Limited Memory, Learning, and Stochastic Choice*

Drew Fudenberg† Giacomo Lanzani‡
Philipp Strack§

October 22, 2025

Abstract

We characterize the long-run behavior of agents who recall only a random subset of their past experiences. When empirical frequencies of actions converge, the limit must be a *stochastic memory equilibrium*. Limited memory rationalizes the stochastic choices of random utility models and explains the effects of reminders on behavior and underreaction to large samples. Allowing for recency and rehearsal effects, along with limited memory, explains correlated prediction errors in forecasts of returns and the equity premium puzzle.

JEL codes: D83, D90

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

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

1 Introduction

People typically remember only a relatively limited number of their past experiences, and what they remember is stochastic: they might recall some things one week and different things the next. For example, when someone is deciding whether to go to the gym, they may recall only a few of the experiences they have had there. Similarly, when deciding between promoting a more skilled subordinate or a harder-working one, a manager may rely on past experiences in which creativity or grit was required. In both cases, the experiences that come to mind can change over time and can depend on the decision maker's past behavior.

We develop a tractable model of limited and stochastic memory that lets us analyze its long-run implications and relate them to many well-established behavioral regularities, such as the stochasticity of choice, the importance of reminders, the effect of skewness in the evaluation of alternatives, underreaction to large samples, and some forms of the equity premium puzzle. In our model, agents are naive: they update their beliefs as if the experiences they remember were the only ones that occurred, and they maximize their expected utility given these beliefs. Because agents only recall a small subset of their experiences, their beliefs and behavior remain stochastic as the number of their experiences goes to infinity. We say that a distribution of actions is a *stochastic 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 stochastic memory equilibrium.

We show that limited memory generates the stochastic choices of a random utility model and links them to the objective environment. In particular, the equilibrium choice probabilities are consistent with Lu [2016]'s information representation of stochastic choice rules. In binary-choice problems where the agent observes a signal about the quality of the chosen action that is affiliated with their prior, the induced stochastic choice rule is monotone in the sense that actions with higher utility are chosen more often. Thus, although limit behavior is stochastic, the environment disciplines the errors, making more costly mistakes less likely.

Moreover, with specific outcome distributions, stochastic memory equilibrium provides a learning foundation for some widely used stochastic choice models. With normally distributed outcomes and a normal prior, stochastic memory equilibrium generates a mixed probit random utility model, where the variance-covariance matrix of the resulting probit accommodates both payoff monotonicity and diminishing sensitivity, as in baseline probit, and it also captures frequency dependence: less frequently chosen actions have noisier perceived values. If instead actions correspond to the choice of vectors of desirable features, and outcomes correspond to situations in which those features proved helpful, we show that the limit frequency is that of 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 valuable. We then illustrate another way in which limited memory links stochastic choice with the objective environment: it leads agents to underweight rare events when probabilities are learned rather than given, as found by Hertwig, Barron, Weber, and Erev [2004]. This underweighting biases the action distribution towards actions with negatively skewed utilities.

In addition to linking random utility models with the objective environment, our limited memory model captures the fact that reminders about past experiences can induce more use of a beneficial action, such as going to the gym, and more so for those who otherwise take the action less frequently (Calzolari and Nardotto [2017] and Augenblick, Jack, Kaur, Masiye, and Swanson [2024]). Moreover, sample-size insensitivity (Kahneman and Tversky [1972]) and underreaction to signals, which are usually attributed to "underinference" from perfectly recalled events (Phillips and Edwards [1966]), can alternatively be explained by limited memory. Moreover, our model, unlike underinference, explains why people perceive uncertainty even when they have very large samples, as found by Kahneman and Tversky [1972].

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

We then expand the model to capture "rehearsal" and "recency" by allowing experiences that occurred or were remembered in the previous period to be more likely to be recalled in the next. This addition makes the model more realistic and lets it fit evidence about the importance of rehearsal and recency. We find that a generalization of our baseline fixed-point condition characterizes the limits of the action distribution: beliefs conditional on the objective history are autocorrelated rather than i.i.d., and the limit action distribution must be consistent with the stationary distribution of the Markov chain of beliefs it induces. Action distributions with this property, which we call *ergodic memory equilibrium*, let us extend Mullainathan [2002]'s analysis of the effect of rehearsal on income forecasts from short-run predictions to the long run and more general functional forms. It also allows us to 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 The psychology literature surveyed by Gershman, Fiete, and Irie [2025] shows that, as assumed by our model, information retrieval rather than information storage is the primary memory bottleneck. Information retrieval has been described informally as stochastic since the early stages of the psychology literature, and the limited nature of memory has been documented at least since Miller [1956]. See Kahana [2012] for a recent textbook exposition. Shadlen and Shohamy [2016] and Sial, Sydnor, and Taubinsky [2024] provide more recent evidence of stochastic memory. d'Acremont, Schultz, and Bossaerts [2013] provides fMRI evidence that the brain maintains a neural representation of accumulated frequencies in memory-related regions, which are then combined with prior knowledge when forming posterior beliefs. This motivates modeling belief updating as operating on retrieved samples of past evidence. Reder [2014] and Duncan and Shohamy [2020] provide evidence of partial or complete unawareness of memory limitations. 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.

In a stochastic memory equilibrium, the agent's 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 an optimal stopping problem in which the agent wants to match a binary action to 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 the analysis of optimal stopping to settings where the agent is uncertain of the payoff difference between the actions, Che and Mierendorff [2019] further extends it to more general signal structures, and Hébert and Woodford [2023] replaces specific signals with a bound on the rate of information flow. Matějka and McKay [2015] and Ke and Villas-Boas [2019] also connect stochastic choice to the objective environment in models in which agents first obtain information and then make a single decision. Lu [2016] and Natenzon [2019] axiomatize stochastic choice due to Bayesian updating; there, the distribution and number of signals are exogenous.

In Osborne and Rubinstein [1998], Wilson [2014], Jehiel and Steiner [2020], and Salant and Cherry [2020], the agent gets an exogenous signal and chooses a single action. Crucially, unlike us, these papers do not allow for feedback between imperfect memory and the agent's data.[1] Danenberg and Spiegler [2025] study the equilibrium action distribution when the precision of the signals about the payoff to each action depends on the probability the action is played, but does not model the agent's period-by-period decisions and information. Gonçalves [2023] defines an equilibrium concept for agents who make a single decision based on optimal sequential sampling from the equilibrium distribution. In all of these papers, aggregate behavior is assumed to be described by a fixed-point condition, as opposed to our approach of proving that a specific fixed-point condition describes the possible limit frequencies.

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

The baseline version of our model, without rehearsal, can be interpreted as a social learning model in which, at each period, different agents draw a subset of past agents' experiences and make a decision. Under this interpretation, the closest paper is Banerjee and Fudenberg [2004], but that paper differs in many ways. For example, there agents do make inferences from the prevalence of each action in their database, and later agents do

[1] Koszegi, Loewenstein, and Murooka [2021] considers a different form of circularity in the memory process, allowing for negative memories to be more frequently recalled when holding more negative beliefs, which induces multiple equilibria with different sustaining beliefs.

not observe the private signals of their predecessors.[2]

There is an extensive psychology literature documenting various forms of the recency effect and proposing explanations for it, see, e.g., the summaries in Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, and Usher [2005] and Erev and Haruvy [2016]. There is also extensive evidence of the importance of rehearsal; see, e.g., the textbook by Kandel et al. [2000]. Mullainathan [2002] analyzes the short-run implications of rehearsal in a specific parametric context but does not study its long-run effects. Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [2023] links limited memory to recurrent errors in inference, while Bordalo, Gennaioli, Lanzani, and Shleifer [2025] shows how memory can impact choice through matching of the current problems with a category of similar past experiences.

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.

The agent knows that the map from actions to probability distributions over outcomes is fixed, but is uncertain about the outcome distributions each action induces. We suppose that 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 that the agent initially thinks are possible. We maintain the following assumption throughout:

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

We assume correct specification to highlight the issues that arise solely from limited memory, but this is not necessary; our results generalize to the case where the agent is misspecified.

Histories, Memory, and Recalled Periods We call action-outcome pairs $(a, y) \in A \times Y$ *experiences*. A period $t \in \mathbb{N}$ *history* is a sequence $h_t = (A \times Y)^t$. We assume that the probability the agent remembers any given past experience at the beginning of period $t + 1$

[2] Wolitzky [2018] extends Banerjee and Fudenberg [2004] to allow for the possibility that only outcomes, but not actions, are observed by the subsequent agents.

is $m_{t+1} = \min\{1, k/t\}$, with $k > 0$. As we will see, k is the expected number of experiences the agent recalls when their sample is large.

After history $h_t = (a_i, y_i)_{i=1}^t$, the *recalled periods* r_t are a random subset of $\{1, ..., t\}$. We assume for now that each past experience has an independent probability of being recalled, so[3]

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

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

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,[4] and that the agent is unaware of their limited memory and naïvely updates their beliefs as if the experiences they remember are the only ones that have occurred. Thus, the agent's beliefs only depend on the number of times each (a, y) pair is recalled, which we call the agent's *database*. We let $\mathcal{D} = \mathbb{N}^{A\times Y}$ denote the set of these databases, so that the posterior probability of any (measurable) $C \subseteq \Theta$ after database $d \in \mathcal{D}$ is[5]

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

Optimal Policies The agent is myopic, with utility function $u : A \times Y \to \mathbb{R}$.[6] Let

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

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

[4]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.

[5]If the agent believes their utility function is subject to very unlikely random shocks that can induce each action as the best reply, and dogmatically believes they recall every experience, they attribute unexplained behavior to the utility shock so Bayes rule coincides with equation (2). Heidhues, Kőszegi, and Strack [2023] study the case where the agent forgets their preference shocks and tries to infer them from their own actions.

[6]Thus, the agent does not consider how what they might learn today would help them make future decisions, and does not consider giving themselves reminders or deliberately distorting their memory manipulations as in Bénabou and Tirole [2002] and following work.

denote the actions that maximize the agent's current period expected utility when their belief is $\nu \in \Delta(\Theta)$.[7] A *Markovian policy* $\pi : \Delta(\Theta) \to A$ is a measurable function that specifies a pure action for every belief. The agent uses an *optimal Markovian policy* π, i.e., for every $\nu \in \Delta(\Theta)$, $\pi(\nu) \in BR(\nu)$. Together, a true data generating process p^* and a Markovian policy function uniquely induce a probability measure over histories, denoted as $\mathbb{P}_\pi$.

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

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

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

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

The definition of limit frequency requires that for every ε ball around α, there is a strictly positive probability that the empirical frequency eventually lies in that ball.[8]

3 Stochastic Memory Equilibrium

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

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

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

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

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

This is a product distribution where the marginal distribution for each action-outcome pair (a, y) is Poisson with mean $k\alpha(a)p_a^*(y)$. We will show that η_α is the limit distribution of databases if the action frequency converges 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 the average memory capacity k.

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

To prove the lemma, Lemma A.1 in Section A.1 uses a law of large numbers for martingale differences to show that the joint frequency of each action and outcome pair (a, y) converges to $\alpha_t(a)p_a^*(y)$; 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.[9]

Limit Distribution of Beliefs The second step is to associate the candidate action distribution with the distribution of beliefs that it induces. Let $F_\alpha^{\mu_0}$ be the distribution of beliefs induced by the distribution η_α of databases and prior μ_0, i.e., for all measurable $\mathcal{C} \subseteq \Delta(\Theta)$,

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

Let $\mathcal{O}$ denote the set of measurable selections from the (mixed) best reply correspondence: i.e., $\rho : \Delta(\Theta) \to \Delta(A)$ is in $\mathcal{O}$ if and only if ρ is measurable and $\rho(\nu) \in \Delta(BR(\nu))$ for all $\nu \in \Delta(\Theta)$. For any $\rho \in \mathcal{O}$, let $\psi_\rho^{\mu_0}$ be the function that maps $\alpha \in \Delta(A)$ to the action distribution generated when the agent uses policy ρ and their beliefs are distributed according to $F_\alpha^{\mu_0}$, as illustrated in Figure 1. Formally,

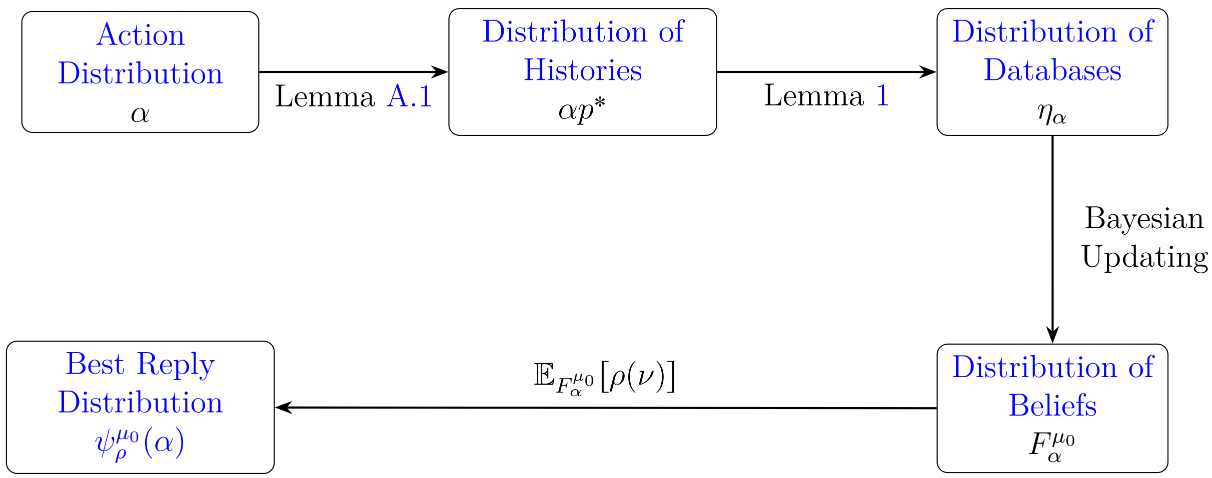
$$\psi_\rho^{\mu_0}(\alpha) = \int_{\Delta(\Theta)} \rho(\nu) dF_\alpha^{\mu_0}(\nu). \tag{4}$$

In words, if action distribution α is played forever, it induces distribution αp^* over histories. This distribution and the memory function m together induce a distribution of databases η_α, and Bayesian updating on each database generates distribution $F_\alpha^{\mu_0}$ over

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

posterior beliefs. Assigning $\rho(\nu)$ to each posterior belief ν generates action distribution $\psi_\rho^{\mu_0}(\alpha)$.

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



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

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

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

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

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

The first step of the proof is the characterization of the limit databases in Lemma 1. The second step of the proof uses the Benaim, Hofbauer, and Sorin [2005] extension of

[10]Example 1 in the Online Appendix shows there can be multiple stochastic memory equilibria; we do not know whether there can be multiple limit frequencies.

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 $F_{\alpha}^{\mu_0}$ (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.

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

The following theorem shows that a stochastic memory equilibrium exists even when there are no limit frequencies.

Theorem 2. *A stochastic memory equilibrium exists.*

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

4 Applications

4.1 Random Utility and Stochastic Choice

This section relates limited memory to the most widely used model of non-deterministic behavior in single-agent problems, namely 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 stochastic memory equilibrium, the agent's beliefs about the expected utility of each action are determined by their random memories. Connecting these concepts helps motivate 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 linear orders on A. A stochastic choice function c has a *random utility representation* $\zeta \in \Delta(\mathcal{P})$ if for all $M \in \mathcal{M}$,[11]

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

To relate our model of memory and learning to random utility and stochastic choice, suppose that at some single time t, an experimenter elicits the agent's choice distribution on each menu, i.e., each subset of A. (As usual, this requires many observations of choice from each menu.) We assume that when confronted with one of these menus, the decision maker breaks ties deterministically in a menu-independent way.[12]

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

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

The next result shows that when the distribution of the agent's actions converges, the agent's menu choices converge to a limit that has a random utility representation. Moreover, the limit empirical distribution of action coincides with the choice distribution induced by that random utility model on the complete action set, and is consistent with Lu [2016]'s information representation.[13] Together, Proposition 1 and Theorem 1 show how the information structure induced by η_α determines the choice probabilities.

Proposition 1. *For every optimal Markovian policy* π *and* $\alpha^* \in \Delta(A)$*, and on* $\mathbb{P}_\pi$*-every sequence of histories such that* $\lim_{t \to \infty} \alpha_t = \alpha^*$*, the observed distribution of actions approaches a random utility representation* ζ *with* $\alpha^*(a) = c_\zeta(a, A)$ *for all* $a \in A$*, and is consistent with Lu [2016]'s information representation.*

All proofs for this section are in Appendix A.3. The proof of Proposition 1 first constructs the target random utility representation. To do this, we associate to every database

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

[12] Formally, we require that for all $M, M' \in \mathcal{M}$ if $a, a' \in BR(\nu|M) \cap M'$ and $\pi(\nu|M) = a$, then $\pi(\nu|M') \neq a'$, where for any $M \subseteq A$, $BR(\nu|M) = \text{argmax}_{a \in M} \int_\Theta \sum_{y \in Y} u(a, y) p_a(y) d\nu(p)$.

[13] In an information representation of stochastic choice, the decision maker maximizes the expectation of a fixed utility function with respect to a distribution over posteriors over states.

d a strict (i.e., antisymmetric) 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 $\mu(\cdot|d)$.[14] The random utility representation is then defined by assigning each ranking the limit probability of the databases that induce it. Since Lemma A.2 guarantees that the distribution over databases converges and the set of menus is finite, this pushforward measure also converges.

4.1.1 Monotonicity

The prior belief of the agent induces a (subjective) joint distribution over the pairs $(\mathbb{E}_p[y_a], y_a)$.[15] The next result shows that in binary actions settings $A = \{a, a'\}$, actions with higher payoffs are played more frequently if these joint distributions are affiliated.[16] For simplicity, we also make the (generically satisfied) assumption that two actions are never indifferent after any observable database.

Proposition 2 (Monotonicity)**.** *Suppose that $\mathbb{E}_p[y_a]$ and y_a are affiliated under the prior. If the objective distribution of y_a increases in the sense of first-order stochastic dominance, keeping fixed the objective distribution of $y_{a'}$, the frequency of a in the stochastic memory equilibria with the highest and lowest values of $\alpha(a)$ both increase.*

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

4.1.2 Specific Signal Structures

For some signal structures, the fixed-point condition defining stochastic memory equilibrium admits an explicit solution, and the equilibria correspond to important stochastic choice models. This section gives two tractable examples, the normal and binomial cases. The

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

[15] The joint probability assigned by the prior to any measurable $C \subseteq \mathbb{R}$ and $c \in \{y_a : y \in Y\}$ is $\int_{\{p:\mathbb{E}_p[y_a]\in C\}} p(\{y \in Y : y_a = c\})d\mu(p)$.

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

section also connects stochastic memory equilibria with the Probit model (also in the normal environment) and the Elimination by Aspects model (Tversky [1972]).

Probit The Probit model (Thurstone [1927]) is a stochastic choice model with two desirable features: payoff monotonicity and diminishing sensitivity.[17] With normal signals, our learning model delivers the related *mixed probit* (Hausman and Wise [1978], Greene [2000]) specification, which also has these two properties.[18] Moreover, the form of mixed probit we obtain also implies that the agent has more precise estimates of the values of actions they take more frequently, as found by Frydman and Jin [2022]. As Strzalecki [2023] (Example 1.15) points out, this frequency dependence is not accommodated by baseline probit.

In a *normal environment*, the payoffs y_a of each action a are i.i.d. normally distributed with means $\bar{y}_a$ and variance σ^2. The agent knows σ^2, and their prior belief is that that $(\bar{y}_1, \ldots, \bar{y}_{|A|})$ are independently normally distributed with mean 0 and variance σ_0^2.[19]

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

In the mixed probit specification generated by our model, the variance σ^2/n_a of the payoff to a is stochastically decreasing (i.e., in the sense of first-order stochastic dominance) in $\alpha(a)$. Hence, the payoffs of more frequently chosen alternatives are more precisely estimated.[20]

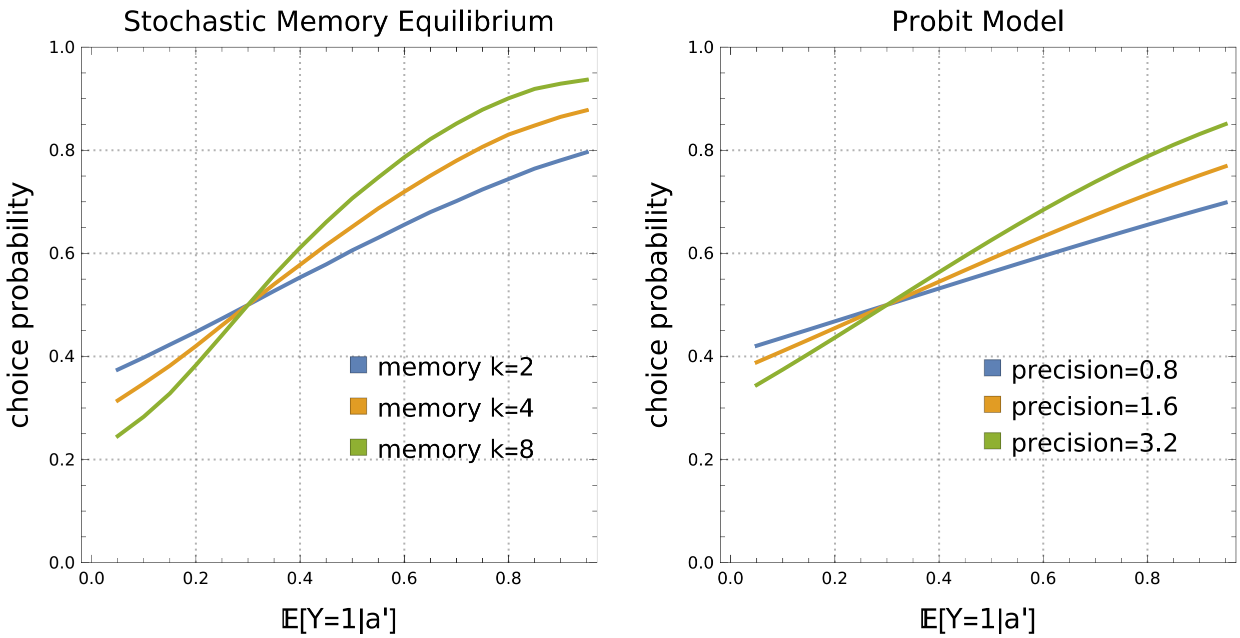
[17] Payoff monotonicity means that more preferable alternatives are more likely to be chosen. Probit satisfies payoff monotonicity when what is subject to the normal shock is the payoff of the alternatives; Thurstone [1927]'s original formulation was a bit different and doesn't necessarily satisfy payoff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

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

[19] Here we allow infinitely many outcomes, but our definitions extend in the obvious way.

[20] Danenberg and Spiegler [2025] makes a similar point about the relation between frequency and precision in a setting with exogenous normally distributed signals and no memory limitations.

Figure 2: **Comparison Limited Memory and exogenous precision probit** Binomial Beta Probability of choosing a' when $A = \{a', a''\}$, $Y = \{0, 1\}$, a symmetric beta prior with $\beta = \gamma = 1$, $k = 2$ (blue), $k = 4$ (orange), $k = 8$ (green), and a'' is known to give outcome 1 with probability .3.



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

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

Assume also that $u(a, y) = y$. A stochastic memory equilibrium is then a solution α to the equation

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

where the expectation is taken with respect to the η_α given by Lemma 1, and we assume that the agent uniformly randomizes over actions when they are indifferent.

With endogenous data, the agent has more precise estimates of the payoff of actions that are more frequently chosen. This effect can be seen in Figure 2. Its left panel contains three different expected numbers of recalled experiences that mimic behavior under three different precisions in the probit model. The effect can also be directly observed from equation (5), which can be rearranged as

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

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

Elimination by Aspects Elimination by aspects (Tversky [1972]) postulates that the agent uses "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 only a single alternative is left.

EBA was designed to capture the following observed violation of IIA: Starting from a situation with two alternatives $\{a, a'\}$, the addition of a third alternative that is closer to a, without dominating or being dominated 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 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 remove significantly more choice probability from $(75, 35)$.

EBA is a random utility model. Because it does not have an axiomatic foundation, it seems particularly useful to give it a foundation based on how people remember and process their information. To see how the memory foundation works in the example above, suppose that outcomes are tasks in which either intelligence or motivation is the key feature driving the hired worker's performance. The manager receives a payoff equal to the worker's skill in the dimension (intelligence or motivation) that is relevant in the current period. Further, the manager is uncertain about the probability that each skill is relevant. In particular, the prior is 50-50 on two DGPs: either there is probability .9 that motivation is the relevant factor or probability .9 that success only depends on the worker's intelligence. At the end of the period, the manager observes which factor was relevant for this period's task. Then, in every limit frequency, the probability that $(75, 35)$ is chosen over $(60, 90)$ equals the probability that the manager recalls more past periods in which intelligence was important. And, as predicted by EBA, the addition of $(78, 25)$ will reduce the probability of $(75, 35)$ but not that of $(60, 90)$.[21]

Note that our model predicts that more frequently relevant aspects are more likely to be used to determine choice, while in Tversky [1972] these probabilities are not restricted. Note also that the relation between our model and EBA is more general than in the example. To see this formally, define an *aspects environment* as one where $A \subseteq \{0,1\}^{\mathcal{A}}$ for some finite set of aspects $\mathcal{A}$, $Y = \mathcal{A}$, $u(a, y) = a_y$, for every $p \in \Theta$, $a, a' \in A$, $p_a = p_{a'}$ and such that the prior is *responsive*: for every database $d(A, y) > d(A, y')$ and $a_y > a'_y$ implies that $\int_\Theta \mathbb{E}_p[u(a, \cdot)]d\mu(p|d) > \int_\Theta \mathbb{E}_p[u(a', \cdot)]d\mu(p|d)$.[22] The interpretation is that actions with more 1's have more desirable features, and the outcome is the period's most important feature.[23]

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

4.2 The Effect of Skewness

Our next example illustrates how limited memory leads agents to frequently overlook the possibility of rare events so that the probability that an action is chosen is not determined by its expected payoff but can also depend on higher-level moments of the utility function, such as variance and skewness. Suppose that the agent chooses between two actions. Action

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

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

[23] Following Tversky [1972], we consider a model 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 $0-1$ vector, where 1 indicates an 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 when outcomes are enriched to encode noisy signals about each attribute's level and whether it is currently relevant.

0 always produces the outcome $c \in (0,1)$, e.g., $\mu(\{p : p_0(c) = 1\}) = 1$. The agent believes two possible data generating processes exist for action 1, p and p'. Under p, which has prior probability q, action 1 generates outcome $1/b$ with probability b and 0 otherwise. Under p', it always generates outcome 0.

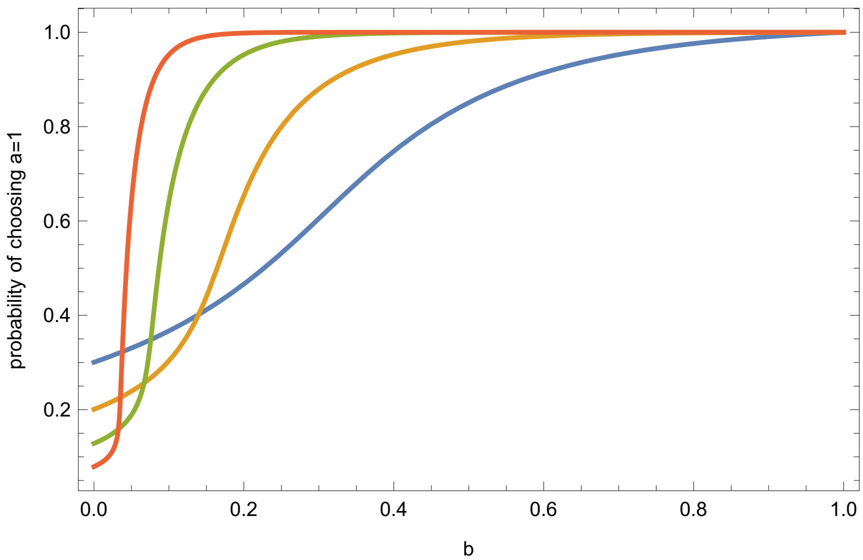
The agent's payoff function is $u(a, y) = y$. Suppose $0.5q/(0.5q + (1 - q)) < c < q$. As the first success reveals that action 1 has an expected payoff of 1, the agent will choose that action whenever they remember a success. The prior expected value of action 1 is q, so the agent will take action 1 if they don't remember any outcomes of the risky arm. When the agent remembers at least one occurrence of $y = 0$ and no occurrences of $y = b$, the posterior expected value associated with action 1 is at most $0.5q/(0.5q + (1 - q))$, so it is optimal for the agent to take the safe action.

This setting generates the equilibrium probability of choosing action 1 displayed in Figure 3 as a function of the value of b under the actual data generating process $p^* = p$. Because rare events are unlikely to be recalled, they will not be present in most databases. Consequently, actions with a given expected payoff tend to be chosen more often when they deliver a good payoff with a high probability than when they deliver a very good payoff more rarely.[24] This is consistent with the evidence in Hertwig, Barron, Weber, and Erev [2004], which also shows that, as our model predicts, this effect is not obtained if agents are told the probabilities instead of learning them from experience. The figure also illustrates that the effect decreases in the memory capacity k, and that in the limit $k \to \infty$, the agent chooses optimally.[25] In addition, the fact that *limited* memory tends to favor actions that are negatively skewed suggests an advantage for a form of *selective* memory that has been often observed: a higher probability of remembering more extreme observations can offset the excessive reliance of frequent and less extreme observations highlighted in Figure 3. It is the interplay between limited memory and selectivity that makes such a boost to extreme outcomes an improvement; with partial but unbounded memory, such a bias can only harm the agent (cf. Fudenberg, Lanzani, and Strack, 2024).

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

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

Figure 3: **Effect of memory size depends on skewness** Probability of choosing $a = 1$ when $\mathbb{E}_{p*}[u(1, y)] = 1$, and $k = 4$ (blue), $k = 8$ (orange), $k = 16$ (green), and $k = 32$ (red).



4.3 The Effect of Reminders

Reminders are informational interventions that have a significant effect on behavior, for example, by helping people recall the positive impact of actions such as attending the gym or saving for the future.[26] Our model can explain how reminders can affect the agent's actions and why reminders about infrequently occurring events have more effect than reminders about frequent ones. Thus, although our memory model predicts choices in line with the random utility model in a fixed environment, its predictions about the effect of changes in the informational environment can be very different.

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

For example, suppose that every period, the agent decides whether or not to go to the gym. If the agent does not engage in the activity ($a = 0$), they know they receive a payoff of 0. If they do engage in the activity, ($a = 1$) they pay an immediate cost $c \in (0, 1)$ and with

[26] See, e.g., Strandbygaard, Thomsen, and Backer [2010] and Karlan, McConnell, Mullainathan, and Zinman [2016].

some unknown probability $p_1(1)$ receive a benefit of 1 (e.g., feeling good about themselves after the workout). The agent's prior belief is that $p_1(1) = p > 0$ with probability $\pi \in (0, 1)$ and $p_1(1) = 0$ with probability $1 - \pi$, and the true probability p^* is equal to p. We assume that $\pi p < c$ so that absent recall of a positive experience, the agent would not engage in the activity, and that $p > c$ so it is beneficial to engage in the activity.

This simple example can also be applied to any activity that provides uncertain benefits for certain costs. Another natural example is saving for future financial hardships, where the reminded benefit corresponds to past experiences where savings helped the agent out of an urgent financial need.

As we prove in Lemma A.5 in the Appendix, the stochastic memory equilibrium with maximal engagement in the activity is given by:[27]

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

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

We have the following Proposition:

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

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

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

Figure 4 illustrates how the frequency of reminders increases the likelihood with which the agent engages in the activity. A natural question is who benefits more from reminders: people who already frequently engage in the activity without reminders or those who do not.

[27] There is always an equilibrium where the agent does not engage in the activity; we select the equilibrium where the agent engages in the activity if such an equilibrium exists. The reason for this equilibrium selection is that if the agent "trembles" and plays each action with at least a small probability ϵ, the stochastic memory equilibrium is unique and converges to the equilibrium where $\alpha(1) > 0$ as $\epsilon \searrow 0$. Thus, by Theorem 1, this equilibrium is also the only possible limit frequency for $\epsilon \searrow 0$. See Lemma A.6 in the Appendix for details.

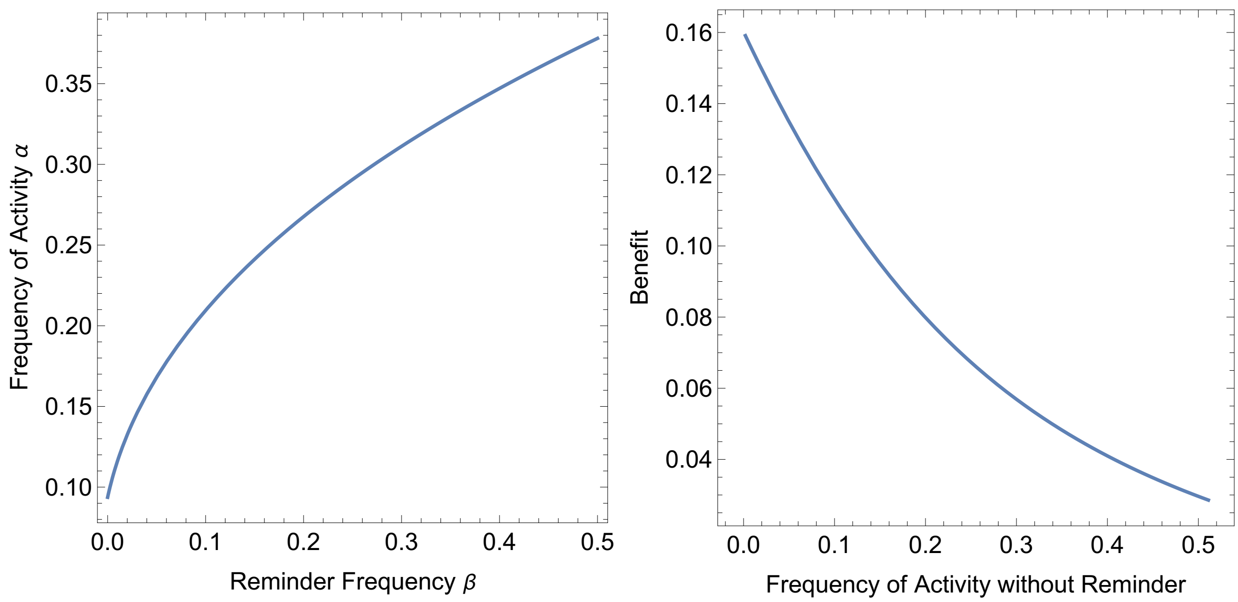


Figure 4: **Reminders increase activity frequency most for agents with low baseline recall.** The left-hand figure shows the effect of reminders on the frequency with which the action is taken when $k = 7$, $p^* = 0.15$. On the right is the effect of reminders as a function of how often the agent engages in the activity without reminders.

Part (ii) of Proposition 4 speaks to that question by showing that an agent for whom p and $p*$ are smaller and thus action $a = 1$ is less attractive benefits more from a higher frequency of reminders. Figure 4 illustrates that the more often a person engages in the activity without a reminder, the less their behavior is affected by a reminder about it. Intuitively, agents who undertake the activity frequently are likely to remember having benefited from doing it in the past, and when (as in our model) one good outcome is enough to induce them to take the action, the reminder is likely to have no effect. This prediction aligns with the evidence presented in Calzolari and Nardotto [2017], which finds that reminders increase gym attendance for individuals with below-median attendance without reminders by 27%. In contrast, reminders have no measurable effect on those with above-median attendance without reminders. It is also consistent with the effect of informational intervention documented in Augenblick, Jack, Kaur, Masiye, and Swanson [2024], which shows that for the same person, informational interventions about one's own spending (and thus the value of saving) are more effective if they involve reminders about atypical expenditures.

Finally, we note that the effect of reminders here is driven by the finiteness of memory, and their effect shrinks as the agent remembers more experiences, i.e., as k becomes large.

Remark. Instead of bringing to mind a random period when the agent played $a = 1$, reminders might lead the agent to recall an experience where the action $a = 1$ was taken and the outcome was good. This would lead to the equilibrium $\alpha_{\beta,p}(1) = 1 + \frac{1}{kp} W\left(-e^{-kp} kp(1-\beta)\right)$. Another alternative formulation is for the agent to observe the (potential) benefit of the action $a = 1$ even if the action $a = 0$ is taken. In either case, our model allows us to describe the effect of reminders.

4.4 Underreaction to Evidence

The agent sometimes relies on a small dataset to make decisions in a stochastic memory equilibrium. 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

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

with $c \in (0, 1)$. Underinference and limited memory both predict underreaction to the data. However, underinference predicts that a sufficiently long sequence of observations always leads beliefs to concentrate around the observed frequency. In contrast, our model predicts that the agent perceives uncertainty even in the limit, in line with Kahneman and Tversky [1972]'s "universal distribution" conditional on large samples. Also, our model suggests that underreaction will be more severe when people are shown data sequentially without being provided written records of past outcomes, while underinference does not.[28]

5 Almost Unlimited Memory

Since real-world memory has finite capacity, the perfect memory of standard models is, at best, an approximation. Here, we examine how well the perfect memory model captures

[28] See Fudenberg and Peysakhovich [2016] and Esponda, Vespa, and Yuksel [2024] for experimental evidence on the effect of providing agents with records and/or summary statistics.

the limits of stochastic memory equilibria as memory capacity k increases and the expected number of recalled experiences grows without bound. For every action distribution α, let

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

be the set of models that maximize the log-likelihood of the true data generating process, where the weights depend on the frequency of each action. These are the models in the support of the agent's prior that exactly match the objective outcome distribution induced by α.[29] Consequently, when the outcome distribution identifies p^*, it is the unique element of $\Theta(\alpha)$.

Definition 3. A *(unitary-belief) self-confirming equilibrium* is an $\alpha \in \Delta(A)$ such that there is a $\nu^\alpha \in \Delta(\Theta(\alpha))$ such that $\alpha \in \Delta(\text{BR}(\nu^\alpha))$.[30]

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

Theorem 3. *Suppose $(\alpha^k)_{k \in \mathbb{N}}$ is a sequence of stochastic memory equilibria, each with memory capacity k and that $\lim_{k \to \infty} \alpha^k = \hat{\alpha}$. Then $\hat{\alpha}$ is a self-confirming equilibrium.*[31]

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

When the data generating process is exogenous and memory is unlimited, the empirical distribution of recalled outcomes converges almost surely. Fudenberg, Lanzani, and Strack

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

[30]This is called a unitary belief because a single belief is used to rationalize all of the actions in the support of α; the heterogeneous-belief version allows each action to be rationalized by different beliefs. See section A.5 for more on heterogeneity.

[31]If the agent is misspecified, they need not learn the path of play, so the limit outcome need not be self-confirming. Moreover, it need not be supportable by unitary beliefs; see Appendix A.5.

[2024] shows that the agent ends up playing the best reply to this distribution. With stochastic memory, there is a positive fraction of periods in which the agent recalls so little that they play a best reply to their prior, though the probability that this occurs becomes smaller and smaller as k goes to infinity. More generally, when the action does influence the outcome distribution, the prior may affect the probability that an agent with unlimited memory converges to a specific self-confirming equilibrium, but the set of self-confirming equilibria is the same for priors that share a common support. This is not the case for stochastic memory equilibrium.

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.[32] Recency is the idea that the agent gives more weight to more recent outcomes. There is an extensive psychology literature documenting various forms of the recency effect and proposing explanations for it; see e.g., Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, and Usher [2005]. As in Mullainathan [2002], we study a very simple form of recency bias, where the previous period's experience is more likely to be remembered, while the periods before that do not receive an extra boost. To model rehearsal, we assume that the experiences recalled in the last period are also more likely to be recalled now.[33]

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

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

where d_t is the database recalled in period t and $r \geqslant 0$ is the weight on rehearsal and recency; $r = 0$ reduces to the baseline model.[34] Thus the recalled periods at time $t+1$

[32] Conlon [2024] finds substantial rehearsal effects, with an estimated baseline probability of recalling an instance of around 47% and an increase in the probability of being recalled due to rehearsal 25%.

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

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

given the previous period's database d_t and experience (a_t, y_t) are distributed as

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

We continue to assume that, upon seeing a database, the agents update their beliefs according to equation (2). Thus, when updating, they are also naive with respect to the effect of rehearsal, which is consistent with the findings in Conlon [2024].

6.1 Ergodic Memory Equilibrium

Limit Distribution of Databases As in the baseline model, the limit distribution of databases is a product of Poisson distributions, but now they depend on the database recalled in the previous period, in addition to the action frequency and the probability of the outcomes given the actions. Thus, we define a Markov chain over databases for each action distribution $\alpha \in \Delta(A)$.

Definition 4. For every $\alpha \in \Delta(A)$ the Markov chain η_α, $\alpha \in \Delta(A)$, has state space $\mathcal{D}$ and Markov kernel $(\eta_{\alpha,d})_{d \in \mathcal{D}} = \left(\sum_{y' \in Y} p^*_{\pi(\mu(\cdot|d))}(y') \eta^{y'}_{\alpha,d}\right)_{d \in \mathcal{D}} \in \Delta(\mathcal{D})^{\mathcal{D}}$ where $\eta^{y'}_{\alpha,d}$ is the product of independent Poisson distributions with parameter for $(a,y) \in A \times Y$ equal to[35]

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

Intuitively, 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, 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 converges to α (Claim 1 in the Appendix).

The first step is to note that at any time, every sub-database 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

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

empty database. A calculation shows the Markov chain is also positive recurrent, which yields the following lemma. (All proofs for this section are in Appendix A.6.)

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

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

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

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

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

Theorem 4. *An ergodic memory equilibrium exists.*

The proof extends that of Theorem 2 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 1.

6.2 Applications of Memory Rehearsal to Finance

Our model of finite expected memory and rehearsal lets us generalize the findings of Mullainathan [2002] about income forecasts beyond the specific parametric structure it assumes. It also lets us provide a novel memory-based explanation of the equity-premium puzzle. 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.[36]

Correlated Prediction Errors The rehearsal memory function (6) generates the same predictions about one-period correlations as Mullainathan [2002], without assuming a specific functional form. First, a high outcome last period boosts the memorability of high realizations, so the forecasting error will be negatively correlated with the most recent information.[37] Second, when the baseline probability of remembering an event is low, and the rehearsal effect is strong, the forecast errors in successive periods are positively correlated for the same reason as in Mullainathan [2002]: memories that are remembered are more likely to be remembered again.

Asset Pricing Suppose that each of a continuum of risk-neutral agents indexed by $x \in [0, 1]$ has a constant per-period amount w to invest. Every period, each agent decides whether to buy, sell, or not trade a unit of a representative equity portfolio and invest the wealth net of the expenditure/revenues from the risky asset in the risk-free asset.[38] 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, and period-specific shock. The risky asset is in net-zero supply and prices p_0 and p_1 are determined by market clearing: $p_1 - p_0 = \text{Median}[\mathbb{E}_x[\theta]]$. The agents' actions impact their payoffs, but all agents observe the sequence of realized prices and returns.

In this setting, the equity premium puzzle is that the observed price of the risky asset is much lower than predicted by the above equation if the distribution of θ were known

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

[37] 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.

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

and equal to that observed in the data, and 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.[39] 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 equals $\mathcal{H}_\alpha$. In particular, even in the long run, the agents will rely on a limited number of observations.

If the prior is symmetric and centered around $\underline{\theta} < \theta$, 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.[40] This explanation is consistent with the empirical evidence, which shows that investors have beliefs that are systematically overpessimistic about stock returns, as opposed to correct beliefs with extreme risk aversion (see, e.g., Dominitz and Manski [2007] for a systematic analysis). It is also consistent with the fact that investors tend not to use available historical sequences of returns but rather rely on the experience they lived through (see Malmendier and Nagel [2016] and the references therein).

7 Conclusion

This paper provides a simple and general model of limited memory that can be applied to many economic problems. In particular, the role of memory in belief formation is important for behavioral economics and macroeconomics, so our work will be useful there. The paper shows how limited memory can provide a foundation for some well-known stochastic choice models. It characterizes how the asymptotic outcomes with limited memory relate to the asymptotic outcomes with unlimited memory capacity.

Our model was designed to highlight the effects of limited and stochastic memory, so it left out many other important aspects of how memories are formed and used. However, our analysis can be extended in several interesting directions. Figure 1 provides a blueprint

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

[40] 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.

for how to do this, namely by modifying one or more of the maps in the composition of functions that define $\psi_\rho^{\mu_0}(\alpha)$. Proceeding counterclockwise from the final step, relaxing subjective expected utility corresponds to replacing BR with some other correspondence in the definition of the allowed policies ρ, and allowing for non-Bayesian updating allows other maps from the distribution of databases to the distribution of beliefs.

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

Finally, we single out one generalization that seems particularly important given the empirical evidence on memory (see, e.g., Kahana, 2012). Memory is subject to significant associativeness effects, with current context cueing memory of experiences that realized in similar situations. Our model can incorporate this by introducing exogenous signals observed before the agent takes an action and a similarity measure over those signals/situations, where the probability of recalling a past experience depends on a measure of how different its context was from the current one. Although associativeness does not change the set of limit points with unbounded memory (Fudenberg, Lanzani, and Strack [2024], we conjecture that under limited memory, the associativeness matters in the long run. Specifically, if the agent believes that similar signals induce similar outcome distributions, the agent would behave as if they were using a kernel estimator, and associative memory would make the kernel used more reliant on similar signals.[41]

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

A Appendix

A.1 Preliminaries

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

Lemma A.1. *We have*

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

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

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

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

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

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

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

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]) $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^{(a,y)} = 0$, $\mathbb{P}_\pi$-a.s.

Note that

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

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

Lemma A.2. *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 $F_\alpha^{\mu_0}$, and $F_{(\cdot)}^{\mu_0}$ is continuous in α.*

Proof. The weak convergence of the distribution of μ_t given h_{t-1} follows from Lemma 1. Let $(\alpha_n)_{n\in\mathbb{N}} \in \Delta(A)^{\mathbb{N}}$ be a sequence converging to α^*, and fix some $\varepsilon > 0$. For every $\alpha \in \Delta(A)$, let $(N_{a,y}^{\alpha})_{(a,y)\in A\times Y}$ be the $|A||Y|$ independent random variables with the same distributions as the marginals of η_α on (a,y). Since this is a finite number of random variables, choose $K\in\mathbb{N}$ such that $\mathbb{P}[\max_{(a,y)\in A\times Y} N_{a,y}^{\alpha^*} > K] < \varepsilon$. Since all the $N_{a,y}^{\alpha^*}$ have Poisson distributions and under a Poisson distribution, the probability of each outcome is continuous in the mean parameter, there is $M\in\mathbb{N}$ 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

$$\left|\int_{\Delta(\Theta)} f(\nu)dF_{\alpha_n}^{\mu_0} - \int_{\Delta(\Theta)} f(\nu)dF_{\alpha}^{\mu_0}\right| < 2\max_{\nu\in\Delta(\Theta)}|f(\nu)|((K+1)|A\times Y|)\varepsilon, \tag{8}$$

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

The proof of Theorem 2 applies a fixed-point theorem to the correspondence $\Psi^{\mu_0} : \Delta(A) \rightrightarrows \Delta(A)$ defined by $\Psi^{\mu_0}(\alpha) = \{\psi_\rho^{\mu_0}(\alpha) : \rho \in \mathcal{O}\}$.

Lemma A.3.

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

A.2 Proofs for Section 3

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

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

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

where for every $x \geqslant y \geqslant 0$, $\binom{x}{y}$ denotes the binomial coefficient between x and y. By way of contradiction, suppose that there is a subsequence of histories h_{t_n}, an $(a,y) \in A \times Y$, a $j \in \mathbb{N}$, and an $\varepsilon > 0$ such that for every $n \in \mathbb{N}$

$$|\mathbb{P}_\pi\left[d_{t_n}(a,y) = j | h_{t_n}\right] - \eta_{\alpha_{t_n}}[d(a,y) = j]| \geqslant \varepsilon \tag{9}$$

In words, this says that the probability of recalling a database where experience (a,y) is recalled j times conditional on history h_{t_n} is more than ε away from the probability that $\eta_{\alpha_{t_n}}$ assigns to those databases. Since $\Delta(A)$ is compact, by restricting to a subsequence, we can also assume that α_{t_n} is converging to some $\alpha \in \Delta(A)$.

Suppose first that $\lim_{t\to\infty} \upsilon_t(a,y)\, t = \infty$. Then 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 $e^{-\lambda_{a,y}} \frac{\lambda^C_{a,y}}{C!}$, where

$$\lambda_{a,y} = \lim_{t\to\infty} \upsilon_t(a,y)\, t\, \frac{k}{t} = \alpha(a)p^*_a(y)\, k.$$

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

Next, if $\lim_{t\to\infty} \upsilon_t(a,y)\, t \neq \infty$, then $\lim_{t\to\infty} \alpha_t(a)p^*_a(y) = 0$, so the marginal of η_α on (a,y) is a Dirac on 0. Thus, by equation (9), the distribution of $d(a,y)$ conditional on h_{t_n} assigns probability bounded away from 0 to some strictly positive $j \in \mathbb{N}$. Moreover, there exists a $C \in \mathbb{N}$ such that for all $t \geqslant C$ the distribution of $d(a,y)$ conditional on h_t is FOSD dominated by the distribution of $d(a,y)$ conditional on h'_t, where $(h'_t)_{t\in\mathbb{N}}$ is an alternative sequence of histories such that $(a'_t, y_t) = (a,y)$ if and only if $t = 2^n$ for some $n \in \mathbb{N}$

(recall that Binomial (n, p) is FOSD-increasing in n; here $N_t' > N_t'$ for large t). However, under the alternative sequence of histories $(h_t')_{t\in\mathbb{N}}$, we have that $\lim_{t\to\infty} \upsilon_t(a, y)\, t = \infty$, and therefore from the previous case, it must converge to a Poisson distribution with parameter $\lim_{n\to\infty} \upsilon_t(a, y)k = 0$, i.e., to a Dirac distribution. But since the only positive-valued distribution that is FOSD-dominated by a Dirac at 0 is the Dirac at 0 itself, we reach a contradiction, and the result follows. □

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

The proofs of Theorems 1 and 5 use a 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], define the continuous-time interpolation of $(\alpha_t)_{t\in\mathbb{N}}$ to be the function $w : \mathbb{R}_+ \to \Delta(A)$

$$w(\tau_t + c) = \alpha_t + c\frac{\alpha_{t+1} - \alpha_t}{\tau_{t+1} - \tau_t}, \quad \forall t \in \mathbb{N}, \forall c \in \left[0, \frac{1}{t+1}\right]. \tag{10}$$

Proof of Theorem 1. We want to show that if α is not a stochastic memory equilibrium, it is not a limit frequency. To do this, we extend Esponda, Pouzo, and Yamamoto [2021a]'s 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 show that (10) converges to a solution of

$$\dot{\boldsymbol{\alpha}}(t) \in \Psi^{\mu_0}(\boldsymbol{\alpha}(t)) - \boldsymbol{\alpha}(t). \tag{11}$$

A solution to (11) with initial point $\alpha_0 \in \Delta(A)$ is a mapping $x : \mathbb{R}_+ \to \Delta(A)$ that is absolutely continuous over compact intervals, with $x(0) = \alpha^*$, and (11) satisfied for almost every t. For every $T \in \mathbb{N}$ and $\alpha_0 \in \Delta(A)$, let $X_{\alpha^*}^T$ be the set of solutions to (11) over $[0, T]$ with initial conditions α^*, and let $X^T = \bigcup_{\alpha_0\in\Delta(A)} X_{\alpha^*}^T$.

Now we show that the continuous-time interpolation of α defined in (10) can, in the long run, be approximated arbitrarily well by a solution to (11). Recall that π is the optimal Markovian policy employed by the agent, and that for any measurable selection ρ from the best response correspondence, $\psi_\rho^{\mu_0}(\alpha) = \int_{\Delta(\Theta)} \rho(\nu) dF_\alpha^{\mu_0}(\nu)$. Define the stochastic process $\tilde{U}_t = t(\alpha_t - \frac{t-1}{t}\alpha_{t-1}) - \psi_\pi^{\mu_0}(\alpha_{t-1}) = \delta_{a_t} - \psi_\pi^{\mu_0}(\alpha_{t-1})$ and let $b_t = \mathbb{E}[\tilde{U}_{t+1}|h_t]$, $U_t = \tilde{U}_t - b_t$.

Then observe that

$$\alpha_{t+1} - \alpha_t - \frac{U_{t+1} + b_{t+1}}{t+1} = \frac{\psi_\pi^{\mu_0}(\alpha_t) - \alpha_t}{t+1}. \tag{12}$$

Since both $\psi_\pi^{\mu_0}(\alpha_t)$ and α_{t+1} are uniformly bounded, U_t is a uniformly bounded martingale difference process. Moreover, by Lemma A.4, b_{t+1} converges to 0 on $\mathbb{P}_\pi$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$ where $\lim_{t\to\infty} \alpha_t = \alpha$, so condition (i) of Proposition 1.3 in Benaim, Hofbauer, and Sorin [2005] is satisfied (setting $\gamma_n = 1/n$ in their formula) by Remark 4.5 in Benaïm and Hirsch [1999]. Condition (ii) is also satisfied because $||\alpha_{t+1} - \alpha_t||_\infty < 1/(t+1)$, w is Lipschitz continuous of order 1, and α_t is uniformly bounded because it takes values in $\Delta(A)$. Therefore, since w is the interpolated process for the stochastic process in equation (12), it is a perturbed solution of (11). Thus, by Theorem 4.2 in Benaim, Hofbauer, and Sorin [2005], on $\mathbb{P}_\pi$ almost every sequence of histories $(h_t)_{t\in\mathbb{N}}$ where $\lim_{t\to\infty} \alpha_t = \alpha$[42]

$$\lim_{t\to\infty} \inf_{\tilde{\boldsymbol{\alpha}}\in X^T} \sup_{s\in[0,T]} ||w(t+s) - \tilde{\boldsymbol{\alpha}}(s)|| = 0 \qquad \forall T \in \mathbb{N}. \tag{13}$$

If α is not a stochastic memory equilibrium, then parts 1 and 4 of Lemma A.3 and the separating hyperplane theorem (see, e.g., Section 14.5 of Royden and Fitzpatrick [2010] for the version used here) guarantee that there exists $f \in \mathbb{R}^A$ with $\alpha \cdot f > \sup_{\bar{\alpha}\in\Psi(\alpha)} \bar{\alpha}\cdot f$. By part 4 of Lemma A.3, $\alpha \cdot f > \max_{\bar{\alpha}\in\Psi(\alpha)} \bar{\alpha}\cdot f$. 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 $\alpha_0 \in B_\varepsilon(\alpha)$ and every solution in $X^T_{\alpha*}$, $\boldsymbol{\alpha}(t)\cdot f$ decreases at rate at least $K/2$ until the solution leaves $B_\varepsilon(\alpha)$. So there exists $T \in \mathbb{N}$ such that for every initial condition $\alpha_0 \in B_\varepsilon(\alpha)$ and every solution in $X^T_{\alpha*}$, the differential inclusion leaves $B_\varepsilon(\alpha)$ by time T, that is,[43]

$$\sup_{\tilde{\boldsymbol{\alpha}}\in X^T_{\alpha*}} \inf\{t : \tilde{\boldsymbol{\alpha}}(t) \notin B_\varepsilon(\alpha)\} \leqslant T \qquad \forall \alpha_0 \in B_\varepsilon(\alpha). \tag{14}$$

To conclude the proof, we will show that α_t does not asymptotically lies in an $\varepsilon/2$ ball around α on any path $(h_t)_{t\in\mathbb{N}}$ where (13) applies. Since the set of samples where (13)

[42]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].

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

does not apply has 0 probability under $\mathbb{P}_\pi$, this implies that α is not a limit frequency. If there is no $\hat{T} \in \mathbb{N}$ such that $w(c) \in B_{\varepsilon/2}(\alpha)$ for all $c > \hat{T}$, $(\alpha_t)_{t\in\mathbb{N}}$ does not converge to α. So let $\hat{T} \in \mathbb{N}$ be such that on the chosen path $(h_t)_{t\in\mathbb{N}}$, $w(c) \in B_{\varepsilon/2}(\alpha)$ for all $c > \hat{T}$ and $\inf_{\tilde{\boldsymbol{\alpha}}\in X^T} \sup_{0\leqslant s\leqslant T} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}(s)|| < \varepsilon/4$, and take $\tilde{\boldsymbol{\alpha}} \in X^T$ with

$$\sup_{0\leqslant s\leqslant T} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}(s)|| < \varepsilon/4. \tag{15}$$

Then (14) implies that the differential inclusion leaves $B_\varepsilon(\alpha)$ at least once between $\hat{T}$ and $\hat{T}+T$, and by (15), α_t must leave $B_{\varepsilon/2}(\alpha)$ at least once between $\hat{T}$ and $\hat{T}+T$. This proves Theorem 1. □

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

A.3 Proofs for Section 4.1

Proof of Proposition 1. Suppose that α_t converges to some α^*. We first construct the target random utility representation ζ and then prove that α_t approaches it. Consider the map $G : \mathcal{D} \to \mathcal{P}$ such that $aG(d)a'$ if and only if $a = \pi(\mu(\cdot|d) \mid \{a, a'\})$, define ζ by $\zeta(P) = \eta_{\alpha^*}(G^{-1}(P))$, and let c_η be the associated stochastic choice. Observe that this binary relation is total by definition, and it is transitive by the assumption that ties are broken in a menu-independent way, so it is indeed a linear order. Because the sets of menus and actions are finite, the proposition can be established by showing that the probabilities that any given action a is chosen from any given menu M converge those of c_ζ.

Fix $M \in \mathcal{M}$ and $a \in M$. Lemma 1 shows that for $\mathbb{P}_\pi$ almost any sequence of histories such that $\mathbb{P}_\pi[\lim_{t\to\infty} \alpha_t = \alpha^*] > 0$, the probability that the number of recalled (a, y) experiences is equal to c for every $c \in \mathbb{N}$ converges to its probability under a Poisson random variable with parameter $\alpha^*(a)p^*_a(y)k$, and these distributions are independent across (a, y) pairs. Therefore, as $\tau \to \infty$, the probability distribution η^{h_τ} over databases recalled at period $\tau+1$ given h_τ converges to η_{α^*}.

To link this with the convergence of the stochastic choice rule, let $D_M(a) = \{d \in \mathcal{D} : \forall a' \in M, aG(d)a'\}$. Because the agent's tiebreaking is menu-independent, the difference between $c_\zeta(a, M)$ and $\mathbb{P}_\pi[a_{\tau+1} = a | h_\tau, M]$ is bounded by $|\eta_{\alpha^*}(D_M(a)) - \eta^{h_\tau}(D_M(a))|$. For every $l \in \mathbb{N}$, this is less than

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

Both addends can be made arbitrarily small by taking l and τ large. This establishes the random utility representation.

To see that the stochastic choice rule admits an information representation, we map our objects to those in Lu [2016]. We identify the states S with Θ, the acts H with actions A, and the outcomes with utility realizations $Z = \{u(a, y) : (a, y) \in A \times Y\}$ (where u is our u). Let the utility function u of Lu [2016] be the identity, with each action mapped to an Anscombe-Aumann act by associating each state p to the lottery that gives probability $p_a(\{y : u(a, y) = \mathfrak{u}\})$ to $\mathfrak{u} \in Z$. Given this mapping, the stochastic choice rule admits a representation with distribution over posteriors $\nu = F_\alpha^{\mu_0}$. Finally, that $\alpha^*(a) = c_\zeta(a, A)$ follows by Theorem 1. □

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

Proof of Proposition 3. We first derive the limit distribution of actions conditional on recalling $n_a \in \mathbb{N}$ experiences for each action $a \in A$. Because the agent's prior is symmetric across the actions and Normal, it is optimal for them to choose the action with the highest posterior mean. Let $\hat{y}_a$ be the average outcome over the n_a recalled experiences where action a was used. When the agent recalls n_a experiences for action a, each is normally distributed with mean $\bar{y}_a$ and variance σ^2, so $\hat{y}_a$ is Normally distributed with mean $\bar{y}_a$ and variance σ^2/n_a. Thus the posterior mean of y_a is $\frac{\hat{y}_a n_a/\sigma^2}{1/\sigma_0^2+n_a/\sigma^2}$, and the induced choice probabilities are equal to those in a Probit model with means $\left(\bar{y}_a \frac{n_a/\sigma^2}{1/\sigma_0^2+n_a/\sigma^2}\right)_{a\in A}$ and a diagonal covariance matrix with entries $\frac{n_a/\sigma^2}{(1/\sigma_0^2+n_a/\sigma^2)^2}$.

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

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

A.4 Proofs for Section 4.3

Lemma A.5. *The choice probabilities of the agent are*

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

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

Proof of Proposition 4. To see part (i), observe that W is an increasing function and hence $\alpha_{\beta,p}$ is increasing in β. By Lemma A.5, an equilibrium where the agent engages in the activity with strictly positive probability absent reminders exists if and only if $pk > 1$, and thus we can assume $pk > 1$ when establishing part (ii). We observe that the marginal effect of reminders is given by

$$\frac{\partial \alpha_{\beta,p}(1)}{\partial \beta} = \frac{1}{k(1-p\beta)} \times \frac{-W\left(-e^{-kp} kp(1-p\beta)\right)}{1 + W\left(-e^{-kp} kp(1-p\beta)\right)}.$$

To simplify notation define $\psi(p) = -e^{-kp}kp(1-p\beta)$. Taking logarithms and differentiating with respect to p yields

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

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

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

As $kp > 1$ and $W \in [-1,0]$ we obtain that

$$\frac{\partial}{\partial p}\log\left(\frac{\partial\alpha_{\beta,p}(1)}{\partial\beta}\right) \leqslant \frac{\beta}{k(1-p\beta)} - \frac{kp\beta}{kp(1-p\beta)} \leqslant \left[1-\frac{1}{p}\right]\frac{\beta}{k(1-p\beta)} < 0\,.$$

Furthermore as W is concave we can bound $W'(\psi) \leqslant W'(0) = 1$ for all $\psi \in [-1/e, 0]$. Thus, $\alpha_{\beta,p}(1)$ is sub-modular in (β, p) this and Lemma A.5 imply (ii). ☐

Lemma A.6. *Without reminders ($\beta = 0$) and an exogenous probability $\epsilon > 0$ of choosing action $a = 1$ there is a unique equilibrium with $\alpha(1) > 0$ if there are multiple equilibria for $\epsilon = 0$. This equilibrium converges to the equilibrium where action 1 is played with strictly positive probability as $\epsilon \searrow 0$.*

A.5 Proofs for Section 5

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

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

Lemma A.7. *Suppose that $\lim_{k\to\infty}\alpha^k = \hat{\alpha}$. For every $\varepsilon > 0$ and $(a, a', y, y') \in A^2 \times Y^2$ with $\hat{\alpha}(a') > 0$ and $p^*_{a'}(y') > 0$, $\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}\left[\left|\frac{d(a,y)}{d(a',y')} - \frac{\hat{\alpha}(a)p^*_a(y)}{\hat{\alpha}(a')p^*_{a'}(y')}\right| > \varepsilon\right] = 0$. Moreover, $\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}\left[d(a', y') - \frac{\hat{\alpha}(a')p^*_{a'}(y')k}{2} < \varepsilon\right] = 0$.*

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

Unitary-data self-confirming equilibrium allows each action in the support of α to be rationalized by a different belief but requires that all beliefs are supported over the likelihood maximizers given the same data. This is more restrictive than heterogeneous-belief self-confirming equilibrium (Fudenberg and Levine [1993]), which only requires that each action a in the support of α is a best response to a belief over the maximizers corresponding to data about the consequences of the particular pure action a. This difference is a consequence of the different origins of the heterogeneity for the two equilibrium concepts.[44]

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

Lemma A.8. *When $p^* \in \Theta$ (as we have assumed here), every unitary-data self-confirming equilibrium is a self-confirming equilibrium.*[45]

Proof of Theorem 3. Let $f(d) \in \Delta(A \times Y)$ denote the empirical joint distribution over action-outcome pairs corresponding to $d \in \mathcal{D}$. By Lemma A.7, for every $M \in \mathbb{N}$ and $\varepsilon > 0$, $\lim_{k\to\infty} \mathbb{P}_{\eta^k_{\alpha^k}} \left[d : \sum_{a\in A, y\in Y} d(a,y) > M, \max_{a\in A, y\in Y} |f(d)(a,y) - \hat{\alpha}(a)p^*_a(y)| < \varepsilon \right]$ is equal to 1. That is, with probability approaching 1, the database is large, and the recalled frequency of pair (a, y) is approximately proportional to $\hat{\alpha}(a)p^*_a(y)$.

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

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

Since the RHS is growing to ∞ as k grows and ε can be arbitrarily small, the agent's

[44] 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.

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

beliefs concentrate on $\Theta(\hat{\alpha})$. It follows from the upper hemicontinuity of the best reply correspondence that $\hat{\alpha}$ is a unitary-data self-confirming equilibrium.

Finally, by Lemma A.8 and Assumption 1, every unitary-data self-confirming equilibrium is a self-confirming equilibrium. □

A.6 Proofs for Section 6

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

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

Moreover, for any $d' \in D'$, there is a simple path of length $\tau \in \mathbb{N}$ connecting the empty database and d', i.e., a finite sequence of distinct databases $(\tilde{d}_0, ..., \tilde{d}_\tau)$ with $\tilde{d}_0$ the empty database, $\tilde{d}_\tau = d'$, and $\eta_{\alpha,\tilde{d}_i}(\tilde{d}_{i+1}) > 0$ for all $i \in \{0, ..., \tau - 1\}$. Let $M = \prod_{i=0}^{\tau-1} \eta_{\alpha,\tilde{d}_i}(\tilde{d}_{i+1})$. Thus the expected time of return to d' when it is the state at time t is bounded from above by

$$\sum_{i=1}^{\infty}(1 - P(\text{return time} \leqslant i)) \leqslant \tau + \sum_{i=1}^{\infty}(1 - P(\text{return time} \leqslant \tau + i))$$

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

so d' is positive recurrent.

Since there is zero probability of leaving D' and all the states in D' are positive recurrent, η_α has a unique invariant distribution (see Theorem 6.5.3 in Durrett [2019]). □

Lemma A.9. *For any sequence of histories $(h_t)_{t\in\mathbb{N}}$ such that $\lim_{t\to\infty} \upsilon_t(a, y) = \alpha_t(a)p^*_a(y)$ for all $(a, y) \in A \times Y$, the distribution of d_t when $d_{t-1} = d'$ converges $\sum_{y'\in Y} p^*_{\pi(\mu(\cdot|d'))}(y')\eta^{y'}_{\alpha_t,d'}$ where $\eta^{y'}_{\alpha_t,d'}$ is the product of independent Poisson distributions with parameter for $(a, y) \in$*

$A \times Y$ equal to

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

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

The proof of this and the next Lemma are in Online Appendix B.1. Let $\Psi_{d'}(\alpha)$ denote the distributions over actions induced by an optimal Markovian mixed policy ρ and random beliefs ν: $\Psi^{\mu_0}(\alpha, d') = \{\int_{\Delta(\Theta)} \rho(\nu) dF^{\mu_0}_{\alpha,d'}(\nu) : \rho \in \mathcal{O}\}$.

Lemma A.10.

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

Proof of Theorem 4. Lemma A.10 shows that every fixed point of $\bar{\Psi} = \int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d') d\mathcal{H}_{(\cdot)}(d')$ is an ergodic memory equilibrium, and that $\bar{\Psi}$ is non-empty valued and convex valued. Together with the closed-graph theorem, the lemma also shows $\bar{\Psi}$ has has a closed graph, so it has a fixed point by the Kakutani fixed point theorem. □

Proof of Theorem 5. The proof has three steps. First, Lemma A.9 shows that the transition matrices over databases converge as $t \to \infty$ and says what the limit is. 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 and that the differential inclusion cannot converge to something that is not an ergodic memory equilibrium, as in the proof of Theorem 1.

Claim 1. *Let $\varepsilon > 0$. There exists a $\delta > 0$ such that if $\limsup_{t\to\infty} ||\alpha_t - \alpha||_\infty \leqslant \delta$ then the distance between the distribution of databases and $\mathcal{H}_\alpha$ converges to be smaller than ε as $t \to \infty$.*

Proof. Lemma A.9 shows that the transition matrices over databases converge as $t \to \infty$ and says what the limit is. Moreover, for every $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that the probability that $(\sum_{y' \in Y} p^*_{\pi(\mu(\cdot|d'))}(y')\eta^{y'}_{\alpha_t,d'})_{t\in\mathbb{N}}$ assigns to the transition from an arbitrary database to the set of databases with K or more experiences is smaller than ε. Create a coarser finite 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, i.e., databases d, d' are pooled if $\min\{\sum_{(a,y)\in A\times Y} d(a,y), \sum_{(a,y)\in A\times Y} d'(a,y)\} > K$ and$\{(a,y) \in A \times Y : d(a,y) = 0\} = \{(a,y) \in A \times Y : d'(a,y) = 0\}$. Transition to the null database always has positive probability in the Markov chain for the restricted process, so the limit matrix is regular, and the result follows by the fact that the iteration of Doeblin matrices that are arbitrarily close to a matrix P becomes arbitrarily close to the ergodic distribution of each of P (cf. Mitrophanov [2005]).[46] 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 (10) can be approximated by $\dot{\boldsymbol{\alpha}}(t) \in \mathbb{E}_{\mathcal{H}_{\boldsymbol{\alpha}_t}}[\Psi(\boldsymbol{\alpha}(t), d)] - \boldsymbol{\alpha}(t)$. The last step parallels the last step of the proof of Theorem 1 with $\int_{\mathcal{D}} \Psi^{\mu_0}(\alpha, d')d\mathcal{H}_{\boldsymbol{\alpha}}$ in place of $\Psi^{\mu_0}(\alpha)$ after observing that $\int_{\mathcal{D}} \Psi^{\mu_0}(\alpha, d')d\mathcal{H}_{\boldsymbol{\alpha}}$ inherits the key properties of Ψ^{μ_0}, as shown by Lemma (A.10); we omit the remaining details. □

References

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

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

Augenblick, N., B. K. Jack, S. Kaur, F. Masiye, and N. Swanson (2024). "Retrieval failures and consumption smoothing: A field experiment on seasonal poverty." *Available at Semantic Scholar CorpusID*, 263850623.

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

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

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

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

Bénabou, R. and J. Tirole (2002). "Self-confidence and personal motivation." *The Quarterly Journal of Economics*, 117, 871–915.

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

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

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

Bordalo, P., N. Gennaioli, G. Lanzani, and A. Shleifer (2025). *A cognitive theory of reasoning and choice*. Tech. rep. National Bureau of Economic Research.

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

Cerreia-Vioglio, S., R. Corrao, and G. Lanzani (2024). "Dynamic opinion aggregation: Long-run stability and disagreement." *Review of Economic Studies*, 91, 1406–1447.

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.

Conlon, J. J. (2024). *Memory Rehearsal and Belief Biases*. Tech. rep. Working paper.

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

Danenberg, T. and R. Spiegler (2025). "Representative sampling equilibrium." *Journal of Political Economy: Microeconomics, Forthcoming*.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Mitrophanov, A. Y. (2005). "Sensitivity and convergence of uniformly ergodic Markov chains." *Journal of Applied Probability*, 42, 1003–1014.

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

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

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

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

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

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

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

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

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

Sial, A. Y., J. R. Sydnor, and D. Taubinsky (2024). "Biased Memory and Perceptions of Self-Control." *Working Paper*.

Strandbygaard, U., S. F. Thomsen, and V. Backer (2010). "A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study." *Respiratory medicine*, 104, 166–171.

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.

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

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

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

Wolitzky, A. (2018). "Learning from Others' Outcomes." *American Economic Review*, 108, 2763–2801.

B Online Appendix

B.1 Omitted Proofs

Proof of Lemma A.3.

1. Since the set of actions is finite, there is at least one measurable selection from the best reply correspondence.

2. $\Delta(A)$ is finite-dimensional and bounded, $\cup_{\rho \in \mathcal{O}} \rho(\nu)$ is closed for every $\nu \in \Delta(\Theta)$, and $\Psi^{\mu_0}(\alpha)$ is the Aumann integral Aumann [1965] of the (mixed) best reply correspondence with respect to the distribution of beliefs $F_\alpha^{\mu_0}$. Therefore, it satisfies the assumptions of case (i) of Theorem 2.1.37 of Molchanov [2017], so it is closed.

3. By Lemma A.2, $F_{(\cdot)}^{\mu_0}$ is continuous in α, and so by Artstein and Wets [1988], Theorem 4.2, Ψ^{μ_0} is upper hemicontinuous.

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

5. This follows immediately from the definition of stochastic memory equilibrium.

□

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

Proof of Lemma A.5. If there is a reminder, the agent takes action 0 only if they do not get a reminder and they do not recall a good outcome, which has probability $\exp -(\alpha * k * p)(1-\beta)$, or they do not recall a good outcome and are given a reminder, but the remainder is about a bad outcome outcome, which has probability $exp - (\alpha * k * p)\beta(1-p)$. Thus, the equilibrium probability that the agent takes action 1 is given by the fixed-point equation

$$\alpha_{\beta,p}(1) = 1 - e^{-kp\alpha_{\beta,p}(1)}(1 - \beta + \beta(1-p)).$$

First, observe that for $\beta > 0$ any solution to the equation

$$\alpha_{\beta,p}(1) = 1 - e^{-k\alpha_{\beta,p}(1)p}(1 - \beta + \beta(1-p)) \tag{16}$$

must satisfy $\alpha_{\beta,p}(1) > 0$ as the right-hand side is strictly positive. For $\beta = 0$ the function $\hat\alpha \mapsto \hat\alpha - \left(1 - e^{-kp\,\hat\alpha}\right)$ has derivative $1 - kp\,e^{-kp\,\hat\alpha}$ and is convex. As for $kp \leqslant 1$, the function crosses 0 from below at $\hat\alpha = 0$ this is the unique solution to (16).

For $kp > 1$ or $\beta > 0$ a positive solution exists, namely $\alpha_{\beta,p}(1) = 1+\frac{1}{kp}W\left(-e^{-kp}kp(1-p\beta)\right)$, which we select by assumption. The function $\hat\alpha \mapsto \hat\alpha - \left(1 - e^{-kp\,\hat\alpha}\right)$ has positive derivative at $\alpha_{0,p}(1)$, so $0 \leqslant 1 - kp\,e^{-kp\,\alpha_{0,p}(1)} = 1 - kp(1-\alpha_{0,p}(1))$. To show that $\alpha_{0,p}(1)$ is increasing in p, observe that by equation (16) it is the solution to $\alpha_{0,p}(1) = 1 - e^{-kp\,\alpha_{0,p}(1)}$. By the implicit function theorem

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

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

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

Proof of Lemma A.6. We note that there are multiple equilibria for $\epsilon = 0$ if and only if $pk > 1$. To see that the equilibrium is unique if with probability $\epsilon > 0$ the agent takes the action $a = 1$ observe that the frequency with which the agent takes the action $a = 1$ must solve $\alpha_p(1) = \epsilon + (1-\epsilon)[1 - e^{-k\alpha_p(1)p}]$.

First, note that any solution of the above equation must satisfy $\alpha > \epsilon$. Observe that the function $\alpha \mapsto \epsilon + (1-\epsilon)[1 - e^{-k\alpha p}] - \alpha$ is concave and thus can have at most 2 roots. At $\alpha = 0$ this function is positive and has derivative $(1-\epsilon)kp - 1$, which as $kp > 1$, is also positive for ϵ small enough. Thus, this equation can only have the single positive solution $\alpha(1) = 1 + \frac{1}{kp}W\left(-e^{-kp}kp(1-\epsilon)\right)$, which converges to the equilibrium where $\alpha(1) > 0$ for $\epsilon = 0$ because W is continuous. □

Proof of Lemma A.7. Choose any $a \in \operatorname{supp}\hat\alpha$. By Chernoff's Theorem (e.g., Theorem 9.3 in Billingsley [2017])

$$\mathbb{P}_{\eta^k_{\alpha^k}}\left[|d(a,y) - k\alpha^k(a)\,p^*_a(y)\,| > \beta k\alpha^k(a)\,p^*_a(y)\right] \leqslant \inf_{c\in\mathbb{R}} M_{d(a,y)}(c)\,e^{-c\beta k\alpha^k(a)p^*_a(y)} \qquad \forall\beta\in(0,1)$$

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

$(0,1)$ we have

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

Since $(\beta-1-\beta\ln\beta)<0$ for all $\beta\in(0,1)$, taking the limit $k\to\infty$ gives

$$\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}\left[\left|d(a,y)-k\alpha^k\left(a\right)p^*_a\left(y\right)\right| > \beta k\alpha^k\left(a\right)p^*_a\left(y\right)\right]=0 \qquad \forall\beta\in(0,1)\,. \tag{17}$$

Analogous steps show that $\lim_{k\to\infty}\mathbb{P}_{\eta^k_{\alpha^k}}[d(\bar{a},y)>k\lambda]=0$ for all $\bar{a}\notin\operatorname{supp}\hat{\alpha}(a)$ and $\lambda>0$.

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

$$\frac{\alpha^k\left(a\right)p^*_a\left(y\right)\left(1+\gamma\right)}{\alpha^k\left(a'\right)p^*_{a'}\left(y'\right)\left(1-\gamma\right)}-\frac{\alpha^k\left(a\right)m\left(a,y\right)p^*_a\left(y\right)}{\alpha^k\left(a'\right)p^*_{a'}\left(y'\right)}<\varepsilon$$

and

$$\frac{\alpha^k\left(a\right)p^*_a\left(y\right)}{\alpha^k\left(a'\right)m\left(a',y'\right)p^*_{a'}\left(y'\right)}-\frac{\alpha^k\left(a\right)p^*_a\left(y\right)\left(1-\gamma\right)}{\alpha^k\left(a'\right)m\left(a',y'\right)p^*_{a'}\left(y'\right)\left(1+\gamma\right)}<\varepsilon.$$

Thus we have

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

Equation (17) 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 (17). □

Proof of Lemma A.8. If $\hat{\alpha}$ is a unitary-data self-confirming equilibrium then for all $a'\in\operatorname{supp}(\hat{\alpha})$ there is $\nu^{a'}\in\Delta(\Theta(\hat{\alpha}))$ such that $a'\in BR(\nu^{a'})$. Moreover, because the agent is correctly specified for all $a',a''\in\operatorname{supp}(\alpha)$ and for every $p\in\operatorname{supp}\nu^{a'}$, $p_{a''}=p^*_{a''}$, so the agent has a single and correct belief $\hat{u}_{a'}$ about the expected payoff of every $a'\in\operatorname{supp}(\alpha)$. And since $\hat{\alpha}$ is a unitary-data equilibrium, $\hat{u}\geqslant\sum_y u(a,y)\nu^{a'}_a$ for all $a'\in\operatorname{supp}(\alpha)$ and $a\in A$. Thus, $\hat{\alpha}$ is a unitary-belief self-confirming equilibrium. □

Proof of Lemma A.9. Given a database d' recalled in period $t-1$, the experience of period $t-1$ is equal to (a,y) with probability 0 if $a \neq \pi(\{\mu(\cdot|d')\})$ and $p^*_{\pi(\{\mu(\cdot|d')\})}(y)$ otherwise. Therefore, with probability $p^*_{\pi(\{\mu(\cdot|d')\})}(y')$ the database at time $t \geqslant k+r$ is distributed as a product of multinomial distributions:

$$\begin{aligned}\mathbb{P}_\pi\left[d_t = d | d_{t-1} = d'\right] = &\prod_{(a,y)\in A\times Y} \binom{\upsilon_t(a,y)t}{d(a,y)} \\ \times &\left(\frac{k + r(\mathbb{1}_{\mathbb{N}\backslash 0}(d'(a,y)) + \mathbb{1}_{\{0\}}(d'(a,y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a))}{t}\right)^{d(a,y)} \\ \times &\left(1 - \frac{k + r(\mathbb{1}_{\mathbb{N}\backslash 0}(d'(a,y)) + \mathbb{1}_{\{0\}}(d'(a,y))\mathbb{1}_{\pi(\{\mu(\cdot|d')\})}(a))}{t}\right)^{\upsilon_t(a,y)t - d(a,y)}.\end{aligned}$$

Suppose first that $\lim_{t\to\infty} \upsilon_t(a,y)\,t = \infty$. Then, 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

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

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 $\lambda(d')_{a,y}$.

Next, suppose that $\lim_{t\to\infty} \upsilon_t(a,y)\,t \neq \infty$. Then there is a $C \in \mathbb{N}$ such that for all $t \geqslant C$ the distribution of $d(a,y)$ conditional on h_t and d' is FOSD dominated by the distribution of $d(a,y)$ conditional on h'_t and d' where $(h'_t)_{t\in\mathbb{N}}$ is an alternative sequence of histories such that $(a'_t, y_t) = (a,y)$ if and only if $t = 2^n$ for some $n \in \mathbb{N}$. This ancillary distribution converges to a Dirac on 0 from the first part of the proof, so the result follows.

Moreover, let $(\alpha_n)_{n\in\mathbb{N}} \in \Delta(A)$ be a sequence converging to α^*, and fix some $\varepsilon > 0$. For every $\alpha \in \Delta(A)$, let $(N^{\alpha^*,y'}_{a,y}(d'))_{(a,y)\in A\times Y}$ be the $|A||Y|$ independent random variables with the same distributions as the marginals of $\eta^{y'}_{\alpha,d'}$ on (a,y). Since all the $N^{\alpha^*,y'}_{a,y}(d')$ have Poisson distributions, there is a $K \in \mathbb{N}$ such that

$$\mathbb{P}\left[\max_{(a,y,y')\in A\times Y^2} N^{\alpha^*,y'}_{a,y}(d') > K\right] < \varepsilon.$$

Let $M \in \mathbb{N}$ be such that $\mathbb{P}[\max_{(a,y)\in A\times Y} N_{a,y}^{\alpha_n,y'}(d') > K] < \varepsilon$ and $|\mathbb{P}[N_{a,y}^{\alpha_n,y'}(d') = c] - \mathbb{P}[N_{a,y}^{\alpha^*,y'}(d') = c]| < \varepsilon$ for all $(a, y, y') \in A \times Y^2$, 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

$$\left|\int_{\Delta(\Theta)} f(\nu)dF^{\mu_0}_{\alpha_n,d'} - \int_{\Delta(\Theta)} f(\nu)dF^{\mu_0}_{\alpha,d'}\right| < 2 \max_{\nu\in\Delta(\Theta)} |f(\nu)|((K+1)|A \times Y^2|)\varepsilon,$$

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

Proof of Lemma A.10. First, observe that for every $\alpha \in \Delta(A)$, $\int_{\mathcal{D}} \Psi^{\mu_0}(\alpha, d')d\mathcal{H}_\alpha(d')$ is the Aumann [1965] integral of the mixed best reply correspondence with respect to the measure $\int_{\mathcal{D}} F^{\mu_0}_{\alpha,d'}d\mathcal{H}_{\boldsymbol{\alpha}'}(d')$.

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.9, $F^{\mu_0}_{(\cdot)}$ 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 $\int_{\mathcal{D}} \Psi^{\mu_0}(\cdot, d')d\mathcal{H}_{(\cdot)}(d')$. Therefore, by Artstein and Wets [1988], Theorem 4.2, Ψ^{μ_0} is upper hemicontinuous.

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

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

□

B.2 Additional Examples

Example 1. *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.