A retrieved-context theory of financial decisions*

Jessica A. Wachter† Michael Jacob Kahana‡

January 8, 2021

Abstract

Studies of human memory indicate that features of an event evoke memories of prior associated contextual states, which in turn become associated with the current event's features. This mechanism allows the remote past to influence the present, even as agents gradually update their beliefs about their environment. We apply a version of retrieved context theory, drawn from the literature on human memory, to explain three types of evidence in the financial economics literature: the role of early life experience in shaping investment choices, occurrence of financial crises, and the impact of fear on asset allocation. These applications suggest a recasting of neoclassical rational expectations in terms of beliefs as governed by principles of human memory.

JEL codes: D91, E03, G02, G11, G12

*We are grateful to Ernst Fehr, Cary Frydman, Nicola Gennaioli, Robin Greenwood, Valentin Haddad, Ulrike Malmendier, James Paron, Nikolai Roussanov, Andrei Shleifer, Alp Simsek, Ansgar Walther, Michael Woodford, and seminar participants at Arizona State University, the California Institute of Technology, Imperial College, Northwestern University, Wharton, University of Toronto, the AFA meetings, the Miami Behavioral Finance Conference, the NBER Behavioral Finance Meeting, the NBER Summer Institute, the Red Rock Finance Conference, and the Sloan/Nomis Workshop on the Cognitive Foundations for Human Behavior for helpful comments.

†Department of Finance, The Wharton School, University of Pennsylvania and NBER; Email: jwachter@wharton.upenn.edu

‡Department of Psychology, University of Pennsylvania; Email: kahana@psych.upenn.edu

1 Introduction

Standard decision-making under uncertainty starts with a probability space and an information structure. The information structure implies that the agent associates a value with every subset of the space and then maximizes expected utility. This is the approach of Savage (1954). The difficulty that agents have in forming beliefs over an entire state space has been formulated in the Ellsberg paradox (Ellsberg, 1961), ambiguity aversion formalized by Gilboa and Schmeidler (1989) and in the alternative representations of choice as a probabilistic selection among a small set of alternatives, due to Luce (1959) and McFadden (2001).[1] The set of possible states of nature is impossibly large and ever-changing. Nonetheless, we as individuals do manage to make decisions under uncertainty.

In this paper, we propose a memory-based model of decision-making under uncertainty. A wealth of data support the idea of a human memory system that maintains a record of associations between experiential features of the environment, and underlying contextual states (Kahana, 2012). This record of associations, together with inference about the current contextual state, constitutes a belief system that could potentially affect any kind of choice under uncertainty. This belief system responds to the current environment through retrieved context. The mechanism of retrieved context is how memory "knows" what information is most relevant to bring forward to our attention at any given time. At the same time, any new experience, and the context itself, is then stored again in the memory system (Howard and Kahana, 2002).

This paper applies these concepts to puzzles in asset pricing and portfolio choice that defy the standard Bayesian paradigm. Chief among these are the result that life experience has near-permanent effects on financial decision making (Malmendier and

[1]The problem of determining the underlying state space continues to be a point of contention in recent literature on ambiguity aversion: see, for example, the debate concerning rectangularity of the model set (Epstein and Schneider, 2003; Hansen and Sargent, 2018).

Nagel, 2011, 2016; Malmendier et al., 2017; Malmendier and Shen, 2018), and that an exogenous cue, such as a horror movie, can influence financial decisions (Guiso et al., 2018). We also apply the framework to understanding the sudden onset of the financial crisis, and to over and under-reaction more generally.

When making a decision, an agent is confronted by certain *features* of the environment. The agent connects features over time through *context*. Context is an internal mental state that endows the agent with an understanding of possibly latent aspects of the environment that are relevant for the decision at hand. We will think of context as assigning probabilities to the underlying states of nature, and at that point proceed in a manner similar to the standard economic approach. A natural benchmark is the Bayesian model in which the agent learns about the unobserved state from observable features. Principles of memory, however, can lead context to evolve in ways that are distinctly non-Bayesian.

Whereas many applications of psychological principles to economic decision making have focused on cognitive biases such as loss aversion and narrow framing (see Barberis (2013)), or on limited attention (see Gabaix (2019)) the literature on human learning and memory offers a different perspective. Three major laws (first articulated by Aristotle) govern the human memory system: similarity, contiguity, and recency: Similarity refers to the priority accorded to information that is similar to the presently active features, contiguity refers to the priority given to features that share a history of co-occurrence with the presently active features, and recency refers to priority given to recently experienced features. All three "laws" exhibit universality across agents, feature types, and memory tasks and thus provide a strong basis for a theory of economic decision-making.

While few economic models explicitly incorporate these laws, there are exceptions. Gilboa and Schmeidler (1995) replace axiomatic expected utility with utility computed using probabilities that incorporate the similarity of the current situation to past situations. Mullainathan (2002) proposes a model in which agents tend to remember those past events which resemble current events, and where a previous recollection increases the likelihood of future recollection. He applies the model to the consumption-savings decision. Nagel and Xu (2018) show that a constant-gain learning rule about growth in dividends can explain a number of asset pricing puzzles; they motivate this learning rule using the memory principle of recency. Recency-bias is present also in models of extrapolative expectations (Barberis et al., 2015) and in natural expectations (Fuster et al., 2010). These models do not employ context-based retrieval, which is the focus of our paper. Bordalo et al. (2019) develop a model based on the geometric similarity of representations in memory. They focus on the the role that similarity in memory representations plays in accounting for the propensity of agents to make large expenditures on housing or durable goods when lower expenditures would appear optimal by standard theory. Their work differs from ours in that we focus on the retrieval of prior contextual states, and we directly model contextual evolution. In their model, as in psychological studies such as Godden and Baddeley (1975), context is embedded in the environment, and thus is static; the feature layer of the environment and the context layer are the same.

The remainder of the paper is organized as follows. Section 2 describes the model and derives general properties. Section 3 describes the psychological and neural basis for the model. Section 4 discusses applications to problems in economics and finance. Section 5 briefly describes alternative approaches. Section 6 concludes.

2 Integrating Memory into Decision Making

Section 2.1 describes the economic setting. Section 2.2 outlines assumptions on memory and shows how Bayesian updating emerges as a limiting case. Section 2.3 derives general properties of the model, temporarily abstracting from some aspects of the decision problem in Section 2.1. Section 2.4 discusses features retrieval, which links memory to the decision problem in Section 2.1. Finally, Section 2.5 establishes temporal contiguity, a key property for the applications that follow.

2.1 Economic setting

We consider an agent who develops memories by experiencing events across time. We represent these events by a discrete-time stochastic process Y_t. We use the notation $\{Y_t\}$ to denote this process and assume that Y_t takes on values in a finite set $\mathcal{Y} = \{y_1, \ldots, y_n\}$. At times, it will be useful to define the physical process for Y_t. To this end, we assume there exists a persistent latent process Z_t taking on values in a finite set $\mathcal{Z} = \{z_1, \ldots, z_m\}$. Let p^Z_{ik} denote the probability of transition from state i to state k: $p^Z_{ik} \equiv \text{Prob}(Z_{t+1} = z_k \,|\, Z_t = z_i)$. Let $p(y_j|z)$ denote the probability that $Y_t = y_j$ conditional on $Z_t = z$ for $j = 1, \ldots, n$. It is as if nature delivers a set of persistent states about which the agent can partially learn through observation. This is a standard set-up in macroeconomics and finance (Hamilton, 1994; Sims, 2003). We depart from the standard set-up in that we do not assume ergodicity, nor do we we assume inference based on knowledge of p^Z or $p(\cdot|z)$

We consider a static decision problem under uncertainty. In each period, the agent makes a choice denoted by π, perhaps subject to constraints. We assume that the agent has a utility function, denoted by V that depends on π and on the outcome of the state of the world next period. An example is a one-period portfolio allocation problem in which π is the allocation to a risky asset. Our benchmark assumption is as follows:

Assumption 1. *The agent solves*

$$\max_{\pi} \mathbb{E}^Y_t \left[V(Y_{t+1}, \pi) \right] \tag{1}$$

where $\mathbb{E}_t^Y$ *is the time-t subjective expectation over* Y_{t+1}.

The expectation $\mathbb{E}_t^Y$ is the subject of our paper. By assuming that the utility function is over outcomes Y_{t+1} we require that only outcomes traced to observables matter to the agent.[2] While a notion of memory requires a multiperiod agent, we focus on the static problem, leaving aside the question of the agent's view of future self as beyond the scope of this paper.

2.2 Human memory

A standard approach to the problem outlined above is to endow an agent with a system of prior beliefs on the joint dynamics $\{Y_t, Z_t\}$. Based on this prior and on the data, the agent can infer a posterior distribution over unknown quantities of interest such as transition probabilities and latent states. Depending on how restrictive one makes the agent's prior distributions, the problem becomes more or less well-identified, with an unavoidable tradeoff between bias and precision. Thus the agent's inference problem is a difficult one.

The literature on human memory offers an alternative approach to the problem of decision-making under uncertainty. Our assumption, formalized below, is that agents use their memory to guide inference. For the remainder of this section and the next, we focus on the problem of memory and beliefs, returning to the actual decision problem in Section 2.4 and Section 4. We draw on the vast literature describing the influence of past experience on present behavior, a topic that has occupied the attention of experimental psychologists for more than a century (Ebbinghaus, 1913; Müller and Pilzecker, 1900; Jost, 1897; Müller and Schumann, 1894; Ladd and Woodworth, 1911; Carr, 1931). Because memories of recent experiences readily come to mind, this early work sought to uncover the factors that lead to forgetting. Experimental findings quickly challenged

[2]Assumption 1 implies context does not play a direct role in utility. The agent is only influenced by, say, mood through what mood tells the agent about the distribution of future features.

the folk assertion that memories decay over time, eventually becoming completely erased. Rather, they revealed that removing a source of interference, or reinstating the "context" of original learning, readily restored these seemingly forgotten memories (McGeoch, 1932; Underwood, 1948; Estes, 1955).

The original view of context was that it consisted of latent, background information unrelated to the present stimulus. Experiments often treated context as an aspect of the external environment. In contrast to this early work, Howard and Kahana (2002) proposed a model in which context became a *mental representation* of what had formerly been thought of as the physical environment. Context went from being an external physical concept to the internal state of the agent. In the Howard and Kahana theory, which we term *retrieved context theory*, the set of psychological (or neural) features that represent a stimulus enter into association this with internal mental state. The database of such associations form the basis for performance in recall, recognition, and categorization tasks. Subsequent work, e.g. Polyn et al. (2009) Lohnas et al. (2015), has thus emphasized the view of context as an internal process, evolving endogenously based on the stimuli which the agent encounters.

In a free recall experiment, the researcher presents subjects with a list of of items, often words, which they can recall in any order. We draw an analogy from this list of items to the features that nature presents to the agent (an analogy that is implicit in the notion of the free recall experiment). Howard and Kahana (2002) model these features (words) as basis vectors in a large n-dimensional space f_t. Their key innovation is the idea of a mental context that links these features through time. Following their work and the subsequent line of research, we model context as a norm-1 vector of length m that evolves based on past context and current features:

$$x_t = (1 - \zeta)x_{t-1} + \zeta x_t^{\text{in}}, \tag{2}$$

where x^{in} (the "in" stands for input) is the aspect of context that arises (is retrieved) from the current environment, and where ζ lies between 0 and 1. Note that context follows a vector auto-regression (VAR).[3]

A defining assumption of retrieved context theory is the manner in which the agent retrieves context from the environment. The agent forms a record of associations between features and the internal context, and at each time t, stores this record in an $m \times n$ matrix M_t (memory). Retrieved context arises from multiplying current features f_t with the memory matrix representing associations as of the previous period:[4]

$$x_t^{\text{in}} \propto M_{t-1} f_t.$$

It is convenient to scale x^{in} so its elements sum to one:

$$x_t^{\text{in}} \equiv \frac{M_{t-1} f_t}{||M_{t-1} f_t||}, \tag{3}$$

where, unless stated otherwise, $\| \cdot \|$ denotes the sum of the elements in the vector.[5] The current state for the agent is then summarized by M_t and x_t.

Associations themselves arise from past associations and the *outer product* between current context and features:

$$M_t = M_{t-1} + x_t f_t^\top. \tag{4}$$

[3]Formally, features f_i are elements of $\mathcal{B}^n \subset \mathbb{R}^n$, where $\mathcal{B}^n$ is a set of basis vectors that spans n-dimensional space. Context is an element of $\mathcal{A}^m \subset \mathbb{R}^m$, for $m \leq n$. That is $\mathcal{A}^m = \{x_t = [x_{1t}, \ldots, x_{mt}]^\top \in \mathbb{R}^m \,|\, \iota^\top x_t = 1\}$, where ι denotes a conforming vector of ones. We will have use for features that are not basis vectors, in which case they are elements of the unit circle in n-dimensional space.

[4]The symbol $\propto$ denotes equality up to multiplication by a positive scalar. Its use implies that we only care about the magnitude of the elements of the vector relative to one another, not in absolute terms.

[5]Because all the vectors we consider have non-negative entries, $\| \cdot \|$ is a valid distance measure; in fact it is distance under the L^1-norm. The memory literature, e.g. Polyn et al. (2009), uses the L^2-norm, with $x_t = \rho_t x_{t-1} + \zeta x_t^{\text{in}}$ and $\rho_t \approx 1 - \zeta$, to maintain x_t on the unit circle.

To complete the model, the agent must possess initial associations M_0. How the agent comes by these associations is beyond the reach of this paper.

To understand the implications of (3) and (4), consider what happens when the agent is cued with features f_t. This cuing will recover all the contexts previously associated with those features. Retrieved context equals:

$$\begin{aligned} x_t^{\text{in}} &= \frac{M_{t-1} f_t}{||M_{t-1} f_t||} \\ &\propto M_0 f_t + \sum_{s=1}^{t} (x_s f_s^\top) f_t \\ &\propto M_0 f_t + \sum_{s=1}^{t} x_s (f_s^\top f_t). \end{aligned} \tag{5}$$

A benchmark case (Assumption 2 below) has $f_s^\top f_t$ equal either to zero or one. It is zero if $f_s \neq f_t$; it is one if $f_s = f_t$. Equation 5 shows that features evoke the past contexts under which they are experienced. A context x_s appears in the sum in (5) if the corresponding f_s equals f_t; otherwise it does not. The interpretation remains valid even when f_s are not orthonormal basis vectors. Even when features are not orthonormal, (5) is an average of past contexts under which similar features were experienced.

According to retrieved context theory, context determines what an agent is most likely to remember. Figure 1 illustrates the mechanism. The current state of context contains a component that overlaps with the contexts of recent experiences, and a retrieved context component that overlaps with items experienced close in time to the just-recalled item(s). The figure illustrates these two effects as spotlights shining down on memories arrayed on the stage of life. Memories are not truly forgotten, but just obscured when they fall outside of the spotlights. Section 3 discusses the implications of this model for memory, and in particular for the temporal contiguity property in the introduction.

The theory developed by Howard and Kahana (2002), Polyn et al. (2009), and Lohnas et al. (2015) among others treats memory as an outcome of a mechanistic process. There is no decision-maker maximizing an objective function, nor is there an underlying probability space on which such an agent would form beliefs. The first step in mapping their framework to decision-making under uncertainty is to formally connect features with observable aspects of the environment:

Assumption 2. *Features are characterized by an $n \times 1$ column vector f_t such that*

$$f_t(j) = \begin{cases} 1 & \textit{if} \quad Y_t = y_j \\ 0 & \textit{otherwise.} \end{cases}$$

That is, $f_t = e_j$, the jth standard basis vector in n-dimensional space, where j corresponds to the state of Y_t.

Unless stated otherwise, we assume features come from the underlying physical environment (and thus are basis vectors) — namely that Assumption 2 holds. It will at times be useful to relax Assumption 2, and consider features arising, for example, in an experiment, and also as part of the agent's recollections. We use the notation e_j to denote the jth standard basis vector in $\mathbb{R}^n$, representing features. When context is a basis vector, we use the notation $\hat{e}_i$ to denote the ith standard basis vector in $\mathbb{R}^m$.

The second step in linking memory to decision-making is to relate the contents of memory to subjective probabilities. Such a link is implicit in the model as defined in the memory literature, and becomes explicit under a simple benchmark in which context is fully observed. In this case the contents of M_t equal Bayesian unconditional probabilities.

Theorem 1. *Assume context corresponds to the underlying state Z_t, and that it is fully observable. Then M_t correctly encodes posterior probabilities of each state.*

If x_t and f_t equal the ith basis vector and jth basis vector respectively, then $x_t f_t^\top$ represents exactly one occurrence of state $Z_t = z_i$ and $Y_t = y_j$. At time t, the ijth entry of M_t consists of an initial value, plus the number of times z_i and y_j occurred together up to time t. Overall, then, elements appear in M_t in proportion to their relative occurrence in the agent's environment. This intuition extends to the case in which x_t (or even f_t) is not a basis vector; memory stores a fraction of an observation. For example, if x_t places weight 1/2 on the first two entries, then $x_t f_t^\top$ is as if the agent has seen "fractional" occurrences of the first two states co-occurring with state j.

Now suppose context is latent and evolves endogenously. Equation 4 still implies that the agent stores joint occurrences of context and features. Supposing that $\zeta = 1$, the agent will retain the incorrect associations, regardless of how much data are observed:

Theorem 2. *Assuming $\zeta = 1$, M_t evolves such that the relative magnitudes of the elements in each column remain invariant while the magnitudes across columns change. That is, the agent correctly estimates the occurence of features, but incorrectly associates features with context.*

In the special case of $\zeta = 1$,

$$x_{t+1} = x_{t+1}^{\text{in}} = \alpha M_t f_{t+1},$$

where α is a positive scalar of proportionality. Consider the formation of the new association:

$$\begin{aligned} M_{t+1} &= M_t + x_{t+1} f_{t+1}^\top \\ &= M_t + \alpha M_t \left(f_{t+1} f_{t+1}^\top \right). \end{aligned} \tag{6}$$

Suppose, for example, that $f_{t+1} = e_1$, the first basis vector. Then $f_{t+1} f_{t+1}^\top$ is the matrix with 1 in the first diagonal element and zero otherwise. Equation 6 takes the first column of M_t and multiplies it by $1+\alpha$, leaving the rest of the matrix unchanged.

Consider the intuition behind Theorem 1: the agent experiences events and stores them in memory. Similar reasoning is at work in Theorem 2: the agent "experiences" joint occurrences of x_t and f_t and stores them in memory (the agent's thoughts become data). However, x_t is not reality; rather it is context that is retrieved based on prior associations from features. The agent nonetheless stores it as an event in memory. The agent correctly learns the frequency of outcomes of $\{Y_t\}$, but matches this frequency incorrectly to the underlying states.

This intuition extends to $\zeta < 1$. Note that columns of M translate directly into which context is retrieved; for this reason, we describe the more general result directly in terms of retrieved context. Rather than being equal at all times, retrieved context decays at a rate that is slower than exponential, and at an ever slower rate as time goes by. The decay takes place in "event time," not in calendar time. For example, consider, as we will in Section 4, the experience of stock market losses. Suppose these take place at a sequence of times $t_1, t_2, \ldots$. For the first loss that the agent experiences, retrieved context is determined by the initial matrix M_0. For the second loss,

$$x^{\text{in}}_{t_2} = \left(1 - \frac{1-\zeta}{2}\right) x^{\text{in}}_{t_1} + \frac{1-\zeta}{2} x_{t_2-1},$$

where x_{t_2-1} is the context just before the loss. We generalize this to an arbitrary number of losses in the following theorem:

Theorem 3. *Consider events occurring at a subsequence of times* $\{t_1, t_2, \ldots, t_\ell, \ldots\}$*. Assumption 2 implies that retrieved context follows a VAR:*

$$x^{\text{in}}_{t_\ell} = \left(1 - \frac{1-\zeta}{\ell}\right) x^{\text{in}}_{t_{\ell-1}} + \frac{1-\zeta}{\ell} x_{t_\ell - 1} \tag{7}$$

for $\ell > 1$. For the first occurrence of the event ($\ell = 1$), $x^{\text{in}}_{t_1} \propto M_0 e_i$, where e_i is the basis vector corresponding to the event.

Assuming that prior to each event, an agent is in a "neutral" context, Then beliefs decay toward this context at a rate that is slower than exponential, and time-scale invariant (decay operates on event time not on calendar time).[6] This is similar to the long-term recency effects modeled by Nagel and Xu (2018).

The next assumption links context directly to subjective probabilities.

Assumption 3. *At time t, the agent assigns the probability $x_t(i)$ to state i, and acts as if this probability is permanent.*

There are two parts to Assumption 3. The first is that the agent's subjective probability of state i at time t is $x_t(i)$. That there should be some link between context and beliefs is highly plausible. The second (namely the permanence of context) pertains to the agent's view of future self. Behind Assumption 3 is the idea that what memory brings to mind is the basis of probabilities that inform financial decisions. Recent experimental work in finance (Gödker et al., 2019; Enke et al., 2020) suggests that this is the case.

Assumption 3 is consistent with Bayesian updating, under the view that M_t contains unconditional probabilities.

Theorem 4. *Consider a Bayesian agent who assumes $\{Z_t\}$ is iid and who interprets M_{t-1} as containing (known) probabilities of joint occurrences of $\{Z_t\}$ and $\{Y_t\}$. Then retrieved context (3) is the conditional probability of Z_t given Y_t.*

While Theorem 4 technically requires an iid process (if Z_t were not iid, then the agent would use additional information from previous Y_t values in inferring the current value), it nonetheless points to the dual role of memory: Besides maintaining a record of

[6]In the memory literature, "neutral" refers to features or contexts without positive or negative emotional valence (Long et al., 2015).

past experiences (Theorem 1), memory also allows for the most important information to come to mind.

Recall that one of the three Aristotelian laws is similarity: better memory for an item that is similar to the just-remembered item. The modern memory literature has concerned itself with similarity since the beginning of the field (see Woodworth (1938) for a survey of the early literature). The multidimensional nature of the model readily incorporates similarity. The features space is spanned by physical features, represented by basis vectors. Virtual experiences, and even thoughts themselves (which are a form of virtual experience) can be close to these physical features in a mathematically well-defined sense. Features that are close in this high-dimensional space will retrieve similar contexts. In this way, a virtual experience, such as a narrative of a stock market crash, or a movie representing danger, can evoke memories similar to the actual physical experience. The following theorem formalizes this property:

Theorem 5 (Similarity)**.** *Features that resemble one another retrieve similar contexts. Namely, consider a series of features vectors* $\{f_t^n\}$ *with* $\lim_{n\to\infty} f_t^n = f_t$*. Then* $\lim_{n\to\infty} x_t^{\mathrm{in},n} = x_t^{\mathrm{in}}$*, with a rate of convergence independent of* t*.*

The series $\{f_t^n\}$ consists of ever more precise reminders of f_t, and hence an ever more effective retrieval cue for context x_t^{in}. Similarity in this model differs from that in Bayesian models, in which information that has similar content results in similar changes to beliefs. In this case, it is the experience that is similar (there may be no information content in the usual sense). In what follows, we apply this theorem to account for experimental evidence on time-varying risk aversion that is lies outside of the Bayesian domain.

Similarity plays an important role in the geometric model of Bordalo et al. (2019) and in the model of Mullainathan (2002), both of which parsimoniously capture the role of associations in memory. Contextual dynamics and temporal contiguity are points of departure from these studies.

2.3 Implications of autoregressive context

Standard inference models in economics imply that agents have equal access to stored information at all points in time. As Section 5 discusses, this assumption, while convenient, is contrary to experimental data. Retrieved context theory concerns itself with the retrieval of information.[7] This section derives implications of retrieved context theory following from the fact that context is a VAR: neglected risk, extrapolative beliefs, and, under different conditions, over and under-reaction.

Assume a time interval $[t, t']$ during which the agent fails to experience features associated with a given state i.[8] If at time t, the agent places zero probability on state i ($x_t(i) = 0$) then, assuming no crisises between t and t', $x_s(i) = 0$ for any $s \in [t, t']$. More interestingly, assume the agent places a positive probability of crisis at time t ($x_t(i) > 0$). For this agent, experiencing sufficiently diverse non-crisis features will lead to an exponential decline in $x_s(i)$. Following Gennaioli and Shleifer (2018), we refer to the phenomena by which a low-probability event is ignored as *neglected risk*.

Theorem 6 (Neglected risk)**.** *Fix a state i and assume that between time t and t', the agent fails to experience features associated with state i. Further assume mutually orthogonality of features experienced between t and t'. Then the probability the agent places on state i decays exponentially: $x_{t'}(i) = (1 - \zeta)^{t-t'} x_t(i)$.*

The assumption of independent features allows us to focus on the exponential decay of context, suppressing the creation of associations in M. If the agent begins the interval in a crisis context, any features nearby in time will become associated with that context.

[7]We consider decisions in which some human agency is involved; modern technology allows for the mechanical retrieval of vast quantities of information, but one still must know, for example, what model to estimate, or what terms for which to search.

[8]Features that are associated with state i are those that retrieve state i. See Appendix B.

As formerly innocuous features of the environment come to be associated with a crisis, the crisis becomes harder to forget. Strictly speaking, for exponential decay to occur, the agent must experience a sequence of novel features.[9]

For comparison, consider a Bayesian agent who begins with the same prior probability of a crisis. Assume that the crisis is observable.[10] Given non-crisis observations, the decline in the posterior mean probability of a crisis is not exponential, but, the Bayesian case, hyperbolic (Appendix C). We illustrate the difference between exponential decay implied by Theorem 6 and Bayesian updating in Panel A of Figure 3 (Section 4.1 provides more detail). It takes many more observations for a Bayesian agent to reach the same level of certainty that a crisis will not occur, and (assuming a nonzero probability of such events), it is much less likely that the agent will in fact reach this level of certainty. More generally, Equation 2 implies that recently-experienced features are over-represented in the agent's context x_t. This mechanism leads to extrapolative beliefs.[11]

Over-weighting of recently-experienced features also implies that the agent slowly updates beliefs in response to new information, but will eventually update too strongly. The model thus accounts for short-term under-reaction and long-term reversal.[12] Assume a subset of outcomes in $\mathcal{Y}$ that occur if and only if the state is i: $Z_t = z_i \iff Y_t \in \mathcal{Y}_i$, for a nonempty subset $\mathcal{Y}_i \in \mathcal{Y}$ (call this the uniqueness condition). Also assume that the agent's associations at time $t-1$ accurately reflect the underlying correlations (call this correct associations). That is, M_{t-1} is block-diagonal, reflecting the physical correlations.

[9]Orthogonal features, are used in the memory laboratory to reset context. These features consist of "distractor tasks," such as asking subjects to solve arithmetic problems, or to view outdoor scenes (Howard and Kahana, 1999; Manning et al., 2016).

[10]Namely, assume that the space of outcomes $\mathcal{Y}$ partitions into outcomes that can occur during a crisis and those that cannot.

[11]See Barberis et al. (2015) for discussion of evidence on extrapolation in financial markets.

[12]For models and for a discussion of the evidence on under-reaction and reversal, see Daniel et al. (1998), Barberis et al. (1998), and Hong and Stein (1999).

Theorem 7 (Short-run under-reaction; long-term reversal)**.** *Consider a state i satisfying the uniqueness condition and correct associations at t. Assume that t is large, or that $\mathcal{Y}_i$ contains many elements. Then*

1. *If $x_t(i)$ is low relative to p^Z_{ii} (features associated with i are novel or ζ low) positive revisions to beliefs about i tend to be followed by further positive revisions.*

2. *If $x_t(i)$ is high relative to p^Z_{ii} (i.e., the agent has repeatedly experienced features associated with i or ζ high) then beliefs tend to reverse.*

For example, if agents receive positive earnings news on a given company, they will update their beliefs, generating an increase in price (thus far, the model behaves the same as in the full information case). However, further high-growth observations (more likely than not if the state is persistent) will cause further updates in beliefs, and hence positive realized returns. Once the belief in the persistent high growth state is sufficiently high, the agent will by surprised by any shift to the low state.[13] As with Theorem 6, we require an additional technical assumption (that either t is large or $\mathcal{Y}_i$ contains many elements). Under-reaction also characterizes models in which information is lost on the front end, e.g. Mullainathan (2002). In our model, under-reaction occurs because of slow-moving context.

To restate the conclusions from Theorems 6 and 7: agents temporarily forget contexts associated with features that they have not seen for a sufficiently long duration. Agents' beliefs take time to adjust to novel circumstances, but then tend to adjust too much. If the agent knew the underlying persistence, he or she could perhaps adjust context. However, true persistence is unknown. In fact, if states are transient, the agent over-reacts, confusing a conditional probability with an unconditional probability. For simplicity, we assume the limiting case of iid states.

[13]The price momentum effect is the finding that stocks with the highest price appreciation measured over the past 12 months outperform those with the lowest price appreciation (Jegadeesh and Titman, 1993). These gains partially reverse one year later. Macroeconomic expectations appear to under-react and then overshoot Angeletos et al. (2020), as do earnings expectations Bordalo et al. (2020).

Theorem 8 (Over-reaction)**.** *For ζ sufficiently high and for iid $\{Z_t\}$, upward revisions to beliefs about state i reverse on average.*

Theorems 7 and 8 both address the predictability of

$$\Delta x_{t+1}(i) = \mathrm{Prob}_{t+1}(Z_{t+1} = z_i) - \mathrm{Prob}_t(Z_{t+1} = z_i), \tag{8}$$

for $\Delta x_{t+1}(i) = x_{t+1}(i) - x_t(i)$, and where $\mathrm{Prob}_t(\cdot)$ corresponds to the agent's subjective probability, conditional on time-t information. Consider a Bayesian agent who infers state i with certainty (as he or she would in the setting of Theorem 7). The Bayesian agent forecasts that in the next period, state i will obtain with probability p^Z_{ii}, so that $\mathrm{Prob}_t(Z_{t+1} = z_i) = p^Z_{ii}$. This is in fact the actual probability of $Z_{t+1} = z_i$. The difference on the right-hand side of (8) is zero on average under Bayesian updating.

Let $\mathbf{E}$ denote the expectation calculated by the econometrician. Revisions in beliefs equal:

$$\begin{aligned} \mathbf{E}_t[\Delta x_{t+1}(i) | \Delta x_t(i) > 0] &\approx \zeta(p^Z_{ii} - x_t(i)) \\ &\approx \zeta\left(p^Z_{ii} - ((1-\zeta)x_{t-1}(i) + \zeta)\right), \end{aligned} \tag{9}$$

because $x^{\mathrm{in}}_{t+1}(i) = 1$. Consider first the case with p^Z_{ii} high and $x_{t-1}(i)$ low (think of it as zero). This is the setting of the first statement of Theorem 7. Beliefs under-react to news because the persistence of state i is greater than the weight of the new information is context. Slow adjustment of context implies that the agent cannot "take in" all the information at once. Now consider p^Z_{ii} high, $x_{t-1}(i)$ high (think of it as one). This is the setting of Theorem 6 and of the second statement of Theorem 7. If the agent has seen sufficient observations consistent with a state, the agent forgets that other states are possible. Beliefs predictably reverse because of mean-reversion in the state. Finally, consider the case of p^Z_{ii} low, and ζ high. This is the setting of

Theorem 8. If the agent puts relatively high weight on new information, or if the state is transitory, then then the agent over-reacts. In the last two cases, the agent forgets that the world may be different next period. Note that our model generates over-reaction from contextual dynamics; this is contrast to Mullainathan (2002) and Bordalo et al. (2019), in which over-reaction through a mechanism akin to similarity, namely that irrelevant information acts as a cue. Our model also incorporates this latter effect, but over-reaction can occur even when the agent's associations are correct.

While optimality of memory lies outside the scope of this study, one might conjecture that slow context evolution protects against over-reaction. If the state is transitory, then it is indeed inappropriate to react strongly. On the other hand, times might call for an instantaneous shift in perspective, as captured by Theorem 8. Understanding the constraints on storage and retrieval is a prerequisite to a study of optimality of memory; (Azeredo da Silveira and Woodford, 2019) offer a theory of such constraints. A constrained-optimal view of memory might shed light on when there is sufficient information for an agent to act as Bayesian models specify, and when the agent instead must rely on the mechanisms that we emphasize here.

2.4 Features retrieval

The previous analysis abstracts from how agents go from probabilities on underlying states in $\mathcal{Z}$ (given by context) to outcomes in $\mathcal{Y}$. While nature supplies $p(y|z)$, agents must learn this distribution from observations. Subjects in memory experiments must also go from context to a specific memory that is then recorded in the laboratory. That is, just as context is retrieved from features, there is an additional step by which features are retrieved from context.

The same associative principles govern features retrieval as govern context retrieval.

Given a basis context vector $\hat{e}_i$ define retrieved features as

$$f_{i,t}^{\text{in}} = \frac{M_{t-1}^{\top}\hat{e}_i}{||M_{t-1}^{\top}\hat{e}_i||}. \tag{10}$$

Because $\| \sum_{i=1}^{m} x_t(i) f_{i,t}^{\text{in}} \| = 1$, the $n \times 1$ vector

$$f_t^{\text{in}} \equiv \sum_{i=1}^{m} x_t(i) f_{i,t}^{\text{in}} \tag{11}$$

is a subjective probability distribution over features. Note that the same interpretation as in (5) applies: the agent retrieves features experienced under similar contexts.

We assume (11) to stay as close as possible to Bayesian updating. Memory research, however, supports a retrieval rule that is closer to winner-take-all ("stronger" features – those with a higher weight in (11) – inhibit the recall of weaker ones). For example, Howard and Kahana (2002) assume a Euclidean distance rule, whereas Polyn et al. (2009) assume a "leaky accumulator" model. Except where noted, assuming a winner-take-all process does not affect our results below.

The existence of retrieved features that are not the same as physical features raises the question: which are encoded in (4)? The most natural assumption would seem to be that it is the physical features, and we use this assumption in establishing temporal contiguity in Section 2.5. However, while economic models assume agents see reality unfiltered (Woodford (2020) discusses exceptions), neurobiology supports a notion of filtering based on memory-driven expectations (Reynolds and Chelazzi, 2004; Makino and Komiyama, 2015). What is retained in each re-remembering is not exactly what occurred, but rather a distorted copy of the original event (Rubin et al., 2008). Experimental evidence supports encoding of retrieved features at study (Greene, 1992; Siegel and Kahana, 2014), and at test (Zaromb et al., 2006; Howard et al., 2009; Miller et al., 2012; Kuhn et al., 2018). Encoding of retrieved features is a form of rehearsal, as discussed by Mullainathan (2002). In our model, however, the effect of rehearsal is permanent as opposed to short-term (see the discussion in Section 3). We will incorporate encoding of retrieved features into the applications (Section 4) that follow.

The process of features retrieval and encoding is how disparate features become "glued" together. As we show in the following section, it is key to understanding temporal contiguity and the jump back in time.

2.5 Temporal Contiguity

Temporal contiguity serves as a fundamental organizing principle of the memory system (Healey et al., 2019). In this section we show, by means of an example, how the model accounts for temporal contiguity effects. We consider a stylized model of the Great Depression, in which an economic collapse follows a financial crisis. Subsequent to the Great Depression, many narratives emphasized the importance of the stock market crash of 1929, runs on financial institutions, and the connection between these purely financial events and the subsequent period of high unemployment and sharply decreasing output and consumption from which the Great Depression received its name. But from where did these narratives arise in the first place?

Let x_{1929} denote context as of 1929, which we refer to as time $t-1$. Let f_{crisis} be features associated with the failure of a financial institution, also occurring at $t-1$. Equation 4 implies that this combination of features and context become associated in memory:

$$M_{t-1} = M_{t-2} + x_{1929} f_{\text{crisis}}^{\top}. \tag{12}$$

Let $f_{\text{depression}}$ be features associated with depression occurring in the next period t. Call $x_{\text{depression}}^{\text{in}}$ the context retrieved by features $f_{\text{depression}}$:

$$x_{\text{depression}}^{\text{in}} \propto M_{t-1} f_{\text{depression}}. \tag{13}$$

Retrieved context $x^{\text{in}}_{\text{depression}}$ is the agent's state of mind when confronted with the observable features of the Great Depression, such as mass unemployment. Calling this state of mind $x^{\text{in}}_{\text{depression}}$ is simply terminology. We make no assumptions on the utility consequences or associations other than what follows from the model. Note that we assume the crisis strictly precedes depression and that we make no assumptions on a prior link between x_{1929} and $x^{\text{in}}_{\text{depression}}$. Features f_{crisis} and $f_{\text{depression}}$ are orthogonal, so that even though M_{t-1} appears in both (13) and (12), the occurrence of the crisis nothing to do with the retrieval of $x^{\text{in}}_{\text{depression}}$.[14]

Context evolution (2) implies that the retrieved depression context combines with previous context, to create the current context:

$$x_{1930} \equiv x_t = (1-\zeta)x_{1929} + \zeta x^{\text{in}}_{\text{depression}}, \tag{14}$$

where x_{1929} is the time-$(t-1)$ context and x_{1930} is the time-t context.[15] Crucially, x_{1930} is a weighted average between $x^{\text{in}}_{\text{depression}}$ and x_{1929}, even if the events leading to the retrieval of $x^{\text{in}}_{\text{depression}}$ had nothing to do with the events of x_{1929}. From (4), it follows that

$$M_t = M_{t-1} + x_{1930} f^{\top}_{\text{depression}}. \tag{15}$$

Thus far, it is clear that $f_{\text{depression}}$ and x_{1930} are associated, as are f_{crisis} and x_{1929}. What is not yet clear is how f_{crisis} relates to $f_{\text{depression}}$.

Suppose f_{crisis} appears at some future time $t' > t$. As described in Theorem 8, there

[14]More precisely, $x^{\text{in}}_{\text{depression}}$ would be retrieved whether or not the crisis had occurred.

[15]For simplicity, we refer to this as context in 1930, even though unemployment rose throughout the early 1930s.

is a jump back in time – the agent retrieves x_{1929}:

$$\begin{aligned} x_{t'}^{\text{in}} &\propto M_{t'-1} f_{\text{crisis}} \\ &\propto \left(M_0 + \sum_{s=1}^{t-2} x_s f_s^\top + x_{1929} f_{\text{crisis}}^\top + \sum_{s=t}^{t'-1} x_s f_s^\top \right) f_{\text{crisis}} \qquad (16) \\ &\propto x_{1929}. \qquad (17) \end{aligned}$$

To obtain (17) from (16), we assume f_{crisis} only appears once – in 1929, and that there are no other prior associations with f_{crisis}. These assumptions simplify the algebra without changing our conclusions, and we relax them in Appendix D. In any case, a financial crisis retrieves x_{1929} – a crisis retrieves the context of the previous crisis. It does not retrieve x_{1930}. And yet the agent believes a depression is imminent. Why is this?

The reason lies with the features retrieved by x_{1929}:[16]

$$\begin{aligned} f_{t'}^{\text{in}} &\propto M_{t'-1}^\top x_{1929} \\ &\propto M_0^\top x_{1929} + \sum_{s=1}^{t'-1} (f_s x_s^\top) x_{1929} \\ &\propto \underbrace{\left(M_0^\top + \sum_{s=1}^{t-2} f_s x_s^\top \right) x_{1929}}_{\text{prior associations}} + f_{\text{crisis}} + \underbrace{\sum_{l=1}^{t'-t} (1-\zeta)^l f_{t+l-1}}_{\text{depression features}} + \cdots \qquad (18) \end{aligned}$$

The re-appearance of the crisis retrieves features associated with x_{1929} prior to the actual crisis (the first term in (18)). They retrieve f_{crisis}, because x_{1929} is the very context under which the crisis was experienced. Most importantly, they retrieve $f_{\text{depression}}$ because x_{1929} was part of context (14) at the time of depression. The time–t term in

[16] To simplify the algebra, we assume that $x_{t'} \approx x_{t'}^{\text{in}}$ as will be the case if crisis features are persistent. We also assume $x_{t'}$ is a basis vector. We relax these assumptions in Appendix D.

(18) equals

$$
\begin{aligned}
f_t x_t^\top x_{1929} &= f_{\text{depression}} x_{1930}^\top x_{1929} \\
&= f_{\text{depression}} \left((1-\zeta) x_{1929} + \zeta x_{\text{depression}}^{\text{in}} \right)^\top x_{1929} \\
&= (1-\zeta) f_{\text{depression}}
\end{aligned}
$$

Thus re-appearance of a crisis retrieves the depression. If the depression features continue for more than one year, as in fact occurred, the crisis context re-instates these as well, with geometrically declining weights.

Because we do not take a stand on whether context x_{1929} is retrieved at some future time, there is a potential for additional terms in (18). That is, x_s^{in}, for $s = t+1, \ldots, t'-1$ might be correlated with x_{1929}. If, for example, a financial crisis occurred again while the agent was already in context x_{1930}, that would lead to additional terms in (18) containing $f_{\text{depression}}$ and would strengthen the associations between depressions and crises. If, on the other hand, a crisis occurred during a context very different from x_{1929}, the additional terms in (18) would be orthogonal to $f_{\text{depression}}$, and the association would weaken.

Equation 18 shows that the probability distribution over future features places weight on $f_{\text{depression}}$. Suppose one were to compare the weight on $f_{\text{depression}}$ retrieved by x_{1929} versus a retrieval of the depression itself (x_{1930}). We find

$$
f_{\text{depression}}^{\text{in}} \propto \underbrace{\left(M_0^\top + \sum_{s=1}^{t-2} f_s x_s^\top \right) x_{1930}}_{\text{prior associations}} + (1-\zeta) f_{\text{crisis}} + \underbrace{\sum_{l=1}^{t'-t} (1-\zeta)^{l-1} f_{t+l-1}}_{\text{depression features}} + \cdots \tag{19}
$$

Note that a depression context, like the x_{1929}, retrieves the crisis, and it retrieves depression, though with slightly different weights. For example, the weight on $f_t = f_{\text{depression}}$ in (18) is $1 - \zeta$ times its value in (19). At the heart of this similarity is

(2), contextual evolution implies that contexts close in time must be similar. Though in this example a crisis need not occur in a depression state, the agent's subjective beliefs are as if it does. In effect, contexts x_{1929} and x_{1930} place similar weights on the underlying states in $\mathcal{Z}$ because of the features they retrieve. Financial crises and depressions become associated simply because they were associated in time. Because context is a vector autoregression, a financial crisis pulls up all events that occurred around 1929, including the Great Depression. A crisis leads the agent to think about depression features, which is to say in the language of traditional economics that the agent gives them a high probability.

It is useful to compare the role of associations in this model to a Bayesian one. Consider the reasoning that lies behind the formation of subjective Bayesian probabilities. The agent conceptualizes states of the world. The agent knows features that occur in each state of the world. There is a concept of events occurring "at the same time" – and yet in reality hardly anything occurs truly contemporaneously. If the agent's conceptualization of the underlying states happens to be correct, then arrival of one of the features is correctly taken as a signal that others will soon occur, and the agent's model of the world moves closer to the truth. The Bayesian set-up is convenient, but fragile. If the agent's prior did not allow for the correct features to signal a change in states (and there are a great many potential signals), there would be no way for the agent to learn.

Indeed, the "signal" argument requires that, in 1929, the agent foresaw the Great Depression, and then prior to 2008, believed that a signal such as one received in 1929 might recur, foreshadowing the next Great Depression. It seems more likely that in 1929, agents did not place some probability on the Great Depression being imminent; had they, events might have unfolded differently. Thus they did not start with priors that the Bayesian analysis would require. Rather, after the Great Depression occurred, they associated the events (note that our analysis greatly reduces the sensitivity of the prior; M_0 is part of the analysis, but does not prevent the agent from forming other associations). The crisis followed by the depression *created an association that simply was not present before*. Then, after years went by and context shifted, the assumption became that great depressions, and for that matter financial crises, were a thing of the past (accorded zero probability), until events sufficiently similar to 1929 made individuals feel that the Great Depression was about to occur all over again. One might argue that the fact that individuals explicitly considered this possibility, whereas they had not in 1929 is what *prevented* the Great Depression from recurring in 2009.

One might object that this temporal association is unrealistic. Clearly agents should be able to distinguish between reminders and actual shifts in distributions. Simply because an event recurs, or a context changes should not change an agent's perspective. The well-known phenomenon of post-traumatic stress disorder, suggests that this idea is not so easily dismissed. In the next section we discuss the psychological and neural basis for contextual retrieval.

3 Psychological and Neural Basis for Contextual Retrieval

Before turning to specific applications, we summarize the psychological and neural evidence for context as an internal state.

In the memory laboratory, researchers create experiences by presenting subjects with lists of easily identifiable items, such as common words or recognizable pictures. Subjects attempt to remember these items under varying retrieval conditions: these include *free recall*, in which subjects recall as many items as they can in any order, *cued recall*, in which subjects attempt to recall a particular target item in response to a cue, and *recognition* in which subjects judge whether or not they encountered a test item on a study list. In each of these experimental paradigms, memory obeys the classic "Laws of Association" which appear first in the work of Aristotle, and later in Hume (1748). The first of these is *recency*: human subjects exhibit better memory for recent experiences, *semantic similarity*: we remember experiences that are most similar in meaning to those we are currently experiencing, and finally, *temporal contiguity*: we remember items that occurred contiguously in time to recently-recalled items. Although quantified in the memory laboratory, each of these phenomena appears robustly in real-world settings, as described below.

A longstanding and persistently active research agenda in experimental psychology seeks to uncover the cognitive and neural mechanisms that could give rise to these regularities. Students of memory have proposed many hypotheses which they have tested in the laboratory. Some striking findings include the fact that recency and contiguity appear regardless of whether you measure memory for list items presented seconds apart or many minutes apart, or for autobiographical memories separated by days or weeks.

What gives rise to the recency and contiguity effects that appear ubiquitously in both laboratory memory experiments and in our daily lives? One influential class of explanations posits the existence of a fixed-capacity memory buffer, better known as "short-term memory." In such models, retrieval involves two stages: first, subjects report the items maintained in the short-term store; next, they search through long-term memory guided by interitem and context-to-item associations, and subject to interference from similar memories. This model accounts for contiguity owing to the strengthening of interitem associations among items that share time in the short-term store (Kahana, 1996).

Although the short-term memory model produces recency and contiguity in immediate recall, it cannot readily explain why similar recency and contiguity effects appear for experiences that are widely separated in time, and thus neither likely to be present in short-term store at the time of recall, or to have occurred together in short-term store (Howard and Kahana, 1999; Healey et al., 2019).[17] Related to accounts based on short-term memory, neurobiological models of association posit that patterns of brain activity can associate with one another when they co-occur within a short time window governing synaptic plasticity (Abbott and Blum, 1996; Kempter et al., 1999). These models, however, also struggle to explain why robust temporal contiguity appear for temporally separated events.

These findings suggested the alternative retrieved context framework that we extend to economic choice behavior. According to this view, context evolves recursively by adding the retrieved past contexts associated with an item, remembered or experienced, to the prior state of context. The retrieved context will bear similarity to contiguously experienced items, generating the contiguity effect. Because retrieval depends on the *relative* similarities among competing items, strong contiguity effects can appear even for items separated by very long intervals. The same is true for recency effects, in both the model and in the data.

Figure 2 illustrates the temporal contiguity effect (TCE) and how it has provided empirical support for the idea of context retrieval. To measure the effect of contiguity on memory retrieval, researchers examine subjects' tendency to successively recall items experienced in proximate list positions. In free recall, this tendency appears as decreasing probability of successively recalling items f_t and f_{t+lag} as a function of lag, conditional on the availability of that transition (Kahana, 1996). This TCE reaches its maximum at $lag = \pm 1$, but also exhibits a forward asymmetry in the form of higher

[17]Within economics, Mullainathan (2002) could be understood as a model of short-term memory. Items are available to be recalled (in the short-term store) or are not. They enter and exit the short-term store with probabilities determined by associations with current events. However, there is no mechanism by which an item from the distant past can evoke another item simply by virtue of being proximate in time. Nagel and Xu (2018) invoke long-term recency to explain economic phenomena, but do not otherwise employ associativeness.

probability for positive as compared with negative lags. Equations 2–4 generate a forward asymmetry in the contiguity effect because recalling an item reinstates both its associated study-list context and its associated pre-experimental context. Whereas the study-list context became associated, symmetrically, with both prior and subsequent list items, the pre-experimental context became associated only with subsequently encoded list items, leading to a forward asymmetric contiguity effect, as seen in the data.[18]

Figure 1A shows that interitem distraction does not disrupt the TCE. Figure 1B-D shows that the TCE appears robustly for both younger and older adults, for subjects of varying intellectual ability, and for both naïve and highly practiced subjects. Figure 1E shows that the TCE appears even for transitions between items studied on distinct lists, despite these items being separated by many other item presentations. Figure 1F-H shows that the TCE also predicts confusions between different study pairs in a cued recall task, in errors made when subjects attempt to recall an individual list item in response to a sequential cue, and in tasks that do not depend on inter-item associations at all, such as picture recognition (see caption for details). Finally, long-range contiguity appears in many real-life memory tasks, such as recalling autobiographical

[18]Consider the example in Section 2.5, and treat the 1929, 1930 episodes as a "study phase." The model predicts that crises better recall economic contractions than the reverse. Let

$$x_{1929} = (1-\zeta)x_{\text{prior}} + \zeta x^{\text{in}}_{\text{crisis}},$$

where $x^{\text{in}}_{\text{crisis}}$ is comprised of the 1929 stock market crash *and any previous financial crises.* These previous financial crises form the "pre-experimental" context associated with a crisis. The depression then becomes associated with this pre-experimental crisis context as well as with the 1929 crash because they are both part of $x^{\text{in}}_{\text{crisis}}$, $x^{\text{in}}_{\text{crisis}}$ is part of x_{1929}, and x_{1929} is a part of x_{1930} which cooccurs with $f_{\text{depression}}$.

In contrast, there are no means by which previous economic contractions can become associated with crises. Consider (14), where $x^{\text{in}}_{\text{depression}}$ consists of the Great Depression and previous economic contractions. Like crises, these previous economic contractions are part of "pre-experimental" context. However, whereas prior crises become associated with depressions, prior depressions do not become associated with crises. Why? Though these depressions are part of $x^{\text{in}}_{\text{depression}}$, they are orthogonal to x_{1929}, which is the means by which crises and depressions are associated. Thus there are two routes in memory by which a crisis recalls a depression: the 1929 crisis itself, and any previous crisis. However, there is only one route by which a large contraction recalls a crisis, and that is the Great Depression.

memories (Moreton and Ward, 2010) and remembering news events (Uitvlugt and Healey, 2019). These findings argue for a general associative memory mechanism, like context retrieval, that requires neither strict temporal proximity, nor specialized mnemonic strategies.

A second source of data in favor of retrieved context arises from neurobiology. The theory implies that brain states representing the context of an original experience reactivate or replay during the subsequent remembering of that experience. Several studies tested this idea using neural recordings. These studies found that in free recall (Manning et al., 2011), cued recall (Yaffe et al., 2014) and recognition memory (Howard et al., 2012; Folkerts et al., 2018) brain activity during memory retrieval resembles not only the activity of the original studied item, but also the brain states associated with neighboring items in the study list. Thus, one observes contiguity both at the behavioral and at the neural level, with these effects being strongly correlated (Manning et al., 2011). Finally, this recursive nature of the contextual retrieval process offers a unified account of many other psychological phenomena including the spacing effect (Lohnas and Kahana, 2014b), the compound cueing effect (Lohnas and Kahana, 2014a), and the phenomena of memory consolidation and reconsolidation (Sederberg et al., 2011).

Memory theory thus indicates that remembering an item involves a jump-back-in-time to the state of mind that obtained when the item was previously experienced. This reinstatement, in turn, becomes encoded with the new experience and also persists to flavor the encoding of subsequently experienced items. The persistence of the previously retrieved contextual states enables memory to carry the distant past into the future, allowing the contextual states associated with an old memory to re-enter one's life following a salient cue and associate with subsequent "neutral" memories. While the original memory is retained in association with its encoding context, the retrieval and re-experiencing of that memory forms a new memory in association with the mixture of the prior and retrieved context.

One might argue that, while retrieved context theory offers a persuasive account of human memory phenomena, memories need not affect behavior, and still less, conscious decisions such as how much to invest in the stock market. While evidence on the role of experience in economic decision-making suggests otherwise, one might still argue that experience operates through a conscious process of attaining knowledge, rather than memory per se. Such a purely rational account would, however, miss important memory phenomena.[19] Evidence shows that agents re-live events from both the remote and recent past, often involuntarily (Rubin and Berntsen, 2009). If memory is a process of knowledge accumulation, it appears to be one that is outside of conscious control. An extreme example of the power of involuntary memory is post-traumatic stress disorder (PTSD), in which a traumatic event is not only "persistently reexperienced" but "causes clinically significant distress or impairment in social, occupational, or other important areas of functioning."[20] Studies show that PTSD is diminished when brain injury, childhood amnesia, or pharmacologically-induced amnesia blunts encoding, indicating that it is primarily a memory disorder (Rubin et al., 2008). As such, it exhibits patterns that are well-accounted for by retrieved context theory (Cohen and Kahana, 2020). Overall, evidence on PTSD suggests that it is best understood in terms of principles that govern "normal" memory functioning (Rubin et al.). In other words, there is no clear line separating trauma-induced and normal memories. It appears that people relive the past involuntarily and unawares, to the extent that they base their behavior on a biased representation of the external environment. In what follows, we show how this idea can account for economic phenomena that are difficult to explain otherwise.

[19]Though such an account would also have to explain why agents base their decisions on their *particular* experience.

[20]See the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision.; APA, 2000, pp. 467–468).

4 Applications

This section describes three applications of our theory. Section 4.1 describes an application to portfolio allocation, illustrating how long-run stock-market experience might influence portfolio choice. Section 4.2 shows how memory dynamics might effect stock prices and interest rates in an otherwise standard macro-finance model in which circumstances lead an agent to recall a rare event such as an economic depression. Lastly, Section 4.3 shows how the model can account for the effects of changes in short-term context, thus explaining observed experimental effects on portfolio choice.

In each of these sections, rather than modeling the full lifetime of an agent's memories, we assume a self-contained decision problem. We follow the memory literature in making an assumption on the matrix M_t prior to the decision problem at hand. Associations represented by M_t are motivated by the temporal contiguity property (Section 2.5). That said, we do not generate the prior M_t within the model. In particular, we require a memory matrix that is sparser than lifetime simulations of the process (2), (4) and (10) would imply. Augmenting the model with costly storage (Azeredo da Silveira and Woodford, 2019) could endogenously generate such sparsity while maintaining the model's ability to account for temporal contiguity. Encoding of retrieved features, generated using a process that downweights low probability items also "organizes" memory, leading to a sparser M_t as noted by Polyn et al. (2009). Wachter and Kahana (2020) show that winner-take-all retrieval and encoding leads to a more organized M_t.

4.1 Retrieved-context theory and the persistence of beliefs

A basic account of the behavior of financial markets and the macroeconomy requires that agents disagree (Lucas, 1975; Grossman and Stiglitz, 1980). Such disagreement poses a problem for standard Bayesian models in which agents begin with possibly different priors, but nonetheless see the same data and take a rational view of others' beliefs beliefs. Recent evidence linking economic decisions to lifetime experience suggests that experience may be the place to look for understanding how disagreement arises and what causes it to persist (Malmendier and Nagel, 2011, 2016). The explanatory power of experience immediately suggests a role for memory.

We focus on the results of Malmendier and Nagel (2011), who show that experienced stock returns affect portfolio decisions, because the departure from rationality is particularly striking. The size of the equity premium – the expected return on stocks over Treasury bills – is discussed in finance textbooks, the media, and in popular books. Provided that the equity premium is positive, participation in the stock market is optimal (Arrow, 1971). Yet a large percentage of households do not participate in equity markets (Campbell, 2016). In what follows, we show how the theory in Section 2 can generate, based on life experience, permanent pessimism regarding stock returns. While survey evidence suggests agents are on average pessimistic (Goetzmann et al., 2017), the aim in this section is not to explain average pessimism, but why some agents remain pessimistic in the face of contrary data. This is what is required to account for non-participation.

4.1.1 The portfolio choice problem

Assume latent states $\mathcal{Z} = \{z_1, z_2\}$, and let $p < 1/2$ denote the unconditional probability of the adverse state z_2, which we refer to as a depression. The investor allocates wealth between a risky asset with net return $\tilde{r}$, and a riskfree bond with zero net return. The agent also receives risky labor income $\tilde{\ell}$. The set $\mathcal{Y}$ thus consists of joint outcomes of labor income and stock returns. Assume that the value of labor income is known given the state: $\ell(z_1) > \ell(z_2) = 0$. Assume that the stock return $\tilde{r}$ takes on values r_g (gain) and $r_l < r_g$ (loss). Assume $\text{Prob}(r_g|z_1) = \frac{1}{2}(1-p)^{-1}$ and $\text{Prob}(r_g|z_2) = 0$ so that the marginal distribution of $\tilde{r}$ is 50% gains and 50% losses. Because gains do not occur in state 2, their probability is slightly elevated under the normal state z_1.

The agent prefers more wealth to less, and is risk averse. Mean-variance utility tractably captures these preferences (Markowitz, 1952). We assume the agent solves:

$$\max_{\pi} \mathbb{E}[(1+\pi\tilde{r}+\tilde{\ell})] - \frac{1}{2}\mathbb{V}(1+\pi\tilde{r}+\tilde{\ell}), \tag{20}$$

where π is the percent allocation to the risky asset. The expectation and the variance in (20) are with respect to the agents' subjective preferences.[21] Setting the derivative of the objective function with respect to π equal to zero leads to

$$\pi = \frac{\mathbb{E}\tilde{r} - \mathrm{Cov}(\tilde{r},\tilde{\ell})}{\mathbb{V}(\tilde{r})}. \tag{21}$$

Stocks are unattractive because they deliver a low return in the negative labor income state.

Normalize the mean of $\tilde{r}$ to 1, implying $r_g = 1+\sigma$, $r_\ell = 1-\sigma$, where σ is the standard deviation of $\tilde{r}$. Let $\tilde{\ell}(z_1) = \ell > 0$ and $\tilde{\ell}(z_2) = 0$. Then $\mathbb{E}\tilde{\ell} = (1-p)\ell$ and $Var(\tilde{\ell}) = p(1-p)\ell^2$. Direct calculation implies

$$\mathrm{Cov}(\tilde{r},\tilde{\ell}) = \mathbb{E}\big[(\tilde{r}-\mathbb{E}\tilde{r})\tilde{\ell}\big] = \frac{1}{2}\sigma\ell - \left(\frac{1}{2}-p\right)\sigma\ell = p\sigma\ell,$$

so that the optimal allocation (21) equals

$$\pi(p) = \frac{1-p\sigma\ell}{\sigma^2}. \tag{22}$$

The greater the probability that the agent assigns to the depression, the less he or she allocates to the risky asset. We next consider how memory impacts the agent's

[21]For convenience, we assume that the agent solves an unconditional problem; that is, the agent chooses a portfolio allocation once and for all, and does not attempt to estimate the current state. In Section 4.1.3, we show to how to formally specify the problem using a conditional expectation.

allocations.

4.1.2 Memory for stock market gains and losses

In the Bayesian benchmark (Theorem 1) the agent retains a perfect memory of gains and losses, and their associations with depressions. We now consider the implications of contextual retrieval and encoding for the agent's beliefs and portfolio allocation. Because both labor income states and stock returns influence utility, Assumption 1 implies that they both must be features of the environment.[22] Table 1 summarizes the features space.

Table 1: Features corresponding to gains, losses, and depressions

Basis Vector	Features	Outcome for wealth
e_1	gain	$1 + \pi r_g + \ell_n$
e_2	loss	$1 + \pi r_l + \ell_n$
e_3	depression	$1 + \pi r_l + \ell_d$

Prior to time t, the agent has associations

$$M_{t-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} - p^* & p^* \end{bmatrix} \tag{23}$$

Loss and depression share a context. Section 2.4 shows how associations of the form (23) arise.[23] First assume $\zeta = 1$, as in Theorem 2. Consider retrieved context in

[22] Under our assumptions, gains cannot occur with a negative outcome of ℓ. There are thus three possible features.

[23] Section 2.5 discusses how events experienced close in time are linked to similar contexts (in that they retrieve similar features, or equivalently, that they are close in terms of the standard distance measure. In order to focus on the main mechanism in these examples, we assume that the agent experiences depressions and stock market losses under the same context. It is reasonable to imagine memory consolidates contexts that are sufficiently close into a single context.

response to a gain ($f_t = e_1$):

$$x_t = x_t^{\text{in}} \propto M_{t-1} e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

By (10), context in turn retrieves features corresponding to a gain:

$$f_t^{\text{in}} \propto M_{t-1}^{\top} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{2} e_1 \propto e_1$$

Now consider retrieved context in response to a loss ($f_t = e_2$):

$$x_t = x_t^{\text{in}} \propto M_{t-1} e_2 = \begin{bmatrix} 0 \\ \frac{1}{2} - p^* \end{bmatrix} \propto \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{24}$$

What does this context imply about the probabilities the agent places on future features? For this, we need to look at retrieved features. It follows from (10), that context (24) retrieves some probability on depression as well as on loss:

$$f_t^{\text{in}} \propto M_{t-1}^{\top} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} - p^* \\ p^* \end{bmatrix} \propto \begin{bmatrix} 0 \\ 1 - 2p^* \\ 2p^* \end{bmatrix} \tag{25}$$

Note that we are generalizing the analysis of Section 2.2 to allow for retrieved features. Combining Theorem 2 together with Theorem 4 implies a Bayesian theory of disagreement over latent states (see also Andreoni and Mylovanov (2012)). However, if agents have differing beliefs only about latent states, but their utility depends ultimately on states that *can* be observed (as in Assumption 1), it does not appear that disagreement has been generated in a meaningful sense.

Having recalled features (25), the agent encodes them with context (24). If encoding were instead to take place with actual features, the form of M_t implies learning with be the same as the Bayesian benchmark (Theorem 1). Memory M_t evolves according to (4), with

$$x_t f_t^\top = \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} e_1^\top & = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \text{if } gain \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} [\, 0 \; 1-2p^* \; 2p^* \,] & = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1-2p^* & 2p^* \end{bmatrix} & \text{if } loss \end{cases} \tag{26}$$

After τ periods of which k are gains:

$$M_{t+\tau} = \begin{bmatrix} \frac{1}{2}t + k & 0 & 0 \\ 0 & (\frac{1}{2} - p^*)t + (1-2p^*)(\tau - k) & p^* t + 2p^*(\tau - k) \end{bmatrix} \tag{27}$$

Note that the relative probability of losses and depressions, $M(2,2)/M(2,3)$, remains the same, regardless of how much experience the agent accumulates.[24] The agent does learn, in particular, about the relative probabilities of stock market losses and gains. In fact, the agent's beliefs will converge to the truth in this regard. However, the agent still overestimates the probability of a depression. We thus recover the result of Theorem 2, this time generalized to allow for features retrieval. Probabilities remain distorted, irrespective of the quantity of data.

Why, intuitively, does the agent fail to update his or her probabilities? The reason is that the agent's memory over-associates a stock market loss with a depression. The appearance of a stock market loss, then reinstates the depression context. This act of recalling the depression context is similar to experiencing the depression. Thus, a high probability of depression remains associated with losses in the agent's mind. Note that if the agent had begun with the correct associations, this process would have produced

[24] Note that $(p^* t + 2p^*(\tau - k))/((\frac{1}{2} - p^*)t + (1-2p^*)(\tau - k)) = p^*(\frac{1}{2} - p^*)$ regardless of k.

the correct probabilities.

The solid line in Figure 3 illustrates that, when $\zeta = 1$, the agent never changes his or her beliefs.[25] In contrast, the Bayesian agent quickly learns that unemployment is very unusual. Though learning the precise value takes time, even with a pessimistic prior, the agent quickly learns enough information to imply investing a substantial portion of wealth in stock.[26]

4.1.3 Generalizing to autoregressive context

This section generalizes the conclusions of the previous section to $\zeta < 1$. Consider a richer features space, but continue to assign e_1 to gains e_2 to losses, and e_3 to losses combined with a depression. For simplicity, assume the following form for initial associations:

$$M_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 1-2p^* & 2p^* & 0 & \cdots \\ 0 & 0 & 0 & \hat{M}_0 & \\ \vdots & \vdots & \vdots & & \end{bmatrix}, \tag{28}$$

where $\hat{M}_0$ represents associations other than gains, losses, and depressions.[27] Many paths of experience could influence the agent's memory for losses. Here, we will focus on one such path. Appendix E gives details of the arguments below.

Assume the agent experiences losses at times $t_1, t_2, \ldots, t_\ell, \ldots$. Also assume that the agent does not experience actual depressions; relaxing this assumption will strengthen our results. Context at the time of the first loss equals $x_{t_1} = (1-\zeta)x_{t_1-1} + \zeta\hat{e}_2$, where $x_{t_1-1}^\top \hat{e}_2 = 0$, and where features retrieved by x_{t_1-1} put no weight on depressions. It

[25]In calculating portfolio choice, we use (22), with p^* taken from (27). Alternatively, we could use the idea of a neutral decision context as defined in the section below.

[26]For the purposes of the figure, $p = 0.02$, $p^* = 0.50$, $\sigma = 1$, and $\ell = 2$.

[27]The column sums in M_0 can be interpreted as the number of observations of each feature in a prior sample. See Appendix A, Lemma A.1. Our assumptions on these do not qualitatively effect the results.

follows from (11) that

$$f_{t_1}^{\text{in}} = \zeta[\,0,\ 1-2p^*,\ 2p^*,\ 0,\ \ldots]^\top + (1-\zeta)f_{t_1-1}^{\text{in}}. \tag{29}$$

The subjective probability of depression (the third element of (29)) equals $2\zeta p^*$. Retrieved features are encoded with context x_{t_1}. Contextual drift will lead this agent to dis-associate the context loss with losses and depressions, thus putting a progressively lower weight on depressions.[28] If, for example, the agent experienced a single loss and saw no other features to remind him or her of a loss or a depression, then the weight on the depression would decay exponentially to zero. However, should another loss arise, even after an arbitrarily long time delay, the agent will recall the depression as if no time has passed.

Theorem 3 shows that retrieved context follows a recursion, assuming that encoding takes place with basis vector features. Appendix E proves an analogous result when encoding is with non-basis features. Let $\tilde{q}_t$ denote the agent's subjective probability of $z_2 \in \mathcal{Z}$, namely, $\tilde{q}_t = x_t(2)$. Assume at least one loss has taken place ($t > t_1$). Appendix E shows

$$f_{t,2}^{\text{in}} = \left(1 - \frac{\tilde{q}_{t-1}(1-\tilde{q}_{t-1})}{1+\sum_{s=1}^{t-1}\tilde{q}_s}\right) f_{t-1,2}^{\text{in}} + \frac{\tilde{q}_{t-1}(1-\tilde{q}_{t-1})}{1+\sum_{s=1}^{t-1}\tilde{q}_s} f_{t-1,2}^{\text{in},\perp}, \tag{30}$$

where

$$f_{t-1,2}^{\text{in},\perp} = (1-\tilde{q}_{t-1})^{-1}\sum_{j\neq 2} x_{t-1}(j) f_{t-1,j}^{\text{in}}.$$

The third element of (30) gives the subjective probability of a depression.[29] Iterating

[28] This conclusion would not necessarily hold under winner-take-all features retrieval. See Section 2.4 and the introduction to Section 4 for further discussion. Winner-take-all would lead retrieved features to continue to place high weight on the depression, assuming that agents' beliefs in a depression were above a threshold. In this respect, it would make it easier for us to derive our main result. However, incorporating winner-take-all features retrieval would complicate the analysis, which is why we do not assume it here.

[29] We make the conservative assumption that $f_{t-1,2}^{\text{in},\perp}$ places zero weight on the depression; if not,

on (30) shows how the subjective probability decays from its initial value of $2\zeta p^*$. As of time T, the subjective depression belief $\tilde{p}_T = f^{\text{in}}_{T,2}(3)$ equals

$$\tilde{p}_T = 2\zeta p^* \prod_{t=t_1}^{T} \left(1 - \frac{\tilde{q}_{t-1}(1-\tilde{q}_{t-1})}{1+\sum_{s=1}^{t-1}\tilde{q}_s}\right). \tag{31}$$

Note that the rate of decay depends on $\tilde{q}_{t-1}(1-\tilde{q}_{t-1}) = (x^\top_{t-1}\hat{e}_2)(1-x^\top_{t-1}\hat{e}_2)$. If x_{t-1} is either orthogonal or proportional to $\hat{e}_2$, then $f^{\text{in}}_{t,2} = f^{\text{in}}_{t-1,2}$. In neither case do beliefs decay, because associations are unchanged. If x_{t-1} has some, but not perfect, overlap with $\hat{e}_2$, then the agent learns new associations with the current context, causing beliefs to decay more quickly. Over time, the rate of decay slows, as captured by $1+\sum_{s=1}^{t-1}\tilde{q}_s$. As the sum increases, the term multiplying $f^{\text{in}}_{t-1,2}$ also tends to increase, causing the process to become more persistent.

We now assume that, following each loss, context decays to a neutral value (one that is not associated with depression features). The most recent loss occurred at $t_\ell = \text{argmin}_{t_j}\{t-t_j; t_j < t\}$, and $\tau = t - t_\ell$, the time elapsed since the loss event. Then

$$x_{t_\ell+\tau} = (1-\zeta)^\tau\zeta\hat{e}_2 + (1-(1-\zeta)^\tau\zeta)\bar{x}, \tag{32}$$

is the required time path of context.[30]

We continue to assume that stock market gains and losses are equally likely.[31] We also assume that gains and losses occur one out of every J periods, capturing the fact that the agent experiences other types of features (greater values of J correspond to slower decay of probabilities). We assume that portfolio choice takes place when the agent is in a neutral decision context, namely a context that is $\frac{1}{2}(\hat{e}_1+\hat{e}_2)$. This

beliefs in depressions will be greater.

[30]We continue to assume that features that are losses retrieve context $\hat{e}_2$. Appendix E explains why this assumption is valid for p^* close to $\frac{1}{2}$. In short: for $p^* \approx \frac{1}{2}$, retrieved features are more like depressions than losses. Thus actual losses retrieve the context associated with the initial loss.

[31]Strictly speaking, because depressions do not occur in the sample we are considering, losses should be slightly less likely; we ignore this effect here.

neutral decision context implies a depression probability of $p^* \prod_{t=t_1}^{T} \left(1 - \frac{\tilde{q}_{t-1}(1-\tilde{q}_{t-1})}{1+\sum_{s=1}^{t-1} \tilde{q}_s}\right)$. Figure 3 shows these values, and the resulting portfolio choice when we average over 1000 individuals and assume $J = 4$ and $p^* = 1/2$. The figure shows that slow decay of beliefs is not special to $\zeta = 1$; incorporating contextual drift allows for the more realistic conclusion that early memories still exercise influence, but that they fade gradually over time.

4.2 Context and the jump back in time: Application to the financial crisis

The failure of Lehman Brothers is widely recognized as a point of inflection in the 2008 financial crisis.[32]

An open question is: why was the failure of Lehman Brothers so pivotal? A growing line of research answers this question by focusing on the importance of financial intermediation to the overall the economy. Brunnermeier and Sannikov (2014) and He and Krishnamurthy (2013) develop models in which the balance sheets of intermediaries contribute to business cycle fluctuations. However, while it may be necessary to have specialized institutions trade certain complicated investments, it is not clear why the failure of a financial institution should be followed by a broad-based stock market decline. Common stocks are not intermediated assets: trading costs for common stocks, already quite low for the past half-century, have only gotten lower (Jones, 2002). Another possibility is that Lehman represented a sunspot that caused a run on other intermediaries, and other forms of debt (Allen and Gale, 2009; Gorton and Metrick, 2012). Unanswered is why this should cause the stock market to crash, as it did in the fall of 2008, when most companies have very low leverage and can fund themselves through retained earnings?[33]

[32]See, for example, French et al. (2010).

[33]Kahle and Stulz (2013) argue that firms dependent on bank-lending were not unduly affected by

Gennaioli and Shleifer (2018) emphasize a third possibility: individuals and financial institutions took on too much debt because they incorrectly extrapolated from a recent low-risk environment. This debt created unstable conditions. The Lehman bankruptcy caused a sudden shift in beliefs by reminding agents of the risks they had forgotten. This account is most in spirit of the discussion here. Indeed, our hypothesis is that the financial crisis was a psychological event caused by the failure of Lehman Brothers.

In the model described below, the failure of an important financial institution in the absence of insurance reminds investors of the Great Depression.[34] Some felt that they had – literally – returned to the Great Depression. Once this feeling entered the discourse, it proved hard to shake. Subsequent events showed that in fact there was no Great Depression; yet the association continued through a greatly renewed interest in financial crises and the macroeconomy, and through even the very name Great Recession.[35]

4.2.1 Endowment and preferences

Assume an endowment economy with identical log utility agents, each with time-discount factor β. Let W_t denote an agent's wealth at time t, and let C_t be consumption. Each agent solves

$$\max_{\pi_t, C_t} (1-\beta)\log C_t + \beta \mathbb{E}_t[\log W_{t+1}] \tag{33}$$

the crisis. Gomes et al. (2019) argue that fluctuations in borrowing conditions are more likely to be affected by investment opportunities than the other way around.

[34] See, for example, the reporting of *The Guardian* on the day's events: https://www.theguardian.com/business/2008/sep/15/marketturmoil.stockmarkets.

[35] The model below is stylized; it cannot capture many interesting features of the Great Recession. One line of research in particular focuses on a channel from asset valuations to the real economy, either through real-business-cycle mechanisms (Gourio, 2012) or New Keynesian mechanisms (Caballero and Simsek, 2020). This literature can be viewed as taking beliefs in a changed regime as given and deriving joint implications for real outcomes and asset markets. Our model is about endogenizing the beliefs.

subject to

$$W_{t+1} = (W_t - C_t)\left[R_{f,t+1} + \pi_t(R_{t+1} - R_{f,t+1})\right], \tag{34}$$

where, π_t denotes the percent allocated to the risky asset and where, because agents are identical, we omit an agent subscript. Each agent trades in a risky asset with gross return R_{t+1} and a riskless asset with gross return $R_{f,t+1}$ (known at time t).[36] We specify the aggregate endowment; asset prices will then equilibrate so that it is optimal for the agent to consume this endowment (Lucas, 1978). Let g denote the growth rate in consumption during normal times, and $\delta \in [0, 1)$ the decline in consumption, should a depression occur:

$$\frac{C_{t+1}}{C_t} = \begin{cases} 1+g & \text{with probability } 1-p \\ (1+g)(1-\delta) & \text{with probability } p \end{cases} \tag{35}$$

The aggregate market is a claim to dividends D_t satisfying $D_{t+1}/D_t = (C_{t+1}/C_t)^{\phi}$, with $\phi > 1$. The assumption of $\phi > 1$ captures the fact that payouts to shareholders fell by far more than consumption during the Great Depression (Longstaff and Piazzesi, 2004).[37] Equilibrium will require that $\pi_t = 1$, and that the dividend claim and the riskfree asset are in zero supply.

We briefly describe equilibrium under full information. The wealth-to-consumption

[36] Under time-consistent beliefs, the assumption of log utility implies that (33) is a recursive formulation of the multiperiod consumption and savings problem (Samuelson, 1969).

[37] This specification implies dividends and consumption are perfectly conditionally correlated. In the data, dividends have greater normal-times volatility than consumption, and they are imperfectly correlated with consumption. Both facts could be introduced into the model by assuming that dividends also are subject to independent shocks. Because these shocks are unpriced, and assuming that we abstract from dividends as features about which the agent learns, they would have a negligible effect on the results of interest. The assumption that normal-times growth in dividends is $(1 + g)^{\phi}$ could similarly be relaxed without affecting the results.

ratio $W_t/C_t = 1/(1-\beta)$. In equilibrium, the riskfree rate equals a constant given by

$$\begin{aligned} R_f &= \mathbb{E}\left[\beta \frac{C_t}{C_{t+1}}\right]^{-1} \\ &= \beta^{-1}(1+g)\left(1+p\left(\frac{1}{1-\delta}-1\right)\right)^{-1}. \end{aligned} \tag{36}$$

From (36), it follows that (in a comparative statics sense) an increase in the depression probability p lowers the interest rate. An increase in the depression probability leads the investor to want to save today. Bond prices rise, and riskfree rates fall.

Let S_t equal the value of the aggregate stock market. In equilibrium,

$$S_t = \mathbb{E}_t\left[\beta \frac{C_t}{C_{t+1}}(S_{t+1}+D_{t+1})\right], \tag{37}$$

Solving for a fixed point yields:

$$\frac{S_t}{D_t} = \frac{\beta(1+p\left((1-\delta)^{\phi-1}-1\right))}{(1+g)^{1-\phi}-\beta(1+p((1-\delta)^{\phi-1}-1))} \tag{38}$$

for all t. For $\phi > 1$, an increase in p lowers the stock price. Realized returns on the stock market equal

$$R^S_{t+1} = \begin{cases} \bar{R}^S & \text{with probability } 1-p \\ \bar{R}^s(1-\delta)^{\phi} & \text{with probability } p \end{cases} \tag{39}$$

where $\bar{R}^S$ is the stock market return during normal times:[38]

$$\bar{R}^S = \beta^{-1}(1+g)\left(1+p\left(\frac{1}{(1-\delta)^{1-\phi}}-1\right)\right)^{-1}. \tag{40}$$

[38] Let $\Phi = \beta(1+g)^{\phi-1}(1+p\left((1-\delta)^{\phi-1}-1\right))$. Then

$$\bar{R}^S = \frac{S_{t+1}/D_{t+1}+1}{S_t/D_t}(1+g)^{\phi} = \frac{\Phi/(1-\Phi)+1}{\Phi/(1-\Phi)}(1+g)^{\phi} = \Phi^{-1}(1+g)^{\phi}$$

The expected return for the stock market equals:

$$\mathbb{E}_t R^S_{t+1} = \bar{R}^S(1 + p((1-\delta)^{\phi} - 1)). \tag{41}$$

Subtracting the log of the riskfree rate (36) from the log of the expected return (41) give the equity premium. For small p (in the continuous-time limit), the equity premium is well-approximated by

$$\log \mathbb{E}_t \left[R^S_{t+1}/R_f\right] \approx p\left((1-\delta)^{-1} - 1\right)\left(1 - (1-\delta)^{\phi}\right).$$

The right hand side is is the negative change in marginal utility multiplied by the change in stock price during depressions.

4.2.2 Features, context, and memory

Now assume that agents update beliefs according to retrieved-context theory. An application to asset pricing immediately raises the question of agent's beliefs about others' beliefs. Here we assume agents have identical initial associations M_0, identical experiences (and thus the same M_t), and that beliefs are common knowledge. Relaxing these assumptions would be desirable and would have interesting implications.

There are three possible outcomes in $\mathcal{Y}$: normal, crisis, and depression. These states are captured by basis features vectors: e_1 (normal), e_2 (crisis), and e_3 (depression). At each time t, the agent observes W_t and f_t (we abstract from wealth as a feature). We conjecture an equilibrium in which the agent possesses the correct mapping between f_t and $r_{f,t+1}$ and f_t and r_t. The agent solves (33–34) under the subjective expectation, formed from memory as Section 2.4 describes.[39] The agent's problem satisfies

[39] The agent retrieves context x_t^{in} from features f_t and memory M_{t-1}. Retrieved context and lagged context x_{t-1} combine to determine x_t through (2). Using $M_{t-1}^{\top}$, the agent retrieves f_t^{in}, which gives the probability distribution over future features.

Assumption 1 in that it depends only on the distribution of future features.[40]

The agent's optimal consumption continues to equal $1/(1-\beta)$ times wealth. Assuming the agent sets the subjective probability of depression $p_t = e_3^\top f_{t+1}^{\text{in}}$, the agent will allocate $\pi_t = 1$ to the consumption claim provided that $r_{f,t+1}$ satisfies (36), and r_{t+1} satisfies (39–40) with $\phi = 1$. The assumption that p_t is permanent (agents do not foresee a change in their own, or others' beliefs) implies that (38) is also satisfied in equilibrium. Table 2 summarizes features and their effects on outcomes of interest.

Table 2: Features corresponding to normal times, financial crises, and depressions

Basis Vector	Features	Consumption claim return	Crisis?
e_1	normal	$\beta^{-1}(1+g)$	No
e_2	crisis	$\beta^{-1}(1+g)$	Yes
e_3	depression	$\beta^{-1}(1+g)(1-\delta)$	Yes

Equilibrium returns on the consumption claim equal growth in consumption, scaled by β^{-1}, and thus depend only on the realization of a depression. While the agent uses current features (such as crises) to form beliefs as in (1), and indeed the value function depends in equilibrium on the riskfree rate and hence on current features, the only *future* feature of interest to the agent is whether or not there will be a depression.

We assume a simple form for the memory matrix at time $t-1$, the period before the crisis:

$$M_{t-1} = \begin{array}{c} \begin{array}{ccc} \text{normal} & \text{crisis} & \text{depression} \end{array} \\ \begin{bmatrix} 1-p^c & 0 & 0 \\ 0 & p^c(1-q) & p^c q \end{bmatrix} \end{array} \tag{42}$$

perhaps generated as in Section 2.4.[41] We assume two underlying states in $\mathcal{Z}$, one generating the financial crisis that is, in the agent's mind, associated with the depression,

[40]We assume the agent is capable of conceiving of the equilibrium and calculating quantities that are directly implied, through algebraic equations, to features (this is required for optimization as well as equilibrium). These are strong assumptions and relaxing them is an interesting topic for future research. We make them here to focus on one departure from the standard model at a time.

[41]As in the previous example, we consolidate two contexts into one.

and one associated with normality. While it is not necessary to take a stand on whether there is a physical association between a depression and a financial crisis, for ease of comparison with the full-information case above, we assume that the true correlation between disasters and crises equals zero, so that the disaster probability is iid.

Assume that there has been a sufficiently long period of normal features e_1. so that the agent exhibits neglected risk: namely context $x_{t-1} = [1, 0]^\top$ (Theorem 6).[42] Though agents neglect the depression state, they have not forgotten it. Representing the failure of Lehman brothers is $f_1 = e_2$, the well-publicized failure of a major financial institution. Retrieved context in response to crisis features equals

$$x_t^{\text{in}} \propto M_{t-1} e_2 \propto [0, 1]^\top. \tag{43}$$

Context therefore equals

$$x_t = (1-\zeta)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \zeta \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1-\zeta \\ \zeta \end{bmatrix}. \tag{44}$$

As in Section 2.4, we calculate features retrieved by each component of context:

$$\begin{aligned} f_{1,t}^{\text{in}} &\propto M_{t-1}^\top \begin{bmatrix} 1 \\ 0 \end{bmatrix} \propto e_1 \\ f_{2,t}^{\text{in}} &\propto M_{t-1}^\top \begin{bmatrix} 0 \\ 1 \end{bmatrix} \propto \begin{bmatrix} 0 \\ 1-q \\ q \end{bmatrix} \end{aligned}$$

[42] Strictly speaking, this requires a large number of neutral features; it is simple to alter this example to include these but it also makes the notation more complicated.

Therefore, applying (44) and (11),

$$f_t^{\text{in}} = \begin{bmatrix} 1-\zeta \\ \zeta(1-q) \\ \zeta q \end{bmatrix}. \tag{45}$$

Equation 45 represents reinstatement of the depression context – the financial crisis reminds the agent of the depression. The probability of a depression rises from 0 to ζq, causing an immediate decline in stock prices (38) and in the riskfree rate (36). The sharp drop between $t = 0$ and $t = 1$ of Figure 4 illustrates this effect on the price-dividend ratio and on the riskfree rate.[43]. Figure 4 also shows a sharply negative return corresponding to this event.[44]

What does the model say about the time path of context, and hence that of prices, interest rates, and stock returns, following the event? We discuss in detail one such possible path. Consistent with events in late 2008, we assume several (specifically, three) observations of crisis features, and then normal features. Assume a sufficiently long prior sample so that updating memory is not first-order.[45] If the agent continues to observe crisis features, context updates as follows:

$$x_{t+1} = (1-\zeta)^2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (\zeta(1-\zeta)+\zeta) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{46}$$

Thus recall of the depression increases, the stock price declines further, and realized returns continue to be negative. Figure 4 shows this continued decline. Thus while the initial drop was an over-reaction relative to the correct iid benchmark (Theorem 8),

[43]When we report the riskfree rate in the figure, we assume a zero lower bound. That is, we assume that for institutional reasons, the observed riskfree rate cannot fall below zero, whereas the true riskfree rate might

[44]We assume $q = 0.5$, $p^c = 0.05$, $\delta = 0.15$, $\beta = .98$, $g = 0$, $\phi = 2$.

[45]The solution shown in Figure 4 assumes a prior sample of 100 years and calculates the exact path of context.

there is also a sense in which it was, in the short run, an under-reaction because prices fall more before they stabilize (Theorem 7). Given sufficient crisis observations, the stock price would stop declining once context reached a steady state of $[1-q, q]^\top$, implying zero weight on normal features. In this example, however, assume that normal features return after three periods, leading to a partial recovery. Returns are positive and high, because they represent good news that there has not been a depression.[46] Prices recover more slowly than they fall. In this model with autoregressive beliefs, this is mainly due to the effects of duration embedded in (38). An increase in the probability of disaster decreases the effective maturity of the stock return, and so any increase in p from this low point has a smaller effect.

This duration explanation cannot, however, account for the fact that the price-dividend ratio asymptotes to a lower level. This is because as depression features are retrieved, not only are they encoded with the crisis context (as in Section 4.1), but they are also encoded with the normal context. The agent then associates this context with depression, so that even a return to a normal context implies a permanently elevated probability of depression. More precisely, the agent associates whatever was in context just before the crisis with depression, even though the depression did not occur. The continued retrieval of the depression context ensures that this is a permanent effect. To the extent that (previously normal features) reminiscent of 2009 re-appear, they will remind agents of the Great Recession, which in turn will recall the Great Depression.[47]

[46]Due to log utility, the equity premium is small in this model, even with a high probability of a rare. Little of realized returns correspond to an equity premium.

[47]While Figure 4 represents a clear departure from a full-information benchmark, another benchmark of interest is one in which a Bayesian agent believes in the existence of two states and that the financial crisis serves as a signal for depression. That is, the Bayesian agent shares the associations M_{t-1}, and thus is as close as possible to the agent we consider. Wachter and Zhu (2019) model Bayesian learning in such a setting. The Bayesian agent does not neglect risk (or at least not to the same degree): there is always some probability on the depression state regardless of how long a period of normalcy occurs. The total decline in price relative to the Bayesian benchmark represents an over-reaction; the Bayesian agent believes the state will revert. However, unlike the Bayesian agent, there is also under-reaction in the sense that prices do fully react immediately to the "news" — they take several periods to respond. The permanent adjustment in prices due to the new associations shown in Figure 4 is absent in the Bayesian model, nor does the Bayesian model explain how an agent came

4.3 Fear and asset allocation

Psychological and neuroscientific research reveals a tight link between memory and emotion. For example, when people are sad or depressed, they tend to recall negative events (Matt et al., 1992; Teasdale and Fogarty, 1979). When people remember an emotionally-valent word (positive or negative) in a list of mixed-valence words, the next word they remember tends to be emotionally congruent (Long et al., 2015; Siddiqui and Unsworth, 2011). Emotion also affects memory for neutral items — subjects exhibit superior recall when tested under a state that is emotionally congruent to the study state (Eich, 1995).

The connection between emotion and memory extend to the financial domain. Guiso et al. (2018) found that after the 2008–209 financial crisis, professional investors required twice the premium to accept a risky bet rather than a sure payoff than before, suggesting a role for fear in decision-making. In this case, the finance professionals' response need not be due to fear; an alternative is that they were materially worse off following the crisis, and that they exhibit risk aversion that increases as wealth falls. To demonstrate that it is the memory and not the wealth change that influences risk aversion, Guiso et al. conduct an experiment in in which subjects are randomly assigned to view a scene from a horror movie. Subjects who viewed the scene required a 50% greater premium to accept the lottery as compared to those that did not. Similar results are found by Cohn et al. (2015), who conduct an experiment in which financial professionals are selected randomly to view a chart of a stock market boom versus a crash. Investors in the boom condition invested significantly more in the risky asset that those in the crash condition. The effect of emotion on financial decisions extends beyond the laboratory. Cuculiza et al. (2020) show that analyst earnings forecasts

to associate crisis and depression in the first place. The initial slow adjustment and the permanent change in prices also differentiates the model from that of Gennaioli and Shleifer (2018). Note that this comparison also assumes the agent begins with conditional probabilities that the present model endogenizes through temporal contiguity.

become more negative upon anniversaries of terrorist attacks (as well as following the attacks themselves). Ramadorai et al. (2020) examine the trading behavior of investors following the outcome of IPO lotteries. Investors who received shares in companies that subsequently performed well not only purchase similar stocks, but trade more in general. While Loewenstein (2000) proposes a model by which emotion influences utility, his model is silent on the connection between emotion and recent experience.

In what follows, we apply our framework to explain findings on fear and aversion to risk. We specifically consider the set-up of Guiso et al. (2018), in which an agent chooses between a risky asset (a lottery) and a sure investment. As in Section 4.1, the agent chooses an allocation π to a risky investment with net return $\tilde{r}$, and $1-\pi$ to a safe investment with net return of zero. Also as in Section 4.1, the agent is subject to a second source of risk, which we can think of as loss of a job, loss of health, or some other financial calamity. The risky investment (capturing the lottery in the experiment) experiences a gain or a loss, each with equal probability:

$$\tilde{r} = \begin{cases} \mu+\sigma & \text{prob. } 1/2 \\ \mu-\sigma & \text{prob. } 1/2. \end{cases}$$

Wealth equals $1+\pi\tilde{r}-\tilde{\ell}$, where $\tilde{\ell}$ is the second source of risk mentioned above:

$$\tilde{\ell} = \begin{cases} 0 & \text{prob. } 1-p \\ \delta & \text{prob. } p. \end{cases}$$

In what follows, we refer to $\tilde{\ell}$ as a human capital shock.

Because $\tilde{r}$ is the outcome of a lottery, $\tilde{\ell}$ and $\tilde{r}$ must be independent. Moreover, $\tilde{r}$ has a well-defined set of outcomes (it obeys the Savage (1954) model). Unlike in Section 4.1, we cannot rely on the agent's misperception of the correlation between the stock return and human capital. While it is possible that the agent has such a misperception, here we assume that agents understand that the outcome of the lottery does not bear on other events.

While the agent is told that the outcome of the lottery is 50/50, the probability p is unknown. The agent determines the more ambiguous probabilities of $\tilde{\ell}$ based on memory. Assume that the agent has log utility over wealth, so that perceived probabilities of $\tilde{\ell}$ influence the utility of the lottery (any power function over wealth would have this property). The agent feels more fearful of a bad outcome, and thus is less likely to take on risk. Whereas the connection between an increased likelihood of a bad outcome and a willingness to take on risk is not irrational, the increased fear of the bad outcome is.

As in Section 4.1.3, we assume the first features vector corresponds to a normal state, whereas the next two correspond to the negative outcomes. There are also a large number of "neutral" features, i.e. neither positive or negative. The matrix M_{t-1} takes the form:

$$M_{t-1} = \begin{array}{c} \begin{array}{ccccc} \text{normal} & \text{fin. crisis} & \text{danger} & \text{other associations} & \end{array} \\ \begin{bmatrix} 1-p_1-p_2 & 0 & 0 & 0 & \cdots \\ 0 & p_1 & p_2 & 0 & \cdots \\ 0 & 0 & 0 & \hat{M}_t & \\ \vdots & \vdots & \vdots & & \end{bmatrix} \end{array}$$

The second context corresponds to a state in which negative events occur. Temporal contiguity could account for the associations between a financial crisis and danger in this context. The relative values of columns indicate how much of each feature the agent has experienced. Given our focus on context and feature retrieval, this will not be important in what follows. Features retrieved by each component of context are:

$$\begin{aligned} f^{\text{in}}_{1,t} &\propto M^{\top}_{t-1}\hat{e}_1 \propto e_1 \\ f^{\text{in}}_{2,t} &\propto M^{\top}_{t-1}\hat{e}_2 \propto (p_1+p_2)^{-1}(p_1 e_2 + p_2 e_3) \end{aligned}$$

Let $\tilde{q}_t = x_t(2)$ denote the subjective probability of the second state (in which $\tilde{\ell} = \delta$), and assume that, prior to the experiment, $\tilde{q}_{t-1} = p_1 + p_2 \equiv p$.

The scene from the movie is a feature that is similar to, but not exactly the same as, danger. That is, $f_t \approx \hat{e}_3$, Assuming f_t is sufficiently close to $\hat{e}_3$, then Theorem 5 applies: retrieved context will be similar to what would occur under actual danger. That is $x^{\text{in}}_t \approx \hat{e}_2$, and $\tilde{q}_t \approx (1-\zeta)\tilde{q}_{t-1} + \zeta$. If the agent had watched this particular scene on a prior occasion, then the approximation $x^{\text{in}}_t \approx \hat{e}_2$ would not hold as well, because the features specific to f_t would have their own associations. Guiso et al. (2018) chose a scene likely to be unfamiliar to subjects.

What matters, ultimately, is retrieved features. It follows from (11) that

$$f^{\text{in}}_t \approx (1-\tilde{q}_t)e_1 + \tilde{q}_t \propto (p_1+p_2)^{-1}(p_1 e_2 + p_2 e_3),$$

so that the probability of danger or depression equals $\tilde{q}_t$. The agent chooses π to maximize:

$$\mathbb{E}^{\ell}\left[\frac{1}{2}\log(1+\pi(\mu+\sigma)-\tilde{\ell}) + \frac{1}{2}\log(1+\pi(\mu-\sigma)-\tilde{\ell})\right], \tag{47}$$

where the outcome $\tilde{\ell} = \delta$ occurs with probability $\tilde{q}_t$ and is 0 otherwise. We assume an excess return $\mu = 4\%$, a standard deviation $\sigma = 20\%$, a prior probability of the negative labor market outcome $p = 2\%$, and a percent decline $\delta = 0.8$, should the outcome occur. As elsewhere, $\zeta = 0.35$. The agent trades off the higher return arising from greater π with greater risk. The higher is the probability of a bad realization, the less risk the agent can afford to take. The experiment reminds agents that such bad realizations can occur.

Figure 5 shows (47) as a function of the allocation π. At an initial level of zero, taking on a small value of risk is optimal (the function is increasing). After a certain level of π, the function begins to fall, representing the curvature in the utility function. When the agent is not fearful, this occurs at 70%. When the agent is fearful, it occurs at 30%. The response of the agent to the experiment cannot be Bayesian: a movie has not changed anything about the outside world. In that sense, the response of risk-taking to viewing a horror movie is a good test of our theory. This, and related results, show that it is possible to manipulate internal mental state (context) in a way that changes decision-making. Context alters agents beliefs about the outside world, and hence agents' decisions.

5 Retrieved Context Theory and Alternative Memory Models

Our theoretical framework for modeling human memory—retrieved context theory—builds upon scholarship going back to the early 20th century. McGeoch's classic (1932) theory of forgetting assumed that memories do not wither away in time, but rather that the retrieval cues used to search memory may either reveal or occlude a particular experience. McGeoch theorized that this retrieval-based interference depends on the state of context, the mental "set" of the rememberer, and the activation of similar "competitor" memories. Estes (1959) and Bower (1972) developed a mathematical foundation for these ideas, positing a VAR context representation that determined the retrievability of items from memory, accounting for the law of recency (Crowder, 1976). These models also allowed for explicit manipulations of context that could alter the memorability of particular experiences, as seen in experimental studies.

Unlike recency effects, which are easy to quantify, the influence of contiguity eluded careful measurement (Murdock, 1974). In studying the order and timing of recall sequences, (Kahana, 1996) introduced a conditional-probability measure of contiguity by computing the likelihood of recalling an item as a function of its contiguity to the just recalled item. Howard and Kahana (1999) theorized that contiguity could arise from retrieval of context, rather than direct interitem associations, and subsequent work provided support for this account (see, Healey et al., 2019, for a review).

Retrieved-context theory provides a unified account of recency and contiguity effects at short and long time scales. As a vector-based model of associative learning, this theory nests earlier work on similarity-based organization in memory and cue-dependent interference effects (Kahana, 2012). Two key aspects of this model make novel predictions not shared by most other memory theories: First, each encoding event involves an internal retrieval whose output modifies memory. Thus, thoughts become memories that influence subsequent retrieval. Second, the recursive definition of context predicts a forward asymmetry in memory retrieval because the contextual states previously associated with an item recombine with the items context and associate with subsequent items. The recursive contextual dynamics imply that when a new event matches, or closely resembles, an earlier experience, the context of that earlier experience will re-embed in memory; analogously, new contexts that resemble old contexts will tend to retrieve old features, which will also re-embed in memory. In the present theory of economic choice, these re-embeddings help to explain persistent disagreement in the face of accumulating evidence.

Whereas modern memory theorists have adopted McGeoch's early emphasis on retrieval processes, other classic models saw memory-guided choice as primarily reflecting the strength of memories established during learning. Consider a repeated event, such that each occurrence supplies an agent with information about the world (e.g., ordering a cappuccino and learning about its price). According to strength theory, the association between a cappuccino and its price continuously updates, and a future reminder of a cappuccino retrieves a single sufficient statistic reflecting the distribution of experienced prices. In this model, one updates the summary statistic and discards the memories of each experienced event. A fundamental problem facing strength theories of memory is defining the unit of memory; at some point a new experience is sufficiently different from an existing "memory" to constitute a new memory, but the model is silent as to how that point is determined.

Unlike strength theory, exemplar theory assumes that the memory system separately records each event (feature vector) in an ever growing matrix of memories (Estes, 1986; Hintzman, 1988; Nosofsky, 1992). The events retain their individuality, including their ordinal position within the series. Aggregation of memory happens at the time of retrieval, when the memory system compares a cue event against all previously stored examples, computing the summary statistic at the time of test. Although a wealth of data favor exemplar over strength-based models, the latter have resisted extinction due to their parsimony and computational efficiency (Murdock, 1985; Wixted, 2007).

Exemplar and strength-based models share a major limitation: they lack any mechanism for associating events that co-occur in a spatiotemporal context. Such associations have long held a central role in philosophical conceptions of the "association of ideas" (Hume) and form the basis for the Aristotelian "Law of Contiguity". Continental philosophers (Herbart, 1834) developed theories based on chained associations that stimulated the earliest experimental work on human memory (Ebbinghaus, 1913). These theories did not posit a specific representation of time, but rather assumed that contiguously experienced events become associated and that the strength of this associations falls as a function of the temporal separation of the events (Solway et al., 2012). The repetition of an item, evoked by an external stimulus or an internal retrieval, triggers retrieval of the item's neighbors as a function of the strength of their association. Chaining theory encounters serious obstacles when two lists of sequentially presented items share an overlapping item, or an overlapping subsequence of items. In this case, one cannot recall either list without suffering catastrophic interference between the competitor items (i.e the items that follow the overlapping items in both lists). Such interference prevents the model from recovering order information, or even accurately recalling a series with repeated or high similar elements (Lashley, 1951; Henson et al., 1996; Kahana and Jacobs, 2000). Despite the failure of many of its predictions, chaining theory retains a powerful appeal based in part on the everyday experience of sequential cuing of memories.

The idea that memories preserve a record of their spatial information, and the complementary observation that spatial cuing offers an aid to learning and retention, formed a centerpiece of the medieval "arts of memory" (Yates, 1966). Soon after the advent of list memory studies in the late 1800s, researchers recognized that subjects often visualized a series of items as occurring within a virtual, mental, space, much as medieval scholars used the "palaces of memory" to commit vast written works to memory (Ladd and Woodworth, 1911). This idea of positional coding as a means of representing ordinal information offered an alternative to the classic idea of chained associations described above. Positional coding models assumed that in learning a series of items, subjects formed associations between items and positions (location on an array, or position within an ordinal series). Later, at the time of test, subjects used positional information, assumed to be a mental primitive, to cue retrieval of items. Modern positional coding theories offer some of the most successful accounts of memory of short, ordered lists: phone numbers and postal codes and the like (Burgess and Hitch, 2006; Brown et al., 2000).

Although one can imagine items in a series as occupying locations in space, they actually occupy locations in time. For early researchers, this raised the question of whether memories retain information about their time of occurrence. Although the notion of time tagging appeared in some of the earliest psychological literature (James,

1890) memory scholars did not take stock of its significance until the emergence of experimental procedures that required subjects to explicitly judge the temporal order of studied items (Yntema and Trask, 1963; Hinrichs and Buschke, 1968) and models developed to explain these data (Bower, 1972). Consistent with the aforementioned notions of spatial and temporal coding of memories, neural recordings in both human and non-human animals have identified individual neurons in the brain that encode spatial information (O'Keefe and Dostrovsky, 1971; Ekstrom et al., 2003), and temporal information (MacDonald et al., 2011; Pastalkova et al., 2008; Umbach et al., 2020). Further, the activity of neural ensembles coding time and place during memory storage predict aspects of subsequent recall even when the recall task requires neither memory for the time or place of occurrence (Miller et al., 2013; Umbach et al., 2020).

Early exemplar models were not designed to match the aforementioned evidence on centrality of time and place; these models defined similarity by perception or meaning. Researchers met the challenge posed by this evidence by adding time and place as features of memory. Brown et al. (2007) proposed an exemplar memory model (termed SIMPLE) that includes an explicit representation of temporal information, such that subjects confuse memories that occurred at similar times in the past. They showed how this model could account for several major list recall phenomena.

An exemplar theory which posits that agents store a complete record of the attributes describing every memory would support optimal choice. But the marked failures of memory in our daily lives and the presumed finite capacity of the memory system pose a serious challenge for such a model. If memory failures primarily reflect inaccurate storage, there must be a massive degree of "lossy" data compression on the front end. A commonly adopted view is that initial processing occurs via a limited capacity short-term storage system; only a tiny fraction of information survives to make it into long-term storage.[48] Contrary to this limited storage view, Gallistel and King

48

(2009) summarize extensive data, and computational arguments, in favor of the brain's ability to store massive amounts of information. Failures of memory, they argue, reflect failures of retrieval rather than storage (see, also, Tulving and Madigan (1970)). Using sensitive indices of retrieval one can show that a once experienced laboratory event can leave a near-permanent record in memory (Kolers and Magee, 1978; Standing, 1973; Brady et al., 2011).

Thus, to build a choice model upon an exemplar theory with complete memory requires a fully-specified model for retrieval. The principles of memory used to motivate our modeling approach–similarity, recency and contiguity–would need to emerge from the hypothesized retrieval process. Most retrieval choice rules will naturally give rise to similarity effects (Kahana, 2012, Chapter 4). Recency effects arise based on the similarity of temporal codes between study and test (recency is another form of similarity). The challenge, however, facing these models is to explain the contiguity effect and its persistence across time scales. The ability to retrieve temporal codes associated with past memories, and use those codes to retrieve subsequent memories, requires substantial machinery that is not part of exemplar models. Although a chaining model, as described above, can produce associations among nearest neighbors, it does not easily account for associations that span multiple items as seen in the data. These findings require something like the contextual retrieval process used in this paper.

To summarize, a half-century of scholarship has shown that memory depends critically on the cues present at the time of retrieval, and that retrieval operates during both learning and recall. Retrieval during learning determines the information present on subsequent learning and recall trials. Scholars no longer see encoding and retrieval as distinct phases of memory; both processes play an important role during the acquisition of new knowledge and the during the recall or recognition of past experiences. The retrieved context framework builds upon classic notions of contextual variability and cue-dependent recall to offer a unified account of the principles of recency, contiguity and similarity, and their persistence across time scales. Kahana (2020) provides a more complete discussion of the relation between retrieved context theory and other models of memory, as well as discussing some of the important open questions to be addressed in the study of memory.

6 Conclusion

What makes us know what we know, perceive what we perceive, think what we think? What makes us the same person when we get up in the morning as we were the day before? What makes life a connection of meaningful events, and not just a random set of stimuli? It is our memories – the experience of our lives that is unique to each individual.

The standard model in economics would have it otherwise. Under this framework, individuals maximize expected utility, seeing the form of the utility function and beliefs about the future as fundamentally stable. These assumption arise from pure reason – under a specific notion of rationality. Although empirical and experimental studies have challenged this framework, its parsimony and the appeal of assuming agents smarter than ourselves have led to its continued use.

In this paper we propose an alternative that is also parsimonious, potentially rational, and based on a centuries-old program in experimental psychology, namely, the study of memory. We show that principles emerging from this program offer a very different set of implications than the standard ones in economics.

First, decisions can be affected by seemingly irrelevant information, making revealed preferences far less stable than they might be otherwise. Second, individuals have trouble processing information, not because of the amorphous idea of lack of attention, but because of a stable internal state. Because we carry with us an internal state, we cannot "take in" our surroundings all at once. This internal state, however, allows us to tag memories in time, with the powerful consequence of temporal contiguity, a property of memory noted since ancient times. Temporal contiguity strings together events, pulling up an entire universe with one memory, and forms a basis for conjectures of causal behavior. Finally, retrieval and encoding of memories implies that our beliefs may not converge, regardless of how many data points we observe. The reason is that the same data comes, for everyone, with its own associations – its own context. This context then triggers an actual perception of different data.

This paper represents a start in connecting memory with decision-making. Many questions remain unanswered. We assumed a single decision-maker solving a static problem (though memory itself is dynamic). A key question pertains to the decision-maker's view not only of future self, but of other selfs in the economy. Like beliefs about the physical world, these may also be contextually-dependent. A second question pertains to the boundaries between the type of recall of probabilities that we consider here, and model-driven decisions. To some extent we all do use models; at what level does the model come into play? Finally, economics concerns itself with maximization under constraints. Perhaps memory evolved to solve some maximization problem, but if so, which one? These are some of the questions we hope will be answered in future work.

A Proofs of results in Section 2.2

The following Lemma clarifies that the elements of M_t can be thought of as proportional to probabilities, with a prior distribution given by M_0. The scaling in (3) implies that the absolute magnitude of the elements in M is irrelevant. The time path of retrieved context, will the same if one employs a modified updating rule for M that scales the sum of its elements to equal one. The following Lemma makes this statement precise.

Lemma A.1. *Let $\{x_t\}_{t=1}^T$ be the time path of context up to time T given $\{f_t\}_{t=1}^T$, M_0, x_0, and updating rule (4). Let $t_0 = \iota^\top M_0 \iota$, namely, the sum of the elements in M_0. Let $\{\tilde{x}_t\}_{t=1}^T$ be a time path of context given the same features, initial condition $\tilde{M}_0 = \frac{1}{t_0} M_0$, $x_0 = \tilde{x}_0$, and updating rule*

$$\tilde{M}_t = \frac{t_0 + t - 1}{t_0 + t} \tilde{M}_{t-1} + \frac{1}{t_0 + t} \tilde{x}_t f_t^\top. \tag{A.1}$$

Then $\tilde{x}_t = x_t$ and $\tilde{M}_t = \frac{1}{t_0 + t} M_t$.

Proof. Assume by induction that $\tilde{x}_{t-1} = x_{t-1}$ and $\tilde{M}_{t-1} = (t_0 + t - 1)^{-1} M_{t-1}$. It follows from (3) that $x_t^{\text{in}} = \tilde{x}_t^{\text{in}}$, implying that $x_t = \tilde{x}_t$. It remains to show that $\tilde{M}_t = (t_0 + t)^{-1} M_t$. By (A.1),

$$\tilde{M}_t = \frac{t_0 + t - 1}{t_0 + t} \tilde{M}_{t-1} + \frac{1}{t_0 + t} x_t f_t^\top \tag{A.2}$$

Recall $\iota^\top M_0 \iota = t_0$. Further recall that $\iota^\top f_t = \iota^\top x_t = 1$. It follows that the elements of the outer product matrix $x_t f_t^\top$ sum to one. Using the induction step and substituting into (A.2) implies

$$\tilde{M}_t = \frac{1}{t_0 + t} M_t,$$

as required. □

It might seem that the updating rule (A.1) contains more information than (4) in that (A.1) includes both the current sample size t and a prior sample size t_0. Lemma A.1 says that this intuition is not correct and that they contain the same information. The reason is that the extra information in (4) is contained in the size of M itself. The sum of the elements in M equals the size of the sample, whereas the sum of the elements of $\tilde{M}$ equals one. The updating rule (A.1) takes information that was previously embedded in M and puts it into the updating rule.

Proof of Theorem 1. Under the assumptions of the theorem, context x_t is such that $x_t(i) = 1$ if $Z_t = z_i$ and 0 otherwise. Consider a Bayesian agent with a Dirichlet prior, with parameters given by M_0. Specifically, let $K = mn$, and let P be the $m \times n$ matrix of prior probabilities p_{ij} over state (z_i, y_j):

$$\text{vec}(P) \sim \text{Dir}(K, \text{vec}(M_0)).$$

Note that the prior mean of p_{ij} is $M_0(i,j)/t_0$, with $t_0 = \iota^\top M_0 \iota$. Suppose the agent observes $\{Y_t, Z_t\}_{t=1}^t$. Let $\hat{Y}_s = (Y_s, Z_s)$. Assume the agent computes a quasi-likelihood function under the assumption that observations are iid:[49]

$$\mathcal{L}(\hat{Y}_1, \ldots, \hat{Y}_t | P) = \prod_{t=1}^{t} l(\hat{Y}_s | P)$$

Each term $l(\hat{Y}_s|P$ is multinomial. The posterior distribution $p(P|\hat{Y}_1, \ldots, \hat{Y}_t, M_0)$ is therefore Dirichlet (Gelman et al., 2004):

$$\text{vec}(P) \,|\, \{Y_t, Z_t\}_{t=1}^t \sim \text{Dir}(K, \text{vec}(M_t)).$$

The mean of the posterior distribution for p_{ij} is then $M_t(i,j)/(t_0 + t)$ as required. □

[49] Alternatively, the agent could use the true likelihood that would allow for autocorrelation. This would necessitate a prior over p_{ij}^Z for all pairs (i,j). The quasi-likelihood function avoids this complication.

Lemma A.2. *Given Assumption 2, the following two invariance conditions are equivalent:*

1. *For any positive integer s, t, if $f_s = f_t$ then $x_s^{\text{in}} = x_t^{\text{in}}$.*

2. *For any basis vector $\bar{f}$, and any non-negative integers s, t, $M_s\bar{f} \propto M_t\bar{f}$.*

Proof. Assume condition 1 above. Assume $f_s = f_t \equiv \bar{f}$. By (3) and condition 1

$$M_{s-1}\bar{f} = M_{s-1}f_s \propto x_s^{\text{in}} = x_t^{\text{in}} \propto M_{t-1}f_t = M_{t-1}\bar{f},$$

proving condition 2.

Now assume condition 2. Let $f_s = f_t = \bar{f}$. Then by (3) and condition 2,

$$x_s^{\text{in}} \propto M_{s-1}f_s \propto M_{t-1}\bar{f} \propto x_t^{\text{in}}.$$

Equality follows because elements of retrieved context must sum to 1. This proves condition 1. □

Note that the proof does not, strictly speaking, require Assumption 2. It holds for any subset of possible basis vectors (assuming we restrict x^{in} accordingly). In what follows, we will mainly be concerned with the case in which the features are basis vectors.

Proof of Theorem 2. Given Lemma A.2 and $\zeta = 1$, it suffices to show that, for any (basis) features vector $\bar{f}$, and any non-negative integer s, t, $M_s\bar{f} \propto M_t\bar{f}$.

For convenience, normalize one of the times to 0. We prove, by induction, that for non-negative integers t:

$$M_t\bar{f} \propto M_0\bar{f}.$$

Clearly the statement holds for $t = 0$. Assume

$$M_{t-1}\bar{f} \propto M_0\bar{f}. \tag{A.3}$$

It follows from (4) that

$$M_t = M_{t-1} + x_t f_t^\top. \tag{A.4}$$

It follows from (2), (3), and $\zeta = 1$ that

$$x_t = (\|M_{t-1}f_t\|)^{-1}M_{t-1}f_t. \tag{A.5}$$

Substituting (A.5) into (A.4) implies

$$M_t = M_{t-1} + (\|M_{t-1}f_t\|)^{-1}M_{t-1}f_t f_t^\top. \tag{A.6}$$

Consider some basis vector $\bar{f}$. It follows from (A.6) that

$$M_t\bar{f} = M_{t-1}\bar{f} + (\|M_{t-1}f_t\|)^{-1}(M_{t-1}f_t)(f_t^\top\bar{f}). \tag{A.7}$$

First assume $\bar{f} \neq f_t$. Then $f_t^\top\bar{f} = 0$, the second term on the right-hand side of (A.7) equals zero and

$$M_t\bar{f} = M_{t-1}\bar{f}.$$

Now assume $\bar{f} = f_t$. Then $f_t^\top\bar{f} = 1$ and

$$M_t\bar{f} = M_{t-1}\bar{f} + (\|M_{t-1}f_t\|)^{-1}M_{t-1}\bar{f} \propto M_{t-1}\bar{f}.$$

Thus for any basis vector $\bar{f}$,

$$M_t\bar{f} \propto M_{t-1}\bar{f} \propto M_0\bar{f},$$

where the second statement of proportionality follows from the induction step (A.3). $\square$

Lemma A.3. *Assume thus far an agent has experienced an event at times* $\{t_1, \ldots, t_\ell\}$*. Then Assumption 2 implies*

$$x_{t_\ell}^{\text{in}} = \frac{1}{\ell}\left(x_{t_1}^{\text{in}} + \sum_{k=1}^{\ell-1} x_{t_k}\right). \tag{A.8}$$

Proof. Without loss of generality, assume the event is represented by basis vector e_1. A direct application of (3) implies that, should this occur at time t:

$$x_t^{\text{in}} \propto M_{t-1}e_1 \tag{A.9}$$

Substituting in from (5), we find:

$$x_t^{\text{in}} \propto M_0 e_1 + \sum_{s=1}^{t-1} x_s(f_s^\top e_1) \tag{A.10}$$

If the agent experienced the event at time $s < t$, $f_s^\top e_1 = 1$; otherwise it is equal to zero (note that the Lemma assumes all features are basis vectors). Therefore:

$$x_t^{\text{in}} \propto M_0 e_1 + \sum_{s \in \{t_1, \ldots, t_\ell\}} x_s$$

Note that the vector on the right hand side has elements summing to ℓ. The result follows from the fact that elements of x_t^{in} must sum to 1. $\square$

Proof of Theorem 3. Consider retrieved context for the ℓth event. A slight rewriting of (A.8) implies

$$
\begin{aligned}
x_{t_\ell}^{\text{in}} &= \frac{1}{\ell}\left(x_{t_1}^{\text{in}} + \sum_{k=1}^{\ell-2} x_{t_k} + x_{t_{\ell-1}}\right) \\
&= \frac{1}{\ell}\left(\left(x_{t_1}^{\text{in}} + \sum_{s=1}^{\ell-2} x_{t_k}\right) + (1-\zeta)x_{t_{\ell-1}-1} + \zeta x_{t_{\ell-1}}^{\text{in}}\right) \qquad \text{(A.11)}
\end{aligned}
$$

The second line follows from (2):

$$
x_{t_{\ell-1}} = (1-\zeta)x_{t_{\ell-1}-1} + \zeta x_{t_{\ell-1}}^{\text{in}}.
$$

The term in the inner parentheses in (A.11) equals $(\ell-1)x_{t_{\ell-1}}^{\text{in}}$. This follows from (A.8), applied to $x_{t_{\ell-1}}^{\text{in}}$. Therefore,

$$
x_{t_\ell}^{\text{in}} = \frac{1}{\ell}\left((\ell-1)x_{t_{\ell-1}}^{\text{in}} + (1-\zeta)x_{t_{\ell-1}-1} + \zeta x_{t_{\ell-1}}^{\text{in}}\right).
$$

Collecting terms in $x_{t_\ell-1}^{\text{in}}$ establishes the result. $\square$

Proof of Theorem 4. Assume $\{Z_t\}$ is iid. Define a matrix P such that $P(i,j) = p(z_i, y_j)$, the joint probability of z_i and y_j. Then $P(i,j) \propto M_{t-1}$, with the constant of proportionality equal to the sum of the elements of M_{t-1}. Suppose $Y_t = y_j$. Then f_t is the jth basis vector and

$$
\begin{aligned}
x_t^{\text{in}} &= \frac{M_{t-1}f_t}{||M_{t-1}f_t||} \\
&= \frac{Pe_j}{||Pe_j||} \\
&= \left(\sum_i p(z_i, y_j)\right)^{-1} \begin{bmatrix} p(z_1, y_j) \\ \vdots \\ p(z_m, y_j) \end{bmatrix}.
\end{aligned}
$$

Note that $\sum_i p(z_i, y_j)$ is simply the unconditional probability of y_j. Thus

$$x_t^{\text{in}}(i) = p(z_i, y_j)\left(\sum_i p(z_i, y_j)\right)^{-1} = p(z_i \,|\, y_j),$$

the conditional probability of $Z_t = z_i$, given $Y_t = y_j$. $\square$

Proof of Theorem 5. It suffices to show that, as a function of features elements, x_t^{in} is uniformly continuous, where we define continuity by the L^1-norm. We first show that unscaled x_t^{in} is uniformly continuous. Define $t_0 = \iota^\top M_0 \iota$. We show that, for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\left\| \frac{1}{t_0 + t}(M_0 + \sum_{s=1}^{t} x_s f_s^\top)(f_t - \hat{f}_t), \right\| < \epsilon \tag{A.12}$$

provided that $\|f_t - \hat{f}_t\| < \delta$.

Standard triangle inequality argument imply that it suffices to show that

$$\begin{aligned} \left\| \frac{1}{t} \sum_{s=1}^{t} x_s f_s^\top (f_t - \hat{f}_t) \right\| &< \epsilon/2 && \text{(A.13)} \\ \left\| \frac{1}{t_0} M_0 (f_t - \hat{f}_t) \right\| &< \epsilon/2 && \text{(A.14)} \end{aligned}$$

For (A.13), note that $f_s^\top (f_t - \hat{f}_t)$ is a scalar, so that

$$\begin{aligned} \left\| \frac{1}{t} \sum_{s=1}^{t} x_s f_s^\top (f_t - \hat{f}_t) \right\| &= \frac{1}{t} \sum_{s=1}^{t} \|x_s\| \|f_s^\top (f_t - \hat{f}_t)\| \\ &\leq \frac{1}{t} \sum_{s=1}^{t} \|x_s\| \|f_s\| \|f_t - \hat{f}_t\| \\ &= \frac{1}{t} \sum_{s=1}^{t} \|f_t - \hat{f}_t\| = \|f_t - \hat{f}_t\|. && \text{(A.15)} \end{aligned}$$

For (A.14), note that

$$\left\|\frac{1}{t_0}M_0(f_t-\hat{f}_t)\right\| = \left\|\frac{1}{t_0}\sum_{j=1}^{m}M_0(i,j)(f_t(j)-\hat{f}_t(j))\right\|$$
$$\leq \left\|\frac{1}{t_0}\sum_{j=1}^{m}M_0(i,j)\right\|\max_j\{|f_t(j)-\hat{f}_t(j)|\}. \qquad \text{(A.16)}$$

It suffices then to choose δ so that the right-hand side of (A.15) and (A.16) are less than $\epsilon/2$.

We now extend this argument to show that x_t^{in} is uniformly continuous.[50] Define x_{*t}^{in} to be unscaled x_t^{in}:

$$x_{*t}^{\text{in}} = \frac{1}{t_0+t}(M_0+\sum_{s=1}^{t}x_s f_s^{\top})f_t$$
$$\hat{x}_{*t}^{\text{in}} = \frac{1}{t_0+t}(M_0+\sum_{s=1}^{t}x_s f_s^{\top})\hat{f}_t.$$

We have shown $\|\hat{x}_{*t}^{\text{in}}-x_{*t}^{\text{in}}\|<\epsilon$. Our aim is to show $\|\hat{x}_t^{\text{in}}-x_t^{\text{in}}\|<\epsilon$. Because we are interested in the limit for a fixed f_t (given t), and therefore a fixed x_{*t}^{in}, it suffices to show, with a suitable adjustment to ϵ, that

$$\|\hat{x}_t^{\text{in}}-x_t^{\text{in}}\|\|x_{*t}^{\text{in}}\|<\epsilon.$$

[50] We treat the previous history of contexts and features, as well as M_0, as fixed. Then x_t^{in} is a function of time and of f_t. We let $\hat{f}_t\to f_t$ and show that the convergence of x_t^{in} does not depend on t. It will, however, depend on the choice of f_t because it depends on the scale of $\frac{1}{t_0+t}(M_0+\sum_{s=1}^{t}x_s f_s^{\top})f_t$ as the subsequent argument makes clear.

Finally note that

$$\begin{aligned}
\|\hat{x}_t^{\text{in}} - x_t^{\text{in}}\|\|x_{*t}^{\text{in}}\| &= \left\| x_t^{\text{in}}\|x_{*t}^{\text{in}}\| - \hat{x}_t^{\text{in}}\|\hat{x}_{*t}^{\text{in}}\| + \hat{x}_t^{\text{in}}\|\hat{x}_{*t}^{\text{in}}\| - \hat{x}_t^{\text{in}}\|x_{*t}^{\text{in}}\| \right\| \\
&= \left\| (x_{*t}^{\text{in}} - \hat{x}_{*t}^{\text{in}}) - \hat{x}_t^{\text{in}}\left(\|\hat{x}_{*t}^{\text{in}}\| - \|x_{*t}^{\text{in}}\|\right) \right\| \\
&\leq \|x_{*t}^{\text{in}} - \hat{x}_{*t}^{\text{in}}\| + \|\hat{x}_t^{\text{in}}\| \left| \|\hat{x}_{*t}^{\text{in}}\| - \|x_{*t}^{\text{in}}\| \right| \\
&= \|x_{*t}^{\text{in}} - \hat{x}_{*t}^{\text{in}}\| + \left| \|\hat{x}_{*t}^{\text{in}}\| - \|x_{*t}^{\text{in}}\| \right| < \epsilon,
\end{aligned}$$

provided that $\|x_{*t}^{\text{in}} - \hat{x}_{*t}^{\text{in}}\| < \epsilon/2$. □

B Proofs of results in Section 2.3

Definition (Associated features). *Features vector $\bar{f}$ and state $Z_t = z_i$ are* associated *at time t if either one of the two conditions hold:*

1. $\hat{e}_i^\top M_0 \bar{f} \neq 0$
2. *There exists an $s \leq t$ such that for $f_s^\top \bar{f} \neq 0$, $x_s(i) \neq 0$.*

If $\bar{f}$ is associated with state i, then the agent has either experienced features $\bar{f}$ in a context that places weight on state i, or initial memory associates $\bar{f}$ with state i.

Definition (Uniquely associated features). *Features vector $\bar{f}$ and state $Z_t = z_i$ are* uniquely associated *at time t if $\bar{f}$ is only associated with state i at t.*

If $\bar{f}$ is uniquely associated with state i, it can only retrieve state i.

Notation. *Let $\Omega_{i,t} \subset \mathcal{B}^n$ denote the set of features uniquely associated with state i at time t. Let $\Omega_{i,t}^\perp \subset \mathcal{B}^n$ denote the set of features not associated with state i at time t.*

Lemma B.1. *1. Features retrieve a context placing weight on i if and only if these features are associated with state i.*

2. Features uniquely associated with state i retrieve only state i (if $f_{t+1} \in \Omega_{i,t}$, then $x^{\text{in}}_{t+1}(i) = 1$).

Proof. Let $\bar{f}$ be features at time $t+1$. Consider retrieved context (3):

$$\begin{aligned} x^{\text{in}}_{t+1} &\propto M_t \bar{f} \\ &\propto M_0 \bar{f} + \sum_{s=0}^{t} x_s (f_s^\top \bar{f}), \end{aligned}$$

it follows that

$$x^{\text{in}}_{t+1}(i) \propto \hat{e}_i^\top M_0 \bar{f} + \sum_{s=0}^{t} x_s(i)(f_s^\top \bar{f})$$

The right hand side is nonzero, if and only if $\bar{f}$ is associated with state i.

Now assume $\bar{f}$ is uniquely associated with state i. Then, for $k \neq i$, $x^{\text{in}}_{t+1}(k) = 0$. Because the elements of the context vector sum to 1, $x^{\text{in}}_{t+1}(i) = 1$. $\square$

Lemma B.2 (Context reset). *For a given integer $\tau > 0$, assume the agent experiences a sequence of features $f_{t+1}, \ldots, f_{t+\tau} \in \Omega^{\perp}_{i,t}$. Also assume the features are orthogonal to one another. Then*

$$x_{t+\tau}(i) = (1-\zeta)^{\tau} x_t(i). \tag{B.1}$$

Thus as $\tau \to \infty$, $x_{t+\tau}(i) \to 0$.

Proof. We prove (B.1) by induction on τ. It holds trivially for $\tau = 0$. Assume (B.1) holds for $\tau - 1$. Given features $f_{t+1}, \ldots, f_{t+\tau}$, it follows from (4) that

$$M_{t+\tau-1} = M_t + x_{t+1} f_{t+1}^\top + \cdots + x_{t+1} f_{t+t+\tau-1}^\top$$

Assume features $f_{t+\tau}$ not associated with i at t and orthogonal to $f_{t+1}, \ldots, f_{t+\tau-1}$. It follows from (3) that

$$
\begin{aligned}
x^{\text{in}}_{t+\tau} &\propto M_{t+\tau-1} f_{t+\tau} \\
&\propto M_t f_{t+\tau} + x_{t+1} f^{\top}_{t+1} f_{t+\tau} + \cdots + x_{t+1} f^{\top}_{t+\tau-1} f_{t+\tau} \\
&\propto M_t f_{t+\tau}
\end{aligned}
$$

Thus, lack of association at time t, $x^{\text{in}}_{t+\tau}(i) = 0$. Recall that we have assumed by induction that $x_{t+\tau-1}(i) = (1-\zeta)^{\tau-1} x_t(i)$. Then (B.1) follows from (2). □

Exponential decay of context, combined with a sufficiently large number of orthogonal features, implies the possibility of context "reset." Context reset is not a mere mathematical construction: novel features are used in the memory laboratory to reset context. These novel features, presumably orthogonal to the features that the agent has recently experienced, are introduced through "distractor tasks" that often involve solving arithmetic problems under time constraints (Howard and Kahana, 1999).

Proof of Theorem 6. See Lemma B.2. □

The sets $\Omega_{i,t}$ and $\Omega^{\perp}_{i,t}$ define subjective associations. It is useful to have notation for the analogous concepts in the physical world.

Notation. *Let $\mathcal{Y}_i \subset \mathcal{Y}$ denote the set of outcomes that can* only *occur in state i. That is:*

$$\mathcal{Y}_i \equiv \{y_j \in \mathcal{Y} : p(y_j \,|\, z_i) > 0 \;\&\; p(y_j \,|\, z_k) = 0, \forall k \neq i\}.$$

Let $\mathcal{Y}^{\perp}_i$ denote the set of outcomes that cannot occur in state i:

$$\mathcal{Y}^{\perp}_i \equiv \{y_j \in \mathcal{Y} : p(y_j | z_i) = 0\}.$$

Definition (Correct associations)**.** *The agent has correct associations with state i if the agent's associations reflect reality. That is: $\Omega_i = \mathcal{Y}_i$, and $\Omega^{\perp}_i = \mathcal{Y}^{\perp}_i$.*

If features unambiguously signal a state, then context shifts in the direction of that state.

Lemma B.3. *Consider a nonempty state i such that $\mathcal{Y}_i^\perp = \mathcal{Y} \setminus \mathcal{Y}_i$. Assume at time $t-1$ that the agent has correct associations with state i. Then*

1. *For $x_{t-1}(i) \in [0,1)$, $x_t(i) > x_{t-1}(i)$ if and only if $Z_t = z_i$.*

2. *For $x_{t-1}(i) = 1$, $x_t(i) = x_{t-1}(i)$ if and only if $Z_t = z_i$.*

Proof. First rewrite (2) as:

$$\Delta x_t(i) = \zeta(x_t^{\text{in}}(i) - x_{t-1}(i)). \tag{B.2}$$

Under the stated assumptions, $f_t \in \Omega_{i,t-1}$ implies $Y_t \in \mathcal{Y}_i$ and $f_t \in \Omega_{i,t-1}^\perp$ implies $Y_t \in \mathcal{Y}_i^\perp$. Moreover, by Lemma B.1, $f_t \in \Omega_{i,t-1}$ implies $x_t^{\text{in}}(i) = 1$ and $x_t^{\text{in}}(i) = 0$ otherwise. It follows that if $x_t^{\text{in}}(i) = 1$, we must have $Y_t \in \mathcal{Y}_i$. If $x_t^{\text{in}}(i) = 0$, we must have $Y_t \in \mathcal{Y}_i^\perp$. Therefore, $x_t^{\text{in}}(i) = 1$ if and only if $Z_t = z_i$. It then follows from (B.2) that

$$\Delta x_t(i) = \begin{cases} \zeta(1 - x_{t-1}(i)) & \text{if } Z_t = z_i \\ -\zeta x_{t-1}(i) & \text{otherwise} \end{cases} \tag{B.3}$$

Suppose first that $x_{t-1}(i) \in [0,1)$. If $Z_t = z_i$, $x_t(i) - x_{t-1}(i) = \zeta(1 - x_{t-1}(i)) > 0$. If $Z_t \neq z_i$, $x_t(i) - x_{t-1}(i) = -\zeta x_{t-1}(i) \leq 0$, establishing the first statement.

Now suppose $x_{t-1}(i) = 1$. If $Z_t = z_i$, then $x_t(i) = x_{t-1}(i) = 1$. If $Z_t \neq z_i$. $x_t(i) < x_{t-1}(i) = 1$, establishing the second statement. □

We use the notation $\mathbf{E}$ to denote the expectation taken under the econometrician's measure, whereas $\mathbb{E}$ is the expectation taken under the agent's subjective probability.

Proof of Theorem 7. We calculate $\mathbf{E}[\Delta x_{t+1}(i) \,|\, \Delta x_t(i) > 0]$, where $\mathbf{E}$ denotes the expectation that the econometrician calculates. We assume throughout that $x_{t-1}(i) <$

1. Then lemma B.3 and the assumptions of the theorem together imply that $\Delta x_t(i) > 0$ if and only if $Z_t = z_i$. We therefore need only calculate $\mathbf{E}[\Delta x_{t+1}(i) \,|\, Z_t = z_i]$.

Given that $Z_t = z_i$, we consider what happens at $t+1$. If $Z_{t+1} \neq z_i$, $f_{t+1} \neq f_t$ by the assumptions of the theorem. We have $y_{t+1} \in \mathcal{Y}_i^\perp$, and $f_{t+1} \in \Omega_{i,t}^\perp$. Thus, $x_{t+1}^{\text{in}}(i) = 0$ by Lemma B.1. If, on the other hand, $Z_{t+1} = z_i$, $y_{t+1} \in \mathcal{Y}_i$, and $f_{t+1} \in \Omega_{i,t-1}$. Recall

$$x_{t+1}^{\text{in}} \propto (M_{t-1} + x_t f_t^\top) f_{t+1}. \tag{B.4}$$

If the features are again novel ($f_{t+1} \neq f_i$), then $x_{t+1}^{\text{in}}(i) = 1$ by Lemma B.1. However, if $f_{t+1} = f_t$, (B.4) becomes

$$x_{t+1}^{\text{in}} \propto M_{t-1} f_{t+1} + x_t.$$

The first term is proportional to $\hat{e}_i$ by assumption. Its magnitude will depend on the magnitude of the elements in M_{t-1}. Let t_0 be the length of the prior sample.[51] Lemma A.1, implies

$$x_{t+1}^{\text{in}} \propto (t + t_0 - 1)\hat{e}_i + x_t$$

and therefore that

$$x_{t+1}^{\text{in}}(i) = (\|(t + t_0 - 1)\hat{e}_i + x_t\|)^{-1}(t + t_0 - 1 + x_t(i)). \tag{B.5}$$

To summarize, conditional on $Z_{t+1} = z_i$, we have

$$x_{t+1}^{\text{in}}(i) = \begin{cases} 1 & \text{if } f_{t+1} \neq f_t \\ (\|(t + t_0 - 1)\hat{e}_i + x_t\|)^{-1}(t + t_0 - 1 + x_t(i)) & f_{t+1} = f_t \end{cases}$$

Let

$$\bar{x}_t^{\text{in}}(i; x_t) = \mathbf{E}_t[x_{t+1}^{\text{in}}(i) | Z_t = Z_{t+1} = z_i] \tag{B.6}$$

[51] The length of the prior sample is the sum of the elements in M_0. See Lemma A.1 for further discussion.

Note that $x_t(i) < \bar{x}_t(i; x_t) < 1$. A large prior sample or many elements of $\mathcal{Y}_i$ give us $\bar{x}_t(i; x_t)$ close to 1.

Substituting (B.6) into (B.2) implies

$$\mathbf{E}_t[\Delta x_{t+1}(i) | Z_{t+1} = Z_t = z_i] = \begin{cases} \zeta(\bar{x}^{\text{in}}(i; x_t) - x_t(i)) & Z_{t+1} = z_i \\ -\zeta x_t(i) & \text{otherwise} \end{cases}$$

Taking the expectation over the possible outcomes of Z_{t+1}, we find:

$$\begin{aligned} \mathbf{E}_t[\Delta x_{t+1}(i) | \Delta x_t(i) > 0] &= \mathbf{E}_t[\Delta x_{t+1}(i) | Z_t = z_i] \\ &= \zeta(p^Z_{ii} \bar{x}^{\text{in}}(i; x_t) - x_t(i)). \end{aligned}$$

The theorem follows. $\square$

Proof of Theorem 8. Because Z_t is iid, $x_t^{\text{in}}(i) = p(z_i \,|\, Y_t)$ (Theorem 4). Applying (B.2):

$$\Delta x_{t+1}(i) = \zeta(p(z_i | Y_{t+1}) - x_t(i)). \tag{B.7}$$

The assumption of iid Z_t implies that Y_t is also iid. Thus,

$$\mathbf{E}_t p(z_i | Y_{t+1}) = \mathbf{E} p(z_i | Y_{t+1}) = p(z_i). \tag{B.8}$$

We now calculate $\mathbf{E}_t x_t(i)$, conditional on an upward revision in beliefs from $t-1$ to t. First note $p(z_i) = \sum_{j=1}^n p(z_i | y_j) p(y_j)$. For $y_j \in \mathcal{Y}_i^{\perp}$, $p(z_i | y_j) = 0$. Because $p(z_i)$ averages these zero terms with terms j such that $p(z_i | y_j) > 0$,

$$\mathbf{E}\left[p(z_i | Y_t) \,|\, Y_t \notin \mathcal{Y}_i^{\perp}\right] > p(z_i).$$

If we further add the provision that Y_t is such that $p(z_i | Y_t) > \bar{x} \geq 0$, we weakly increase the expectation on the left hand side above. That is,

$$\begin{aligned}\mathbf{E}\left[p(z_i|Y_t)\,|\,p(z_i|Y_t)>\bar{x}\right] &= \mathbf{E}\left[p(z_i|Y_t)\,|\,Y_t\notin\mathcal{Y}_i^{\perp}\ \&\ p(z_i|Y_t)>\bar{x}\right]\\ &\geq \mathbf{E}\left[p(z_i|Y_t)\,|\,Y_t\notin\mathcal{Y}_i^{\perp}\right]\\ &> p(z_i),\end{aligned}\tag{B.9}$$

because $p(z_i|y_j)>\bar{x}$ implies $p(z_i|y_j)>0$ which implies $y_j\notin\mathcal{Y}_i^{\perp}$.

By definition, an upward revision in expectations about state i at time t occurs if and only if $x_t(i)>x_{t-1}(i)$, which in turn occurs if and only if $x_t^{\text{in}}(i)>x_{t-1}(i)$. Finally, recall $p(z_i\,|\,Y_t)=x_t^{\text{in}}(i)$. Putting these pieces together:

$$\begin{aligned}\mathbf{E}\left[x_t^{\text{in}}(i)|\Delta x_t(i)>0\right] &= \mathbf{E}\left[p(z_i|Y_t)\,|\,\Delta x_t(i)>0\right]\\ &= \mathbf{E}\left[p(z_i|Y_t)\,|\,p(z_i|Y_t)>x_{t-1}(i)\right]\\ &> p(z_i),\end{aligned}\tag{B.10}$$

by (B.9). It then follows from (B.7) that

$$\begin{aligned}\mathbf{E}\left[\Delta x_{t+1}(i)|\Delta x_t(i)>0\right] &= \zeta\left(p(z_i)-\mathbf{E}[(1-\zeta)x_{t-1}(i)+\zeta x_t^{\text{in}}(i)|\Delta x_t(i)>0]\right)\\ &= \zeta\left(p(z_i)-\zeta\mathbf{E}\left[x_t^{\text{in}}(i)|\Delta x_t(i)>0\right]\right)-(1-\zeta)\mathbf{E}\left[x_{t-1}(i)|\Delta x_t(i)>0\right]\end{aligned}$$

For ζ sufficiently large, the first term $p(z_i)-\zeta\mathbf{E}\left[x_t^{\text{in}}(i)|\Delta x_t(i)>0\right]$ is negative by (B.10), whereas the second term is small. □

C Bayesian updating from rare events

Consider a Bayesian agent learning about the probability of a rare event from observations of the event. In the terminology of Section 2.1, we assume for the purpose of this section that Z_t is iid, and that there are two outcomes of $\{Z_t\}$, and that the set $\mathcal{Y}$ partitions into outcomes possible in one of these states and those possible in the other. This is of course the same as saying that Z_t is observable.

Let p denote the probability of the rare event. The agent has prior

$$p \sim \text{Beta}(p^*\tau + 1, (1 - p^*)\tau + 1), \tag{C.1}$$

for $p^*, \tau \geq 0$. This prior corresponds to beliefs if the agent had begun with a prior that is uniform on $[0, 1]$ and observed a sample of length τ, of which there were $p^*\tau$ occurrences of the rare event. The density function corresponding to (C.1) is given by

$$f(p) \propto p^{p^*\tau}(1 - p)^{(1-p^*)\tau},$$

where the constant of proportionality does not depend on p.

Assume T years of data. For concreteness, we will call the rare event a crisis. Conditional on the probability p, the likelihood of exactly N occurrences of the event equals

$$\mathcal{L}(N \text{ crises} \,|\, p) = \binom{T}{N} p^N (1 - p)^{T-N}. \tag{C.2}$$

Therefore the posterior distribution equals

$$\begin{aligned} f(p \,|\, N \text{ crises}) &\propto \mathcal{L}(N \text{ crises}|p) f(p) \\ &\propto p^{N+p^*\tau}(1 - p)^{T+\tau-(N+p^*\tau)} \end{aligned}$$

where once again we have ignored terms that do not depend on p. This is proportional to the Beta density, so

$$p \,|\, N \text{ crises} \sim \text{Beta}(N + p^*\tau + 1, T + \tau - (N + p^*\tau) + 1).$$

It follows from properties of the Beta distribution that the posterior mean equals

$$\mathbb{E}[p \,|\, N \text{ crises}] = \frac{N + p^*\tau + 1}{T + \tau + 2}. \tag{C.3}$$

The posterior mean depends on the sample path. Figure 3 shows the average posterior mean, assuming the likelihood (C.2):

$$\begin{aligned} \mathbb{E}_{\#\text{crises}}\left[\mathbb{E}[p \,|\, N\text{crises}]\right] &= \int \frac{N + p^*\tau + 1}{T + 2} \mathcal{L}(N \text{ crises} \,|\, p)\, dN \\ &= \frac{pT + p^*\tau + 1}{T + \tau + 2} \end{aligned}$$

where we have used the fact that, conditional on p, N has a binomial distribution, and therefore $\mathbb{E}[N \,|\, p] = pT$. The figure corresponds to the case of $\tau = 0$, however the results are very similar for $\tau > 0$, provided that the actual sample is large relative to the prior sample.

D Generalizing results in Section 2.5

Consider the setting of Section 2.5: the agent experiences a crisis at time $t - 1$ (say, 1929), followed by a depression at time t. Section 2.5 considers the effect of re-appearance of crisis features at time $t' > t$ under two simplifying assumptions, (1) x_{1929} is a basis vector and (2) $x^{\text{in}}_{t'} = x_{t'}$. In this appendix we relax these assumptions. Given the relative uniqueness of the 1929 stock market crash and the Great Depression we fix ideas by assuming that f_{crisis} and $f_{\text{depression}}$ were novel events, and that prior associations through M_0 were sufficiently weak as to be negligible. We assume that depression features last for k periods. Because our primary interest is the jump-back-in time and not details of associations per se, we assume that $f_{\text{depression}}$ did not reoccurred prior to t', and that the context retrieved by the *original* depression features is not associated with the context in 1929.[52]

The reappearance of crisis features, $f_{t'} = f_{\text{crisis}}$ implies $x^{\text{in}}_{t'} = x_{1929}$ as in Section 2.5. Context at time t' is now:

$$x_{t'} = (1-\zeta)x_{t'-1} + \zeta x_{1929}. \tag{D.1}$$

Assume that t' represents the first appearance of crisis features, so that $x^{\top}_{t'-1}x_{1929} = 0$. Let $w_{1929,i}$ be the projection of x_{1929} onto basis vector i, so that

$$x_{1929} = \sum_{i=1}^{m} w_{1929,i}\hat{e}_i \qquad w_{1929,i} \geq 0. \tag{D.2}$$

with $\sum_{i=1}^{m} w_{1929,i} = 1$.

From (10), it follows that features retrieved by (basis) context vector $\hat{e}_i$ at time t' equal

$$f^{\text{in}}_{i,t'} = \alpha_{i,t'} M^{\top}_{t'-1}\hat{e}_i, \qquad i = 1, \ldots, m, \tag{D.3}$$

where $\alpha_{i,t'} = ||M^{\top}_{t'-1}\hat{e}_i||^{-1}$ ensures that the elements of $f^{\text{in}}_{i,t'}$ sum to 1. According to (11), we calculate the features retrieved by context at time t' by taking the weighted sum of (D.3). Because context at time t' is itself a weighted sum of prior context and retrieved context, we can consider each of these terms separately.

Define the notation

$$f^{\text{in}}_{1929,t'} \equiv \sum_{i=1}^{m} w_{1929,i} f^{\text{in}}_{i,t'}.$$

By (11), $f^{\text{in}}_{1929,t'}$ are the features retrieved by $x^{\text{in}}_{t'} = x_{1929}$. The subjective probability of a depression implied by $f^{\text{in}}_{1929,t'}$ equals the inner product $(f_{\text{depression}})^{\top} f^{\text{in}}_{1929,t'}$. Note that $f_{\text{depression}}$ is the basis vector representing physical depression features, so the inner product is simply the entry of $f^{\text{in}}_{1929,t'}$ that corresponds to depressions among all the

[52]This would occur if depression features were novel between t and $t+k-1$. We disregard the small effect of learning in this initial period.

elements of $\mathcal{Y}$. The inner product equals

$$(f_{\text{depression}})^\top f^{\text{in}}_{1929,t'} = \sum_{i=1}^{m} w_{1929,i} (f_{\text{depression}})^\top f^{\text{in}}_{i,t'}.$$

Using (D.3):

$$\begin{aligned}
(f_{\text{depression}})^\top f^{\text{in}}_{i,t'} &= \alpha_i (f_{\text{depression}})^\top M^\top_{t'-1} \hat{e}_i && \text{(D.4)} \\
&= \alpha_i (f_{\text{depression}})^\top \left(M_0^\top \hat{e}_i + \sum_{s=1}^{t'-1} f_s x_s^\top \right) \hat{e}_i && \text{(D.5)} \\
&= \alpha_i \sum_{s=1}^{t'-1} (f_{\text{depression}})^\top f_s)(x_s^\top \hat{e}_i) && \text{(D.6)} \\
&= \alpha_i (x_t + \cdots + x_{t+k-1})^\top \hat{e}_i && \text{(D.7)} \\
&= \alpha_i (\sum_{l=1}^{k} (1-\zeta)^l x_{1929} + x^\perp_{1929})^\top \hat{e}_i, && \text{(D.8)}
\end{aligned}$$

where $x^\perp_{1929}$ is a vector orthogonal to x_{1929}. Equation D.6 follows from the lack of prior associations for depression features through M_0. Equation D.7 follows from the assumption of no depression features after time $t + k - 1$, and (D.8) follows from the orthogonality of retrieved context and x_{1929} between t and $t + k - 1$. Because $x^\top_{1929} \hat{e}_i = w_{1929,i}$,

$$(f_{\text{depression}})^\top f^{\text{in}}_{i,t'} = \left(\frac{1}{\zeta} - 1 \right) \left(1 - (1-\zeta)^k \right) \sum_{i=1}^{m} \alpha_i w^2_{1929,i} \tag{D.9}$$

In the case of basis x_{1929}, $w_{1929,i}$ equals 1 for exactly one i and is otherwise 0. The weight α_I is determined by how common the context x_{1929} is. If uncommon, then α_i simply equals $\left(\frac{1}{\zeta} - 1 \right) \left(1 - (1-\zeta)^k \right)$ and x_{1929} retrieves a probability of 1.

Note however, that total context is not $x^{\text{in}}_{t'} = x_{1929}$ but rather (D.1). Under the assumption of orthogonality, the probability of a depression goes from zero (features retrieved by $x_{t'-1}$ to ζ multiplied by (D.9).

E Proofs for Section 4.1

This Appendix contains proofs generalizing the results in Section 4.1 to $\zeta < 1$. We first derive a recursion, analogous to that in Theorem 3, characterizing features retrieved by a fixed context $\hat{e}_i$. The following is a simple extension of (5) to retrieved features.

Lemma E.1. *Features retrieved by context $\hat{e}_i$ as of time t satisfy:*

$$f_{t,i}^{\text{in}} = \left(\iota^\top M_0^\top \hat{e}_i + \sum_{s=1}^{t-1} (x_s^\top \hat{e}_i)\right)^{-1} \left(M_0^\top \hat{e}_i + \sum_{s=1}^{t-1} f_s (x_s^\top \hat{e}_i)\right) \tag{E.1}$$

where f_s can be physical or retrieved features, and where ι is the $n \times 1$ vector of ones.

Proof. Combining (10) with (4) implies

$$\begin{aligned} f_{t,i}^{\text{in}} &\propto M_0^\top \hat{e}_i + \sum_{s=1}^{t} (f_s x_s^\top) \hat{e}_i \\ &\propto M_0^\top \hat{e}_i + \sum_{s=1}^{t} f_s (x_s^\top \hat{e}_i) \end{aligned}$$

which is analogous to (5) for retrieved features. To construct the normalizing constant, pre-multiply (E.2) with ι, and recall that $\iota^\top f_s = 1$ for all s. □

The following is an extension of Theorem 3 to features retrieval.

Theorem E.2. *Let $k_i = \|M_0^\top \hat{e}_i\|$. Assume retrieved features are encoded with context. Then retrieved features obey the following recursion*

$$f_{t,i}^{\text{in}} = \left(1 - \frac{(x_{t-1}^\top \hat{e}_i)(1 - x_{t-1}^\top \hat{e}_i)}{k_i + \sum_{s=1}^{t-1} x_s^\top \hat{e}_i}\right) f_{t-1,i}^{\text{in}} + \frac{(x_{t-1}^\top \hat{e}_i)(1 - x_{t-1}^\top \hat{e}_i)}{k_i + \sum_{s=1}^{t-1} x_s^\top \hat{e}_i} f_{t-1,i}^{\text{in},\perp}, \tag{E.2}$$

where

$$f_{t-1,i}^{\text{in},\perp} = (1 - x_{t-1}^\top \hat{e}_i)^{-1} \sum_{j \neq i} x_{t-1}(j) f_{t-1,j}^{\text{in}}, \tag{E.3}$$

namely $f^{\text{in},\perp}_{t-1,i}$ *represents features retrieved at* $t-1$ *by context elements other than* i*. The initial condition is* $f_{0,i} \propto M_0^\top \hat{e}_i$*.*

Proof. We apply (E.1), setting $k_i = \iota^\top M_0^\top \hat{e}_i$ and $f_s = f^{\text{in}}_s$:

$$f^{\text{in}}_{t,i} = \left(k_i + \sum_{s=1}^{t-1} (x_s^\top \hat{e}_i)\right)^{-1} \left(M_0^\top \hat{e}_i + \sum_{s=1}^{t-1} f^{\text{in}}_s (x_s^\top \hat{e}_i)\right) \tag{E.4}$$

We rewrite (E.4), using the same recursive reasoning as in the proof of Theorem 3:

$$f^{\text{in}}_{t,i} = \left(k_i + \sum_{s=1}^{t-1} (x_s^\top \hat{e}_i)\right)^{-1} \underbrace{\left(M_0^\top \hat{e}_i + \sum_{s=1}^{t-2} f^{\text{in}}_s (x_s^\top \hat{e}_i)\right)}_{(1+\sum_{s=1}^{t-2} x_s^\top \hat{e}_i) f^{\text{in}}_{t-1,i}} + \left(k_i + \sum_{s=1}^{t-1} x_s^\top \hat{e}_i\right)^{-1} f^{\text{in}}_{t-1}(x_{t-1}^\top \hat{e}_i). \tag{E.5}$$

Note that we apply (E.1) at $t-1$ to conclude $M_0^\top \hat{e}_i + \sum_{s=1}^{t-2} f^{\text{in}}_s (x_s^\top \hat{e}_i) = (1+\sum_{s=1}^{t-2} x_s^\top \hat{e}_i) f^{\text{in}}_{t-1,i}$.

Retrieved features at time $t-1$ are a weighted average of those retrieved by $\hat{e}_i$, and those retrieved by the other elements of context.

$$f^{\text{in}}_{t-1} = f^{\text{in}}_{t-1,i}(x_{t-1}^\top \hat{e}_i) + f^{\text{in},\perp}_{t-1,i}(1 - x_{t-1}^\top \hat{e}_i), \tag{E.6}$$

where $f^{\text{in},\perp}_{t-1,i}$ is as defined in (E.3). Combining (E.5) and (E.6) shows that $f^{\text{in}}_{t,i}$ is a weighted average of $f^{\text{in}}_{t-1,i}$ and $f^{\text{in},\perp}_{t-1,i}$. Moreover, the coefficient multiplying $f^{\text{in},\perp}_{t-1,i}$ must equal $\left(k_i + \sum_{s=1}^{t-1} x_s^\top \hat{e}_i\right)^{-1} (1 - x_{t-1}^\top \hat{e}_i)(x_{t-1}^\top \hat{e}_i)$. Because the elements of $f^{\text{in}}_{t,i}$ must sum to 1, it follows that the coefficient on $f^{\text{in}}_{t-1,i}$ equals one minus this quantity, as shown in (E.2). $\square$

The following Lemma generalizes Lemma A.3 to encoding of retrieved features. We assume that only the physical event triggers encoding of features that are non-orthogonal to the event. For example, if the event is a stock market loss, we disregard events that were not losses, but nonetheless reminded the agent of losses, as second-order.

Lemma E.3. *Assume the agent experiences an event at $\{t_1, \ldots, t_\ell\}$, and that features vectors are otherwise orthogonal to the event. Then retrieved context in response to $f_{t_\ell} = e_i$ is proportional to*

$$x^{\text{in}}_{t_\ell} \propto M_0 e_i + \sum_{k=2}^{\ell-1} x_{t_k} (f_{t_k}^\top e_i) \tag{E.7}$$

Note that when features are basis vectors, (E.7) reduces to (A.8).

Proof of Lemma E.3. By (5) and the fact that t_1 is the first occurence of the event:

$$x^{\text{in}}_{t_1} \propto M_0 e_i.$$

Moreover,

$$M_{t_1 - 1} e_i = M_0 e_i$$

Assume by induction that (E.7) holds for the $(\ell-1)$st occurrence of the event:

$$x^{\text{in}}_{t_{\ell-1}} \propto M_0 e_i + \sum_{k=2}^{t_{\ell-2}} x_{t_k} (f_{t_k}^\top e_i). \tag{E.8}$$

and that

$$M_{t_{\ell-1}-1} e_i = M_0 e_i + \sum_{k=2}^{t_{\ell-2}} x_{t_k} (f_{t_k}^\top e_i). \tag{E.9}$$

By (2), context equals

$$x_{t_{\ell-1}} = (1-\zeta) x_{t_{\ell-1}-1} + \zeta x^{\text{in}}_{t_{\ell-1}}$$

and memory is updated as:

$$M_{t_{\ell-1}} = M_{t_{\ell-1}-1} + x_{t_{\ell-1}} f_{t_{\ell-1}}^{\top}. \tag{E.10}$$

where it is not necessary to take a stance on whether $f_{t_{\ell-1}} = e_i$ or features retrieved by $x_{t_{\ell-1}}$.

By (3),

$$x_{t_\ell}^{\text{in}} \propto M_{t_\ell - 1} e_i.$$

By definition, $t_{\ell-1}$ is the occurrence of the event just before the occurence at t_ℓ. Because of this and additional assumptions of the theorem, all features between $t_{\ell-1}$ and t_ℓ are orthogonal to e_i. Therefore we can ignore terms in $M_{t_\ell - 1}$ that occur after $t_{\ell-1}$, and

$$x_{t_\ell}^{\text{in}} \propto M_{t_{\ell-1}} e_i$$

Substituting in from (E.10):

$$x_{t_\ell}^{\text{in}} \propto (M_{t_{\ell-1}-1} + x_{t_{\ell-1}} f_{t_{\ell-1}}^{\top}) e_i.$$

Substituting in from (E.9):

$$x_{t_\ell}^{\text{in}} \propto M_0 e_i + \sum_{k=2}^{t_{\ell-2}} x_{t_k} (f_{t_k}^{\top} e_i) + x_{t_{\ell-1}} (f_{t_{\ell-1}}^{\top} e_i).$$

Because we can ignore terms in $M_{t_\ell - 1}$ that occur after $t_{\ell-1}$:

$$M_{t_\ell - 1} e_i = M_{t_{\ell-1}} e_i$$

It follows from (E.9) and (4) that

$$M_{t_{\ell-1}} e_i = M_0 e_i + \sum_{k=2}^{t_{\ell-2}} x_{t_k} (f_{t_k}^\top e_i) + x_{t_{\ell-1}} (f_{t_{\ell-1}}^\top e_i),$$

completing the proof. $\square$

References

Abbott, L. F. and Blum, K. I. (1996). Functional significance of long-term potentiation for sequence learning and prediction. *Cerebral Cortex*, 6(3):406–416.

Allen, F. and Gale, D. (2009). *Understanding Financial Crises*. Clarendon lectures in finance. Oxford University Press.

Andreoni, J. and Mylovanov, T. (2012). Diverging opinions. *American Economic Journal: Microeconomics*, 4(1):209–232.

Angeletos, G.-M., Huo, Z., and Sastry, K. A. (2020). *Imperfect Macroeconomic Expectations: Evidence and Theory*. NBER Macroeconomic Annual. University of Chicago Press.

Arrow, K. J. (1971). *Essays in the Theory of Risk-Bearing*. Markham Publishing Co, Chicago.

Azeredo da Silveira, R. and Woodford, M. (2019). Noisy memory and over-reaction to news. *AEA Papers and Proceedings*, 109:557–61.

Barberis, N., Greenwood, R., Jin, L., and Shleifer, A. (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115(1):1–24.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, page 37.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, 27(1):173–196.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2020). Expectations of fundamentals and stock market puzzles. Working Paper 27283, National Bureau of Economic Research.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2019). Memory, attention, and choice. NBER Working Paper No. 23256.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In Melton, A. W. and Martin, E., editors, *Coding Processes in Human Memory*, chapter 5, pages 85–121. John Wiley and Sons, New York.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, 11(5):4–4.

Brown, G. D. A., Neath, I., and Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3):539–576.

Brown, G. D. A., Preece, T., and Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107(1):127–181.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Burgess, N. and Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55:627–652.

Caballero, R. J. and Simsek, A. (2020). A risk-centric model of demand recessions and speculation*. *The Quarterly Journal of Economics*, 135(3):1493–1566.

Campbell, J. Y. (2016). Restoring rational choice: The challenge of consumer financial regulation. *American Economic Review*, 106(5):1–30.

Carr, H. A. (1931). The laws of association. *Psychological Review*, 38:212–228.

Cohen, R. T. and Kahana, M. J. (2020). Retrieved-context theory of memory in emotional disorders. Working paper, University of Pennsylvania.

Cohn, A., Engelmann, J., Fehr, E., and Marechal, M. A. (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *The American Economic Review*, 105(2):860–885.

Crowder, R. G. (1976). *Principles of learning and memory*. Lawrence Erlbaum and Associates, Hillsdale, NJ.

Cuculiza, C., Antoniou, C., Kumar, A., and Maligkris, A. (2020). Terrorist attacks, analyst sentiment, and earnings forecasts. forthcoming, *Management Science*.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance*, 53(6):1839–1885.

Ebbinghaus, H. (1885/1913). *Memory: a contribution to experimental psychology*. Teachers College, Columbia University, New York.

Eich, E. (1995). Searching for mood dependent memory. *Psychological Science*, 6(2):67–75.

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425:184–187.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Enke, B., Schwerter, F., and Zimmermann, F. (2020). Associative memory and belief formation. working paper, Harvard University.

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62:145–154.

Estes, W. K. (1959). *Component and pattern models with Markovian interpretations. Studies in Mathematical Learning Theory*. Stanford University Press, Stanford, CA.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, 18(4):500–549.

Folkerts, S., Rutishauser, U., and Howard, M. (2018). Human episodic memory retrieval is accompanied by a neural contiguity effect. *Journal of Neuroscience*, 38(17):4200–4211.

French, K. R., Baily, M. N., Campbell, J. Y., Cochrane, J. H., Diamond, D. W., Duffie, D., Kashyap, A. K., Mishkin, F. S., Rajan, R. G., Scharfstein, D. S., Shiller, R. J., Shin, H. S., Slaughter, M. J., Stein, J. C., and Stulz, R. M. (2010). *The Squam Lake Report: Fixing the Financial System*. Princeton University Press.

Fuster, A., Laibson, D., and Mendel, B. (2010). Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives*, 24(4):67–84.

Gabaix, X. (2019). Behavioral inattention. In Bernheim, D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics*, volume 2. Elsevier.

Gallistel, C. R. and King, A. P. (2009). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. Wiley-Blackwell.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL.

Gennaioli, N. and Shleifer, A. (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press, Princeton, NJ.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

Gilboa, I. and Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*, 110(3):605–639.

Godden, D. R. and Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, 66:325–331.

Gödker, K., Jiao, P., and Smeets, P. (2019). Investor Memory. Working paper, Maastrich University.

Goetzmann, W. N., Kim, D., and Shiller, R. J. (2017). Crash beliefs from investor surveys. NBER Working Paper No. 22143.

Gomes, J. F., Grotteria, M., and Wachter, J. A. (2019). Cyclical dispersion in expected defaults. *Review of Financial Studies*, 32(4):1275–1308.

Gorton, G. and Metrick, A. (2012). Securitized banking and the run on repo. *Journal of Financial Economics*, 104(3):425–451.

Gourio, F. (2012). Disaster risk and business cycles. *American Economic Review*, 102(6):2734–2766.

Greene, R. L. (1992). *Human memory: Paradigms and paradoxes*. Lawrence Erlbaum and Associates, Hillsdale, New Jersey.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408.

Guiso, L., Sapienza, P., and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, 128(3):403–421.

Hamilton, J. D. (1994). *Time Series Analysis*. Oxford University Press, Princeton, NJ.

Hansen, L. P. and Sargent, T. J. (2018). Structured uncertainty and model misspecification. Working paper, New York University and University of Chicago.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.

Healey, M. K., Long, N. M., and Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 26(3):699—720.

Henson, R., Norris, D. G., Page, M. P. A., and Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology*, 49A:80–115.

Herbart, J. F. (1834). Lehrbuch zur Psychologie. In Kehrbach, K., editor, *Sämtliche Werke*, volume 4, pages 376–377. August Wilhelm Unzer, Königsberg, 2nd edition.

Hinrichs, J. V. and Buschke, H. (1968). Judgment of recency under steady-state conditions. *Journal of experimental psychology*, 78(4p1):574.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in multiple-trace memory model. *Psychological Review*, 95:528–551.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184.

Howard, M. W., Jing, B., Rao, V. A., Provyn, J. P., and Datey, A. V. (2009). Bridging the gap: Transitive associations between items presented in similar temporal contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35:391–407.

Howard, M. W. and Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4):923–941.

Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3):269–299.

Howard, M. W., Viskontas, I. V., Shankar, K. H., and Fried, I. (2012). Ensembles of human MTL neurons "jump back in time" in response to a repeated stimulus. *Hippocampus*, 22:1833–1847.

Hume, D. (1748). *An Enquiry Concerning Human Understanding.* Retrieved from Project Gutenberg: http://www.gutenberg.org/files/9662/9662-h/9662-h.htm.

James, W. (1890). *The principles of psychology.* Henry Holt and Co, Inc., New York, NY, US.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48:65–91.

Jones, C. M. (2002). A century of stock market liquidity and trading costs. Working paper, Columbia University.

Jost, A. (1897). Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 16:436–472.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24(1):103–109.

Kahana, M. J. (2012). *Foundations of Human Memory.* Oxford University Press, New York, NY.

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, 71:107–138.

Kahana, M. J. and Jacobs, J. (2000). Inter-response times in serial recall: Effects of intraserial repetition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26:1188–1197.

Kahle, K. M. and Stulz, R. (2013). Access to capital, investment, and the financial crisis. *Journal of Financial Economics*, 110(2):280–299.

Kempter, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E*, 59:4498–4514.

Kolers, P. A. and Magee, L. E. (1978). Specificity of pattern-analyzing skills in reading. *Canadian Journal of Psychology*, 32:43–51.

Kuhn, J. R., Lohnas, L. J., and Kahana, M. J. (2018). A spacing account of negative recency in final free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(8):1180–1185.

Ladd, G. T. and Woodworth, R. S. (1911). *Elements of physiological psychology: A treatise of the activities and nature of the mind from the physical and experimental point of view.* Charles Scribner's Sons, New York, NY.

Lashley, K. (1951). The problem of serial order in behavior. In *Cerebral Mechanisms in Behavior*. Wiley, New York.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review*, 90(2):426–432.

Lohnas, L. J. and Kahana, M. J. (2014a). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40(1):12–24.

Lohnas, L. J. and Kahana, M. J. (2014b). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(3):755–764.

Lohnas, L. J., Polyn, S. M., and Kahana, M. (2015). Expanding the scope of memory search: Intralist and interlist effects in free recall. *Psychological Review*, 122(2):337–363.

Long, N. M., Danoff, M. S., and Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin and Review*, 22(5):1328–1333.

Longstaff, F. A. and Piazzesi, M. (2004). Corporate earnings and the equity premium. *Journal of Financial Economics*, 74:401–421.

Lucas, R. E. (1975). An equilibrium model of the business cycle. *Journal of Political Economy*, 83(6):1113–1144.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 46:1429–1445.

Luce, R. D. (1959). *Individual choice behavior*. Wiley, New York.

MacDonald, C., Lepage, K., Eden, U., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737–749.

Makino, H. and Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. *Nature neuroscience*, 18(8):1116–1122.

Malmendier, U. and Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1):373–416.

Malmendier, U. and Nagel, S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1):53–87.

Malmendier, U., Nagel, S., and Yan, Z. (2017). The making of hawks and doves: Inflation experiences on the FOMC. NBER Working Paper No. 23228.

Malmendier, U. and Shen, L. S. (2018). Scarred consumption. NBER Working Paper No. 24696.

Manning, J., Hulbert, J., Williams, J., Piloto, L., Sahakyan, L., and Norman, K. (2016). A neural signature of contextually mediated intentional forgetting. *Psychonomic Bulletin & Review*, 23:1534–1542.

Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences, USA*, 108(31):12893–12897.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Matt, G. E., Vázquez, C., and Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, 12(2):227–255.

McFadden, D. (2001). Economic choices. *American Economic Review*, 91(3):351–378.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, 39:352–70.

Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., Hefft, S., Merkow, M., Polyn, S. M., Jacobs, J., Kahana, M. J., and Schulze-Bonhage, A. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science*, 342(6162):1111–1114.

Miller, J. F., Weidemann, C. T., and Kahana, M. J. (2012). Recall termination in free recall. *Memory & Cognition*, 40:540–550.

Moreton, B. J. and Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, 17(4):510–515.

Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3):735–774.

Müller, G. E. and Pilzecker, A. (1900). Experimental contributions to memory theory. *Zeitschrift für Psychologie Eganzungsband*, 1:1–300.

Müller, G. E. and Schumann, F. (1894). Experimentelle Beiträge zur Untersuchung des Gedächtnisses. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 6(80-191):257–339.

Murdock, B. B. (1974). *Human memory: Theory and data.* Lawrence Erlbaum and Associates, Potomac, MD.

Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, 13:511–521.

Nagel, S. and Xu, Z. (2018). Asset pricing with fading memory. Working paper, University of Chicago and University of Michigan.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In Ashby, F. G., editor, *Multidimensional models of perception and cognition*, pages 363–394. Lawrence Erlbaum and Associates, Hillsdale, New Jersey.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34:171–175.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321:1322 – 1327.

Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1):129–156.

Ramadorai, T., Anagol, S., and Balasubramaniam, V. (2020). Learning from noise: evidence from india's ipo lotteries. forthcoming, *Journal of Financial Economics*.

Reynolds, J. H. and Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27(1):611–647.

Rubin, D. C. and Berntsen, D. (2009). The frequency of voluntary and involuntary autobiographical memories across the lifespan. *Memory & Cognition*, 37(5):679–688.

Rubin, D. C., Berntsen, D., and Johansen, M. K. (2008). A memory-based model of posttraumatic stress disorder: Evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review*, 115(4):985–1011.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–46.

Savage, L. (1954). *The Foundations of Statistics*. Wiley, New York.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., and Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, 18(3):455–468.

Siddiqui, A. P. and Unsworth, N. (2011). Investigating the role of emotion during the search process in free recall. *Memory & Cognition*, 39(8):1387–1400.

Siegel, L. L. . and Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(3):755–764.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Solway, A., Murdock, B. B., and Kahana, M. J. (2012). Positional and temporal clustering in serial order memory. *Memory & Cognition*, 40(2):177–190.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly journal of experimental psychology*, 25(2):207–222.

Teasdale, J. D. and Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, 88(3):248–257.

Tulving, E. and Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, 21:437–484.

Uitvlugt, M. G. and Healey, M. K. (2019). Temporal proximity links unrelated news events in memory. *Psychological Science*, 30(1):92–104.

Umbach, G., Kantak, P., Jacobs, J., Kahana, M., Pfeiffer, B. E., Sperling, M., and Lega, B. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proceedings of the National Academy of Sciences of the United States of America.*, 117(45):28463–28474.

Underwood, B. J. (1948). 'Spontaneous recovery' of verbal associations. *Journal of Experimental Psychology*, 38:429–439.

Wachter, J. A. and Kahana, M. J. (2020). Associative learning and representativeness. Working paper, University of Pennsylvania.

Wachter, J. A. and Zhu, Y. (2019). Learning with rare disasters. Working paper, University of Pennsylvania.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1):152–176.

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12(1):579–601.

Woodworth, R. S. (1938). *Experimental Psychology.* H. Holt and Company, New York.

Yaffe, R. B., Kerr, M. S., Damera, S., Sarma, S. V., Inati, S. K., and Zaghloul, K. A. (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences*, 111(52):18727–8732.

Yates, F. A. (1966). *The art of memory.* Routledge and Kegan Paul, London, England.

Yntema, D. B. and Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 2:65–74.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., and Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4):792–804.

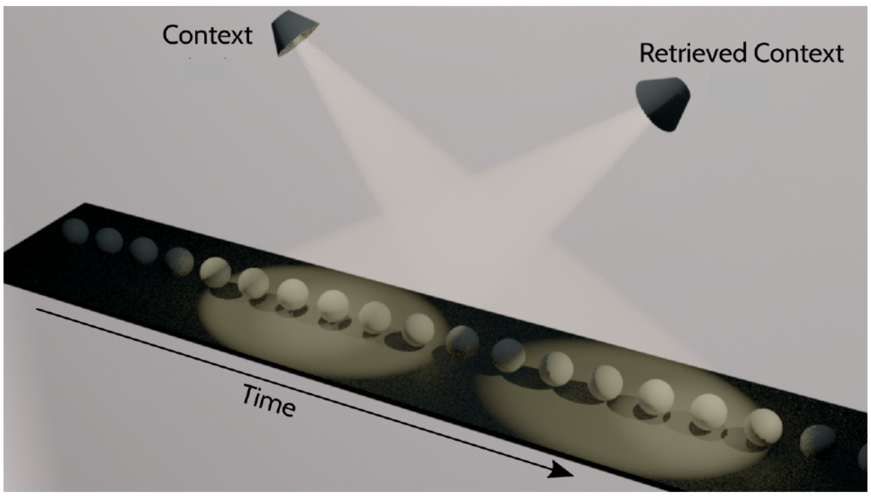


Figure 1: Retrieved Context and the spotlights of memory. In this illustration, memories appear as circles on the stage of life. All experiences that enter memory, as gated by perception and attention, take their place upon the stage. Context serves as a set of spotlights, each shining into memory and illuminating its associated features. The prior state of context illuminates recent memories, whereas the context retrieved by the preceding experience illuminates temporally and semantically contiguous memories. Due to the recursive nature of context and the stochastic nature of retrieval, the lamps can swing over time and illuminate different sets of prior features.

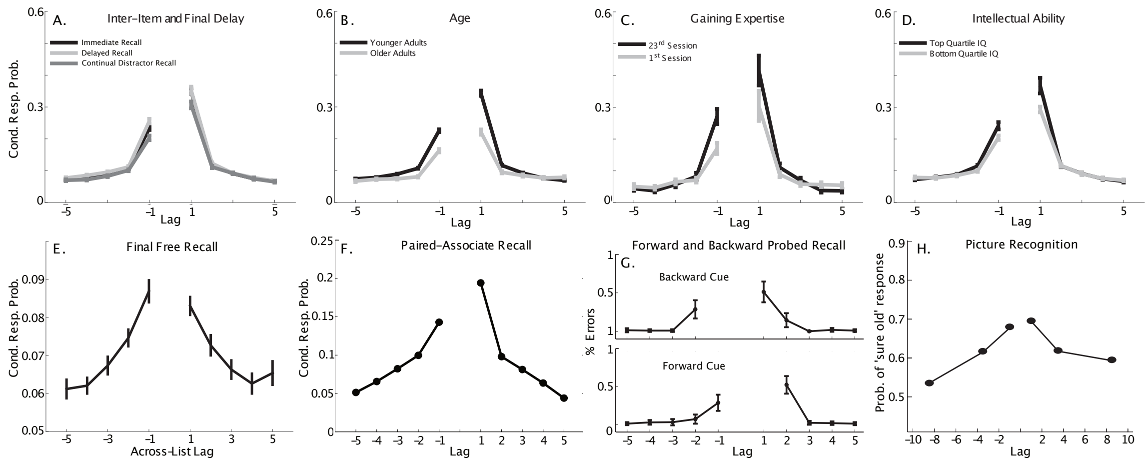
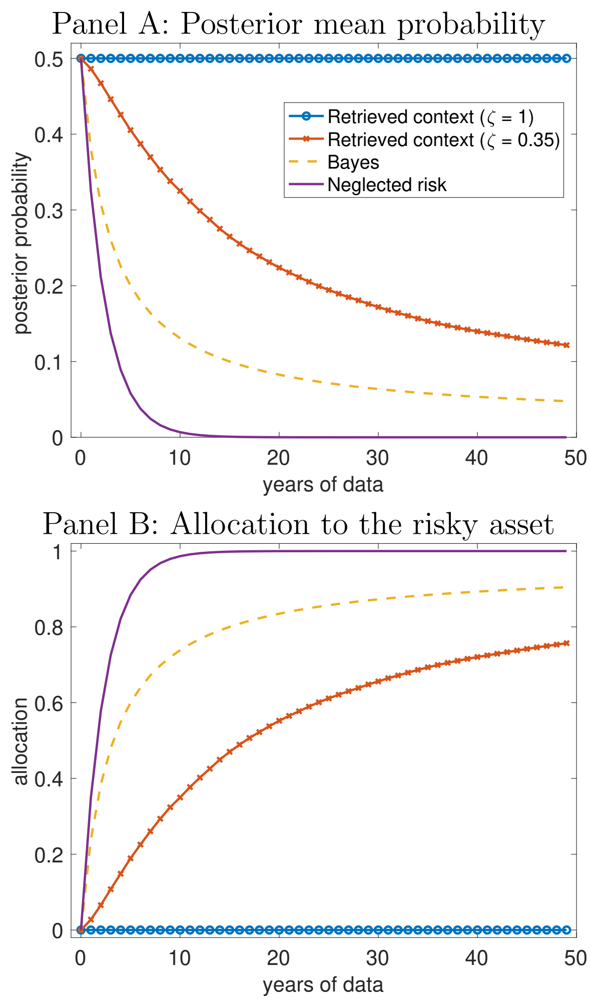


Figure 2: **Universality of Temporal Contiguity. A.** When freely recalling a list of studied items, people tend to successively recall items that appeared in neighboring positions. This temporal contiguity effect (TCE) appears as an in increase in the conditional-response probability as a function of the lag, or distance, between studied items (the lag-CRP). The TCE appears invariant across conditions of immediate recall, delayed recall, and continual-distractor recall, where subjects perform a demanding distractor task between each of the studied items. **B.** Older adults exhibit reduced temporal contiguity, indicating impaired contextual retrieval **C.** Massive practice increases the TCE, as seen in the comparison of 1st and 23rd hour of recall practice. **D.** Higher-IQ subjects exhibit a stronger TCE than individuals with average IQ. **E.** The TCE is not due to inter-item associations as it appears in transitions across different lists, separated by minutes, in a delayed final test given to subjects who studied and recalled many lists. **F.** The TCE appears in conditional error gradients in cued recall, where subjects tend to mistakenly recall items from pairs studied in nearby list positions. **G.** When probed to recall the item that either followed or preceded a cue item, subjects occasionally commit recall errors whose distribution exhibits a TCE both for forward and backward probes. **H.** The TCE also appears when subjects are asked to recognize previously seen travel photos. When successive test items come from nearby positions on the study list, subjects tendency to make high confidence "old" responses exhibits a TCE when the previously tested item was also judged old with high confidence. This effect is not observed for responses made with low confidence. Healey et al. (2019) provide references and descriptions of each experiment.

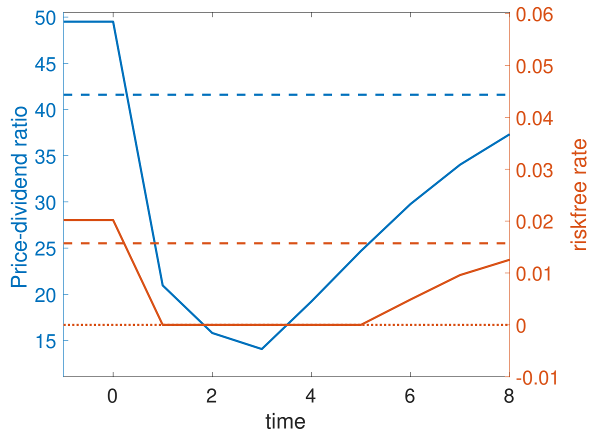
Figure 3: Posterior probability and asset allocation as a function of sample length



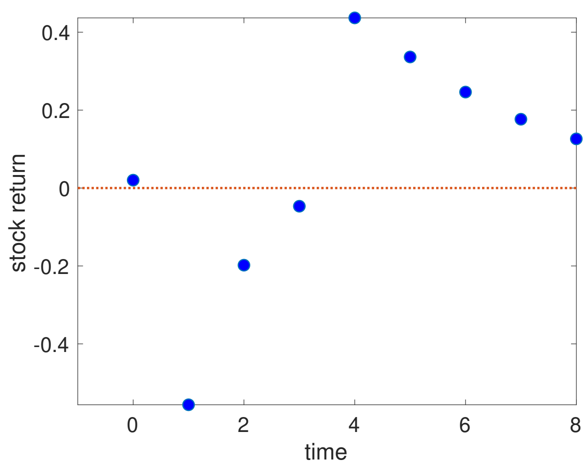
Notes: The figure shows posterior mean of the probability of a depression (Panel A) and the resulting asset allocation (Panel B) for the model presented in Section 4.1. 'Retrieved context ($\zeta = 1$)' is the retrieved context model when the agent places no weight on prior context. 'Retrieved context ($\zeta = 0.35$)' corresponds to when the agent places a weight of 1-0.35 on the prior context. 'Bayes' corresponds to the mean posterior probability when a Bayesian learns about a rare event from occurrences. 'Neglected risk' corresponds to exponential decay of beliefs after observing a rare event, as described in Section 2.3.

Figure 4: Response of prices and returns to a financial crisis

Panel A: Price-dividend ratio and riskfree rate

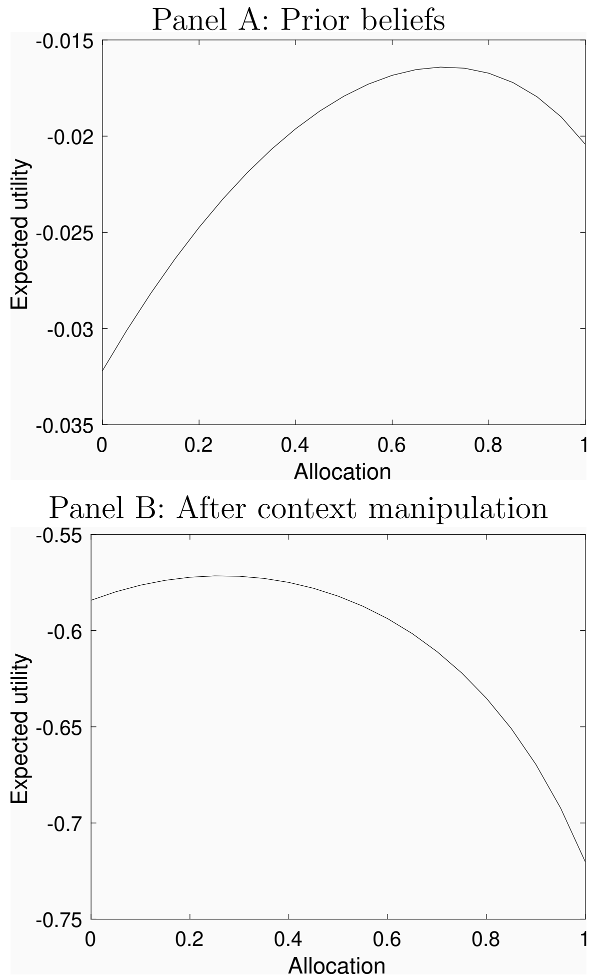


Panel B: Realized equity returns



Notes: Panel A shows the time path of the price-dividend ratio and of the short-term interest rate in response to a period of calm, followed by a financial crisis (in the model of Section 4.2). The dashed lines show full-information values. Panel B shows realized stock returns. The agent observes three periods of crisis features followed by normal features. The figure shows the maximum of the model-implied riskfree rate and zero. Returns are in annual terms.

Figure 5: Expected utility under context manipulation



Notes: This figure shows expected utility as a function of allocation to the risky asset under the model of Section 4.3. Panel A shows utility prior viewing a scene from a horror movie. Panel B shows utility after context has been manipulated by viewing the scene. In Panel B, curvature of the utility function has increased, so that the optimal allocation to the risky asset falls.