are independently inverse Gamma distributed with shape parameter α_0 and rate parameter β_0. The posterior belief after l observations is then again inverse Gamma distributed with shape parameter $\alpha_0 + l$ and rate parameter $\beta_0 + \sum_{r=1}^{l} y_{ra}$. As this inverse Gamma distribution has mean $\frac{\beta_0 + \sum_{\tau=1}^{l} y_{\tau a}}{\alpha_0 + l - 1}$,[33] the posterior mean is monotonically increasing in the average empirical payoff, so it is optimal to choose the action for the empirical average is highest. Note that for any two actions $a \neq a'$, the probability that action a is chosen over a' thus equals

$$\mathbb{P}\left[\sum_{\tau=1}^{l} y_{\tau a'} \leqslant \sum_{\tau=1}^{l} y_{\tau a}\right].$$

We have that in the case of $l = 1$ and two alternatives, the choice probability replicates those of the Luce model

$$\mathbb{P}\left[y_{\tau a'} \leqslant y_{\tau a}\right] = \frac{1/\bar{y}_{a'}}{1/\bar{y}_{a'} + 1/\bar{y}_a} = \frac{\bar{y}_a}{\bar{y}_a + \bar{y}_{a'}}.$$

We note that the sum of l i.i.d. exponential random variables with mean $\bar{y}_a$ is Erlang distributed with shape parameter l and rate $1/\bar{y}_a$. Thus, the probability that for l remembered experiences the action a has a higher average outcome than the action a' is thus given by

$$\begin{aligned}\mathbb{P}\left[\sum_{\tau=1}^{l} y_{\tau a'} \leqslant \sum_{\tau=1}^{l} y_{\tau a}\right] &= 1 - \frac{(2l)!}{(l!)^2}\left|\beta_{-\bar{y}_{a'}/\bar{y}_a}(l, 1-2l)\right| \\ &= 1 - \frac{(2l)!}{(l!)^2}\int_{-\bar{y}_{a'}/\bar{y}_a}^{0} |t|^{l-1}(1-t)^{-2l}dt.\end{aligned}$$

[33] See page 670 of Schervish [2012].

For $l = 2$, we get that

$$\mathbb{P}\left[\sum_{\tau=1}^{l} y_{\tau a'} \leqslant \sum_{\tau=1}^{l} y_{\tau a}\right] = \left(\frac{\bar{y}_a}{\bar{y}_a + \bar{y}_{a'}}\right)^3 \left(1 + 3\frac{\bar{y}_{a'}}{\bar{y}_a}\right).$$

B.3 Multiple Limited Memory Equilibria

Example 2. *Suppose that* $A = \{0,1\}$ *and* $u(0,y) = \frac{2}{3}$, $u(1,y) = y$ *where* $Y = \{0,1\}$. *Let* $p_1^*(1) = 0.9$ *and* $k = 2$, *with the prior about the probability of* 1 *under action* 1 *beta* $(1,2)$. *There are two equilibria,* α', α'' *with* $\alpha'(0) = 1$ *and* $\alpha''(1) = 0.45$, *where the second fixed point is found using the Mathematica program available at* https://www.dropbox.com/scl/fi/w1tzstdynepbd0nz7nnnj/multipleNew.nb?rlkey=5dnjsi2me6injxs62m79n9hl6&dl=0.