Memory and Beliefs in Financial Markets: A Machine Learning Approach

August 31, 2025

Abstract

This paper explores the role of memory in shaping belief formation of financial market participants. We estimate a structural machine learning model of memory-based belief formation applied to consensus earnings forecasts of sell-side stock analysts. The estimated model reveals significant recall distortions compared to a benchmark model trained to fit realized earnings revisions. Specifically, analysts over-recall distant historical episodes most of the time, when recent events are more useful for forming forecasts than those in the distant past, but under-recall them during crisis times, when history helps to interpret unusual events. We document two potential driving forces behind these distortions. First, analyst memory overweights the importance of past earnings and forecasts. Second, analysts are more likely to selectively forget past positive events. Our model of analyst recalls strongly predicts their earnings forecast revisions and errors, as well as stock returns, which suggests that distorted recalls might contribute to mispricing of assets in financial markets.

Keywords: Memory, Machine Learning, Belief Formation, Mispricing, Analyst

1 Introduction

"In response to current events, people often reach for historical analogies, and this occasion was no exception. The trick is to choose the right analogy." - Bernanke (2015)

Belief formation is crucial to asset pricing, as people make investment decisions based on their beliefs about the future state of firms and the economy. But how do people form these beliefs? Psychology literature suggests that memory plays a central role in belief formation, as current events often trigger the retrieval of similar past experiences, which then serve as references for current beliefs (Kahana, 2012). To connect memory with financial markets, a growing body of studies has developed theories (Mullainathan, 2002; Bordalo, Gennaioli and Shleifer, 2020; Wachter and Kahana, 2024) and provided evidence through surveys and lab experiments (Bordalo et al., 2022; Jiang et al., 2023; Enke, Schwerter and Zimmermann, 2024; Gödker, Jiao and Smeets, 2022; Graeber, Roth and Zimmermann, 2024). However, there remains little empirical evidence directly from financial market data that shows how memory shapes the belief formation of market participants and impacts asset pricing.[1]

To fill this gap, we provide new empirical evidence through two essential components of memory studies: mental context and recall. Mental context represents agents' current perception of a firm's state (e.g., as promising or struggling). Recall represents a past firm event similar to the current one, triggered by the mental context. Both components are essential but latent, so we develop a novel empirical approach to extract them by estimating a structural memory model applied to consensus analyst earnings forecasts.[2]

Our empirical approach involves two steps. In the first step, we estimate analyst mental context using a structural memory model. At its core is a seminal machine learning memory model - Long Short-Term Memory (LSTM), which simulates analysts' memory systems.[3] LSTM is a neural network with a dynamic memory cell and control gates that regulate the cell's updating processes. This structure allows the memory cell to store long-term

[1] Jiang et al. (2023) use survey and transaction data to show that memory influences investors' trading. Charles (2024) studies how memory-based attention impacts stock prices around earnings announcements.

[2] We study analyst forecasts as they are often used as proxies for investor beliefs (Brunnermeier et al., 2021).

[3] The theoretical foundation of LSTM dates back to Hochreiter and Schmidhuber (1997). With advances in computational power, it gained widespread application in empirical data tasks from the 2010s. LSTM achieves notable success in fields such as natural language processing and time series prediction, where it captures sequential dependencies effectively. This capability has also paved the way for more advanced generative AI models. Additionally, LSTM closely resembles state-of-the-art memory models in psychology (Kahana, 2012) that underpin Wachter and Kahana (2024)'s theoretical work in finance. These models have not been tested with empirical data.

generalized information about a firm, such as whether analysts perceive it as promising or struggling over time. Then, in our structural memory model, analyst belief formation involves three stages. First, analysts observe external features (e.g., GDP growth and debt-to-asset ratio). Second, they process these features along with past experiences and the current memory cell to form a mental context through their memory system (LSTM). Finally, they update their forecasts for the firm's future cash flows based on this mental context.

In this three-stage belief formation process, we observe external features and analyst forecasts in the data, then we can estimate the model to infer latent analyst mental context. For our empirical analysis, we use monthly data from 1990 to 2020, covering around 1,500 firms. For external features, we use a high-dimensional set of public signals including firm characteristics and macroeconomic variables. For analyst beliefs, we use consensus analyst forecasts for firms' one-year-ahead earnings per share (EPS). Thus, in this paper, we mainly study the analysts' collective memory for each firm in this paper, instead of individuals' memory.

In the second step of our empirical approach, we extract analyst recalls cued by the estimated mental context. Specifically, for each past experience, we assess the probability that analysts retrieve it from their memory database, which includes all firms in history and in the same industry as the current firm. The closer the mental context of a past experience to that of the current event, the higher its probability of being recalled, following the standard methods in psychology and economics literature (Kahana, 2012; Bordalo et al., 2023).

Our novel empirical approach offers three key advantages. First, with its embedded dynamic memory and mental context structure, our approach can capture and test well-established facts of human memory documented in psychology and finance (Kahana, 2012; Wachter and Kahana, 2024; Charles, 2024): recency, temporal contiguity, and semantic similarity.[4] By incorporating dynamic structures, our framework expands on other memory models (e.g., Bordalo, Gennaioli and Shleifer, 2020) that define recalls with static external features and allows these memory principles to be represented and tested more directly. Second, financial market participants often face high-dimensional, non-stationary variables with non-linear interactions. LSTM, as a machine learning method, is designed for such challenges, making it more suitable than parsimonious memory models (Wachter and Ka-

[4]Recency: people tend to refer to recently experienced events when retrieving memories; temporal contiguity: people recall events that occurred close in time to the recalled event; semantic similarity: people access events most similar to their current experiences.

hana, 2024; Bordalo, Gennaioli and Shleifer, 2020). Third, unlike surveys and experiments limited to specific settings (e.g., Enke, Schwerter and Zimmermann, 2024; Bordalo et al., 2022), our approach estimates analyst recalls for each firm across decades. This provides large cross-sectional samples over extended periods, which are essential for empirical asset pricing studies.

We next provide novel and unique empirical evidence demonstrating how memory shapes the belief formation in financial markets across four dimensions: the extracted analyst recalls, the distortions in these recalls, the driving factors behind these distortions, and the asset pricing implications of both the analyst recalls and distortions.

The extracted analyst recalls validate our empirical approach, as key principles of human memory, such as the recency effect and temporal contiguity effect, significantly appear in analyst recalls. Specifically, analysts generally focus on recent episodes and tend to recall events that occurred close together in time. However, past experiences can sometimes outweigh recent episodes; for instance, during the COVID pandemic in 2020, analysts paid more attention to the 2008 global financial crisis than to recent quarters. Our approach proposes new disciplines for rigorously modeling the impact of past experiences over different time periods.[5]

While forming beliefs based on recalls can be rational, deviations from full rationality occur if analysts recall wrong historical episodes. To examine the potential distortions in analyst recalls, we define a benchmark memory model by training the LSTM on realized earnings revisions rather than consensus forecast revisions.[6] Using this trained benchmark model, we then derive benchmark recalls cued by its mental context. Analyst recalls show significant distortions compared to these benchmark recalls. First, most of the time, analysts over-recall distant episodes, favoring long-term memory even when recent events are more relevant for optimal forecasts. Second, during crises, when historical episodes offer particularly valuable insights, analysts under-recall these episodes.

Why are analyst recalls distorted? We identify two main drivers: encoding errors and selective forgetting, both rooted in psychology and economics literature. First, we find that analysts do not encode external features into their memory optimally. By decomposing the memory cells and mental context for both analysts and the benchmark, we compute variable

[5]This finding offers new insights into the constant-gain framework, which assumes a monotonically declining impact of past experiences over time. The literature on experience effect and extrapolation (Malmendier and Nagel, 2011; Nagel and Xu, 2022; Barberis et al., 2015) commonly relies on this framework.

[6]This benchmark model generates optimal forecasts, with predictive power for realized earnings revisions comparable to other non-memory baseline models, such as logistic regression.

importance using 12-month rolling window regressions of memory cells and mental context on external features. Then the variable importance is the reduction in R^2 when features are set to zero one at a time in the regressions. This decomposition reveals two key findings. First, analyst memory cells and mental context overweight past earnings and forecasts, whereas the benchmark mainly focuses on other firm characteristics. Second, analysts fail to adjust feature weights in their mental context sufficiently during crises. This results in under-recalling distant yet useful episodes, leading to underreaction in their earnings forecasts. These findings provide a memory-based foundation and empirical evidence for the literature on encoding errors (e.g., Woodford, 2020; Frydman and Nunnari, 2021; Frydman and Jin, 2022; Drugowitsch et al., 2016).

In addition to encoding errors, we document that selective forgetting also distorts analyst recalls away from the optimal benchmark. In a counterfactual analysis, we remove the forget gate from our LSTM models for both the analysts and the benchmark. Then we elicit counterfactual recalls cued by mental context without selective forgetting. Removing selective forgetting would improve analyst forecasts, as the predictive power of their recalls for realized earnings revisions increases by 68%. Conversely, without selective forgetting, the benchmark recalls' performance in predicting realized earnings revisions declines by 40%. This evidence suggests that analysts do not use the selective forgetting channel effectively, contributing to recall distortions. Specifically, without selective forgetting, analyst recalls align more closely with the optimal benchmark recalls, exhibiting a stronger recency effect and focusing more on positive events.

Next, we show that analyst recalls and recall distortions have substantial predictive power for asset pricing by presenting four significant novel implications.[7] First, we document that our model of analyst recalls predicts future stock returns. In response to current events, when analysts recall positive (negative) events, they tend to form optimistic (pessimistic) beliefs, resulting in positive (negative) stock returns. We quantify recall positivity or negativity using the recalled revision. This is defined as the forecast revision that analysts should make (the difference between realized EPS and the lagged forecast) if the corresponding actual EPS is known, or as the consensus analyst forecast revision from the recalled episode if the actual EPS has not yet been announced. We then sort stocks into portfolios based on the

[7] We present the asset pricing implications through out-of-sample tests. For asset pricing quantities at time t, our memory-based variables are predictions from memory models trained solely on information available prior to t and are measured independently of future stock prices or returns. Our results thus support the common assumption in asset pricing literature that analyst forecasts represent investor beliefs (Brunnermeier et al., 2021).

analyst recalled revisions. Analysts significantly revise down forecasts for firms in the low recalled revisions group and revise up forecasts for firms in the high group. The long-short portfolio yields a significant positive monthly risk-adjusted return of 0.41%.

Second, we demonstrate that memory can serve as a powerful microfoundation for disagreement. Assuming analysts share a common memory database, disagreement arises from randomness in the cued recall process. When analysts retrieve different past experiences with varied future prospects (measured by recalled revisions), they form differing beliefs, leading to abnormal trading volume. We regress the disagreement (Diether, Malloy and Scherbina, 2002), and abnormal trading volume (Cookson and Niessner, 2020) on the dispersion of analyst recalls (defined as the weighted standard deviation of analyst recalled revisions). Regression results indicate that the dispersion of analyst recalls significantly captures both disagreement and trading volume in financial markets.

Third, we examine the economic impact of distorted analyst recalls, defined as the difference between analyst and benchmark recalled revisions. We show that recall distortion is a powerful predictor of forecast errors and contributes to asset mispricing. A regression of analyst EPS forecast errors on recall distortion yields a significantly positive coefficient, indicating that overly positive (negative) recalls lead to over-optimistic (over-pessimistic) forecasts. Even when competing with other predictors of forecast errors (Coibion and Gorodnichenko, 2015; Bordalo et al., 2024), recall distortion remains significant. Taking this a step further, if investors follow analyst beliefs and deviate from rational beliefs as captured by our benchmark recalls, recall distortion should reflect asset mispricing. In a rational market, we would expect a reversal of the abnormal returns caused by recall distortion. To test this, we sort stocks into quintile portfolios by recall distortion. The first quintile contains stocks with overly positive analyst recalls and the fifth quintile contains those with overly negative recalls. Initially, a long-short strategy yields negative returns, reaching a cumulative return low of -0.5% within two months. Subsequently, the rational market corrects the mispricing, leading to a reversal, and the effects of recall distortion vanish within six months.

Fourth, our models of analyst and benchmark recalls offer new insights into asset pricing anomalies, particularly short-term reversal, where a stock's prior-month return is negatively associated with its next-month return. Our memory framework suggests that when investors pay excessive (insufficient) attention to the prior month, they are more (less) likely to mirror their previous actions. For instance, if investors pushed up a stock's price last month, they are more (less) likely to continue doing so. This memory-based attention can offset (amplify) short-term reversal. Using analyst recalls as a proxy for memory-based investor attention, we define biased attention as the difference in probabilities of recalling last month between analysts and the benchmark. Double-sorted portfolios—first by investor biased attention, then by prior-month returns—reveal that, in the group of stocks receiving excessive attention, the short-term reversal strategy yields a monthly risk-adjusted return of -0.58%, indicating a momentum pattern instead. Conversely, short-term reversal is amplified in the group of stocks where investors are inattentive to the prior month's situation. The double-sorted long-short strategy between these groups earns a monthly risk-adjusted return of 1.12%, over three times the magnitude of the standard short-term reversal strategy. Additionally, we observe a reversal in cumulative returns within five months, as the strategy exploits investor belief distortions that are eventually corrected by the market.

The contribution of this paper to the literature is three-fold. First, it contributes to the literature on applications of human memory in economics and finance. Existing research primarily develops theories connecting memory and financial markets (Bordalo, Gennaioli and Shleifer, 2020; Wachter and Kahana, 2024; Nagel and Xu, 2022; Bordalo et al., 2023). We are the first to empirically extract firm-specific recalls over decades by estimating a structural memory-based model. Unlike survey- or experiment-based methods (Bordalo et al., 2022; Jiang et al., 2023; Enke, Schwerter and Zimmermann, 2024; Gödker, Jiao and Smeets, 2022; Graeber, Roth and Zimmermann, 2024), our approach yields such a large panel of recalls suited for empirical asset pricing. We provide new insights into memory's role in belief formation, offering empirical evidence consistent with established memory principles (Kahana, 2012; Charles, 2024), as well as key memory channels such as encoding of external features (Woodford, 2020; Frydman and Nunnari, 2021; Frydman and Jin, 2022; Drugowitsch et al., 2016) and selective forgetting (Walters and Fernbach, 2021; Gödker, Jiao and Smeets, 2022). By using a dynamic memory structure, our approach allows these channels and patterns to be identified in ways that are not readily captured in frameworks without such dynamics (Bordalo, Gennaioli and Shleifer, 2020; Charles and Sui, 2024). Additionally, we introduce a novel memory-based framework to model the impact of past experiences, expanding beyond the commonly applied constant-gain approach (Malmendier and Nagel, 2011; Barberis et al., 2015; Nagel and Xu, 2022).[8]

Second, this paper advances asset pricing research by providing new evidence on how memory-based belief formation shapes asset pricing dynamics. Belief formation is central to asset prices (Brunnermeier et al., 2021), and our memory-based approach offers new insights

[8]This paper also contributes to the literature on experience effects by allowing for the study of multiple types of experiences in high-dimensional settings, providing a richer foundation than the reduced-form methods that typically examine single experiences (Malmendier and Nagel, 2011, 2016).

into return predictability (Barberis, Shleifer and Vishny, 1998; Bordalo et al., 2019, 2024; Cui, De la O and Myers, 2024), disagreement and trading volume (Hong and Stein, 2007; Atmaz and Basak, 2018; Cookson and Niessner, 2020; Liao, Peng and Zhu, 2022), as well as mispricing and asset pricing anomalies (Barberis, 2018; De la O, Han and Myers, 2023; Da, Liu and Schaumburg, 2014; Lehmann, 1990; Jegadeesh, 1990; Bouchaud et al., 2019).

Third, this paper adds to the growing literature applying machine learning in finance and economics. Much research leverages machine learning's predictive power to forecast stock returns and economic outcomes like EPS, GDP growth, and inflation (Gu, Kelly and Xiu, 2020; Chen, Pelger and Zhu, 2023; van Binsbergen, Han and Lopez-Lira, 2023; Bianchi, Ludvigson and Ma, 2022; Chen, Kelly and Xiu, 2024; Lopez-Lira and Tang, 2023). Though this predictive power often operates as a black-box. This paper instead opens the black-box, using models rooted in neuroscience and psychology to capture how beliefs form among financial market participants.[9] It also inspires a new research avenue of applying machine learning in behavioral finance. While similar concepts are emerging, such as Barberis and Jin (2023)'s theoretical application of reinforcement learning to investor behavior, we are the first to apply such approaches to empirical data.

The remainder of the paper is organized as follows. Section 2 presents a framework of memory-based belief formation. Section 3 introduces the details about the structural machine learning memory model applied to consensus analyst earnings forecasts. Section 4 explores estimated analyst recalls and examines their relationship with beliefs and investor trading. Section 5 investigates recall distortions, underlying drivers, and asset pricing implications. Finally, Section 6 concludes.

2 Cued Recall and Belief Formation

We present how cued recall shapes beliefs in financial markets. At time t, the analyst evaluates how she should change her belief $F_{i,t}[EPS_{i,t+l}]$ about firm i's future earnings per share (available at time $t+l$) which is denoted as $EPS_{i,t+l}$, from her last period's belief $F_{i,t-1}[EPS_{i,t+l}]$. Let $\Delta F_{i,t} = F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}]$ denote the change in her belief,

[9] Mullainathan and Spiess (2017) note challenges in interpreting machine learning models, related to the Rashomon effect that refers to the existence of many models that fit the data equally well but rely on different features, especially when inputs are highly correlated. We address this by focusing on higher-level patterns, such as the similarity of estimated vectors, rather than individual parameters, providing stable summaries across models.

i.e., forecast revision.[10] The process of forming $\Delta F_{i,t}$ involves two steps. First, she retrieves similar experiences from her memory database through the cued recall process, where each experience is represented as a tuple of firm and time, i.e., (firm, month). Second, she projects her beliefs about the current prospects of a firm based on the retrieved experiences. For example, when forming beliefs about firm i in month t, the analyst's retrieved experience may be firm j in month τ (prior to month t), then her belief about firm i is influenced by her belief about firm j in month τ. We next formalize these steps, following the approach outlined in Bordalo et al. (2023) and Jiang et al. (2023).

2.1 Step 1: Cued Recall

Cued recall is the process of searching for similar past experiences from the contents of memory. Context is commonly used as a retrieval cue. We take the agent's internal mental state as the context (e.g., her perceived state of firms or the economy), following Wachter and Kahana (2024). In this setting, the human memory system processes external features and transforms them into an internal mental context, allowing the same set of external features to be interpreted differently by the same person as her memory evolves. Thus, the mental context defined in this paper is dynamic, in contrast to another strand of research which uses static external features as context.[11] The dynamic structure of context can produce fundamental principles of human memory, such as the temporal contiguity effect, which states that people tend to recall events that happen contiguously in time. The temporal contiguity effect plays an important role in explaining certain investor behaviors (Wachter and Kahana, 2024)[12] and it has been supported by the empirical evidence in a specific setting of earnings announcements (Charles, 2024). However, the memory models that take static external features are not able to capture the temporal contiguity effect (Bordalo, Gennaioli and Shleifer, 2020). We formally define context and introduce its dynamics in Section 3. For simplicity, we broadly define context in this section as a representation of relevant information

[10] We focus on the change in belief to study how analysts react to new information.

[11] For example, Bordalo, Gennaioli and Shleifer (2020) use environmental features such as location as context to define cued recall in a setting of economic decision making and Jiang et al. (2023) show that today's return acts as a powerful context cue for investors' recall.

[12] For example, in their theoretical work, Wachter and Kahana (2024) provides a memory explanation for the narratives that depression would come right after seeing the financial crisis. Specifically, they show that in investor memory, the Great Depression in 1930 came right after the stock market crash of 1929. Then the re-appearance of a financial crisis today retrieves their memory of the crisis in 1929, as well as the memory of the depression since the state of financial crises and depressions are associated in time in their memory, then the temporal contiguity effect elicits all events happened around 1929, even though the features of crisis and depression are assumed orthogonal.

about each experience. We denote the context for each firm j at time τ as a K-dimensional vector $c_{j,\tau}$.

The past experiences are retrieved in the process of cued recall according to their similarity to the current event. When the past experience is more similar to the current event, it is more likely to be retrieved. We measure the similarity between two experiences (i, t) and (j, τ) by the similarity function S, following standard methods in the memory literature (Kahana, 2012)[13]

$$S(c_{j,\tau}, c_{i,t}) = \exp\left(-\|c_{j,\tau} - c_{i,t}\|_2\right). \tag{2.1}$$

Two experiences are similar to each other if the Euclidean distance between their corresponding context vectors is small. The similarity reaches a maximum of 1 when two context vectors are exactly the same. Then, when encountering event (i, t), each past experience (j, τ) may be retrieved with a probability proportional to the similarity between (i, t) and (j, τ), i.e.,

$$p(c_{j,\tau}) \propto S(c_{j,\tau}, c_{i,t}). \tag{2.2}$$

2.2 Step 2: Belief Projection

In the second step, analysts revise their forecasts based on their retrieved experience. This process is similar to the step of "simulation" in Bordalo et al. (2022). We model the analysts mostly rely on the experience that first gets retrieved.[14] Specifically, when forming forecast revision $\Delta F_{i,t}$ for firm i in month t, suppose one analyst k retrieves experience (j, τ) that is sampled from the probability distribution (2.2). Then her forecast revision $\Delta F_{i,t}^k$ is positively associated with her recalled revision. Due to equivalence under a positive linear transformation, we represent forecast revision $\Delta F_{i,t}^k$ as:

$$\Delta F_{i,t}^k = rr_{j,\tau}, \tag{2.3}$$

[13] In general, the similarity function can be defined as

$$S(c_{j,\tau}, c_{i,t}) = \exp\left(-\xi\|c_{j,\tau} - c_{i,t}\|_\gamma\right),$$

where γ is the distance metric, with $\gamma = 2$ denotes the Euclidean norm, and $\xi \geqslant 0$ measures how quickly similarity decays with the distance. In our empirical tests, we take the most common case where $\gamma = 2$ and $\xi = 1$.

[14] This framework uses a single recall as a baseline. It can be easily extended to multiple recalls, such as a similarity-weighted average of all past experiences (Bordalo et al., 2023), or more complex cases where the probability of retrieving an experience depends on prior recalls (Kahana, 2012).

and $rr_{j,\tau}$ is the recalled revision, defined as

$$rr_{j,\tau} = \begin{cases} EPS_{j,\tau+l} - F_{j,\tau-1}[EPS_{j,\tau+l}] & \text{if } \tau + l < t \\ F_{j,\tau}[EPS_{j,\tau+l}] - F_{j,\tau-1}[EPS_{j,\tau+l}] & \text{if } \tau + l \geqslant t \end{cases} \tag{2.4}$$

where $EPS_{j,\tau+l}$ denotes the realized EPS at time $\tau + l$ and $F_{j,\tau-1}[EPS_{j,\tau+l}]$ denotes the analyst forecast at time $\tau - 1$. Intuitively, we model that the analyst formulates the forecast revisions at time t, by learning from what revisions she should have made in the recalled episode if the realized earnings are available at time t. Otherwise, she takes her own forecast revision as the reference. When $rr_{j,\tau} > 0$, the analyst recalls a positive event, as either she revised up her forecasts, or she learns she should revise up her forecast in the recalled episode. If $rr_{j,\tau} < 0$, she recalls a negative event.

To study the impact of memory on the aggregate market, we further assume an infinite number of analysts, and they share the same memory database. Specifically, when forming forecast revision for firm i at time t, the analysts sample from the memory database $M_{i,t}$ which consists of past experiences that are associated with all the firms in the same industry as firm i before time t. Formally, $M_{i,t} \equiv \{(j, \tau)\}, \forall j, \tau$, such that Industry($j$) = Industry($i$) and $\tau < t$. Then the probability distribution (2.2) is refined as[15]

$$p(c_{j,\tau}) = \frac{S(c_{j,\tau}, c_{i,t})}{\sum_{(m,s)\in M_{i,t}} S(c_{m,s}, c_{i,t})}. \tag{2.5}$$

The consensus forecast revision $\Delta F_{i,t}$ is an average of individual forecast revisions $\Delta F^k_{i,t}$,

$$\Delta F_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) \times \Delta F^k_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) \times rr_{j,\tau}, \tag{2.6}$$

For the view of an aggregate market or a representative analyst, Equation (2.6) also reconciles with the idea of availability heuristic (Tversky and Kahneman, 1974) that people will be biased towards the instances that are easier to recall. Experience that is more similar to the current episode has more weight in formulating today's consensus forecast revisions in our setting.

[15]In practice, the number of analysts covering a firm is limited. We restrict the size of memory database $M_{i,t}$ to five by selecting the five past experiences with highest similarity. This choice aligns with typical scenarios where around five analysts cover a firm. Additionally, Miller (1956) highlighted that the capacity for processing information is limited to approximately seven items due to constraints in working memory. Similarly, Earhard (1967) and Roberts (1972) show that the number of recalls cannot increase without limit, contradicting the typical assumption of rational investors. Our results remain robust to variations in the size of the memory database, as demonstrated in Appendix A6.

3 Estimation of Mental Context and Recall

In this section, we introduce a machine learning approach to estimating mental context and recalls from a structural memory-based model using empirical data. In the framework of belief formation based on cued recall, it is context that shapes the analyst recalls. Thus, we first introduce mental context in a dynamic memory structure, the long short-term memory model (LSTM). Then, we develop LSTM to fit the analysts' belief formation process. Lastly, we introduce the data and the model training process.

3.1 The Structure of LSTM

LSTM is a type of neural network with a chain-like structure designed to handle sequence data. Hochreiter and Schmidhuber (1997) first introduced its theoretical foundation in machine learning literature. At the same time, the architecture of an LSTM unit naturally parallels the state-of-the-art memory model in psychology and economics by Wachter and Kahana (2024), which also builds on three core components. First, the external features $X \in \mathbb{R}^J$ represent public information such as macroeconomic conditions and firm fundamentals. Second, the memory cell $m \in \mathbb{R}^K$ functions as long-term memory, accumulating and carrying information across time. Third, the context vector $c \in \mathbb{R}^K$ serves as the current state representation, summarizing both the most recent inputs and the information retrieved from memory. Here, K represents the dimension of the memory and context space, and J denotes the number of external features. The flow of information is controlled by three gating mechanisms: a forget gate, an input gate, and an output gate. Based on the new external features and prior context state, the gates selectively determine which past information to discard from the memory, what new information to store, and what part of the updated memory to use for the current context. This structure not only enables the LSTM to learn patterns in sequential data and capture dependencies over time, but also aligns naturally with models of human memory.

To better conceptualize the three core components in the framework of analysts forming beliefs, we provide illustrative examples. The external features X are observable signals, such as consumption growth, the firm's sales growth, past stock returns, and the book-to-market ratio. Both the memory cell m and the context state c are latent and need to be estimated. The memory cell m stores analysts' accumulated associations between past features and the contexts in which they were experienced. The context state c represents the analysts' internal mental context, reflecting their perceived current state of the firm conditional on memory. In other words, c can be interpreted as a summary of the most relevant information from memory and current features needed for this period's earnings forecasts. For example, consider a manufacturing firm during Covid. Analysts' memory may include the experience that the firm's food equipment line provided a buffer during the 2008 financial crisis. When the new shock occurs, the context draws on this memory and emphasizes macroeconomic indicators and the sales growth of that product line, guiding earnings forecasts under current conditions.

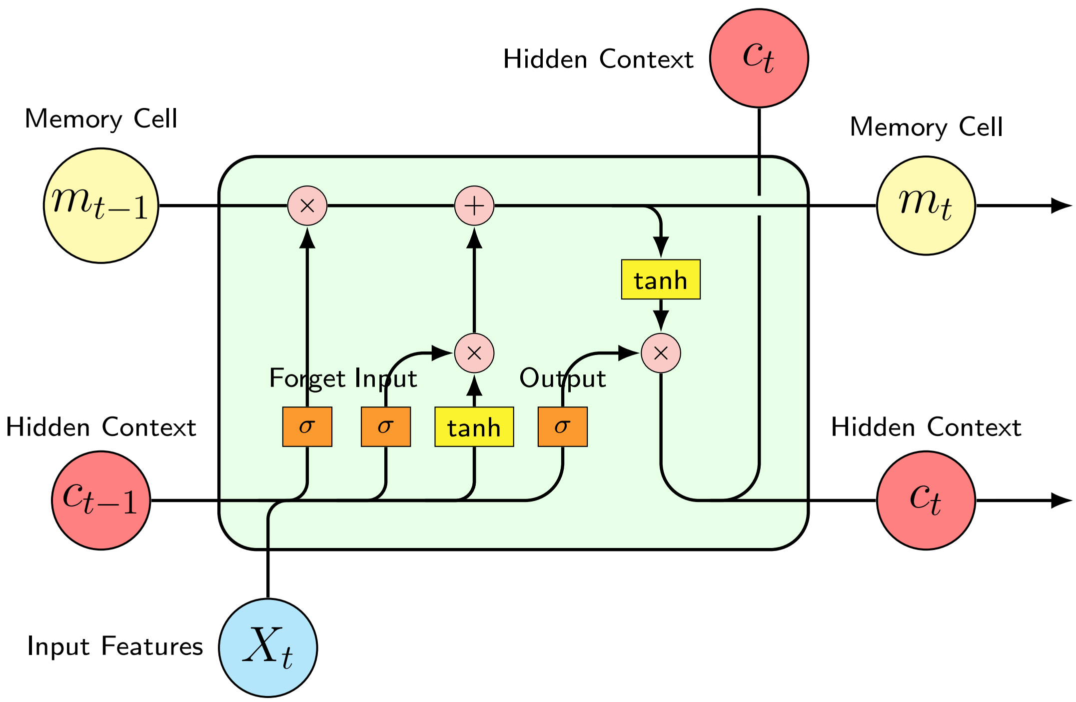


Figure 1: The Structure of a LSTM Unit

Figure 1 illustrates the structure of one LSTM unit. The three gates are shown in orange boxes, each regulated by a sigmoid activation function. The sigmoid function, denoted as σ, applies an element-wise operation that outputs values between zero and one, thereby controlling how much information passes through the gates. The operators "×" and "+" shown in pink circles are the element-wise multiplication and addition, respectively. "tanh" in yellow boxes refers to the hyperbolic tangent activation function.[16]

[16]The incorporation of these activation functions (tanh and sigmoid) has several advantages: first, they introduce the non-linearity to the structure so that it fits more closely to empirical economic and financial data (Teräsvirta, 2006); second, they further standardize the variables, thus reduce the impact of outliers; third, they improve the accuracy and efficiency of model estimation (Dubey, Singh and Chaudhuri, 2022; Jagtap, Kawaguchi and Karniadakis, 2020).

The memory process takes the following steps. First, the forget gate determines the amount of information to retain in the memory cell, this serves as the channel of selective forgetting. Specifically, an output of zero from the sigmoid function indicates that the information should be completely erased, whereas an output of one signals that it should be fully retained,

$$\text{forget}_t = \sigma(W_c^f c_{t-1} + W_x^f X_t + w_0^f). \tag{3.1}$$

Second, the input gate determines the extent to which new information is stored into the memory cell. If the sigmoid function outputs zero, the information is deemed irrelevant and is not stored; conversely, an output of one indicates that the information is fully stored in the memory cell. Before being stored, the information undergoes transformation by the *tanh* function, which scales it to a range between -1 and 1, denoted as $\tilde{x}_t$:

$$\tilde{x}_t = tanh(W_c^{\tilde{x}} c_{t-1} + W_x^{\tilde{x}} X_t + w_0^{\tilde{x}}) \tag{3.2}$$

$$\text{input}_t = \sigma(W_c^i c_{t-1} + W_x^i X_t + w_0^i). \tag{3.3}$$

Third, the memory cell is updated according to the outputs of the forget and input gates as follows:

$$m_t = \text{forget}_t \times m_{t-1} + \text{input}_t \times \tilde{x}_t, \tag{3.4}$$

by erasing contents from the memory cell and adding in new valuable information to it.

Finally, the output gate determines which information should be elicited from the updated memory cell:

$$\text{out}_t = \sigma(W_c^o c_{t-1} + W_x^o X_t + w_0^o). \tag{3.5}$$

Then, the final output, the current mental context c_t, is chosen by the output gate from the updated memory cell through the *tanh* function:

$$c_t = \text{out}_t \times tanh(m_t). \tag{3.6}$$

LSTM replicates the core idea of the retrieved-context memory model in Wachter and Kahana (2024) that the context vector c_t should be dynamic and evolving endogenously, going beyond just the static features of the physical environment. This is evident from two aspects of the model's structure. The dynamics of the memory m_t and context c_t are analogous to those in the retrieved-context theory. In both frameworks, the current context is a function of the prior context and new information retrieved from a long-term memory store. This memory is then updated by accumulating new associations between the current context and observed external features. This creates a feedback loop where memory shapes the interpretation of new information, and that interpretation updates memory.

Second, the gating mechanisms that control the flow of information are a function of both the prior context c_{t-1} and the new external features X_t. This structure ensures that the process of memory updating is context-dependent. The model learns not just what to remember, but how to interpret and store it based on its current state. For example, a reported 11% revenue growth (X_t) can be interpreted very differently based on the analyst's prevailing perceived state of the firm (c_{t-1}). An analyst focused on a firm's past difficulties may view the growth rate as evidence of recovery, while another who sees the firm in a high-growth state may regard the same number as a sign of slowing momentum.

The whole structure of LSTM is recurrent, as each LSTM unit is connected over time. The information obtained and updated in the memory cell and mental context from the previous period will be used for the next period's information processing and updates of memory cells and mental context. We show this recurrent structure in Section 3.3.

Next, we discuss why LSTM is a good model for studying belief formation processes in financial markets.

3.2 Why LSTM?

The benefit of applying LSTM to study belief formation processes in financial markets is twofold. First, LSTM provides a psychologically grounded model of human memory, building on well-established frameworks in psychology and economics (e.g., Wachter and Kahana, 2024). Like the memory model of Wachter and Kahana (2024), LSTM captures the three basic laws and key facts of the human memory system (Howard and Kahana, 2002; Kahana, 1996): recency, temporal contiguity, and semantic similarity. Recency means that people refer to recently experienced events when accessing memory. Temporal contiguity means that people tend to recall an event that occurred contiguously in time to presently-recalled events. Semantic similarity means that people are more likely to access the events that are most similar to that they are experiencing. In LSTM, the mental context c evolves according to the association of external stimuli, inner memory processes, and previous mental context. Howard and Kahana (2002) presents that this autoregressive structure, embedded in the memory cell m, supports recency, while the combination of autoregressive context and the memory cell establishes a channel for temporal contiguity. Semantic similarity naturally arises from contextually cued recall, see Section 2.

Building on Wachter and Kahana (2024), LSTM allows for additional flexibility in studying belief formation. It can handle high-dimensional inputs beyond the basis vectors used in their framework, making it suitable for richer empirical analysis. LSTM also includes multiple structural memory channels, which enable more comprehensive counterfactual exercises. For example, the forget gate provides a way to control how past information fades from the memory cell, which can be used to study selective forgetting and its impact on analyst beliefs (Walters and Fernbach, 2021; Gödker, Jiao and Smeets, 2022), see Section 5.3.2.

Second, analysts need to process high-dimensional and non-stationary financial variables that may involve complex functional forms. LSTM offers advantages common to machine learning methods in handling these challenges, such as feature selection and dimension reduction (Nagel, 2021). For example, in the memory model of Bordalo, Gennaioli and Shleifer (2020), context includes a broad set of external environmental variables, equally weighted, which are then used to generate cued recall for decision-making. While this provides a useful starting point, analysts may naturally assign different weights to features in forming beliefs. LSTM accommodates this by allowing its gates to filter and weight information differently. As we show in Section 5.3.1, analysts place varying weights on external features in their memory cell, and feature importance evolves over time.

Additionally, LSTM can capture underlying dynamics from non-stationary variables. For instance, Chen, Pelger and Zhu (2023) shows that LSTM can extract hidden states from non-stationary, cyclical macroeconomic variables, and Bianchi, Ludvigson and Ma (2022) uses LSTM to provide unbiased benchmarks for forecasts of GDP growth and inflation.

In machine learning and neuroscience, LSTM is recognized as a relatively simple yet fundamental model. While more complex models exist, we apply LSTM in this paper to demonstrate that the memory mechanisms it captures show strong predictive and explanatory power for analyst beliefs and asset pricing patterns. Our approach can also be adapted to explore other specific memory or neuroscientific channels by replacing the basic LSTM model with more advanced models, such as those introduced by Vaswani et al. (2017), Weston, Chopra and Bordes (2014), and Graves et al. (2016)

3.3 LSTM and Analyst Belief Formation

Now we adapt LSTM to the setting of analyst belief formation.

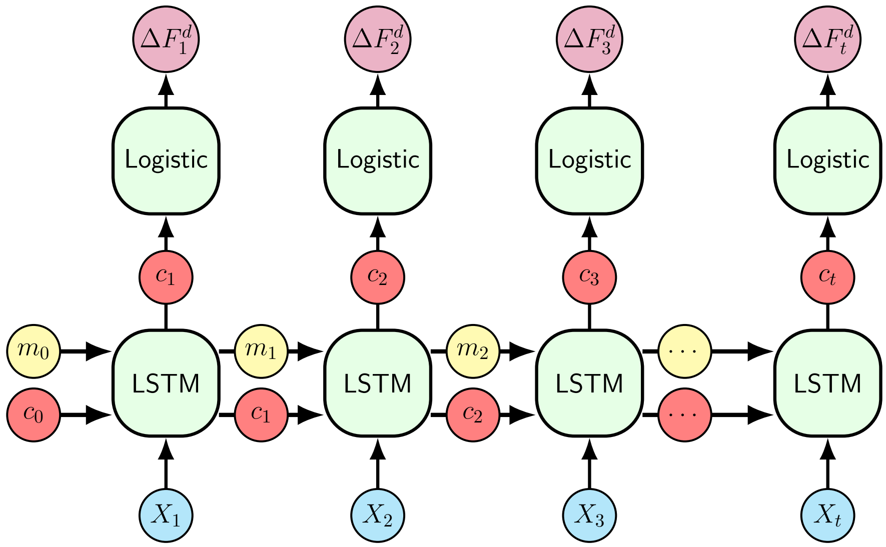


Figure 2: Whole Structure

Figure 2 presents the whole structure of how the analysts form beliefs for one firm as time progresses from left to right. For simplicity, the subscript i representing the firm is omitted in the figures, for example, $X_{i,t}$ is shown as X_t. In this structure, $X_{i,t} \in \mathbb{R}^M$ denotes the public signals (external features) that the analysts use to form EPS forecast revisions at time t for firm i. We introduce the details of $X_{i,t}$ in Section 3.4. $\Delta F^d_{i,t}$ denotes the direction of revision on EPS forecasts at time t for firm i:

$$\Delta F^d_{i,t} = \begin{cases} -1 & \text{if } F_{i,t}[EPS_{i,t+l}] < F_{i,t-1}[EPS_{i,t+l}], \\ 0 & \text{if } F_{i,t}[EPS_{i,t+l}] = F_{i,t-1}[EPS_{i,t+l}], \\ 1 & \text{if } F_{i,t}[EPS_{i,t+l}] > F_{i,t-1}[EPS_{i,t+l}], \end{cases} \tag{3.7}$$

where $F_{i,t}[EPS_{i,t+l}]$ is the time t analyst forecast of firm i's future earnings per share $EPS_{i,t+l}$ that is available at time $t+l$, and $F_{i,t-1}[EPS_{i,t+l}]$ is the time $t-1$ analyst EPS forecast. $\Delta F^d_{i,t} = -1$, if the analysts revise down the forecasts from last period; $\Delta F^d_{i,t} = 1$, if the analysts revise up the forecasts; and $\Delta F^d_{i,t} = 0$, if the analysts do not change the forecasts.[17]

The LSTM units in the whole structure is recurrent. The memory cell and the mental context at period $t-1$ are used as the input of the LSTM cell at period t. After receiving new information $X_{i,t}$, the analysts update memory cell $m_{i,t}$ and mental context c_t according to the last period's memory cell and mental context $m_{i,t-1}$ and $c_{i,t-1}$. Then the analysts decode the mental context vectors $c_{i,t}$ and decide the direction of forecast revision $\Delta F^d_{i,t}$ by

[17]Note, $\Delta F^d_{i,t}$ is different from $\Delta F_{i,t}$ defined in Section 2, as $\Delta F^d_{i,t}$ is the direction of forecast revisions. We opt to fit the LSTM model to directions of forecast revisions, since the training efficiency on continuous forecast revisions is compromised by their relatively small magnitudes.

a logistic function of $c_{i,t}$.

Both external features $X_{i,t}$ and the direction of forecast revisions $\Delta F^d_{i,t}$ are observable, thus we can estimate the analyst memory cell $m_{i,t}$ and mental context $c_{i,t}$ by feeding external features $X_{i,t}$ into the whole structure and fitting revision directions $\Delta F^d_{i,t}$.

Additionally, we provide a benchmark memory by training LSTM to fit the direction of realized EPS revisions which are analogously defined as shown in Equation (3.7):

$$\Delta RE^d_{i,t} = \begin{cases} -1 & \text{if } EPS_{i,t+l} < F_{i,t-1}[EPS_{i,t+l}], \\ 0 & \text{if } EPS_{i,t+l} = F_{i,t-1}[EPS_{i,t+l}], \\ 1 & \text{if } EPS_{i,t+l} > F_{i,t-1}[EPS_{i,t+l}]. \end{cases} \tag{3.8}$$

By comparing $F_{i,t-1}[EPS_{i,t+l}]$ with $EPS_{i,t+l}$, $\Delta RE^d_{i,t}$ is interpreted as the correct forecast revisions that the analysts should have made conditional on their forecasts in the preceding period. This trained benchmark memory model provides real-time, unbiased forecasts for realized earnings revisions, following the concept of machine learning benchmark forecasts proposed by van Binsbergen, Han and Lopez-Lira (2023). We later demonstrate that this benchmark memory model serves as a reference for rational market beliefs. It not only captures biased analyst beliefs but also predicts asset mispricing in financial markets.

Finally, from the estimated mental context $c_{i,t}$, we identify cued recalls (firm j in month τ) based on the probability distribution (2.5). We have two sets of recalls: one is cued by the mental context estimated by training LSTM on the directions of analyst forecast revisions ΔF^d, which we define as analyst recalls. The other is cued by the mental context estimated by training LSTM on the directions of realized EPS revisions ΔRE^d, which we define as benchmark recalls. Benchmark recalls represent the historical episodes analysts should have focused on to make accurate forecast revisions.

3.4 Data

We use an extensive set of monthly public signals as external features X, following van Binsbergen, Han and Lopez-Lira (2023). These include financial ratios from WRDS, other firm-specific fundamentals from COMPUSTAT, macroeconomic variables from the Federal Reserve Bank of Philadelphia, and earnings-related variables from I/B/E/S. Table A1 provides the full list of 79 external features. The sample period spans January 1990 to December 2020, and on average the sample contains around 1,500 firms per month. Industries are defined according to the Fama-French 49 industry portfolios. To reduce the impact of extreme values, all variables are winsorized at the 2.5% level in the cross-section at each time point. For detailed explanations and data processing, see van Binsbergen, Han and Lopez-Lira (2023). To avoid look-ahead bias, all variables in $X_{i,t}$ are publicly announced and available to analysts before month t, with most released during month $t-1$.

Our approach is flexible and can incorporate additional features. Machine learning techniques naturally handle high-dimensional data and feature selection, allowing us to include many features and let the model determine which are most relevant for analyst memory and forecasts. While private information could also be incorporated, it is not available in our data. Therefore, we focus on consensus forecasts, which reduce noise from individual analysts and minimize the influence of unavailable private information. This approach allows us to study the collective memory for each firm.

Analyst forecasts for one-year-ahead EPS are obtained from I/B/E/S. We take the mean of individual forecasts to construct consensus forecasts. Specifically, in Equations (3.7) and (3.8), $F_{i,t}[EPS_{i,\tau+l}]$ and $F_{i,t-1}[EPS_{i,\tau+l}]$ represent the consensus forecasts. The framework and empirical methods can also be applied to other cash flows, such as long-term earnings growth or dividends. While differences between short-term and long-term forecasts, and across cash flow types, are beyond the scope of this paper, understanding how memory affects these outcomes is an interesting avenue for future research.

3.5 Training, Validation, and Testing

We train a single memory model on pooled data from all firms and the entire market history. This model represents the memory of a representative analyst, allowing us to study collective memory at the firm level. Our aggregation approach is justified by extensive evidence from cognitive science, which finds that the core dynamics of memory are highly consistent across individuals and are not artifacts of averaging dissimilar strategies (Healey and Kahana, 2014). Our method is thus conceptually similar to the use of representative agent models in asset pricing, we find that aggregate memory helps explain and predict stock market outcomes.

We design the model training process, sample splitting, and performance evaluation following Gu, Kelly and Xiu (2020). Details are provided in Appendix A2. The base training sample spans January 1990 to December 2004, while the validation sample covers January

2005 to December 2006.[18] The validation sample is used to determine the hyperparameter K, the dimension of memory cell and mental context vectors. We select the optimal K as $K = 10$ based on performance in the validation sample. The test sample spans January 2007 to December 2020.

Beyond the base training sample, we employ the recursive scheme (expanding window) to train the rest of the samples and evaluate the performance, following Gu, Kelly and Xiu (2020). Specifically, after selecting the optimal hyperparameter K, we first train the model on data from January 1990 to December 2004 and conduct out-of-sample analysis for 2005. Then, we expand the training sample to include data from 2005, re-train the model starting from the previous version, and perform out-of-sample analysis for 2006. This process is repeated annually until 2020. This recursive scheme offers two key advantages. First, it allows the model to adapt to the changing economic and financial environment, improving estimate accuracy. Second, it simulates analysts' evolving cognition, capturing how they update their recognition process with new information. This reflects how analysts' perceptions of a firm's state at a specific historical moment can shift as they review that event again later, with more experience and an evolving understanding.

Table 1 presents the out-of-sample performance of the analyst LSTM model and the benchmark LSTM model, covering both the validation and test samples from 2005 to 2020. For comparison, we also include a baseline logistic regression, which directly regresses $\Delta F^d_{i,t}$ and $\Delta RE^d_{i,t}$ on $X_{i,t}$, using the same recursive scheme and dataset as the LSTM models. For predicting analyst forecast revisions (ΔF^d), the analyst LSTM model significantly outperforms the baseline logistic regression, with a prediction accuracy gap of over 8%, and performs notably better than the benchmark LSTM model. For realized earnings revisions (ΔRE^d), the benchmark LSTM model and logistic regression show similar performance, both substantially outperforming the analyst LSTM model. These results yield two main findings. First, the superior performance of the analyst LSTM model over logistic regression in predicting analyst beliefs, alongside the benchmark LSTM model's comparable performance to logistic regression for realized earnings revisions, supports our later finding that analysts are more influenced by long-term memory and experiences than optimal. This highlights the importance of memory channels in modeling analyst belief formation. Second, the benchmark LSTM model demonstrates that memory models can achieve optimal forecast revisions, with predictive power as strong as non-memory models like logistic regression, while the analyst

[18] All variables are standardized following common practice in the literature: each variable is adjusted by subtracting the mean and dividing by the standard deviation calculated from the base training sample. The same mean and standard deviation are then used to scale the validation and test samples.

LSTM model underperforms. This suggests that, despite the memory model's accuracy potential, analysts may form their memory suboptimally, with certain memory processes not functioning as effectively as they could. We explore this further in Sections 5.3.1 and 5.3.2.

Table 1: Out-of-sample prediction accuracy of LSTM and logistic regression

Model	Analyst Forecast Revision (ΔF^d)	Realized Earnings Revision (ΔRE^d)
LSTM (Analyst)	**56.68%**	35.31%
LSTM (Benchmark)	37.37%	**59.52%**
Logistic	48.21%	58.46%

This table shows the average out-of-sample prediction accuracy of LSTM and logistic regression for both the analyst forecast revisions (ΔF^d, see (3.7)) and realized earnings revisions (ΔRE^d, see (3.8)) over the year 2005 to 2020. We apply the recursive scheme to evaluate the out-of-sample performance.

4 Analyst Recalls, Beliefs, and Trading

In this section, we demonstrate analyst recalls using the mental context estimated by the LSTM model, as outlined in Section 3.3. As a validation, we show that our model of analyst recalls has stronger predictive power for analyst EPS forecast revisions than other memory frameworks. We then present evidence that the estimated analyst recalls also capture investor beliefs, as they strongly predict stock returns and trading volume in financial markets.

4.1 Analyst Recalls

We first demonstrate the analyst recalls cued by the mental context c estimated from the LSTM model, which we refer to as LSTM recalls.

Figure 3 displays the analyst LSTM recalls. The darker blue gradients indicate that, at the current time x (corresponding to the columns), experiences from time y (corresponding to the rows) are more likely to be recalled by analysts. Specifically, at each time x, we identify the top 5 historical episodes most likely to be retrieved by analysts for each firm in our sample, using the similarity function (2.1) and probability distribution (2.5). Figure 3 then shows the count of historical episodes from time y that analysts recall when analyzing all firms at the current time x.[19]

[19]In all figures related to recalls, we present results using the top 5 recalls. The relative frequencies shown are consistent regardless of the number of recalls selected. We provide robustness checks using the top 1 and top 10 recalls in Appendix A6.

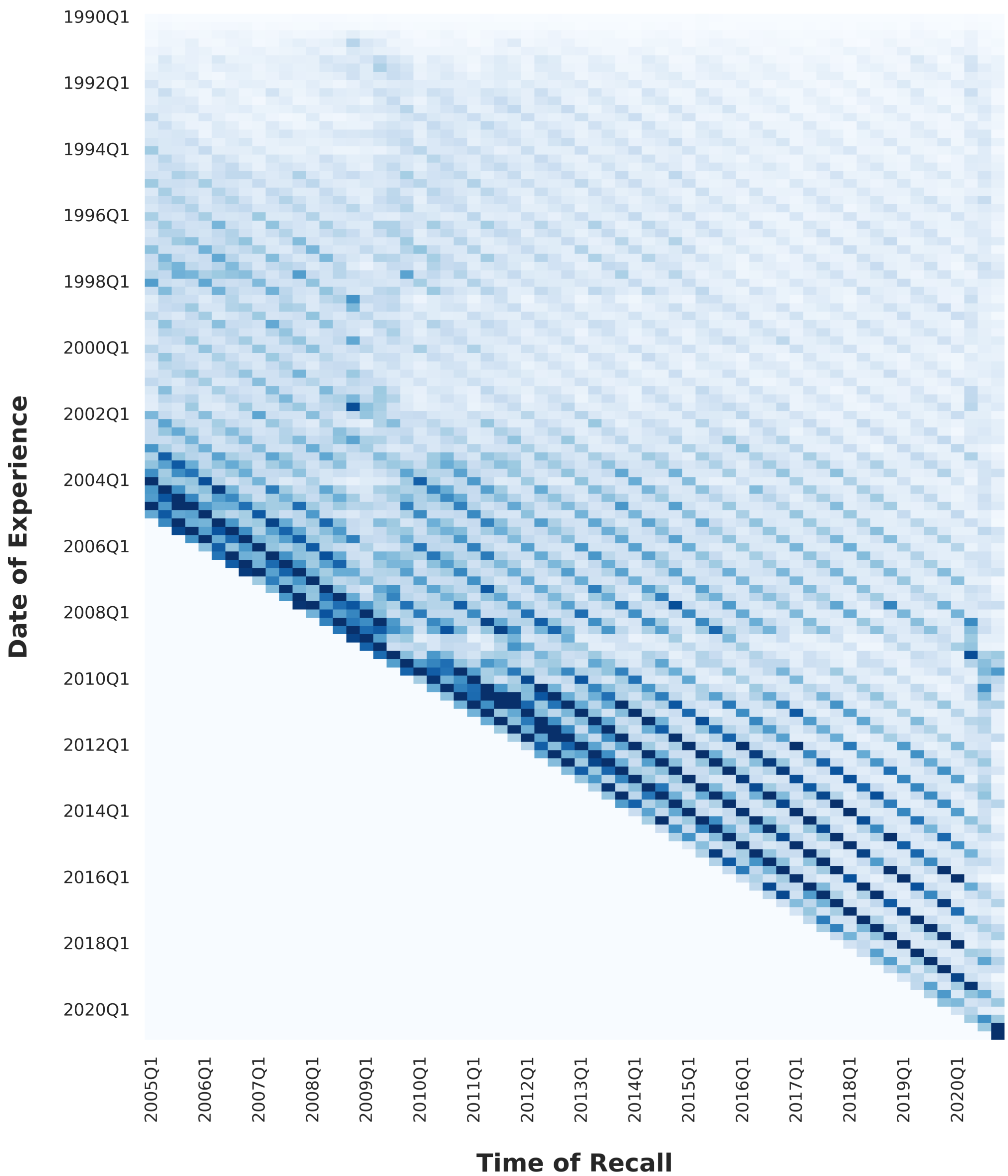


Figure 3: The analyst LSTM recalls cued by the mental context estimated from LSTM

The findings are summarized as follows. First, we observe the well-established facts of human memory: the recency effect and the temporal contiguity effect. Specifically, analysts tend to focus on recent episodes, typically the last quarter or the same quarter in the previous year. They also tend to recall episodes that occurred closely in time.[20]

[20]We formally examine the existence and significance of the temporal contiguity effect with a simulation study that is designed to deal with the nature of similarity between two adjacent vectors of economic and financial variables (external features) in time in Appendix A3.

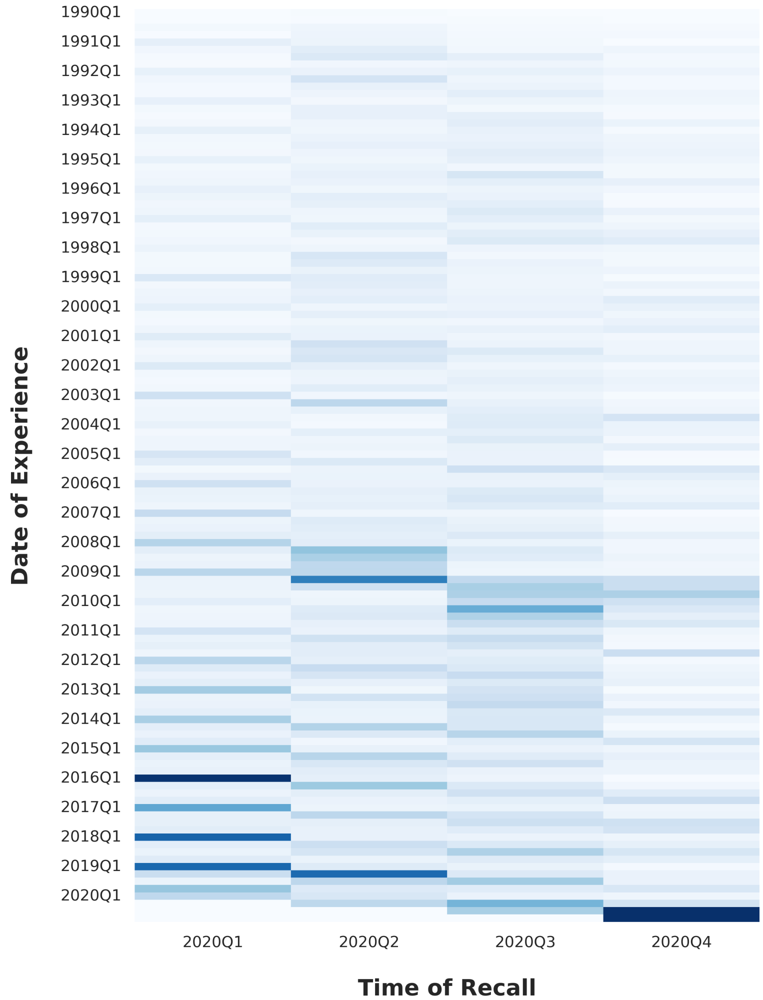


Figure 4: The analyst recalls during COVID

Second, consistent with the theoretical memory literature (Wachter and Kahana, 2024) and the experience effect literature (Malmendier and Nagel, 2011), we find that long-term memory and distant experiences are generally significant, sometimes even more so during certain periods. For example, during the COVID-19 pandemic in 2020, analysts frequently recalled the 2008 global financial crisis (GFC), and its impact even overshadowed more recent episodes. Specifically, Figure 4 illustrates these recalls in 2020. In the second quarter of 2020, analysts recalled the market crash period of the GFC. By the third quarter, rather than simply shifting their recall forward by one quarter, analysts quickly shifted their attention to the GFC recovery period, which aligns with the timing of enforcement of the economic stimulus policies. By the fourth quarter of 2020, analysts seldom recalled the GFC, considering the COVID pandemic unprecedented and concentrating instead on recent events. This overall recall pattern aligns with survey evidence in (Jiang et al., 2023), which shows that investors are more likely to recall both recent and salient episodes.

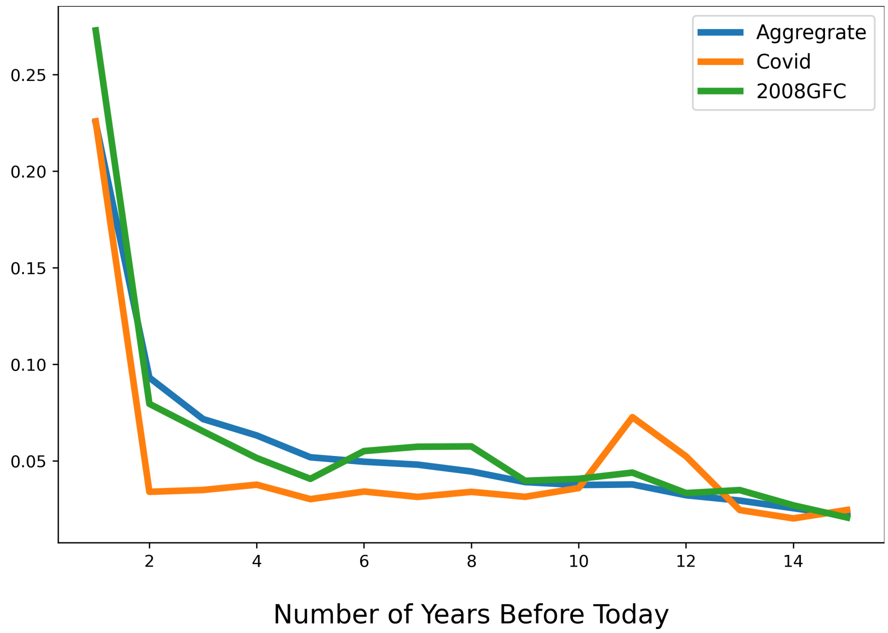


Figure 5: The Frequency of time difference between recall and recalled experience

Additionally, for each period, we calculate the frequency of the time difference between when a recall occurs and when the recalled experience took place. Figure 5 presents the average results for all periods from 2005 to 2020, along with specific averages for the 2008 GFC and the COVID-19 pandemic. Consistent with assumptions in the experience effect and extrapolation literature (Malmendier and Nagel, 2011; Nagel and Xu, 2022; Barberis et al., 2015), we find that, on average, the impact of past experiences decreases as the time difference grows. However, this assumption does not hold during certain periods. For instance, during the COVID-19 pandemic, the impact of the 2008 GFC remains prominent, as shown by the 11th point on the orange line. These findings, derived through our neuroscience-based approach, suggest a more robust framework for modeling the impact of past experiences across different periods. Our estimated analyst recalls also address an essential question from the extrapolation literature: "How far back do people look when forming beliefs?" (Barberis, 2018).

For comparison, we present recalls cued directly by external features X, which we term as naive recalls, in Appendix A4. Specifically, we evaluate the similarity between the current event (firm i, month t) and a past experience (firm j, month τ) as follows:

$$S = \exp\left(-\|X_{j,\tau} - X_{i,t}\|_2\right), \tag{4.1}$$

where we use the set of external features X for firm i in month t to find the most similar sets of historical external features. This follows the memory framework proposed by Bordalo, Gennaioli and Shleifer (2020), which implicitly assumes that all external features are equally important and time-invariant—assumptions that may not hold for analysts, as discussed in Section 5.3.1.

Naive recalls are more concentrated, with recalls clustering around a few specific historical moments at each point in time. They also exhibit a stronger recency effect, as the impact of long-term memory (distant episodes) is negligible except for the COVID pandemic. In contrast, LSTM recalls are more distracted, with analysts recalling episodes that occurred close together in time, and displaying additional temporal linkages such as cyclicality. It is the temporal contiguity effect that contributes to these temporal linkages in recalls. Without a dynamic memory structure like LSTM, which processes external features over time, recalls based solely on external features cannot replicate this well-established principle of human memory. In the next section, we demonstrate that LSTM recalls indeed outperform naive recalls in capturing analyst beliefs.

4.2 Recall and Beliefs

To assess the impact of recalls on current analyst beliefs, we examine the predictive power of predicted analyst recalls on analyst forecast revisions. To obtain the predicted analyst recalls for firm i in month t, we proceed as follows: first, we use the LSTM model trained on data prior to month t. Then, we take the external features $X_{i,t}$ (consisting of public signals available before month t, as defined in Section 3.4), along with the previous mental context $c_{i,t-1}$ and memory cell $m_{i,t-1}$ as inputs for the LSTM model. This generates the predicted current mental context $c_{i,t}$, which we then use to obtain recalls cued by $c_{i,t}$, representing the predicted analyst recalls.

We examine the link between recalls and current beliefs shown in Equation (2.6) empirically with the following predictive regression,

$$\Delta F_{i,t} = \beta \times RR_{i,t} + \varepsilon_{i,t}, \tag{4.2}$$

where $\Delta F_{i,t}$ is the analyst consensus EPS forecast revisions scaled by the stock price $P_{i,t-1}$ at $t-1$,

$$\Delta F_{i,t} = \frac{F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}]}{P_{i,t-1}}. \tag{4.3}$$

$RR_{i,t}$ denotes the average recalled revisions. We analyze two specifications of $RR_{i,t}$: $RR^A_{i,t}$ from the predicted analyst recalls by LSTM and $RR^N_{i,t}$ from the external features cued recalls,

$$RR^A_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) \times rr_{j,\tau}/P_{j,\tau-1} \tag{4.4}$$

$$RR^N_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(X_{j,\tau}) \times rr_{j,\tau}/P_{j,\tau-1}, \tag{4.5}$$

where $rr_{j,\tau}$ denotes the recalled revision in the specific recalled episode (j, τ) defined in Equation (2.4). The probability of any past experience being retrieved is proportional to the similarity function (2.1) for $RR^A_{i,t}$, and proportional to the similarity function (4.1) for $RR^N_{i,t}$.

Table 2 presents the regression results. Column (1) shows that analyst LSTM recalled revisions significantly predict their forecast revisions, while Column (2) indicates that naive recalled revisions lack similar predictive power. The strong performance of the LSTM recalled revisions suggests that our model of analyst recalls based on LSTM effectively captures key aspects of analyst belief formation processes. In Column (4)-(6), we include three control variables: lagged forecast revision $(F_{i,t-1}[EPS_{i,t+l}] - F_{i,t-2}[EPS_{i,t+l}])/P_{i,t-1}$, lagged earnings growth (the difference between the realized earnings for the last fiscal year and the realized earnings for the fiscal year before last, scaled by stock price $P_{i,t-1}$), and lagged forecast $(F_{i,t-1}[EPS_{i,t+l}]/P_{i,t-1})$. All three control variables are statistically significant and contribute additional predictive power. However, their inclusion neither reduces the predictive ability of the LSTM recalled revisions nor improves the performance of the naive recalled revisions. Thus, the addition of control variables does not alter our main findings.

4.3 Recall and Trading Activities

We have shown that analyst recalls capture their beliefs. These recalls are measured independently of future stock prices or returns. The literature often takes analyst forecasts as proxies for investor beliefs (Brunnermeier et al., 2021).[21] If analyst beliefs can indeed represent investor beliefs, we should expect our model of analyst recalls to also capture trading activities in financial markets. Therefore, we examine whether our predicted analyst recalls are associated with stock returns and trading volume.

[21]For example, van Binsbergen, Han and Lopez-Lira (2023) and Bouchaud et al. (2019) provide empirical evidence that investors tend to follow analyst expectations.

Table 2: Analyst forecast revisions and recalled revisions

	Forecast Revisions ($\Delta F_{i,t}$)					
	(1)	(2)	(3)	(4)	(5)	(6)
LSTM recalled revision RR^{A}	0.059***		0.060***	0.061***		0.061***
	(0.015)		(0.013)	(0.013)		(0.011)
Naive recalled revision RR^{N}		0.006	-0.005		0.010	-0.001
		(0.013)	(0.012)		(0.012)	(0.011)
Lagged forecast revision				0.127**	0.139***	0.127**
				(0.043)	(0.043)	(0.043)
Lagged earnings growth				0.013**	0.014**	0.013**
				(0.005)	(0.005)	(0.005)
Lagged forecast				-4.682***	-4.557***	-4.679***
				(0.782)	(0.709)	(0.756)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277,410	277,410	277,410	277,410	277,410	277,410
R-squared	0.053	0.047	0.053	0.070	0.064	0.070
Within R-squared	0.006	0.000	0.006	0.025	0.018	0.025

This table presents results for regressions of the form

$$\Delta F_{i,t} = \beta \times RR_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t},$$

where $\Delta F_{i,t} = (F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}])/P_{i,t-1}$ denotes analysts forecast revisions scaled by the stock price $p_{i,t-1}$ and $RR_{i,t}$ denotes the recalled revisions. We analyze two specifications of $RR_{i,t}$, $RR^A_{i,t}$ and $RR^N_{i,t}$. $RR^A_{i,t}$ denotes the analyst LSTM recalled revisions when cued by the mental context c that is estimated from LSTM $RR^A_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) * rr_{j,\tau}/P_{j,\tau-1}$, with the probability distribution is proportional to the similarity function (2.1). $RR^N_{i,t}$ denotes the naive recalled revisions when cued by the external features X, $RR^N_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(X_{j,\tau}) * rr_{j,\tau}/P_{j,\tau-1}$, with the probability distribution is proportional to the similarity function (4.1). Columns (1) to (3) report the results without control variables while Columns (4) to (6) report the results with control variables: lagged forecast revision $(F_{i,t-1}[EPS_{i,t+l}]-F_{i,t-2}[EPS_{i,t+l}])/P_{i,t-1}$, lagged earnings growth (the difference between the realized earnings for the last fiscal year and the realized earnings for the fiscal year before last, scaled by stock price $P_{i,t-1}$), and lagged forecast $(F_{i,t-1}[EPS_{i,t+l}]/P_{i,t-1})$. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

4.3.1 Recall and Stock Returns

First, we link our predicted analyst recalled revisions RR^A with stock returns by generating long-short portfolios and examining whether these portfolios yield excess returns. We hypothesize that positive (negative) recalled revisions predict future market optimism, leading to positive (negative) short-term stock returns. We drop stocks with prices lower than \$5 each month to alleviate the impact of small and illiquid stocks and the noise of the market microstructure. Our sample thus contains on average around 1,360 stocks in each month.

Table 3: Portfolios sorted on the analyst recalled revisions

	1	2	3	4	5	5-1
Panel A: Risk-adjusted Returns						
Mean	-0.06	0.17	0.13	0.26	0.35	0.41
t-stat	-0.64	1.68	1.46	2.84	3.11	3.01
Panel B: Forecast Revisions						
Mean	-0.14	-0.04	-0.01	0.04	0.17	0.31
t-stat	-3.04	-1.52	-0.29	3.43	5.14	7.27

This table reports the time-series average of risk-adjusted returns based on the Carhart four factor model (Carhart, 1997) (Panel A) and analyst forecast revisions ΔF (Panel B) on value-weighted portfolios formed on the quintiles of the analyst recalled revisions (RR^A). The returns are in percentage. The sample period is from January 2007 to December 2020. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags.

We report the value-weighted risk-adjusted returns of portfolios sorted on the quintiles of analyst recalled revisions RR^A, relative to the Carhart four-factor model Carhart (1997) in Panel A of Table 3. A long-short portfolio yields a monthly return spread of 0.41% per month with t-statistic above 3. In Panel B of Table 3, we find that the analysts revise down their forecasts for firms in the low recalled revisions group (the first quintile) but revise up their forecasts for firms in the high recalled revisions group (the fifth quintile). The difference in the revisions between the long leg group and the short leg group is significant with t-statistic of 7.27. Additionally, we show the returns on the long-short strategy based on the analyst recalled revisions RR^A can not be explained by leading asset pricing models in Table A9. These findings provide evidence supporting our hypothesis that analyst recalls can predict market beliefs and thus exhibit return predictability.

4.3.2 Recall, Disagreement and Trading Volume

Beyond return predictability, investor disagreement and trading volume are also central to asset pricing studies. However, less is known empirically about why there is disagreement among investors and thus trading volume. Literature on disagreement typically identifies two main sources: different information sets, and different models applied to the same information set (Hong and Stein, 2007; Malmendier and Nagel, 2011; Cookson and Niessner, 2020). We provide a new memory-based explanation for disagreement other than these two sources. The idea is that even with the same model and the same information set (memory database), different agents may retrieve distinct past experiences due to the inherent randomness in the process of cued recall.[22]

In our memory-based framework, disagreement arises under two conditions: (1) there is randomness in the process of cued recall, meaning that analysts may retrieve different past experiences from the same memory database; (2) these diverse past experiences must vary in status, so that analysts retrieving different experiences will learn distinct signals.

To illustrate the idea, let's consider three cases. Suppose analysts are now thinking of the current event (firm i, month t). In the first case, firm j in month τ is extremely similar to (i, t), then every analyst would recall (j, τ), and form beliefs $\Delta F_{i,t}$ according to the recalled revision $rr_{j,\tau}$, resulting in no disagreement. In the second and third cases, both (j_1, τ_1) and (j_2, τ_2) are similar to (i, t), and each has a 50% chance of being retrieved. In the second case, $rr_{j_1,\tau_1} = rr_{j_2,\tau_2}$, so regardless of which experience is recalled, analysts form the same beliefs for (i, t), and there is no disagreement. In the third case, however, rr_{j_1,τ_1} is positive and rr_{j_2,τ_2} is negative. Here, the 50% of analysts who recall (j_1, τ_1) form optimistic beliefs for (i, t), while the other 50% who recall (j_2, τ_2) form pessimistic beliefs, leading to disagreement between the two groups. These cases highlight the importance of both conditions: the first case emphasizes the role of randomness in recall, the second demonstrates why variability in recalled experiences is necessary, and the third shows that disagreement emerges when both conditions are met.

If investor beliefs align closely with analyst beliefs, then disagreement among analysts should reflect disagreement among investors, leading to higher trading volume (Atmaz and Basak, 2018; Shalen, 1993). Therefore, we hypothesize that the dispersion of analyst recalls should positively predict both disagreement and trading volume in the market.Specifically, we empirically test this hypothesis using the following regression specifications:

$$\text{Disagreement}_{i,t} = \beta \times \sigma(rr)_{i,t} + \varepsilon_{i,t}, \tag{4.6}$$

and

$$\text{AbVol}_{i,t} = \beta \times \sigma(rr)_{i,t} + \varepsilon_{i,t}, \tag{4.7}$$

[22]Memory can also contribute to disagreement by creating unique information sets for each agent. Based on personal past experiences, agents form different memory databases, leading to distinct information sets or perceptions of the current event (Bordalo et al., 2022).

where $\sigma(rr)_{i,t}$ is the weighted standard deviation of the recalled revisions for firm i in month t,[23]

$$\sigma(rr)_{i,t} = \sqrt{\sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) \times \left(rr_{j,\tau}/P_{j,\tau-1} - RR^A_{i,t}\right)^2}. \tag{4.8}$$

The standard deviation measures the dispersion of recalled revisions, capturing the second condition for our memory-based disagreement. To account for the first condition, randomness in the cued recall process, we refine this measure by incorporating weights based on the probability of each past experience being recalled.

Following Diether, Malloy and Scherbina (2002), we take the analyst forecast dispersion as our disagreement measure Disagreement$_{i,t}$, which is the standard deviation of each individual analyst earnings forecast for firm i in month t scaled by the last period's stock price. AbVol denotes the abnormal log trading volume for firm i in month t, following Cookson and Niessner (2020), it is the difference between the log volume in month t and the average log volume from month $t-12$ to $t-2$. Following Cookson and Niessner (2020), we include both firm and month fixed effects, and add a set of controls including the lagged abnormal trading volume, return and volatility to the regressions.

Table 4 reports the results testing the link between recalls, disagreement, and trading volume. Columns (1)-(3) show that the dispersion of recalled revisions positively and significantly predicts disagreement, though some samples are lost because calculating Disagreement$_{i,t}$ requires at least two analysts making forecasts for the firm in that period. Columns (4)-(6) provide evidence that the dispersion of recalled revisions also significantly predicts trading volume. These findings support our hypothesis that the dispersion of analyst recalled revisions positively predicts both disagreement and trading volume in the market. They further suggest that memory provides a powerful microfoundation for disagreement.

5 Analyst Recall Distortions

Forming beliefs based on the recalled episodes may be rational. Deviation from full rationality comes from the possibility that the recalls are distorted, i.e., analysts do not retrieve the right historical episodes. In this section, we demonstrate that analyst recalls are indeed distorted, and then discuss the asset pricing implications and driving factors behind these recall distortions.

[23] Similar to the measure of average recalled revisions $RR^A_{i,t}$, $\sigma(rr)_{i,t}$ is our model predicted variable, as it is also derived using information before month t, so the regressions (4.6) and (4.7) are also predictive.

Table 4: Dispersion of recalled revisions, disagreement, and trading volume

	Disagreement$_{i,t}$			Abnormal volume AbVol$_{i,t}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\sigma(rr)_{i,t}$	0.019***	0.017***	0.008***	0.089***	0.089***	0.038**
	(0.004)	(0.003)	(0.002)	(0.028)	(0.027)	(0.014)
$\sigma(rr)_{i,t-1}$		0.014***	0.006**		0.001	-0.004
		(0.003)	(0.002)		(0.017)	(0.019)
Disagreement$_{i,t-1}$			0.540***			
			(0.017)			
AbVol$_{i,t-1}$						0.539***
						(0.010)
$Ret_{i,t-1}$	-0.009***	-0.010***	-0.012***	-0.044	-0.044	-0.086**
	(0.002)	(0.002)	(0.002)	(0.052)	(0.052)	(0.035)
$Ret_{i,t-2}$	-0.008***	-0.008***	-0.004***	-0.028	-0.028	0.004
	(0.001)	(0.001)	(0.001)	(0.039)	(0.039)	(0.019)
$Ret_{i,t-3}$	-0.008***	-0.007***	-0.004***	0.033	0.033	0.055**
	(0.001)	(0.001)	(0.001)	(0.044)	(0.044)	(0.025)
$\sigma(Ret)_{i,t-1}$	0.078***	0.077***	0.039***	-0.133	-0.133	-0.573***
	(0.012)	(0.012)	(0.006)	(0.160)	(0.160)	(0.110)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	101,438	101,438	59,160	194,444	194,444	194,444
R-squared	0.435	0.437	0.609	0.175	0.175	0.412

This table presents results for regressions of the form

$$\text{Disagreement}_{i,t} = \beta \times \sigma(rr)_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t},$$

and

$$\text{AbVol}_{i,t} = \beta \times \sigma(rr)_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t},$$

where $\sigma(rr)_{i,t}$ is the weighted standard deviation of recalled revisions for firm i in month t as shown in Equation (4.8); Disagreement$_{i,t}$ measures disagreement, following Diether, Malloy and Scherbina (2002) it is the analyst forecast dispersion - the standard deviation of each individual analyst earnings forecast scaled by the last period's stock price; AbVol is the abnormal log trading volume for firm i in month t, following Cookson and Niessner (2020), it is the difference between the log volume in month t and the average log volume from month $t-12$ to $t-2$. Recalls are found based on the similarity measure shown in (2.1). $Ret_{i,t-1}$, $Ret_{i,t-2}$, $Ret_{i,t-3}$ are the lagged monthly stock return for firm i. $\sigma(Ret)_{i,t-1}$ is the standard deviation of stock return for firm i within a 12-month rolling window $t-11$ to t. The sample period is from January 2007 to December 2020. The reported recall dispersion coefficients are presented as the true values multiplied by 1000 for display convenience. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

5.1 Recall Distortions

We first define benchmark recalls, which are the episodes analysts should recall to form accurate beliefs. To determine these, we train the same LSTM memory model on realized earnings revisions (ΔRE^d) instead of analyst consensus forecast revisions (ΔF^d), allowing us to obtain the benchmark mental context. We then follow a similar memory retrieval process as used for analysts, using the benchmark mental context vectors as the cue to generate the benchmark recalls.

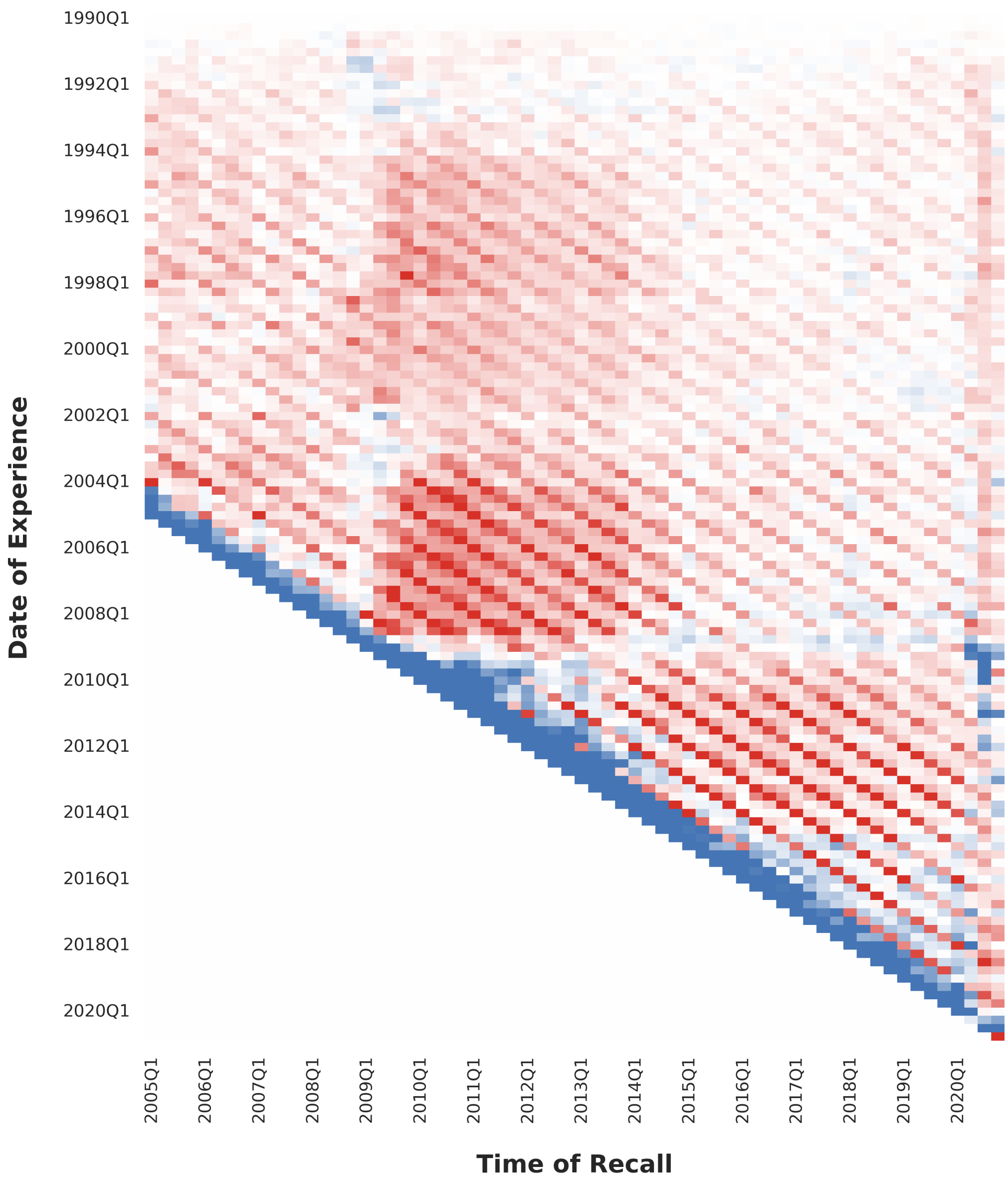


Figure 6: The difference between the analyst recalls and benchmark recalls

Figure 6 presents the difference between the benchmark recalls and analyst recalls. Historical times that are more likely to be recalled by the benchmark are shaded in darker blue, while those more likely to be recalled by analysts are shaded in darker red. Specifically, at each time x (corresponding to the columns), we compute the top 5 historical episodes most likely to be retrieved by analysts and the benchmark for each firm. Figure 6 then displays the difference in the number of episodes from time y (corresponding to the rows) that analysts recall versus the benchmark recalls across all firms at current time x.

Compared with analyst recalls, benchmark recalls in each period are more concentrated. Generally, the benchmark focuses more on recent episodes, with distant past experiences largely deemed irrelevant, while analysts are more influenced by long-term memory. Furthermore, analyst recalls sometimes are wrong. For instance, from 2010 to 2014, analysts frequently recalled episodes from the boom period before 2008, while the benchmark recalls seldom referenced that period. Another example is from the COVID-19 pandemic, when the benchmark heavily focused on the 2008 GFC, analyst recalls were more dispersed.

Additionally, analysts exhibit weaker recency effects during regular periods and stronger recency effects during crises, compared with the benchmark. This suggests that analysts do not fully respond to environmental changes. In regular times, environmental shifts are represented by new input features, prompting the optimal benchmark to quickly adjust to recent, similar scenarios. In contrast, analysts tend to focus on distant episodes that are less relevant to the current situation. During crises, when macroeconomic conditions change dramatically, the benchmark rapidly shifts focus, whereas analysts react insufficiently, remaining anchored in recent experiences. This underreaction in recalls is consistent with analysts' underreaction in their EPS forecasts, as shown in Section 5.3.1. Moreover, the gap between analyst and benchmark recalls provides a memory-based measure of inattention, aligning with the concept of measuring inattention as deviation from an optimal action (Gabaix, 2019).

We further develop a formal statistical test to examine whether the analyst recalls are different from the benchmark recalls. We define the recall distortion as:

$$\Delta R_{i,t} = RR^{\mathrm{A}}_{i,t} - RR^{\mathrm{B}}_{i,t} \tag{5.1}$$

where $RR^{\mathrm{A}}_{i,t}$ denotes the analyst recalled revisions and $RR^{\mathrm{B}}_{i,t}$ denotes the benchmark recalled revisions, as defined in Equation (4.4) based on different mental context vectors for analysts and the benchmark. The null hypothesis that analyst recalls are no different from benchmark recalls corresponds to a mean recall distortion ΔR of zero, i.e., $\overline{\Delta R} = 0$. A paired t-test yields a t-statistic of -4.83,[24] indicating that, overall, analyst recalls are significantly different from benchmark recalls. In other words, analyst recalls are indeed distorted.

[24] We cluster the standard errors at both the industry and year levels.

5.2 Recall Distortions and Asset Pricing Implications

We document that the analysts incorrectly retrieve past experiences. Additionally, we show in Table 1 that our benchmark memory model has significantly stronger predictive power for realized earnings revisions. In this section, we present the economic consequences of deviating from the benchmark memory model. First, we show that the biased analyst recalls relative to the benchmark recalls can predict the errors in analyst beliefs (forecast errors), and then the stock returns. Second, we connect the errors in analyst recalls to asset pricing anomalies.

5.2.1 Forecast Error

We document in Table 2 that our predicted analyst recalls are strong predictors of market beliefs, but these recalls are biased against the proposed benchmark recalls. If indeed, our proposed benchmark recalls can predict future realized earnings, we should expect the biased analyst recalls to predict analyst forecast errors. To test this, we first run the following predictive regression:

$$e_{i,t} = \beta \times \Delta R_{i,t} + \varepsilon_{i,t}, \tag{5.2}$$

where the forecast error $e_{i,t}$ is defined as the difference between the consensus forecasts and realized earnings, scaled by the stock price $P_{i,t-1}$, $e_{i,t} = (F_{i,t}[EPS_{i,t+l}] - EPS_{i,t+l})/P_{i,t-1}$. This regression is predictive because the forecast error $e_{i,t}$ is only available at time $t + l$, and the part of $e_{i,t}$, $F_{i,t}[EPS_{i,t+l}]$ is the consensus forecast made at time t, while we get the regressor $\Delta R_{i,t}$ using all the information before t, the same as how we compute the recalled revisions $RR_{i,t}$ in Equation (4.4).

To further examine whether the biased analyst recalls serve as a powerful predictor of forecast errors, we compete the predicted recall distortion with other well-known predictors of forecast errors documented in literature (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019; Bordalo et al., 2024). Specifically, we include the following controls as introduced in regression (4.2): the lagged forecast revision $(F_{i,t-1}[EPS_{i,t+l}] - F_{i,t-2}[EPS_{i,t+l}])/P_{i,t-1}$,[25]

[25]For example, Bordalo et al. (2019) run the following forecast error predictability regression:

$$EPS_{i,t+l} - F_{i,t}[EPS_{i,t+l}] = \beta_0 + \beta_1(F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}]) + \beta_2 F_{i,t-1}[EPS_{i,t+l}] + \varepsilon_{i,t},$$

thus one of their predictors, forecast revision $F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}]$ is known only at time t. We aim to predict the forecast errors using all the information before time t, thus we take the lagged forecast revision as the control instead of the current forecast revision.

lagged earnings growth (the difference between the realized earnings for the last fiscal year and the realized earnings for the fiscal year before last, scaled by stock price $P_{i,t-1}$), and lagged forecast ($F_{i,t-1}[EPS_{i,t+l}]/P_{i,t-1}$).

We additionally consider two other predictive regression specifications. First, we focus on the magnitude and explore whether the more severe recall distortion predicts larger forecast errors, then thus we take the absolute values for both the regressor ΔR and dependent variable e:

$$|e_{i,t}| = \beta \times |\Delta R_{i,t}| + \varepsilon_{i,t}. \tag{5.3}$$

Second, instead of focusing on the magnitude, we check the direction of the errors. Specifically, we define forecast direction errors $e^d_{i,t}$ as an ordinal variable with 5 different values $-2, -1, 0, 1, 2$:

$$e^d_{i,t} = sign(\Delta F^{\mathrm{d}}_{i,t}) - sign(\Delta RE^{\mathrm{d}}_{i,t}), \tag{5.4}$$

then we run an ordered logit model to estimate the impact of recall distortion on forecast direction errors.[26]

Table 5 reports the estimation results of the three regressions. In all three regressions, we find a strongly significant positive relationship between recall distortion and forecast errors. Columns (3) and (4) demonstrate that larger recall errors predict larger forecast errors, while Columns (5) and (6) show that analysts form over-optimistic (or over-pessimistic) beliefs when recalling overly positive (or negative) episodes. Columns (1) and (2) combine the specifications from Columns (3) to (6), and indicate that analyst recall distortion consistently predicts forecast errors. Furthermore, Columns (2), (4), and (6) demonstrate that even when competing with other forecast error predictors, recall distortion remains significant, and the magnitude of the coefficients remains relatively unchanged.

[26] The ordered logit model can be written as

$$\begin{aligned} e^{d*}_{i,t} &= \beta \times \Delta R_{i,t} + \varepsilon_{i,t} \\ e^d_{i,t} &= \begin{cases} -2 & \Delta F^{d*}_{i,t} \leqslant \mu_{-2} \\ -1 & \mu_{-1} < e^{d*}_{i,t} \leqslant \mu_{-2} \\ 0 & \mu_0 < e^{d*}_{i,t} \leqslant \mu_{-1} \\ 1 & \mu_0 < e^{d*}_{i,t} \leqslant \mu_1 \\ 2 & \mu_1 < e^{d*}_{i,t} \end{cases} \end{aligned} \tag{5.5}$$

where $e^{d*}_{i,t} \in (-\infty, \infty)$ represents a latent variable that captures the degree of over-optimism in analyst beliefs, expressed in real numbers. Four cut-off points, μ_{-2}, μ_{-1}, μ_0, and μ_1, are estimated alongside other parameters.

Table 5: The analyst forecast errors and recall distortion

	Forecast error $e_{i,t}$		Forecast error $\|e_{i,t}\|$		Forecast direction error $e^d_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear	Linear	Linear	Linear	Ologit	Ologit
$\Delta R_{i,t}$	0.041***	0.039***			0.024***	0.023***
	(0.009)	(0.009)			(0.003)	(0.003)
$\|\Delta R_{i,t}\|$			0.125***	0.094***		
			(0.025)	(0.019)		
Lagged forecast revision		0.008		0.032**		-0.020**
		(0.019)		(0.014)		(0.008)
Lagged earnings growth		-0.004		0.010**		-0.002
		(0.002)		(0.003)		(0.001)
Lagged forecast		-1.255**		2.679***		-0.389*
		(0.501)		(0.543)		(0.226)
Firm fixed effect	Yes	Yes	Yes	Yes	No	No
Month fixed effects	Yes	Yes	Yes	Yes	No	No
Observations	277,410	277,410	277,410	277,410	277,487	277,487
R-squared	0.139	0.145	0.267	0.301		

This table presents results for regressions of the form

$$e_{i,t} = \beta \times \Delta R_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t \varepsilon_{i,t}$$

in Columns (1) - (4) and the form

$$\Pr(e^d_{i,t} = j) = \Phi\left(\kappa_j - \beta \times \Delta R_{i,t}\right) - \Phi\left(\kappa_{j-1} - \beta \times \Delta R_{i,t}\right)$$

in Columns (5)-(6). The dependent variable $e_{i,t}$ is analyst forecast error, and $e^d_{i,t} \in \{-2,-1,0,1,2\}$ is the direction of analyst forecast error for firm i at time t. The independent variable $\Delta R_{i,t}$ is analyst recall distortion which is predicted by the model with all the information available before time t as defined in Equation (5.1). We add three controls in Column (2), (4) and (6): lagged forecast revision $(F_{i,t-1}[EPS_{i,t+l}] - F_{i,t-2}[EPS_{i,t+l}])/P_{i,t-1}$, lagged earnings growth (the difference between the realized earnings for the last fiscal year and the realized earnings for the fiscal year before last, scaled by stock price $P_{i,t-1}$), and lagged forecast $(F_{i,t-1}[EPS_{i,t+l}]/P_{i,t-1})$. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

5.2.2 Subsequent Stock Returns

The asset pricing results in Section 2.1 provide evidence that analyst recalls effectively predict investor beliefs. If our proposed benchmark recalls serve as an unbiased reference for rational beliefs, which investors deviate from, a rational market should correct the mistake, thus we should observe a reversal in portfolio returns generated by the analyst recall distortion.

To test the prediction, we sort stocks into quintile portfolios based on our predicted recall distortion ΔR. The first quintile group consists of stocks with overly positive analyst recalls relative to the benchmark recalls (with average ΔR equals to 2.80), while the fifth quintile

Table 6: Analyst forecast revisions after the portfolio formation

Month	1	2	3	4	5	6
Mean	-18.12	-2.42	2.69	-3.81	-1.38	1.08
t-stat	-6.73	-1.52	1.54	-1.12	1.20	0.57

This table reports reports the gaps in the average analyst forecast revisions ΔF between the fifth quintile group and the first quintile group in the months after the portfolio formation Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags.

group contains stocks with overly negative analyst recalls (with average ΔR equals to -4.74). We long stocks in the fifth quintile group and short stocks in the first quintile group. We present the value-weighted cumulative returns of this long-short strategy in Figure 7, the shaded region shows the 95% confidence intervals. Table 6 reports the gaps in the average analyst forecast revisions between the fifth quintile group and the first quintile group in the months after the portfolio formation. In addition to the prediction of a return reversal for this long-short portfolio, as indicated in Table 6, we should observe a negative initial return for the long-short portfolio as the gap in analyst forecast revisions is extremely negative at the beginning. Then the portfolio return may not reverse immediately, as in the second month, the gap in analyst forecast revisions is still negative. Then the gap gradually vanishes.

Figure 7: The cumulative returns on long-short portfolio sorted on recall distortion

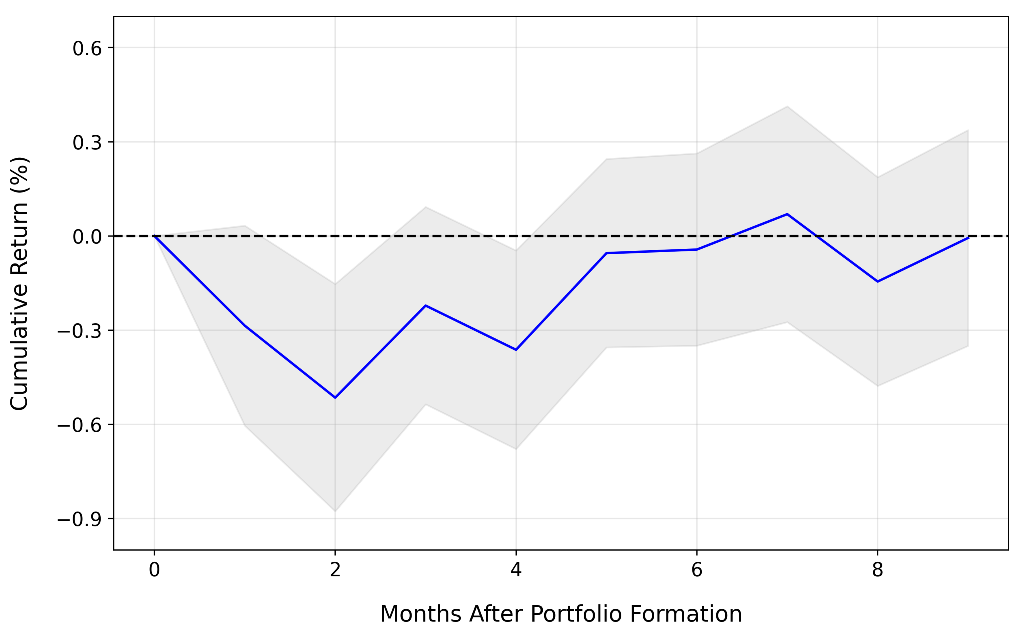


Figure 7 provides supporting evidence for all these predictions as the cumulative returns of the long-short strategy are negative in the first months and reach a maximum of -0.51% around two months after portfolio formation. Then the returns quickly reverse and remain stable around zero after six months.

5.2.3 Memory-Based Attention and the Short-Term Reversal

So far we have been focusing on one property of the recalls, positive or negative, measured by the forecast revisions and the realized revisions in the recalled episodes. However, we can learn more about the connection between analyst recalls and financial markets by examining other characteristics of the recalls. For instance, the temporal aspects of the analyst recalls help identify the specific historical moments that analysts are paying attention to today. Taking it a step further, the same temporal aspects of the benchmark recalls can serve as a reference point, indicating whether analysts exhibit excessive or insufficient attention to specific historical moments. This can help us better understand asset pricing anomalies that are linked to temporal characteristics, such as the short-term reversal.

Short-term reversal describes the phenomenon that a stock's return over the last month is negatively associated with its subsequent return (Lehmann, 1990; Jegadeesh, 1990). The link between recall and short-term reversal is as follows. When investors exhibit excessive (insufficient) attention to the prior month, they are more (less) likely to follow their last month's actions. For example, when investors push up the stock price last month and they pay excessive attention to last month, they will tend to push up the stock price again this month, this would offset the short-term reversal effect. Then if analyst beliefs are representative of investor beliefs, we hypothesize that when analysts pay excessive (insufficient) attention to last month, the short-term reversal effect will be weaker (stronger). In addition, such biased attention reflects investor belief distortion, as it contrasts with the benchmark recalls. Consequently, in a rational market, we should observe a reversal of the hypothesized pattern in the following months.

We first follow Jegadeesh (1990) and re-examine the standard short-term reversal strategy in our sample. We sort stocks into teriles[27] on their prior-month returns, then long stocks in the first tercile group (past losers) and short stocks in the third tercile group (past winners).

[27] Jegadeesh (1990) sort stocks into decile portfolios, however, the equal-weighted risk-adjusted returns of the long-short strategy based on decile portfolios are not statistically significant in our sample, for instance, the CAPM alpha is 0.29% with t-statistic of 1.11. Sorting stocks into tercile portfolios can further ensure that we have enough stock in each portfolio when performing the double sort on memory-based attention and prior-month returns.

Our sample period is from January 2007 to December 2020. Table 7 reports the performance of the standard short-term reversal strategy. The equal-weighted raw return is 0.48% per month which is lower than the monthly raw return of 2.49% (sample period of 1934-1987) and 0.67% (sample period of 1982-2009) documented in Jegadeesh (1990) and Da, Liu and Schaumburg (2014), respectively. The risk-adjusted return based on the Fama-French three-factor model (FF3) is 0.34% per month, and it is similar to the FF3 alpha of 0.33% shown in Da, Liu and Schaumburg (2014).

Table 7: Performance of the standard short-term reversal trading strategy

	Excess Return		CAPM		FF3		Carhart	
	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Intercept	0.48	4.12	0.29	2.09	0.34	2.16	0.34	2.16
MKT-R$_f$			0.25	4.11	0.22	3.69	0.22	3.70
SMB					0.07	0.84	0.07	0.78
HML					0.07	0.64	0.07	0.82
MOM							-0.01	-0.13

This table presents the equal-weighted raw returns and risk-adjusted returns of the standard short-term reversal trading strategy. The strategy sorts stocks into terciles according to prior-month returns, then long past losers and short past winners. The risk-adjusted returns are based on the CAPM, the Fama-French three-factor model (FF3) (Fama and French, 1993), and the Carhart four-factor model Carhart (1997). The sample period spans January 2007 to December 2020. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags.

We define analyst memory-based attention to the last month as follows:

$$a^A_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) \times \mathbb{1}_{\tau=t-1}, \tag{5.6}$$

where $a^A_{i,t}$ evaluates the likelihood that analysts recall episodes from the previous month. We define benchmark memory-based attention to the last month, $a^B_{i,t}$, analogously by using the mental context predicted by the benchmark memory model. Analyst biased attention is then defined as the difference between analyst attention and benchmark attention, $\Delta a_{i,t} = a^A_{i,t} - a^B_{i,t}$. Δa is positive when analysts pay excessive attention to the last month and negative when they pay insufficient attention to it.

To examine the link between analyst biased attention to the last month and short-term reversal, we first sort stocks into terciles based on Δa. The median values of Δa for the first, second, and third terciles are 0.18, -0.11, and -0.30, respectively, indicating that analysts pay excessive attention to the last month in the first tercile group and insufficient attention in the second and third tercile groups. Next, within each Δa tercile, we further sort firms into terciles based on prior-month returns, with the first tercile containing stocks with low prior-month returns and the third tercile containing stocks with high prior-month returns.

Table 8: Portfolios sorted on biased investor memory-based attention and short-term reversal

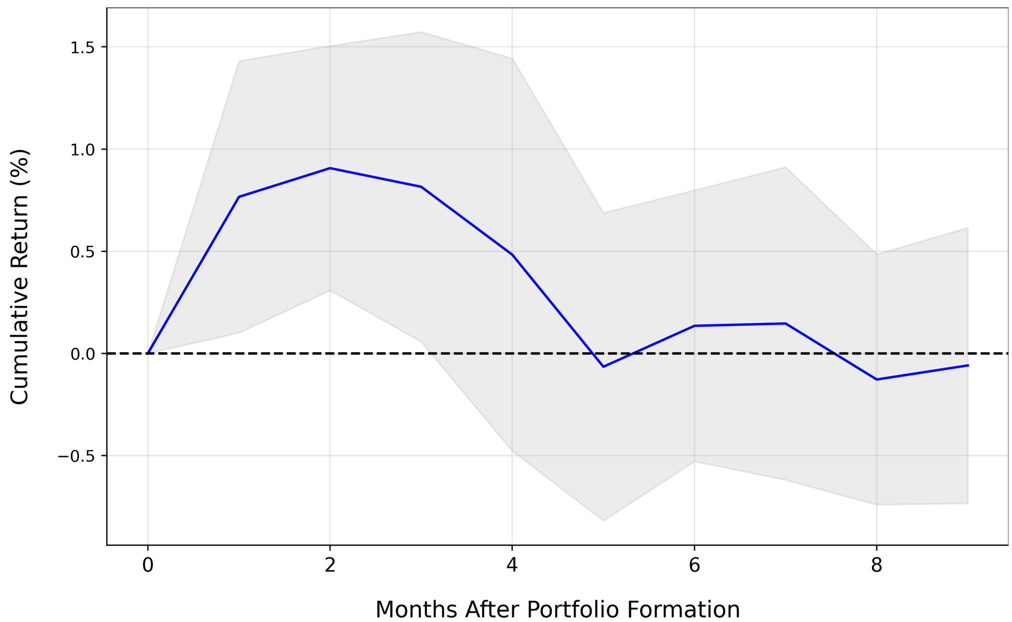
	Short-term reversal			
Attention	1	2	3	3-1
1	0.80	0.69	0.22	-0.58
t-stat	3.11	4.15	0.92	-1.79
2	0.39	0.45	0.73	0.34
t-stat	3.70	5.34	6.04	2.39
3	0.35	0.38	0.86	0.51
t-stat	2.16	3.33	5.04	2.11
3-1	-0.44	-0.32	0.65	1.09
t-stat	-1.56	-1.96	2.96	3.88

This table reports the time-series average of risk-adjusted returns (in percent) on equal-weighted portfolios sorted first on biased investor memory-based attention to prior month Δa then on short-term reversal. The risk-adjusted returns are based on the Carhart four-factor model Carhart (1997). The sample period is January 2007 to December 2020. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags.

Table 8 presents risk-adjusted returns based on the Carhart four-factor model (Carhart, 1997) for these equal-weighted double-sorted portfolios. We find that the monthly alpha of the short-term reversal strategy is -0.58% for stocks where analysts exhibited the most excessive attention to prior-month information.The memory effect entirely overrides the short-term reversal effect. It creates a momentum effect: stocks with higher returns last month continue to earn higher returns this month. Because investors overly believe that current market conditions mirror those of the previous month and replicate last month's trading behavior. On the contrary, the monthly alpha of the short-term reversal strategy is 0.51% with t-statistic of 2.11 for the stocks that analysts are inattentive to last month's situation. Compared to the standard short-term reversal strategy with an alpha of 0.34%, the reversal effect is amplified by excluding stocks most likely to exhibit momentum. The difference in short-term reversal performance between the two Δa groups is highly significant; the double-sorted long-short portfolio earns a 1.09% risk-adjusted return per month, with a t-statistic of 3.88, more than tripling the standard short-term reversal strategy's risk-adjusted return. Figure 8 further shows that the biased attention measure Δa captures investor belief distortion, as the cumulative returns of the double-sorted long-short portfolio reverse within just

5 months.

Figure 8: The cumulative returns on double-sorted long-short portfolio on biased investor memory-based attention and short-term reversal



5.3 Driving Factors of Analyst Recall Distortions

So far, we have examined the economic consequences of analyst recall distortions through several asset pricing tests. But why do analysts make these suboptimal recalls? To explore this question, we propose two potential driving forces: first, analysts do not optimally encode external features into their memory; second, analysts mistakenly forget certain episodes.

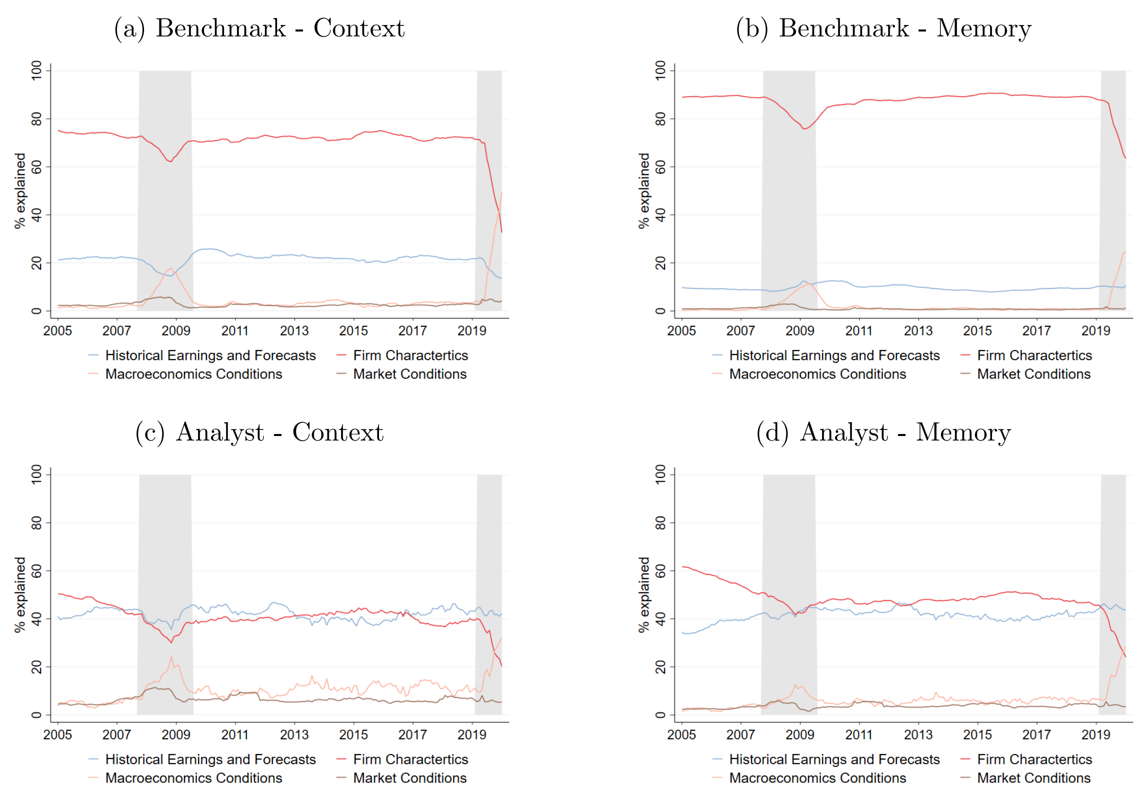
5.3.1 Encoding Errors

We investigate the first potential driving force by examining the focus of the analyst, the benchmark memory cell, and the mental context. Both the memory cell and mental context play essential roles in memory models, e.g., acting as cues for recall. Understanding how these elements manifest in empirical data provides insights into underlying memory mechanisms and can guide future memory modeling. We are the first to explore this.

Both memory cells and mental context are latent in LSTM, so we interpret these components by examining how their variable importance, denoted $\omega_{j,t}$ for each external feature j at time t, changes over time, following Gu, Kelly and Xiu (2020) and Kelly, Pruitt and

Su (2019). To derive $\omega_{j,t}$, we follow these steps. First, for each of the K dimensions of the memory cell (m) and mental context (c), we perform a linear regression on all external features using a 12-month rolling window. Next, we set feature j to zero, keeping the other features and coefficients unchanged, and calculate the reduction in R^2. We then compute the average reduction in R^2 across all K dimensions as $\omega_{j,t}$. To better illustrate variable importance, we categorize the 79 external features into four groups and calculate the sum of $\omega_{j,t}$ for all features j within each group, referred to as group variable importance.[28] The four groups are (1) firm characteristics other than earnings-related variable ω_t^C (Part 1 and Part 2 of Table A1 except for the monthly stock return and P/E ratio related variables such as pe_exi, pe_inc, peg_trailing, capei); (2) macroeconomics conditions ω_t^M (Part 3 of Table A1); (3) market conditions ω_t^R (the firms' stock return over the previous month); (4) historical earnings and forecasts ω_t^F (Part 4 of Table A1 and pe_exi, pe_inc, peg_trailing, capei).

Figure 9: Decomposition of the mental context and memory cells



(a) Benchmark - Context

(b) Benchmark - Memory

(c) Analyst - Context

(d) Analyst - Memory

Figures 9a - 9d present time-series plots of the group variable importance for the benchmark, analyst mental context, and memory cells. The findings are threefold. First, variable

[28]Focusing on group variable importance, rather than individual variables, provides a concrete example of mitigating instability associated with the Rashomon effect (Mullainathan and Spiess, 2017). By aggregating over correlated inputs, our results are less sensitive to which particular feature a model emphasizes.

importance is time-varying across both analysts and the benchmark. For example, during recessions, macroeconomic variables become more important, while firm characteristics and historical earnings and forecasts become less relevant. This pattern aligns with the intuition that recessions have market-wide, systematic effects and supports the prediction of limited attention theory (Kacperczyk, Van Nieuwerburgh and Veldkamp, 2016), which suggests that investors allocate more attention to aggregate news during downturns.[29] However, this behavior may not indicate bias, as the optimal benchmark similarly shifts attention to macroeconomic news in downturns. The observed time-varying variable importance is also consistent with evidence from textual analysis of analyst reports (Bastianello, H Décaire and Guenzel, 2025; Ke, 2025).

Second, we observe that variable importance for the mental context is more volatile than that for the memory cell. This aligns with the interpretation that mental context represents short-term generalized information, capturing the perceived high-frequency dynamics of the firm, while the memory cell stores long-term generalized information, reflecting the perceived firm's low-frequency regime.

Third, the analyst and the benchmark focus on different aspects, i.e., for the benchmark, firm characteristics play the most important role in both the mental context and memory cell, but for analysts, they overweight the importance of the historical earnings-related variables. The finding that analysts do not put proper weights on external features in their memory provides empirical evidence and memory foundation for the models of encoding errors (early noise). The literature on encoding errors examines the economic impact of noisy internal representation of the presented data (e.g., Woodford, 2020; Frydman and Nunnari, 2021; Frydman and Jin, 2022; Drugowitsch et al., 2016). The evidence that analysts overweight their own past forecasting decisions reconciles with the sticky expectation (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019), the self-herding bias (Hirshleifer et al., 2019), and the confirmation bias (Lord, Ross and Lepper, 1979; Barberis, 2003). Additionally, the reliance on past earnings growth is consistent with fundamental extrapolation (Barberis, Shleifer and Vishny, 1998; Barberis et al., 2015). The observed time-varying variable importance and distinct weighting of external features in analyst memory contrast with the model of Bordalo, Gennaioli and Shleifer (2020), which assumes all external features are equally important and time-invariant in the recall process. This evidence supports our choice of a dynamic memory and mental context model to describe analyst behavior.

[29] Kwan, Liu and Matthies (2022) shows that institutional investors focus on aggregate news during economic downturns, as evidenced by daily internet news reading data.

The differences in the composition of analyst and benchmark mental context and memory cells suggest that distortions in analyst recalls may arise because analysts do not process and encode information as efficiently as the optimal benchmark.

By examining the differences in decomposition between the analyst and benchmark mental contexts, we can analyze how analysts perceive and react to changes in the external economic environment compared to the optimal benchmark. To do this, we use a double-differential approach.

Specifically, first, we define $\Delta\omega$ as the time-series changes in variable importance,

$$\Delta\omega_t = \sqrt{\frac{\left(\omega_t^F - \omega_{t-1}^F\right)^2 + \left(\omega_t^M - \omega_{t-1}^M\right)^2 + \left(\omega_t^R - \omega_{t-1}^R\right)^2 + \left(\omega_t^C - \omega_{t-1}^C\right)^2}{4}}$$

where ω_t^F, ω_t^M, ω_t^R, ω_t^C represent the group variable importance of historical earnings and forecasts, macroeconomic conditions, market conditions, and firm characteristics, respectively, as shown in Figure 9a for the benchmark ($\Delta\omega^{\mathrm{b}}t$) and Figure 9c for the analysts ($\Delta\omega^{\mathrm{a}}t$).

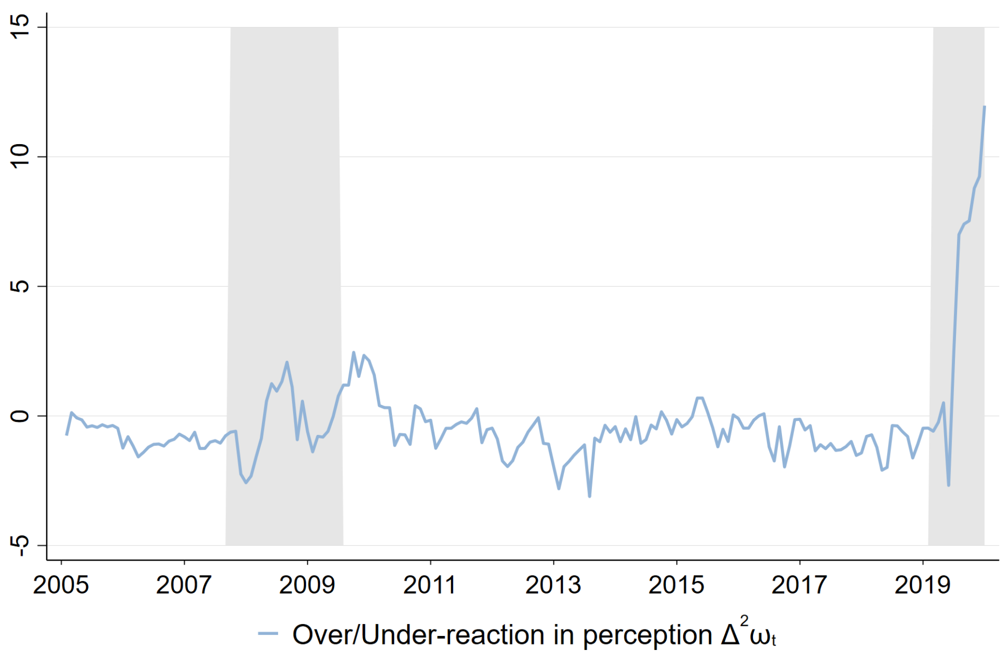


Figure 10: Difference in perceived changes in external economic environment between the analysts and the benchmark

Second, we define the difference between the analyst and the benchmark time-series changes in variable importance as $\Delta^2\omega_t$,

$$\Delta^2\omega_t = \Delta\omega_t^{\mathrm{b}} - \Delta\omega_t^{\mathrm{a}} \tag{5.7}$$

which represents the extent to which analysts perceive changes in the external economic environment differently from the benchmark. When analysts deviate from the benchmark perception, i.e., $\Delta^2\omega_t \neq 0$, they do not respond to changes optimally. We interpret this difference as a misreaction in perception. Specifically, when $\Delta^2\omega_t > 0$, analysts underreact to external changes, and when $\Delta^2\omega_t < 0$, they overreact. Figure 10 illustrates the timing and magnitude of these deviations.

Figure 10 shows that during normal periods, analysts' perceptions of external economic changes are generally aligned with the benchmark. However, during crises, their perceptions diverge. In the 2008 GFC, the difference in perceived changes in variable importance was initially negative but later turned positive, indicating that analysts first overreacted to economic environment changes and then underreacted. Similarly, during the COVID-19 pandemic, analysts significantly underreacted to this systematic change, failing to fully recognize the crisis's potential severity. This observation reconciles the findings in Landier and Thesmar (2020) that analyst consensus forecasts show a smooth downward trend that contrasts with the dramatic stock price movement in response to the salient COVID-19 crisis. This pattern suggests underreaction in analyst forecasts both in the short-term and the long-term expectations.

We further investigate how encoding errors bias analyst recalls and beliefs by examining the impact of misreaction in perceptions. Specifically, we link misreaction in perceptions to misreaction in analyst forecasts and the recency effect observed in their recalls.

First, we define the analysts' misreaction in forecasts $\Delta React_{i,t}$, following Coibion and Gorodnichenko (2015), it captures the relation between the forecast errors and forecast revisions:

$$\Delta React_{i,t} = \begin{cases} \mathrm{sign}(\Delta F_{i,t} - \Delta RE_{i,t}) & \text{if } \Delta F_{i,t} > 0 \\ \mathrm{sign}(\Delta RE_{i,t} - \Delta F_{i,t}) & \text{if } \Delta F_{i,t} < 0 \\ -1 & \text{if } \Delta F_{i,t} = 0 \end{cases}$$

where $\Delta F_{i,t} = (F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}])/P_{i,t-1}$ denotes the analyst forecast revisions and $\Delta RE_{i,t} = (EPS_{i,t+l} - F_{i,t-1}[EPS_{i,t+l}])/P_{i,t-1}$ denotes the realized EPS revisions. When analysts revise their forecasts upwards ($\Delta F_{i,t} > 0$), if $\Delta F_{i,t} - \Delta RE_{i,t} < 0$, this indicates that the analyst forecast $F_{i,t}[EPS_{i,t+l}]$ at time t lies between the realized EPS ($EPS_{i,t+l}$) and the lagged forecast $F_{i,t-1}[EPS_{i,t+l}]$ from $t-1$. This suggests that analysts have insufficiently revised their forecasts upwards, indicating underreaction. Conversely, if $\Delta F_{i,t} - \Delta RE_{i,t} > 0$ when forecasts are revised upwards, this reflects overreaction. Underreaction and overreaction are similarly defined when analysts revise their forecasts downwards ($\Delta F_{i,t} < 0$). If analysts do not revise their forecasts ($\Delta F_{i,t} = 0$) but the actual EPS differs from the lagged forecast, this suggests that analysts should have revised their forecasts, indicating an underreaction. In summary, $\Delta React_{i,t} = -1$ indicates analyst underreaction in forecasts, while $\Delta React_{i,t} = 1$ indicates overreaction. We exclude cases where $\text{sign}(\Delta F_{i,t}) \neq \text{sign}(\Delta RE_{i,t})$, as this suggests analysts revised their forecasts in the wrong direction. Misreaction is best defined when forecast revisions align with the correct direction.[30] To make it align with misreaction in analyst perceptions $\Delta^2 \omega_t$ which is a time-series variable at a monthly frequency, we take the average of $\Delta React_{i,t}$ across firms for each month:

$$\Delta React_t = \frac{\sum_i \Delta React_{i,t}}{\#\ of\ firms\ at\ month\ t} \tag{5.8}$$

The second measure is the biased recency effect shown in analyst recalls, which captures the extent to which analysts recall recent episodes over distant ones relative to the benchmark recalls. Being analogous to the analyst memory-based attention to last month defined in (5.6), we define the analyst recency effect as how likely analyst recalls fall into the last 12 months,[31]

$$Recency^{\text{A}}_{i,t} = \sum_{(j,\tau) \in M_{i,t}} p(c_{j,\tau}) \times \mathbb{1}_{t-\tau \leqslant 12}. \tag{5.9}$$

We similarly define the benchmark recency effect as $Recency^{\text{B}}_{i,t}$ with the mental context predicted by the benchmark memory model. Then we define the time-series biased analyst recency effect relative to the benchmark as the average difference between the benchmark recency effect and analyst recency effect across firms for each month:

$$\Delta Recency_t = \frac{\sum_i Recency^{\text{B}}_{i,t} - Recency^{\text{A}}_{i,t}}{\#\ of\ firms\ at\ month\ t} \tag{5.10}$$

Table 9 presents OLS regression results examining the relationship between misreaction in analyst forecasts, the biased recency effect, and analysts' misreaction to changes in the external economic environment. Column (1) shows that the analysts' misreaction in their forecasts are associated with their differing perceptions of changes in the external economic environment relative to the optimal benchmark. Specifically, when analysts' perceptions

[30] Coibion and Gorodnichenko (2015), Bordalo et al. (2019), and Bouchaud et al. (2019) do not exclude these cases and include them in regression analysis. We provide robustness checks in Appendix A6.

[31] We can alternatively define the recency effect as how likely the recalls fall into the past three years or five years, but the following patterns and conclusions remain unchanged as shown in Appendix A6.

Table 9: Misreaction in analyst perceptions, forecasts and recency effect

	Misreaction	Biased Recency Effect	
		Normal	Crisis
	(1)	(2)	(3)
$\Delta^2\omega$	-0.010**	0.017**	-0.018***
	(0.005)	(0.008)	(0.004)
Observations	180	144	36
R-squared	0.028	0.039	0.223

This table presents the OLS regression results of the misreaction in analyst forecasts and biased recency effect on the analysts' misreaction in perception of changes inexternal economic environment. The independent variable, $\Delta^2\omega_t$, is defined in Equation (5.7). The dependent variables, misreaction $\Delta React_t$ and biased recency effect $\Delta Recency_t$ are defined in Equations (5.8), (5.10), respectively. The sample period is from January 2007 to December 2020, with 2008, 2009, and 2020 are classified as crisis periods, while the remaining years are considered normal periods. The reported t-statistics are robust to heteroskedasticity.

underreact ($\Delta^2\omega_t > 0$), their forecasts also tend to underreact relative to true EPS. As part of their encoding errors, when analysts are slow to adjust the weights of external variables in the encoding process, they develop distorted perceptions, leading to misreactions in their forecasts.

This finding is also consistent with our observations on the biased analyst recency effect. In Columns (2) and (3) of Table 9, we find that during normal times, analysts think the recent episodes are less important when they underreact in perceptions of changes in the external economic environment. However during crisis times, analysts pay more attention to recent episodes as their perceptions underreact further. This result suggests that the patterns shown in Figure 6, where analysts tend to underweight recent periods during normal times but overemphasize them during crisis periods, may stem from analysts' insufficient reaction to external economic changes.

5.3.2 Selective Forgetting

For the second potential driving force, we study the role of selective forgetting in the analyst and the benchmark belief formation processes. Literature has shown that forgetting affects investors' decision-making in surveys and experiments. For instance, Walters and Fernbach (2021) and Gödker, Jiao and Smeets (2022) argue that selective forgetting, as a memory bias, leads investors to have distorted recalls, fostering overconfidence and biased beliefs.[32] In this section, we provide more comprehensive and empirical evidence of selective forgetting's role in belief formation.

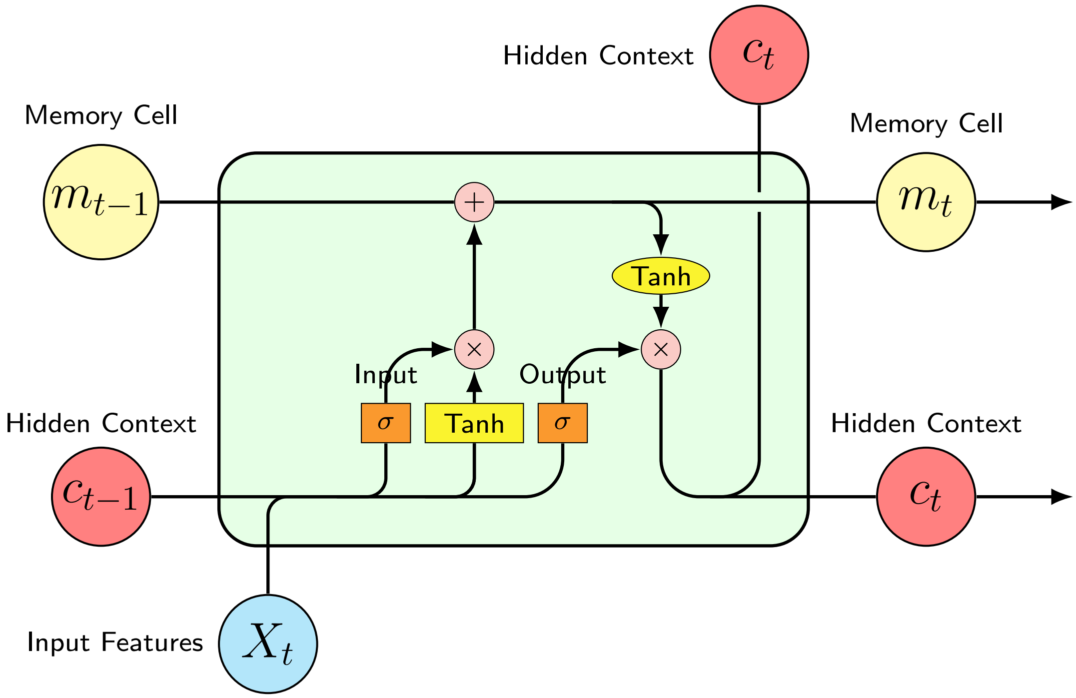


Figure 11: The Structure of LSTM without Forget Gate

Specifically, we perform a counterfactual analysis by blocking the channel of selective forgetting in the original model. Then we examine how analyst recalls change and whether the changes improve their EPS forecasts without selective forgetting. In the full LSTM, the forget gate determines what content to erase from the memory cell. Here, we remove the forget gate, as illustrated in Figure 11. In this modified memory structure, the memory cell updates only by inputting new information, without selective forgetting:

$$m_t = m_{t-1} + input_t \times \tilde{x}_t, \tag{5.11}$$

meaning we set $forget = \mathbf{1}$ in the memory cell updating process in the full LSTM (3.4). In this configuration, analysts can no longer selectively forget information. However, this does not imply full retention of past information. Instead, they remain subject to limited memory capacity and passive forgetting, as new information gradually overrides older content. We

[32] Walters and Fernbach (2021) provide evidence of selective forgetting, showing that participants are more likely to forget consequential losing trades than winning trades and thus recall losing trades less readily than winning ones. Gödker, Jiao and Smeets (2022) report similar findings, with individuals over-remembering positive investment outcomes and under-remembering negative ones.

then take the originally trained full LSTM for analysts and the benchmark, replacing the memory update process (3.4) with the version without the forget gate (5.11), while keeping all other components and parameters unchanged. This modification allows us to extract the new mental context and recalls, which we define as counterfactual recalls.

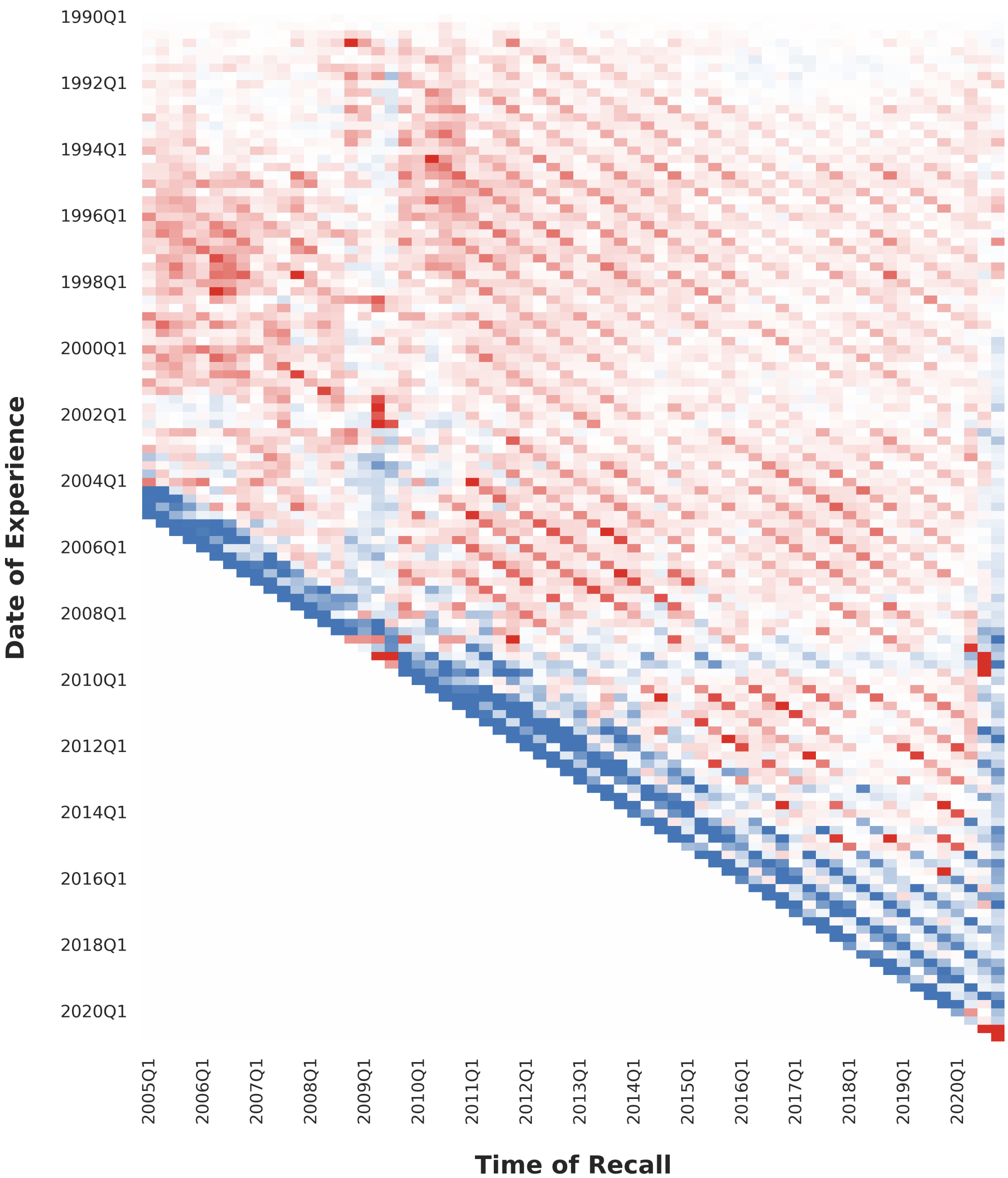


Figure 12: The difference between the analyst recalls and counterfactual recalls

Figure 12 shows the difference between analyst recalls with selective forgetting (as presented in Section 4.1) and the counterfactual recalls obtained by blocking the forget gate. Analogous to the comparison between analyst recalls and benchmark recalls in Figure 6, the darker blue areas highlight historical moments that analysts under-recall, while the darker red areas indicate episodes that analysts over-recall relative to the counterfactual. From Figure 12, we find that without the channel of selective forgetting, analysts would have recalls that are more similar to the benchmark recalls as shown in Figure 6. For example, without selective forgetting, recalls exhibit a stronger recency effect, generally resembling the benchmark recalls. These counterfactual recalls place greater emphasis on recent events, particularly those occurring within the past 12 months.

We formally test the differences in recalls with and without selective forgetting in Table 10, yielding the following findings. First, in Panel A, we observe that recalls without selective forgetting display a stronger recency effect *Recency* (5.9), with the magnitude doubling from 0.052 to 0.104. Additionally, the average recalled revisions RR (4.4) increase from -0.707 to -0.625. This difference of 0.082 is statistically significant (with a t-statistic of 6.69), indicating that recalls are more positive without selective forgetting.

Second, we find that the recalls without selective forgetting are more similar to the benchmark recalls. Specifically, without selective forgetting, the analyst recalled revisions is -0.625, which is closer to the benchmark recalled revision of -0.434 than the recalls with selective forgetting. Additionally, the recency effect in the recalls without selective forgetting is 0.104, which is more aligned with the recency effect of 0.177 implied by the benchmark recall, compared with the recalls with selective forgetting. This evidence suggests that selective forgetting contributes to analyst recall distortions, distorting analysts away from optimal forecasting decisions. In Appendix Table A3, we demonstrate that removing selective forgetting enhances the predictive power of recalled revisions for realized earnings revisions, with the within R^2 for analyst models increasing by 68%.

Third, selective forgetting affects analysts and the benchmark differently. Panel B of Table 10 shows that, without selective forgetting, the benchmark recalls become more pessimistic and exhibit a reduced recency effect, which is in sharp contrast with its impact on analysts. While analysts would improve their forecasts without selective forgetting, the benchmark's performance worsens. Table A3 indicates that removing selective forgetting reduces the predictive power of the benchmark model, with the within R^2 dropping by 40%. This suggests that selective forgetting is not inherently a source of biased beliefs; rather, it has the potential to support optimal belief formation. However, analysts do not utilize this mechanism optimally, leading to distortions in their recalls.

In summary, we find that selective forgetting plays an important role in analyst belief formation processes, but it distorts analyst memory and recalls away from the optimal

Table 10: Paired t-tests between LSTM models with and without selective forgetting

Model	Recency Effect	Recalled Revisions
	(1)	(2)
Panel A: Analysts		
LSTM with forget gate	0.052	-0.707
LSTM without forget gate	0.104	-0.625
Difference	0.051	0.082
t-statistics	(4.75)	(6.69)
Panel B: Benchmark		
LSTM with forget gate	0.177	-0.434
LSTM without forget gate	0.160	-0.488
Difference	-0.017	-0.054
t-statistics	(-4.72)	(-4.27)

This table presents results of the paired t-tests for recency effect (*Recency* (5.9)) and recalled revisions (*RR* (4.4)) between the full LSTM model and the LSTM model without the forget gate (counterfactual). Panel A reports the results for the analyst recalls and panel B reports the results for the benchmark recalls. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level.

benchmark.

6 Conclusion

This paper is the first to extract mental context and recalls in financial markets for a large cross-section of firms over several decades. We achieve this by empirically estimating a structural, memory-based model of analyst EPS forecast revisions. Using this extensive panel, we provide novel evidence linking memory and recall with belief formation in financial markets. Specifically, we find that well-established memory principles, such as the recency and temporal contiguity effects, play a significant role in shaping analyst recalls and beliefs. Our analysis reveals that analyst recalls deviate from an optimal benchmark memory model, which would make optimal earnings forecast revisions. Analysts rely too heavily on long-term memory and underreact to economic changes, over-recalling episodes in distant past during normal periods while under-recalling them in crises. These distortions arise from suboptimal encoding and selective forgetting in analysts' memory systems.

Moreover, we generate extensive new asset pricing predictions grounded directly in memory. Our model of analyst recalls and distortions significantly predicts key patterns in asset pricing, including stock returns, trading volume, forecast errors, mispricing, and anomalies such as short-term reversal.

Our detailed investigation of analyst memory, particularly recalls, provides new insights for theoretical modeling and empirical research on memory in financial markets. Moreover, our approach is well-suited for complex real-world settings, where agents face high-dimensional, non-stationary variables with nonlinear interactions. It can be readily applied to study the impact of memory in other contexts, such as CEO memory and firm decisions or memory in FX forecasts. With high-quality individual-level data, our framework could also be extended to examine heterogeneity in agents' memory or other cognitive systems and their implications for asset pricing, leveraging the strong neuroscientific foundations of certain machine learning models.

Finally, this paper demonstrates that some machine learning models are not black boxes: when rooted in neuroscience and psychology, they can be used to study investor behavior. This insight opens a new avenue for applying machine learning in behavioral finance.

Appendix

Appendix A1 External Features

The following table reports all 79 public signals as external features X. For detailed data processing, such as imputation, see van Binsbergen, Han and Lopez-Lira (2023).

Table A1: Input Features

Part 1. Firm Fundamentals−WRDS Financial Ratios			
Variable	Definition	Variable	Definition
Accrual	Accruals/Average Assets	adv_sale	Advertising Expenses/Sales
aftret_eq	After-tax Return on Average Common Equity	aftret_equity	After-tax Return on Total Stockholders Equity
aftret_invcapx	After-tax Return on Invested Capital	at_turn	Asset turnover
bm	Book/Market	capei	Shillers Cyclically Adjusted P/E ratio
capital_ratio	Capitalization Ratio	cash_debt	Cash Flow/Total Debt
cash_lt	Cash Balance/Total Liabilities	cash_ratio	Cash Ratio
cfm	Cash Flow Margin	curr_debt	Current Liabilities/Total Liabilities
curr_ratio	Current Ratio	debt_asset	Total Debt/Total Assets
debt_at	Total Debt/Total Assets	debt_capital	Total Debt/Capital
debt_ebitda	Total Debt/EBITDA	debt_invcap	Long-term Debt/Invested Capital
divyield	Dividend Yield	dltt_be	Long-term Debt/Book Equity
dpr	Dividend Payout Ratio	efftax	Effective Tax Rate
equity_invcap	Common Equity/Invested Capital	evm	Enterprise Value Multiple
fcf_ocf	Free Cash Flow/Operating Cash Flow	gpm	Gross Profit Margin
GProf	Gross Profit/Total Assets	int_debt	Interest/Average Long-term Debt
int_totdebt	Interest/Average Total Debt	intcov	After-tax Interest Coverage
intcov_ratio	Interest Coverage Ratio	inv_turn	Inventory Turnover
invt_act	Inventory/Current Assets	lt_ppent	Total Liabilities/Total Tangible Assets
npm	Net Profit Margin	ocf_lct	Operating CF/Current Liabilities
opmad	Operating Profit Margin After Depreciation	opmbd	Operating Profit Margin Before Depreciation

Variable	Definition	Variable	Definition
pay_turn	Payables Turnover	pcf	Price/Cash flow
pe_exi	P/E (Diluted, Excl. EI)	pe_inc	P/E (Diluted, Incl. EI)
PEG_trailing	Trailing P/E to Growth ratio	pretret_earnat	Pre-tax Return on Total Earning Assets
pretret_noa	Pre-tax return on Net Operating Assets	profit_lct	Profit Before Depreciation/Current Liabilities
ps	Price/Sales	ptb	Price/Book
ptpm	Pre-tax Profit Margin	quick_ratio	Quick Ratio (Acid Test)
RD_SALE	Research and Development/Sales	rect_act	Receivables/Current Assets
rect_turn	Receivables Turnover	roa	Return on Assets
roce	Return on Capital Employed	roe	Return on Equity
sale_equity	Sales/Stockholders Equity	sale_invcap	Sales/Invested Capital
sale_nwc	Sales/Working Capital	short_debt	Short-Term Debt/Total Debt
totdebt_invcap	Total Debt/Invested Capital		

Part 2. Other Firm Fundamentals

Variable	Definition	Variable	Definition
asset_g	Growth Rate in Total Assets	invest_g	Growth Rate in Capital Expenditure
sales_g	Growth Rate in Sales	return	Monthly Stock Return

Part 3. Macroeconomic Variables

Variable	Definition	Variable	Definition
con_g	Log Difference of Consumption in Goods and Services	IPT_g	Log Difference of Industrial Production Index
GDP_g	Log Difference of Real GDP	unemployment	Unemployment Rate

Part 4. Earnings-Related Variables

Variable	Definition	Variable	Definition
Realized_EP_ANN	Realized Annual Earnings from Last Period/Stock Price from Last Month	Realized_EP_QTR	Realized Quarter Earnings from Last Period/Stock Price from Last Month
AF_EP_lag	Mean Analyst Forecast from Last Period /Stock Price from Last Month	NUMEST_lag	Number of Forecasts from Last Period
Realized_ANN_g	Growth Rate in Realized Annual Earnings	Realized_QTR_g	Growth Rate in Realized Quarter Earnings
AF_g_lag	Lag 1 Growth Rate in Mean Analyst Forecast	Maturity	Months to Fiscal End Date/12

Appendix A2 Model Training and Choice of Hyperparameters

The model training process is as follows: We use the Adam algorithm for stochastic gradient optimization (Kingma and Ba, 2014) with default hyperparameters. To prevent overfitting, we implement early stopping with a patience of 5. The batch size is set to 10,000, and the loss function is the negative log-likelihood, as the final decision function is logistic regression.

The only hyperparameter that requires tuning is K, the dimension of the latent memory cell and mental context vector. To determine the optimal K, we select a set of candidate values, train the model on the training sample, and evaluate the model's performance with each candidate K on the validation sample.

Table A2: Model performance with different choices of the hyperparameter K

Part I. Analyst Forecast Revisions (ΔF^d)				
K	5	10	15	20
Training	59.58%	61.42%	62.26%	62.04%
Validation	54.22%	**55.45%**	55.15%	54.72%
Part II. Realized Earnings Revisions (ΔRE^d)				
K	5	10	15	20
Training	68.68%	71.22%	72.33%	71.74%
Validation	**59.58%**	58.95%	57.73%	57.58%

This table presents the model performance for fitting and predicting the direction of analyst forecast revision (ΔF^d, see (3.7)) and realized earnings revision (ΔRE^d, see (3.8)) in the base training and validation samples with different choices of dimension of the memory cells and mental context vectors (K). The training sample is from January 1990 to December 2004. The validation sample is from January 2005 to December 2006.

Table A2 presents the model's performance in fitting and predicting the direction of analyst forecast revisions (ΔF^d, see (3.7)) and realized earnings revisions (ΔRE^d, see (3.8)) in the base training and validation samples, using various dimensions for the memory cells and mental context vectors (K). Based on prediction accuracy in the validation sample, $K = 10$ is optimal for analyst forecast revisions (ΔF^d) and $K = 5$ for realized earnings revisions (ΔRE^d), though the differences between models are marginal.

Given our primary interest in analyst beliefs, we select $K = 10$ for both models of analyst forecast revisions (ΔF^d) and realized earnings revisions (ΔRE^d) throughout the paper. Using the same K for both models allows for a more consistent comparison between analyst forecasts and benchmark beliefs, minimizing the influence of differing hyperparameters.

Appendix A3 Temporal Contiguity

LSTM has the unique feature that it stores the long-term memory. Combined with the autoregressive structure of the context state, LSTM can demonstrate temporal contiguity — one of the fundamental principles of the human memory system. Temporal contiguity refers to the observation that, when people recall an event, they also tend to recall other temporally successive events. In psychology experiments, Kahana (1996) first documents this tendency: after recalling an item from a specific serial position, people often recall the next item from a neighboring position. Two properties of the temporal contiguity effect are also noted: a forward asymmetry, where recalls are more likely to proceed in a forward direction than a backward one, and time-scale invariance, meaning the contiguity effect remains significant even for events recalled from the distant past (a key reason for incorporating long-term memory).

Beyond the psychology literature, Wachter and Kahana (2024) discuss the temporal contiguity effect in a theoretical financial context, though they do not provide empirical evidence on its significance or existence. In this section, we go further and present empirical evidence demonstrating the importance of temporal contiguity in the belief formation processes of analysts.

To provide clear empirical evidence of temporal contiguity, it is necessary to separate the temporal contiguity effect from the intrinsic similarity (autoregressive) between adjacent vectors of economic and financial variables over time. To achieve this, we design a simulation study that eliminates correlations among input features by decomposing them into orthogonal space. This simulation uses the same recursive model as outlined in Section 3.5, while retaining the previously estimated parameters in the consensus forecast revisions ΔF^d model. This approach ensures that the findings from the simulation study directly reflect the processes involved in analyst belief formation.

The design of the simulation study is as follows. In each round of the simulation, first, we generate a set of input vectors:

$$X_t^{\text{sim}} = \mathbb{E}\left[X\right] + 10u_t \times \sigma\left(X\right), \quad t = 1, 2, ..., T,$$

where $\mathbb{E}\left[X\right]$ and $\sigma\left(X\right)$ represent the time-series mean and the standard deviation (the square root of diagonal elements of the covariance matrix) of the original input features X, respectively. The symbol “×” denotes element-wise product operator. u_t are randomly drawn, mutually orthogonal vectors with L^2 norms of 1. The inner product of any two simulated input vectors, X_t^{sim}, is a fixed number (the square of the L^2 norm of the feature expected value $\|\mathbb{E}[X]\|^2$). This approach maximally removes correlations between input features while keeping the simulated inputs close to the true distribution (the multiplier 10 on u_t serves for this purpose as well).[33] We simulate $T = 70$ periods for X_t^{sim}.[34]

Second, in period $T+1$, we duplicate the simulated input vector from a randomly selected period τ, so that $X_{T+1}^{\text{sim}} = X_\tau^{\text{sim}}$. The index τ is randomly chosen from the interval $(10, T-10)$, excluding the first and last 10 periods to minimize any potential influence from the primacy and recency effects.

Third, we use the empirically estimated model to simulate all mental context $\{c_t^{\text{sim}}\}_{t=1}^{T+1}$ based on the set of simulated input features $\{X_t^{\text{sim}}\}_{t=1}^{T+1}$, then use c_{T+1} as the cue to search for the recalls. Next, we study the similarity around τ to examine the temporal contiguity.[35] The similarity of mental context $S(c_{\tau+l}, c_{T+1})$, as defined in (2.1) (with $\gamma = 2$ and $\xi = 1$), serves as a measure of the likelihood that the episode in period $\tau + l$ will be recalled when using c_{T+1} as the cue.

If the temporal contiguity effect exists, it is expected to observe that similarity decreases as the distance from K increases with a forward asymmetry. In other words, temporal contiguity implies the following predictions:

$$\begin{aligned}
&\forall\ 0 < i < j, S(c_{\tau+i}, c_{T+1}) > S(c_{\tau+j}, c_{T+1}),\\
&\forall\ 0 < i < j, S(c_{\tau-i}, c_{T+1}) > S(c_{\tau-j}, c_{T+1}),\\
&\forall\ i > 0, S(c_{\tau+i}, c_{T+1}) > S(c_{\tau+i}, c_{T+1}).
\end{aligned} \tag{6.1}$$

Figure A1 displays the estimated average temporal contiguity effect of our trained analyst LSTM model using simulated data. The simulation is run 10000 times. The value $S(c_\tau, c_{T+1})$ is omitted from the figure, as it is the highest (confirming the model's ability to retrieve the correct recall) and is not directly relevant to examining temporal contiguity.

[33] Our simulation results are robust to the choice of inner product. If the simulated input vectors are made completely orthogonal to each other (inner products equal to zero), i.e.,

$$X_t^{\text{sim}} = 10u_t,$$

the temporal contiguity effect still holds.

[34] Since the dimension of the original input features X is 79, we can only maximally generate 79 mutually orthogonal vectors.

[35] The process is similar to the standard memory experiment that the participants are shown T different words successively, and then asked to recall one of the word (the τ-th) they have studied. Next, the participants are asked to make free recalls. See for example, Healey, Long and Kahana (2019).

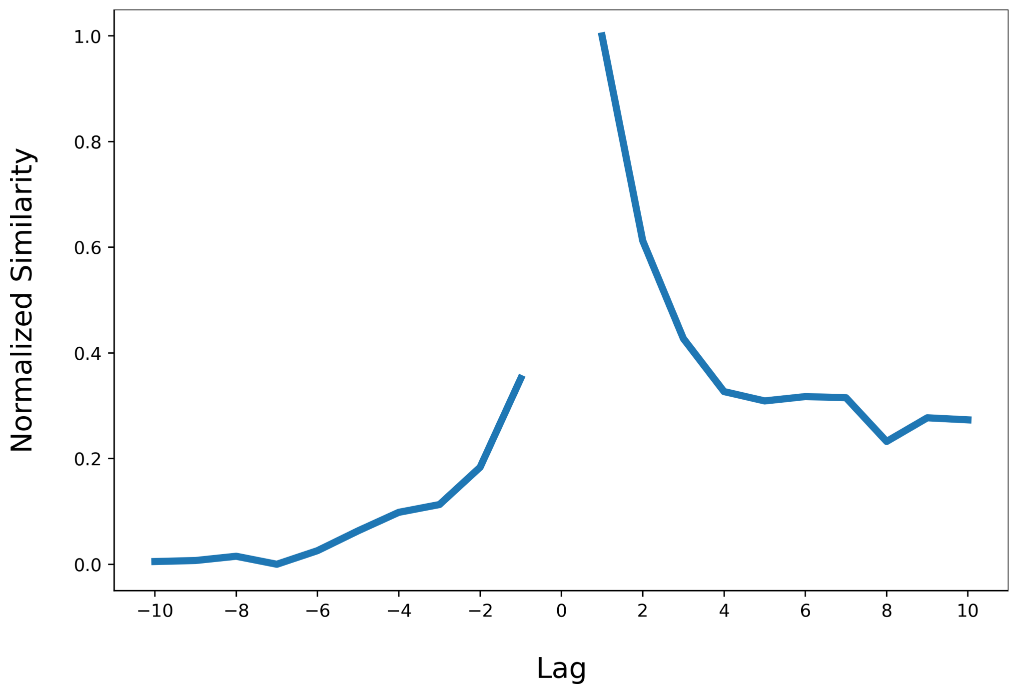


Figure A1: The Temporal contiguity effect in the analyst LSTM model

The X-axis represents l, the positive or negative lag relative to τ, while the Y-axis shows $S(c_{\tau+l}, c_{T+1})$, which is min-max normalized. The figure indicates that similarity is highest immediately after recall ($l = 1$) and decreases as the absolute value of the lag increases. Additionally, the similarity is generally lower for negative lags compared to positive ones.

Overall, the similarity pattern in Figure A1 aligns with the conditions in (6.1) implied by temporal contiguity. This finding supports that our LSTM model empirically demonstrates the fundamental memory principle - the temporal contiguity effect. It also underscores the importance of incorporating temporal contiguity in modeling analyst belief formation processes. These findings highlight the need for a memory model, like LSTM, which can effectively capture temporal contiguity to represent analyst beliefs accurately.

We argue that the LSTM model incorporates both long-term memory and an autoregressive context structure, and that these two channels should work together to produce temporal contiguity. However, these channels may not always be active, as they can diminish depending on model parameters and the empirical data used for training. For example, if the model parameters cause the forget gate to consistently clear the memory cell, effectively blocking the long-term memory channel, the LSTM reduces to a Recurrent Neural Network (RNN), which does not produce temporal contiguity. We examine this scenario next.

We present the simulation results for the temporal contiguity effect using an RNN model.

The simulation design is analogous to that used for the LSTM model, with the only difference that we now employ the trained RNN model in place of the LSTMx. The RNN model is also recursively trained to fit the direction of consensus forecast revisions, ΔF^d, as described in Section 3.5.

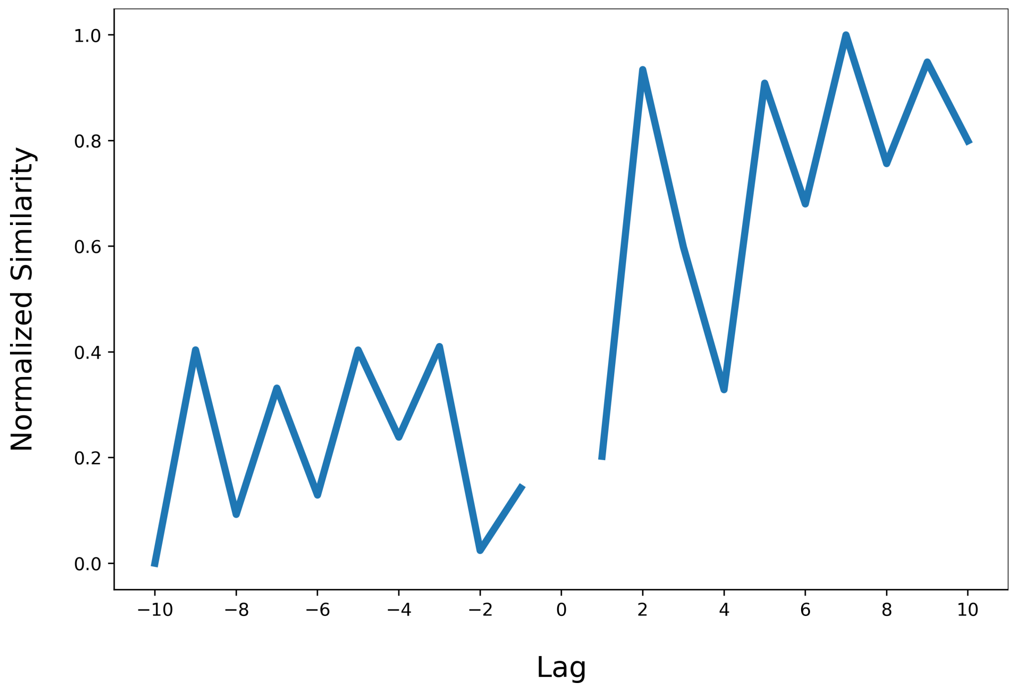


Figure A2: Temporal Contiguity of RNN

Figure A2 reports that RNN does not produce the temporal contiguity effect. The similarity is not consistent with the conditions (6.1) implied by the temporal contiguity effect. However, as indicated in Figure A1, the temporal contiguity effect should be significant in analyst beliefs. Howard and Kahana (2002) document that the channel of long-term memory is essential in generating the temporal contiguity effect. Lack of this channel disables RNN to produce the temporal contiguity effect. This also makes RNN inferior to model the analyst belief formation processes, compared to LSTM.

In sum, our simulation study using the empirically estimated model indicates that both long-term memory and autoregressive context structure are active and jointly contribute to generating temporal contiguity within the investor belief formation processes.

Appendix A4 Naive Recalls

In this appendix, we first present the recalls when cued by the external features X. We standardize the external features as described in Section 3.4 and refer to these recalls as naive recalls. At each point in time x (shown in the columns), we calculate the top 5 historical episodes most likely to be retrieved for each firm in our sample, based on the similarity function (4.1). Figure A3 displays the number of historical episodes from time y (shown in the rows) that are retrieved when analyzing all firms at current time x.

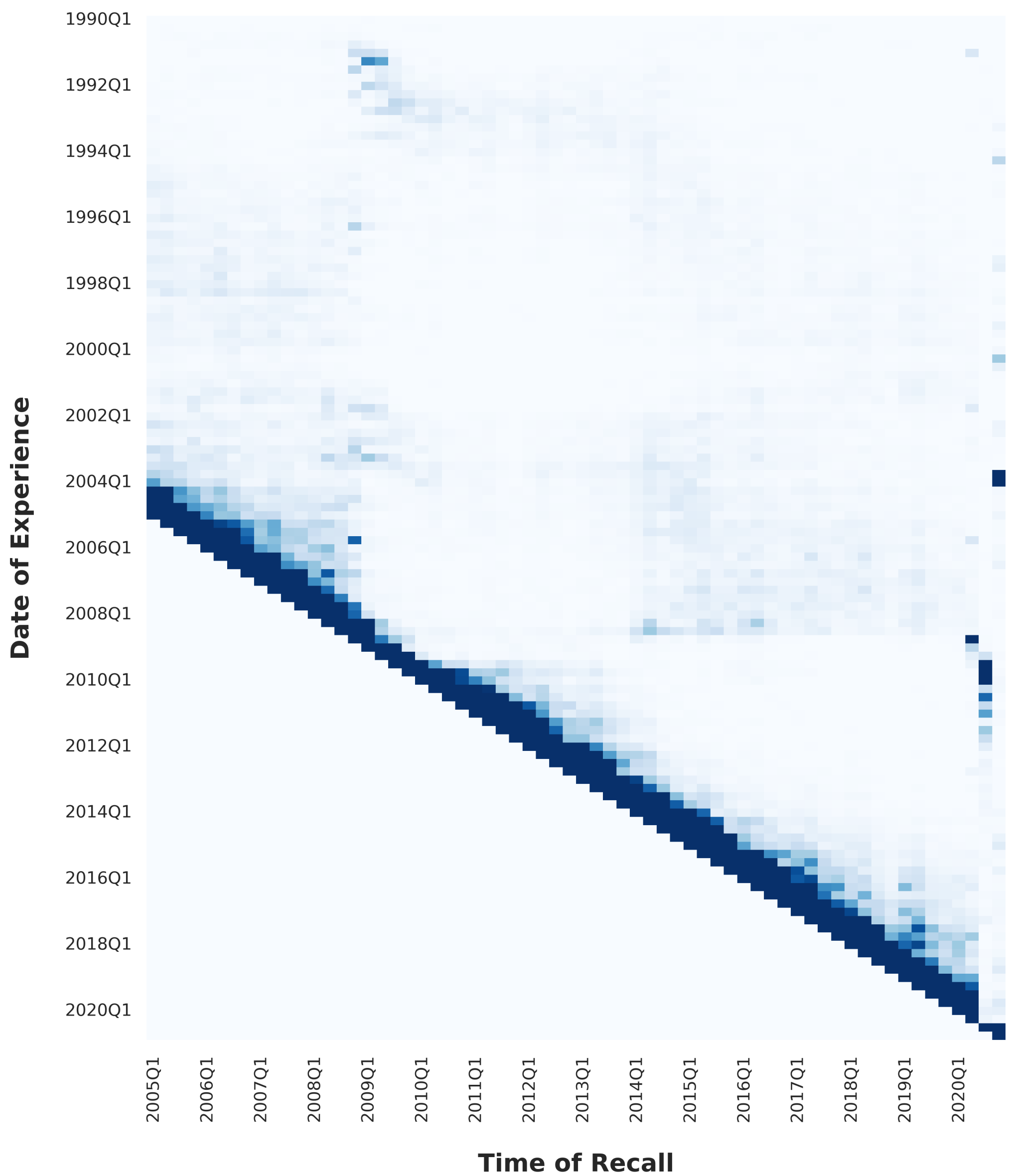


Figure A3: The naive recalls that are cued by external features

Appendix A5 Selective Forgetting

In this appendix, we examine the economic consequences of selective forgetting by comparing the predictive power for realized EPS revisions between two sets of LSTM models: one with the forget gate and one without.

Table A3: Realized EPS revision and recall revisions

	$\Delta RE_{i,t+l}$			
	(1)	(2)	(3)	(4)
Benchmark recalled revisions w/ forget $RR^{\mathrm{B}}_{i,t}$	0.284***			
	(0.081)			
Benchmark recalled revisions w/o forget $RR^{\mathrm{B/F}}_{i,t}$		0.319***		
		(0.065)		
Analyst recalled revisions w/ forget $RR^{\mathrm{A}}_{i,t}$			0.110***	
			(0.017)	
Analyst recalled revisions w/o forget $RR^{\mathrm{A/F}}_{i,t}$				0.203***
				(0.024)
Firm fixed effect	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Observations	277,410	277,410	277,410	277,410
R-squared	0.239	0.233	0.262	0.267
Within R-squared	0.019	0.011	0.011	0.019

This table presents results for regressions of the form

$$\Delta RE_{i,t+l} = \beta \times RR_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t \varepsilon_{i,t}.$$

The dependent variable $\Delta RE_{i,t+l}$ is firm i's realized EPS revision which is available in month $t+l$, i.e., $\Delta RE_{i,t+l} = \frac{EPS_{i,t+l} - F_{i,t-1}[EPS_{i,t+l}]}{P_{i,t-1}}$. Independent variables $RR_{i,t}$ denotes the recalled revisions, as defined in (4.4). We analyze four specifications of $RR_{i,t}$, $RR^B_{i,t}$, $R^{B/F}_{i,t}$, $RR^A_{i,t}$, and $RR^{A/F}_{i,t}$. $RR^B_{i,t}$ denotes the benchmark LSTM recalled revisions, $RR^{B/F}_{i,t}$ denotes the benchmark recalled beliefs generated by the counterfactual LSTM model without the forget gate, and $RR^A_{i,t}$ and $RR^{A/F}_{i,t}$ are the corresponding recalled revisions for analysts. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

We first assess the predictive power of the benchmark memory models, with and without the forget gate. Table A3 presents the results. Comparing the within R^2 in Columns (1) and (2), we find that recalled revisions generated by the benchmark model with the forget gate have greater predictive power than those from the counterfactual model without the forget gate, although both coefficients are highly statistically significant. This suggests that removing the forget gate impairs the belief formation process in the benchmark model and that the benchmark effectively uses the forget gate to selectively disregard irrelevant episodes in memory.

However, this result does not hold for the analyst LSTM models. Comparing the within R^2 in Columns (3) and (4) of Table A3, we find that recalled revisions without selective forgetting outperform those with selective forgetting in predicting realized EPS. This suggests that selective forgetting can be detrimental, rather than beneficial, to analysts in their belief formation processes.

For both analysts and the benchmark, selective forgetting is essential due to limited memory capacity; otherwise, memory would be overwhelmed and past information would be lost. Ideally, selective forgetting should remove unnecessary information while retaining useful experiences. However, our findings indicate that while selective forgetting benefits the benchmark model, it does not operate optimally for analysts. These results suggest that selective forgetting distorts analyst memory and recalls, moving them further from the optimal benchmark.

Appendix A6 Robustness Checks

In the main analyses, we limit the memory database size $M_{i,t}$, denoted as N, to the five most similar past experiences. In this appendix, we provide robustness checks for different choices of N as well as additional necessary robustness checks.

Table A4 shows that recalled revisions positively predict forecast revisions without fixed effects and across different choices of N.

Table A5 illustrates that the dispersion of recalled revisions positively predicts analyst disagreement and abnormal trading volume, excluding crisis periods and across different choices of N.

Table A6 confirms that recall distortion positively predicts analyst forecast errors using different choices of N.

Table A7 tests the robustness of Table 9 using alternative measures of misreactions and recency effects, while Table A8 provides robustness checks for Table 10 with alternative measures of recency effects and recalled revisions.

Table A9 demonstrates that the returns on the long-short strategy based on analyst recalled revisions RR^{A} cannot be explained by leading asset pricing models.

Table A10 reports that the returns on the long-short strategy based on analyst recalled revisions RR^{A} remain effective across different sample periods and choices of N.

Table A11 presents the returns of double-sorted strategies on biased memory-based attention and prior month stock returns, verifying its validity across different sample periods and choices of N.

Figure A4 shows that the cumulative returns on the long-short portfolio sorted by recall distortion are robust across varying sample periods and choices of N.

Figure A5 reports cumulative returns on the double-sorted portfolio based on biased memory-based attention and prior month returns, consistent across different sample periods and choices of N.

Figure A6 displays analyst recalls and their comparison to benchmark recalls across different values of N.

Overall, these robustness checks confirm that our main results hold consistently across varying values of N, time periods, model specifications, and measures.

Table A4: Analyst forecast revisions and recalled revisions - robustness check

	Forecast Revisions ($\Delta F_{i,t}$)					
	Without Fixed Effects			Different Number of Recalls		
				N=1	N=3	N=10
	(1)	(2)	(3)	(4)	(5)	(6)
LSTM recalled revision RR^{A}	0.078***		0.073***	0.017***	0.043***	0.093***
	(0.019)		(0.014)	(0.005)	(0.010)	(0.014)
Naive recalled revision RR^{N}		0.028*	0.007	0.012	0.004	-0.003
		(0.016)	(0.010)	(0.008)	(0.010)	(0.012)
Lagged forecast revision			0.154***	0.132***	0.130***	0.121**
			(0.051)	(0.042)	(0.043)	(0.043)
Lagged earnings growth			0.011**	0.013**	0.013**	0.013**
			(0.005)	(0.005)	(0.005)	(0.005)
Lagged forecast			-2.328***	-4.611***	-4.648***	-4.730***
			(0.615)	(0.746)	(0.750)	(0.772)
Firm fixed effect	No	No	No	Yes	Yes	Yes
Month fixed effects	No	No	No	Yes	Yes	Yes
Observations	277,487	277,487	277,487	277,410	277,410	277,410
R-squared	0.012	0.002	0.025	0.067	0.069	0.073
Within R-squared	0.012	0.002	0.025	0.021	0.023	0.028

This table presents the robustness check results for Table 2 without fixed effects in Columns (1)-(3) and using different number of the most similar past experiences (N) in Columns (4)-(6). The regressions follow the form

$$\Delta F_{i,t} = \beta \times RR_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t},$$

where $\Delta F_{i,t} = (F_{i,t}[EPS_{i,t+l}] - F_{i,t-1}[EPS_{i,t+l}])/P_{i,t-1}$ denotes analysts forecast revisions scaled by the stock price $P_{i,t-1}$ and $RR_{i,t}$ denotes the recalled revisions. We analyze two specifications of $RR_{i,t}$, $RR^A_{i,t}$ and $RR^N_{i,t}$. $RR^A_{i,t}$ denotes the analyst LSTM recalled revisions when cued by the mental context c that is estimated from LSTM $RR^A_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(c_{j,\tau}) * rr_{j,\tau}/P_{j,\tau-1}$, with the probability distribution is proportional to the similarity function (2.1). $RR^N_{i,t}$ denotes the naive recalled revisions when cued by the external features X, $RR^N_{i,t} = \sum_{(j,\tau)\in M_{i,t}} p(X_{j,\tau}) * rr_{j,\tau}/P_{j,\tau-1}$, with the probability distribution is proportional to the similarity function (4.1). Columns (1) to (2) report the results without control variables while Columns (3) to (6) report the results with control variables: lagged forecast revision $(F_{i,t-1}[EPS_{i,t+l}] - F_{i,t-2}[EPS_{i,t+l}])/P_{i,t-1}$, lagged earnings growth (the difference between the realized earnings for the last fiscal year and the realized earnings for the fiscal year before last, scaled by stock price $P_{i,t-1}$), and lagged forecast $(F_{i,t-1}[EPS_{i,t+l}]/P_{i,t-1})$. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

Table A5: Dispersion of recalled revisions, disagreement, and trading volume - robustness check

	Disagreement$_{i,t}$			Abnormal volume AbVol$_{i,t}$		
	Without Crisis	N=3	N=10	Without Crisis	N=3	N=10
	(1)	(2)	(3)	(4)	(5)	(6)
$\sigma(rr)_{i,t}$	0.005*	0.008***	0.009***	0.057***	0.036**	0.046***
	(0.003)	(0.002)	(0.002)	(0.018)	(0.013)	(0.012)
$\sigma(rr)_{i,t-1}$	0.005**	0.009***	0.006**	-0.024	0.003	0.010
	(0.002)	(0.002)	(0.002)	(0.021)	(0.019)	(0.013)
Disagreement$_{i,t-1}$	0.474***	0.540***	0.539***			
	(0.019)	(0.017)	(0.017)			
AbVol$_{i,t-1}$				0.521***	0.539***	0.539***
				(0.008)	(0.010)	(0.010)
$Ret_{i,t-1}$	-0.009***	-0.012***	-0.012***	-0.096**	-0.086**	-0.086**
	(0.001)	(0.002)	(0.002)	(0.040)	(0.035)	(0.035)
$Ret_{i,t-2}$	-0.003***	-0.004***	-0.003***	-0.003	0.004	0.005
	(0.001)	(0.001)	(0.001)	(0.020)	(0.019)	(0.020)
$Ret_{i,t-3}$	-0.004***	-0.004***	-0.004***	0.043	0.055**	0.055**
	(0.001)	(0.001)	(0.001)	(0.024)	(0.025)	(0.025)
$\sigma(Ret)_{i,t-1}$	0.037***	0.039***	0.038***	-0.707***	-0.574***	-0.576***
	(0.005)	(0.006)	(0.006)	(0.107)	(0.110)	(0.110)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	46,461	59,160	59,160	158,424	194,444	194,444
R-squared	0.593	0.608	0.609	0.384	0.412	0.412

This table presents the robustness check results for Table 4 using different number of the most similar past experiences. Columns (1) to (3) present the regression results of disagreement and columns (4) to (6) present the regression results of trading volume. Columns (1) and (4) present the regression results excluding the crisis periods, specifically the years 2007, 2008, and 2020. Columns (2) and (5) present the regression results using three most similar past experiences ($N = 3$). Columns (3) and (6) present the regression results using ten most similar past experiences ($N = 10$). The regressions follow the form

$$\text{Disagreement}_{i,t} = \beta \times \sigma(rr)_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t},$$

and

$$\text{AbVol}_{i,t} = \beta \times \sigma(rr)_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t + \varepsilon_{i,t},$$

where $\sigma(rr)_{i,t}$ is the weighted standard deviation of recalled revisions for firm i in month t as shown in Equation (4.8); Disagreement$_{i,t}$ measures disagreement, following Diether, Malloy and Scherbina (2002) it is the analyst forecast dispersion - the standard deviation of each individual analyst earnings forecast scaled by the last period's stock price; AbVol is the abnormal log trading volume for firm i in month t, following Cookson and Niessner (2020), it is the difference between the log volume in month t and the average log volume from month $t-12$ to $t-2$. Recalls are found based on the similarity measure shown in (2.1). $Ret_{i,t-1}$, $Ret_{i,t-2}$, $Ret_{i,t-3}$ are the lagged monthly stock return for firm i. $\sigma(Ret)_{i,t-1}$ is the standard deviation of stock return for firm i within a 12-month rolling window $t-11$ to t. The sample period is from January 2007 to December 2020. The reported recall dispersion coefficients are presented as the true values multiplied by 1000 for display convenience. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

Table A6: The analyst forecast errors and recall distortion - robustness check

	Forecast error $e_{i,t}$		Forecast error $\lvert e_{i,t}\rvert$		Forecast direction error $e^d_{i,t}$	
	N=3	N=10	N=3	N=10	N=3	N=10
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear	Linear	Linear	Linear	Ologit	Ologit
$\Delta R_{i,t}$	0.033***	0.038***			0.018***	0.027***
	(0.008)	(0.007)			(0.003)	(0.004)
$\lvert\Delta R_{i,t}\rvert$			0.095***	0.095***		
			(0.023)	(0.019)		
Lagged forecast revision	-0.008	-0.007	-0.032**	-0.032**	0.020**	0.020**
	(0.019)	(0.019)	(0.014)	(0.014)	(0.008)	(0.009)
Lagged earnings growth	0.004	0.004	-0.010**	-0.010**	0.002	0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)
Lagged forecast	1.252**	1.251**	-2.679***	-2.677***	0.390*	0.383*
	(0.499)	(0.502)	(0.540)	(0.552)	(0.225)	(0.226)
Firm fixed effect	Yes	Yes	Yes	Yes	No	No
Month fixed effects	Yes	Yes	Yes	Yes	No	No
Observations	277,410	277,410	277,410	277,410	277,487	277,487
R-squared	0.145	0.145	0.302	0.300		

This table presents the robustness check results for Table 5 using different number of the most similar past experiences. Columns (1), (3), and (5) present the regression results using three most similar past experiences ($N = 3$). Columns (2), (4), and (6) present the regression results using ten most similar past experiences ($N = 10$). The regressions follow the form

$$e_{i,t} = \beta \times \Delta R_{i,t} + \theta \times Z_{i,t} + \gamma_i + \eta_t \varepsilon_{i,t}$$

in columns (1) - (4) and the form

$$\Pr(e^d_{i,t} = j) = \Phi\left(\kappa_j - \beta \times \Delta R_{i,t}\right) - \Phi\left(\kappa_{j-1} - \beta \times \Delta R_{i,t}\right)$$

in columns (5)-(6). The dependent variable $e_{i,t}$ is analyst forecast error, and $e^d_{i,t} \in \{-2, -1, 0, 1, 2\}$ is the direction of analyst forecast error for firm i at time t. The independent variable $\Delta R_{i,t}$ is analyst recall distortion which is predicted by the model with all the information available before time t as defined in Equation (5.1). We also include three control variables: lagged forecast revision $(F_{i,t-1}[EPS_{i,t+l}] - F_{i,t-2}[EPS_{i,t+l}])/P_{i,t-1}$, lagged earnings growth (the difference between the realized earnings for the last fiscal year and the realized earnings for the fiscal year before last, scaled by stock price $P_{i,t-1}$), and lagged forecast $(F_{i,t-1}[EPS_{i,t+l}]/P_{i,t-1})$. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level, and reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

Table A7: Misreaction in analyst perceptions, forecasts and recency effect - robustness check

	Misreaction		Biased Recency Effect			
			24 months		36 months	
	$\Delta React_t$ incl. sign($\Delta F_{i,t}$) $\neq$ sign($\Delta RE_{i,t}$)	$\Delta React_t$ excl. $\Delta F_{i,t} = 0$	Normal	Crisis	Normal	Crisis
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta^2\omega$	-0.022***	-0.020***	0.021**	-0.019***	0.020*	-0.019***
	(0.006)	(0.006)	(0.010)	(0.004)	(0.010)	(0.004)
Observations	180	180	144	36	144	36
R-squared	0.099	0.163	0.033	0.251	0.024	0.295

This table presents the robustness check results for Table 9 with alternative measures of misreactions and biased recency effects. The independent variable, $\Delta^2\omega_t$, is defined in Equation (5.7). The dependent variables, misreaction $\Delta React_t$ and biased recency effect $\Delta Recency_t$ are similar to the definition in Equations (5.8), (5.10), respectively. We examine two variations of $\Delta React_t$: including the cases when sign($\Delta F_{i,t}$) $\neq$ sign($\Delta RE_{i,t}$) in Column (1), and excluding the cases when $\Delta F_{i,t} = 0$ in Column (2). Columns (3) and (4) measure recency effects by analysts' recalls falling within the last 24 months. Columns (5) and (6) measure recency effects by analysts' recalls falling within the last 36 months. The sample period is from January 2007 to December 2020, with 2008, 2009, and 2020 are classified as crisis periods, while the remaining years are considered normal periods. The reported t-statistics are robust to heteroskedasticity.

Table A8: Paired t-tests between LSTM models with and without selective forgetting - robustness check

Model	Recency Effect		Recalled Revisions		
	24 months	36 months	N=1	N=3	N=10
	(1)	(2)	(3)	(4)	(5)
Panel A: Analysts					
LSTM with forget gate	0.292	0.360	-0.687	-0.706	-0.708
LSTM without forget gate	0.446	0.520	-0.548	-0.604	-0.647
Difference	0.153	0.160	0.139	0.102	0.061
t-statistics	(8.66)	(9.28)	(8.36)	(7.56)	(5.14)
Panel B: Benchmark					
LSTM with forget gate	0.585	0.636	-0.330	-0.403	-0.467
LSTM without forget gate	0.570	0.620	-0.412	-0.460	-0.515
Difference	-0.015	-0.017	-0.081	-0.057	-0.048
t-statistics	(-1.72)	(-1.81)	(-5.56)	(-3.91)	(-4.06)

This table presents the robustness check results for Table 10 with alternative measures of recency effects and recalled revisions. Columns (1) and (2) present the paired t-tests of recency effects ($Recency$ (5.9)) for 24 motnhs and 36 months. Columns (3) to (5) present the t-tests of recalled revisions (RR (4.4)) using different numbers of the most similar past experiences (N). Panel A reports the results for the analyst recalls between the full LSTM model and the LSTM model without the forget gate (counterfactual) and panel B reports the results for the benchmark recalls. The sample period is from January 2007 to December 2020. Standard errors are clustered at both the industry and year level.

Table A9: Time-series tests with common asset pricing models

	CAPM		FF3		Carhart	
	Coef	t-stat	Coef	t-stat	Coef	t-stat
Intercept	0.55	2.97	0.41	2.55	0.41	3.01
MKT-R_f	-0.16	-2.56	-0.10	-1.89	-0.05	-0.84
SMB			-0.04	-0.36	-0.01	-0.11
HML			-0.24	-3.28	-0.10	-1.88
MOM					0.21	2.89

This table presents the regression of returns (in percent) on the long-short portfolio sorted with the analyst recalled revisions (RR^A), on the CAPM, the Fama-French three-factor model (FF3) (Fama and French, 1993), and the Carhart four-factor model Carhart (1997). The sample period is from 2007 to 2020. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags.

Table A10: Performance of long-short portfolios sorted on the analyst recalled revisions

	CAPM	FF3	Carhart
Panel A: Different sample periods			
2007 - 2019	0.48	0.38	0.36
	(2.91)	(2.37)	(3.09)
2010 - 2019	0.49	0.44	0.35
	(2.81)	(2.81)	(2.09)
2010 - 2020	0.58	0.50	0.47
	(3.65)	(3.30)	(2.76)
Panel B: Different number of past experiences in M			
$N = 1$	0.42	0.29	0.29
	(3.10)	(2.03)	(2.54)
$N = 3$	0.55	0.42	0.42
	(3.35)	(2.67)	(3.44)
$N = 10$	0.49	0.36	0.37
	(2.44)	(1.94)	(2.79)

This table reports the robustness check for Table 3 under different sample periods (Panel A) and using different number of the most similar past experiences (N) in memory database M to form the analyst recalled revisions (RR^A) (Panel B). The table shows the time-series average of risk-adjusted returns on value-weighted long-short portfolios formed on the quintiles of the analyst recalled revisions (RR^A). The risk-adjusted returns are based on the CAPM, the Fama-French three-factor model (FF3) (Fama and French, 1993), and the Carhart four-factor model Carhart (1997). The returns are in percentage. For Panel B, the sample period is from January 2007 to December 2020. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. The t-statistics are reported in parentheses.

Table A11: Performance of short-term reversal strategies in different groups of biased investor memory-based attention

Attention	1	2	3	3-1
Panel A: Different sample periods				
2007 - 2019	-0.59	0.36	0.55	1.14
	(-1.74)	(2.69)	(2.38)	(3.78)
2010 - 2019	-0.51	0.35	0.60	1.11
	(-1.35)	(2.22)	(2.93)	(2.88)
2010 - 2020	-0.53	0.32	0.53	1.07
	(-1.52)	(2.06)	(2.51)	(2.97)
Panel B: Different number of past experiences in M				
$N = 3$	-0.53	0.41	0.40	0.93
	(-1.13)	(2.64)	(1.74)	(2.16)
$N = 10$	-0.35	0.38	0.43	0.77
	(-1.12)	(2.65)	(1.80)	(2.39)

This table reports the robustness check for Table 8 under different sample periods (Panel A) and using different number of the most similar past experiences (N) in memory database M to form the biased investor memory-based attention Δa (Panel B). The table shows the performance of short-term reversal strategies in different tercile groups of the biased investor memory-based attention Δa. The risk-adjusted returns are based on the Carhart four-factor model Carhart (1997). The returns are in percentage. For Panel B, the sample period is from January 2007 to December 2020. Standard errors are adjusted for heteroskedasticity and autocorrelations up to 12 lags. The t-statistics are reported in parentheses.

Figure A4: The cumulative returns on long-short portfolio sorted on recall distortion - robustness check

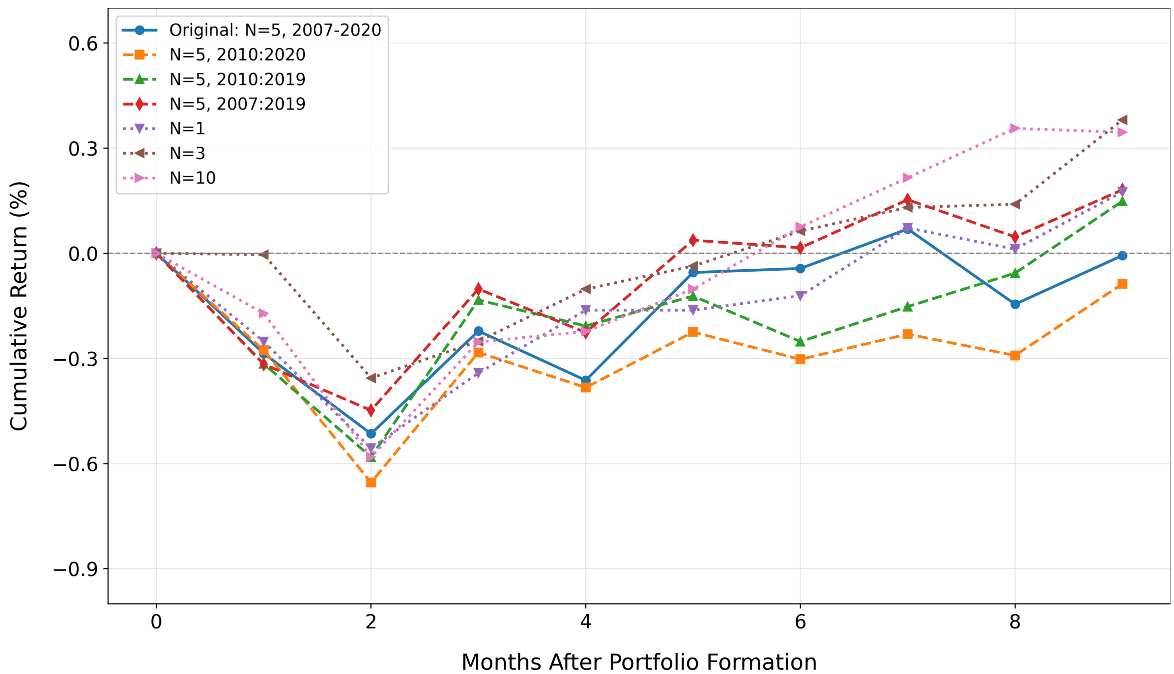


Figure A5: The cumulative returns on double-sorted long-short portfolio on biased investor memory-based attention and short-term reversal - robustness check

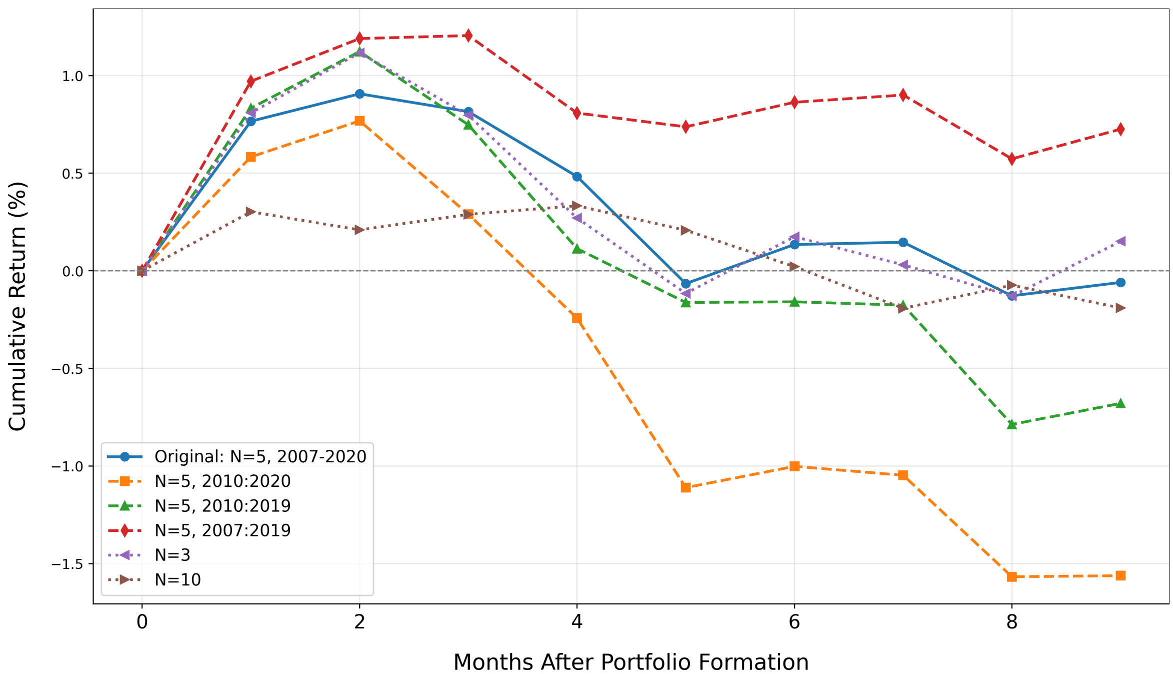
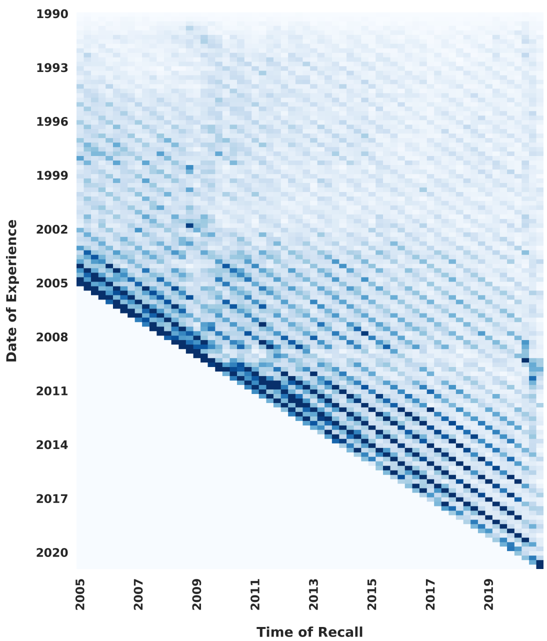
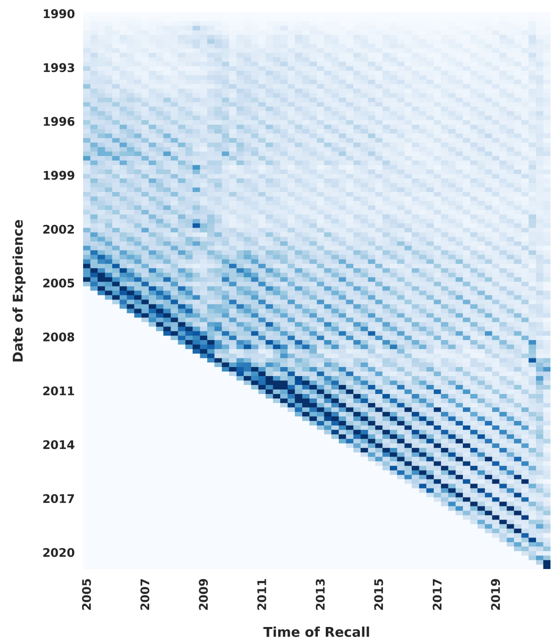


Figure A6: Analyst recalls and comparison with benchmark recalls with $N = 1$ and $N = 10$

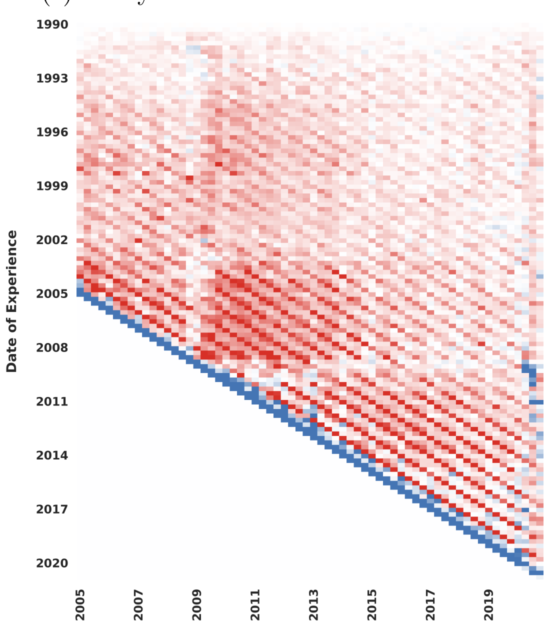
(a) Analyst recalls $N = 1$



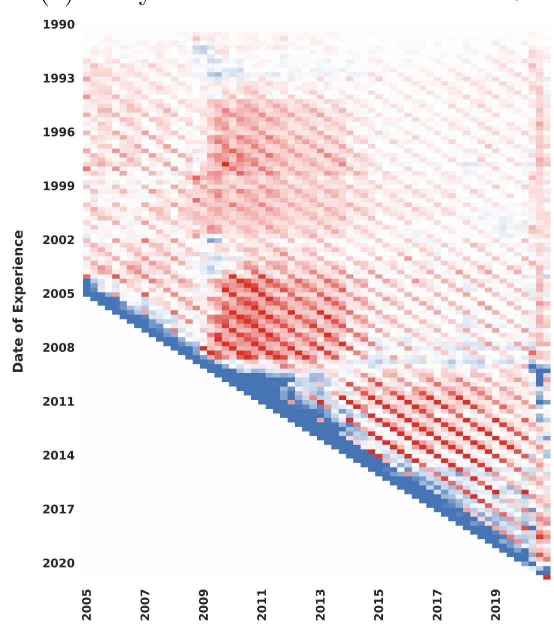
(b) Analyst recalls $N = 10$



(c) Analyst vs Benchmark recalls $N = 1$



(d) Analyst vs Benchmark recalls $N = 10$



References

Atmaz, Adem, and Suleyman Basak. 2018. "Belief dispersion in the stock market." *The Journal of Finance*, 73(3): 1225–1279.

Barberis, Nicholas. 2003. "A Survey of Behavioral Finance." *Handbook of the Economics of Finance*, 1.

Barberis, Nicholas. 2018. "Psychology-based models of asset prices and trading volume." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 1, 79–175. Elsevier.

Barberis, Nicholas, and Lawrence Jin. 2023. "Model-free and Model-based Learning as Joint Drivers of Investor Behavior." *Working paper*.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A model of investor sentiment." *Journal of Financial Economics*, 49(3): 307–343.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015. "X-CAPM: An extrapolative capital asset pricing model." *Journal of financial economics*, 115(1): 1–24.

Bastianello, Francesca, Paul H Décaire, and Marius Guenzel. 2025. "Mental models and financial forecasts." *Working paper*.

Bernanke, Ben S. 2015. *The Courage to Act.* New York:W.W.Norton & Company.

Bianchi, Francesco, Sydney C Ludvigson, and Sai Ma. 2022. "Belief distortions and macroeconomic fluctuations." *American Economic Review*, 112(7): 2269–2315.

Bordalo, Pedro, Giovanni Burro, Katherine B Coffman, Nicola Gennaioli, and Andrei Shleifer. 2022. "Imagining the future: memory, simulation and beliefs about COVID." *Working Paper*.

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer. 2023. "Memory and probability." *The Quarterly Journal of Economics*, 138(1): 265–311.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, Attention, and Choice." *The Quarterly Journal of Economics*, 135(3): 1399–1442.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2019. "Diagnostic expectations and stock returns." *Journal of Finance*, 74(6): 2839–2874.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2024. "Belief overreaction and stock market puzzles." *Journal of Political Economy*, 132(5): 1450–1484.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar. 2019. "Sticky expectations and the profitability anomaly." *The Journal of Finance*, 74(2): 639–674.

Brunnermeier, Markus, Emmanuel Farhi, Ralph SJ Koijen, Arvind Krishnamurthy, Sydney C Ludvigson, Hanno Lustig, Stefan Nagel, and Monika Piazzesi. 2021. "Perspectives on the Future of Asset Pricing." *The Review of Financial Studies*, 34(4): 2126–2160.

Carhart, Mark M. 1997. "On persistence in mutual fund performance." *The Journal of finance*, 52(1): 57–82.

Charles, Constantin. 2024. "Memory Moves Markets." *Working Paper.*

Charles, Constantin, and Pengfei Sui. 2024. "Marketwide Memory." *Working paper.*

Chen, Luyang, Markus Pelger, and Jason Zhu. 2023. "Deep learning in asset pricing." *Management Science.*

Chen, Yifei, Bryan T Kelly, and Dacheng Xiu. 2024. "Expected returns and large language models." *Working paper.*

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 105(8): 2644–2678.

Cookson, J Anthony, and Marina Niessner. 2020. "Why don't we agree? Evidence from a social network of investors." *The Journal of Finance*, 75(1): 173–228.

Cui, Tingyue, Ricardo De la O, and Sean Myers. 2024. "The Subjective Belief Factor." *Ricardo and Myers, Sean, The Subjective Belief Factor (August 23, 2024).*

Da, Zhi, Qianqiu Liu, and Ernst Schaumburg. 2014. "A closer look at the short-term return reversal." *Management Science*, 60(3): 658–674.

De la O, Ricardo, Xiao Han, and Sean Myers. 2023. "The Cross-section of Subjective Expectations: Understanding Prices and Anomalies." *Working paper.*

Diether, Karl B, Christopher J Malloy, and Anna Scherbina. 2002. "Differences of opinion and the cross section of stock returns." *Journal of Finance*, 57(5): 2113–2141.

Drugowitsch, Jan, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin. 2016. "Computational precision of mental inference as critical source of human choice suboptimality." *Neuron*, 92(6): 1398–1411.

Dubey, Shiv Ram, Satish Kumar Singh, and Bidyut Baran Chaudhuri. 2022. "Activation functions in deep learning: A comprehensive survey and benchmark." *Neurocomputing*, 503: 92–108.

Earhard, Marcia. 1967. "Cued recall and free recall as a function of the number of items per cue." *Journal of Verbal Learning and Verbal Behavior*, 6(2): 257–263.

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann. 2024. "Associative memory, beliefs and market interactions." *Journal of Financial Economics*, 157: 103853.

Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics*, 33(1): 3–56.

Frydman, Cary, and Lawrence J Jin. 2022. "Efficient coding and risky choice." *The Quarterly Journal of Economics*, 137(1): 161–213.

Frydman, Cary, and Salvatore Nunnari. 2021. "Cognitive imprecision and strategic behavior." *Working paper.*

Gabaix, Xavier. 2019. "Behavioral inattention." In *Handbook of Behavioral Economics: Applications and Foundations 1.* Vol. 2, 261–343. Elsevier.

Gödker, Katrin, Peiran Jiao, and Paul Smeets. 2022. "Investor memory." *Working paper.*

Graeber, Thomas, Christopher Roth, and Florian Zimmermann. 2024. "Stories, statistics, and memory." *The Quarterly Journal of Economics*, qjae020.

Graves, Alex, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. 2016. "Hybrid computing using a neural network with dynamic external memory." *Nature*, 538(7626): 471–476.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical asset pricing via machine learning." *The Review of Financial Studies*, 33(5): 2223–2273.

Healey, M Karl, and Michael J Kahana. 2014. "Is memory search governed by universal principles or idiosyncratic strategies?" *Journal of Experimental Psychology: General*, 143(2): 575.

Healey, M Karl, Nicole M Long, and Michael J Kahana. 2019. "Contiguity in episodic memory." *Psychonomic bulletin & review*, 26(3): 699–720.

Hirshleifer, David, Yaron Levi, Ben Lourie, and Siew Hong Teoh. 2019. "Decision fatigue and heuristic analyst forecasts." *Journal of Financial Economics*, 133(1): 83–98.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long short-term memory." *Neural computation*, 9(8): 1735–1780.

Hong, Harrison, and Jeremy C Stein. 2007. "Disagreement and the stock market." *Journal of Economic Perspectives*, 21(2): 109–128.

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

Jagtap, Ameya D, Kenji Kawaguchi, and George Em Karniadakis. 2020. “Adaptive activation functions accelerate convergence in deep and physics-informed neural networks.” *Journal of Computational Physics*, 404: 109136.

Jegadeesh, Narasimhan. 1990. “Evidence of predictable behavior of security returns.” *The Journal of Finance*, 45(3): 881–898.

Jiang, Zhengyang, Hongqi Liu, Cameron Peng, and Hongjun Yan. 2023. “Investor Memory and Biased Beliefs: Evidence from the Field.” *Working paper.*

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. “A rational theory of mutual funds’ attention allocation.” *Econometrica*, 84(2): 571–626.

Kahana, Michael J. 1996. “Associative retrieval processes in free recall.” *Memory & cognition*, 24(1): 103–109.

Kahana, Michael Jacob. 2012. *Foundations of human memory.* OUP USA.

Kelly, Bryan T, Seth Pruitt, and Yinan Su. 2019. “Characteristics are covariances: A unified model of risk and return.” *Journal of Financial Economics*, 134(3): 501–524.

Ke, Shikun. 2025. “Analysts’ Belief Formation in Their Own Words.” *Working paper.*

Kingma, Diederik P, and Jimmy Ba. 2014. “Adam: A method for stochastic optimization.” *arXiv preprint arXiv:1412.6980.*

Kwan, Alan, Yukun Liu, and Ben Matthies. 2022. “Institutional Investor Attention.” *Working Paper.*

Landier, Augustin, and David Thesmar. 2020. “Earnings expectations during the COVID-19 crisis.” *The Review of Asset Pricing Studies*, 10(4): 598–617.

Lehmann, Bruce N. 1990. “Fads, martingales, and market efficiency.” *The Quarterly Journal of Economics*, 105(1): 1–28.

Liao, Jingchi, Cameron Peng, and Ning Zhu. 2022. “Extrapolative bubbles and trading volume.” *The The Review of Financial Studies*, 35(4): 1682–1722.

Lopez-Lira, Alejandro, and Yuehua Tang. 2023. “Can chatgpt forecast stock price movements? return predictability and large language models.” *Working paper.*

Lord, Charles G, Lee Ross, and Mark R Lepper. 1979. “Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.” *Journal of Personality and Social Psychology*, 37(11): 2098.

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

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

Miller, George A. 1956. "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review*, 63(2): 81.

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

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives*, 31(2): 87–106.

Nagel, Stefan. 2021. *Machine learning in asset pricing.* Princeton University Press.

Nagel, Stefan, and Zhengyang Xu. 2022. "Asset pricing with fading memory." *The Review of Financial Studies*, 35(5): 2190–2245.

Roberts, William A. 1972. "Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses." *Journal of Experimental Psychology*, 92(3): 365.

Shalen, Catherine T. 1993. "Volume, volatility, and the dispersion of beliefs." *The The Review of Financial Studies*, 6(2): 405–434.

Teräsvirta, Timo. 2006. "Forecasting economic variables with nonlinear models." *Handbook of economic forecasting*, 1: 413–457.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *science*, 185(4157): 1124–1131.

van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira. 2023. "Man versus machine learning: The term structure of earnings expectations and conditional biases." *The Review of Financial Studies*, 36(6): 2361–2396.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." *Advances in Neural Information Processing Systems*, 30.

Wachter, Jessica A, and Michael Jacob Kahana. 2024. "A retrieved-context theory of financial decisions." *The Quarterly Journal of Economics*, 139(2): 1095–1147.

Walters, Daniel J, and Philip M Fernbach. 2021. "Investor memory of past performance is positively biased and predicts overconfidence." *Proceedings of the National Academy of Sciences*, 118(36): e2026680118.

Weston, Jason, Sumit Chopra, and Antoine Bordes. 2014. "Memory networks." *arXiv preprint arXiv:1410.3916.*

Woodford, Michael. 2020. "Modeling imprecision in perception, valuation, and choice." *Annual Review of Economics*, 12: 579–601.