ISSN 0956-8549

Marketwide Memory

By

Constantin Charles

Pengfei Sui

FINANCIAL MARKETS GROUP DISCUSSION PAPER NO. 931

June 2025

Marketwide Memory *

Constantin Charles[†]

Pengfei Sui[‡]

This version: October 25, 2024

Abstract

We propose a novel measure that allows us to study memory associations in financial markets over the course of several decades. Using our measure, we construct memory-based beliefs and show that they can explain return expectations from surveys as well as higher-moment beliefs implied by stock options and the VIX. We also show that memory associations drive trading decisions of individual investors: when investors are more likely to recall a past positive (negative) trading experience with a stock, they are more (less) likely to repurchase that stock. Our measure builds on two well-established regularities of associative recall: similarity and interference. For each point in time, it captures the probability that a representative investor recalls past episodes of a stock. Without any further assumptions, our measure generates signature patterns of cued recall, such as the recency effect and the recall of past crises during extreme episodes. We validate our measure using actual recall patterns extracted from transcripts of corporate events, like earnings calls, and show that our measure predicts which historical periods are mentioned during these events. Overall, our results show that theories of human memory can be broadly applied in financial markets.

*We thank Kai Barron, Stefano Cassella, Cary Frydman, Lawrence Jin, Eben Lazarus, Cameron Peng, Jun Pan, Cameron Peng, Tim Pleskac, Rui Shen, Johannes Wohlfart, Zhengyang Xu and Hongjun Yan, as well as participants at the Chicago Booth Conference in Behavioral Finance and Decision-Making, Stockholm School of Economics, Fudan University, and the University of Zurich for helpful comments. We thank Terrance Odean for providing the trading data.

†London School of Economics. Contact: c.charles@lse.ac.uk

‡The Chinese University of Hong Kong, Shenzhen. Contact: psui@cuhk.edu.cn

1 Introduction

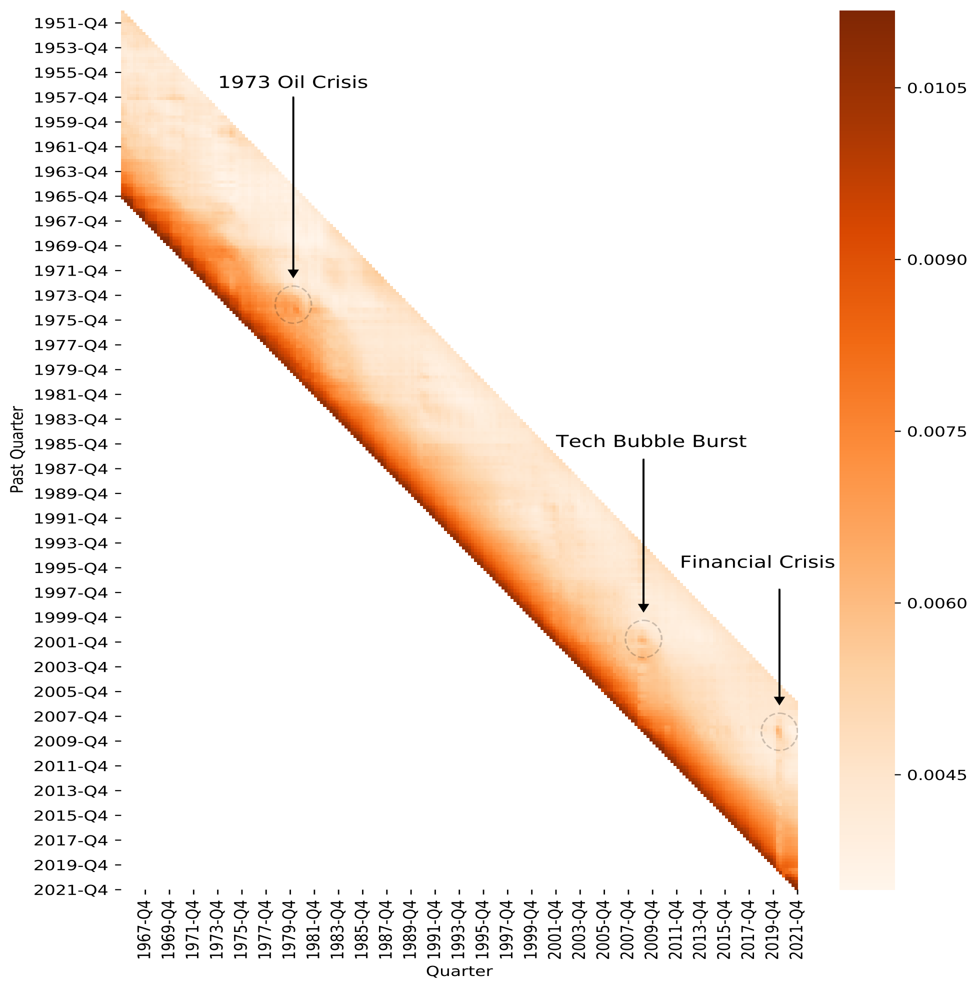

Figure 1. This figure displays the recall probability that our measure assigns to the past 60 quarters (= 15 years) for each quarter from 1966 to 2021. The x-axis indicates the quarter in which recall occurs, while the y-axis indicates the probability weight assigned to past quarters. A darker shade of red indicates a higher probability weight.

In recent years, economic theorists have argued that incorporating aspects of human memory into economic models may be useful in explaining a variety of empirical phenomena, including experience effects à la Malmendier and Nagel (2011), belief over- and underreaction, and return extrapolation (Bordalo et al. (2020); Bordalo et al. (2023); Wachter and Kahana (2023)).

Several studies aimed at testing these models have established links between memory and investor beliefs, investor behavior, and asset prices. These studies employ a range of different techniques, including experimental methods (Enke et al. (2024); Gödker et al. (2024); Graeber et al. (2024)), survey-based approaches (Andre et al. (2022); Jiang et al. (2023)), and the use of unique empirical settings (Charles (2022); Charles (2024)).

In this paper, we build on the links established in these studies, but take a more comprehensive view of memory in financial markets. We estimate the memory associations that a representative investor *would* have if experimentally well-established regularities in recall also operate in the field. By making this assumption, our approach allows us to study memory at a much larger scale (in the cross-section of listed stocks) and for much longer time periods (multiple decades) than prior studies. We find that our estimated memory associations (i) match actual recall patterns extracted from transcripts of corporate events, (ii) explain return expectations from investor surveys as well as higher moment beliefs implied by stock options and the VIX, and (iii) drive trading behavior of individual investors. Overall, our results suggest that theories of human memory can be broadly applied to financial markets.

In estimating these memory associations, we are guided by a simple conceptual framework that builds on Kahana (2012), Bordalo et al. (2020), and Bordalo et al. (2023). To build intuition for our approach, consider an investor who evaluates a stock at the height of the Covid-19 Pandemic. This evaluation of the stock, in the context of a global pandemic, serves as a cue that leads the investor to recall similar past experiences (e.g., the 2008 Financial Crisis). We represent each period as a high-dimensional vector that comprises the stock's features in that period, such as its price, return, trading volume, profitability, capital structure, etc., as well as broader features of the economic environment, like GDP growth and the inflation rate. We then calculate the cosine similarity of the current period's vector (= the cue) with historical vectors (= past experiences). We assume that, all other things equal, the investor is more likely to recall past periods for which the historical vector is more similar to the current vector. This captures the force of similarity, a well-established regularity in associative recall (Kahana (2012)).

It is possible that many historical vectors are similar to the current period's vector. These historical periods all compete for retrieval, making it less likely that the investor recalls any particular historical period. Thus we assume that, all other things equal, the investor is less likely to recall a particular historical period if many historical vectors are similar to the current period's vector. This captures the force of interference, another well-established regularity in associative recall (Kahana (2012)).

By combining the forces of similarity and interference, our measure yields the probability that a representative investor recalls a particular historical period when cued with the current period's vector. We make two simplifying assumptions when calculating this probability. First, we calculate similarity and interference stock-by-stock, effectively assuming that investors narrowly frame each stock. This assumption is supported by previous work showing that investors' thinking is quite "local" (Barberis (2013)). Second, we use a rolling look-back window over the past 15 years for each stock. Thus, for each stock in each month, our measure yields a distribution of the recall probability over the previous 180 months.

Figure 1 above gives a visual representation of the recall probabilities generated by our measure for each quarter from 1966-2021. The x-axis indicates the quarter in which the hypothetical recall occurs, while the y-axis indicates the probability weight that our measure puts on past quarters (averaged across stocks). These probability distributions display signature patterns of cued recall. In particular, they display a strong recency effect, where recent months receive a higher weight than distant months. These results highlight that memory is one way to microfound the recency effect discussed in prior studies (Barberis et al. (2015); Cassella and Gulen (2018); Jin and Sui (2022); Nagel and Xu (2022)). In our framework, we do not impose any direct assumptions related to recency. Rather, the principles of similarity and interference naturally generate the recency effect.

In addition to generating a strong recency effect, our measure also places high probability weights on past crises if recall occurs during turbulent economic times. For instance, if recall occurs during the 2020 COVID-19 Pandemic, our measure assigns a high recall probability to the 2008 Financial Crisis. These spikes in the recall probability of past crises are a key distinguishing feature of our measure and cannot be captured by approaches that focus only on recency or extrapolation (e.g., Barberis et al. (2015); Cassella and Gulen (2018); Jin and Sui (2022); Nagel and Xu (2022)).

A third distinguishing feature of our approach is that it generates recall probabilities specific to each stock, allowing both the strength of the recency effect and the spikes in recalling past crises to vary across different stocks. We show that this cross-sectional variation in recall probabilities helps explain the heterogeneous extrapolation patterns across stocks documented by Da et al. (2021). In further tests, which we discuss in more detail below, we show that this cross-sectional variation is also useful in explaining investors' beliefs and trading decisions at the stock-level.

Before applying our measure to these settings, we first validate it by demonstrating that it generates recall patterns that closely align with *actual* recall patterns observed in financial markets. We collect transcripts of corporate events, such as earnings calls, and extract the historical periods mentioned during these events. Our measure strongly predicts which historical periods are mentioned and outperforms a simple extrapolation model. For example, the 2008 Financial Crisis is heavily discussed during corporate events occurring in the COVID-19 pandemic. A simple extrapolation model, where the weights on past periods decay exponentially, does not predict these recall patterns, but our measure does. More generally, when we control for extrapolative recall in our tests, our measure continues to predict actual recall while the coefficient on the extrapolation model is insignificant (though positive). These results provide strong evidence that our measure accurately and comprehensively captures the stock-level memory associations of participants in financial markets. Having validated our measure, we proceed to apply it to two different financial market settings.

In our first setting, we show that memory-based return expectations explain return expectations elicited from investor surveys. We construct memory-based return expectations for the aggregate market in each month by weighting historical monthly returns of the S&P 500 Index with the recall probabilities assigned to these past months by our measure. We find that memory-based return expectations are strongly correlated (corr = 0.67; rank corr = 0.68) with return expectations from the Gallup investor survey (Greenwood and Shleifer (2014)). In regression tests, we estimate that a one standard deviation increase in the memory-based return expectation (corresponding to a 3.9 percentage point higher annualized return) is associated with a 14.5 percentage point increase in the number of investors who are bullish about the stock market over the next 12 months.

One potential concern with this result is that our memory-based measure might just be a proxy for recent returns, which are known to predict survey expectations (Greenwood and Shleifer (2014)). To directly address this concern, we decompose the memory-based return expectation into expectations from recent memory (most recent 12 months) and distant memory (more than 12 months in the past), and show that both components independently predict survey expectations. We find similar results when we use a five-year cutoff to distinguish between recent and distant memory. Importantly, expectations from distant memory continue to predict survey expectations when we directly benchmark our measure against an extrapolative model, either by controlling for cumulative returns over the past 12 months or by controlling for an exponentially-weighted average of returns over the past five years. These results show that our measure captures more than a simple model of extrapolation.

In a direct test of the key distinguishing feature of our measure, we find that expectations from distant memory have the most explanatory power during recessions. Recessions are exactly the periods in which distant memory captures memories of past crises, such as the 2000 Tech Bubble and the 2008 Financial Crisis. Since recall of past crises is elevated during recessions, these results show how the principle of similarity – the underlying driver of our measure – allows our measure to draw on distant experiences precisely when the current economic environment resembles these distant experiences.

Our tests linking memory-based and survey-based return expectations are limited to time-series variation, since survey expectations are typically elicited for the stock market as a whole, but not for individual stocks. Thus, these tests do not harness the full power of our measure, since they ignore the rich cross-sectional variation generated by our measure. Indeed, our measure yields a full recall probability distribution for each stock-month, allowing us to construct memory-based perceived volatility for each stock in each month. We define memory-based volatility as the standard deviation of monthly returns over the past 180 months, where each historical return is weighted with its associated recall probability. We then link the resulting memory-based volatility for each stock with implied volatility from stock options. We find that memory-based volatility derived from our measure can explain cross-sectional variation in implied volatility, establishing a link between memory and higher-moment beliefs in financial markets. The explanatory power of memory-based volatility survives even when we control for a large set of alternative predictors of option-implied volatility.

In our second setting, we show that the memory associations captured by our measure can explain trading behavior of individual investors. These tests allow us to validate our measure in a different setting, and to test the role of memory not only in the cross-section of stocks, but also in the cross-section of investors.

In our tests, we revisit previously documented evidence that investors are more likely to repurchase a stock if they previously sold the stock for a gain (Strahilevitz et al. (2011)). We hypothesize that memory strongly modulates this effect. In particular, we test whether investors are more likely to repurchase a stock that they previously sold for a gain if the probability of recalling the month in which that gain was realized is higher. We find strong evidence in support of this hypothesis. For the same realized gain (loss), investors are more (less) likely to repurchase the stock if the probability of recalling the month in which the gain (loss) was realized is higher.

In terms of magnitude, for a realized gain of 6.90% (sample average), an increase in the recall probability from 0 to 10% translates to an increase in the repurchase probability of 0.11 percentage points, which is about 22% of the unconditional probability of a repurchase (0.5%). This estimate comes from our preferred specification, which includes extremely tight fixed effects. In particular, we include stock-by-month fixed effects, which control for stock-level information revealed in the current month that might be driving repurchasing behavior across all investors. We also include account-by-month-by-liquidation-month fixed effects. These fixed effects not only control for characteristics of the investor in the current month that might be driving the propensity to repurchase, such as investor wealth or sophistication, but also rule out the possibility that we are merely picking up a horizon effect, where investors simply repurchase recently-liquidated stocks. These fixed effects also fully soak up (potentially investor-specific) extrapolative beliefs.

In further tests, we use each investor's history of portfolio holdings to estimate investor-specific memory associations. Using these investor-specific memory associations in the above tests increases the strength and precision of the estimated effects. We also show that an increase in investor-specific similarity increases the probability of repurchase, while an increase in investor-specific interference decreases the probability of repurchase. Finally, we show that strongly-encoded memories have a larger effect on future repurchasing decisions. We proxy for the encoding strength of a memory using the attention that the investor devoted to the portfolio in the month in which a past gain or loss was realized. Summarizing, our results show that memory is an important mechanism that can explain how past experiences affect future trading decisions (Kaustia and Knüpfer (2008); Choi et al. (2009); Strahilevitz et al. (2011)).

Overall, we provide strong evidence that theories of human memory can be broadly applied in financial markets. As such, our paper contributes to a growing literature arguing that theories of human memory may have broad applications in financial markets (Gilboa and Schmeidler (1995); Mullainathan (2002); Hirshleifer and Welch (2002); Gennaioli and Shleifer (2010); Bordalo et al. (2020); Bodoh-Creed (2020); Da Silveira et al. (2020); Nagel and Xu (2022); Bordalo et al. (2023); Wachter and Kahana (2023); Voigt (2023); Bordalo et al. (2024)). Compared to previous empirical work in this literature, our approach allows us to study memory at a much larger scale than prior studies (Charles (2022); Goetzmann et al. (2022); Colonnelli et al. (2023); Jiang et al. (2023); Enke et al. (2024); Gödker et al. (2024); Graeber et al. (2024); Charles (2024)). In a related paper, Chen and Huang (2023) use a machine learning approach to back out underlying memory associations from analyst forecasts. In contrast, we build on established regularities of associative recall to construct memory-based beliefs.

Our paper also contributes to the literature on beliefs in asset pricing (Greenwood and Shleifer (2014); Barberis et al. (2015); Giglio et al. (2021); Da et al. (2021); Brunnermeier et al. (2021); Adam and Nagel (2023)). We show that memory can explain return expectations from surveys as well as higher moments implied by stock options. Finally, our paper relates to the large literature on investor behavior (for an overview, see Barber and Odean (2013)). We show that memory associations drive the trading behavior of individual investors.

2 Conceptual Framework

This section presents a conceptual framework that guides our empirical approach. The framework builds on Bordalo et al. (2020) and Bordalo et al. (2023). We begin by introducing our first assumption: the forces of associative memory govern the recall of past experiences. We then derive memory-based beliefs, by introducing our second assumption: investors sample past experiences from memory when forming beliefs.

2.1 Cued Recall

Consider a representative investor who actively gathers experiences and stores them in the memory database E. We assume that there are $H > 1$ experiences in the memory database. An experience e

fundamentally captures the historical episodes to which the investor has been exposed. Consequently, it can be represented by $M > 1$ features. For instance, an experience of a stock consists of stock-level features, such as the return of the stock during the experience, as well as of broader, contextual features of the economic state, such as the macroeconomic environment during the experience.

Suppose the investor faces a question Q about stock l.[1] The question brings to mind both current features of stock l as well as broader features of the current economic state (the context). Together, the stock-level and contextual features comprise the cue κ_t, which triggers the recall of past experiences. We formalize the nature of this recall in our first assumption:

Assumption 1. The probability that the investor recalls experience e when faced with cue κ_t is given by:

$$r(e, \kappa_t) = \frac{S(e, \kappa_t)}{\sum_{e' \in E} S(e', \kappa_t)}, \tag{1}$$

where S is a symmetric function.

The numerator in the above expression captures the force of similarity: the investor is more likely to recall experience e if it is more similar to the cue κ_t. The denominator in the above expression captures the force of interference: the investor cannot fully control the recall process and all experiences $e \in E$ compete for retrieval. Thus, if there are many experiences that are similar to the cue, the probability of recalling the focal experience e is lower.

2.2 Memory-Based Beliefs

Suppose now that Q is a probabilistic question, such as *"What is the expected return of stock l?"*. This question requires the investor to construct an underlying distribution of returns, which leads to our second assumption:

Assumption 2. When evaluating a probabilistic question Q, the investor samples past experiences e from the memory database E and assigns a sampling-based probability to each retrieved experience.

[1] We intentionally keep the nature of the question abstract here, since this section focuses on the mechanics of cued recall. In the next section, when we discuss memory-based beliefs, we consider more specific questions.

This sampling from memory yields a subjective probability measure. Using the sampling-based probability and the historical return of each experience, the investor calculates the expected return of stock l.

We now discuss several implications that follow from our assumptions. First, our assumption that associative memory governs the recall of past experiences implies that when estimating the expected return of stock l, the investor places a higher probability weight on historical returns from experiences that are more similar to the cue. This assumption also implies that if there are many experiences that are similar to the cue, the probability weight placed on the historical return from experience e is lower.

Second, the cue and the memory database jointly determine the subjective probability distribution derived from memory. For the same set of experiences in memory, different cues generally result in different probability distributions, and consequently different expected returns.

Third, different databases E result in different probability distributions, which in turn lead to different expected returns. Interference plays a pivotal role for this result. To see this, consider two extreme cases. For the first case, assume that all experiences in E are identical to the cue. In this scenario, memory retrieval is essentially frequency-based, and the recall probability $r(e, \kappa_t)$ is equal to $\frac{1}{H}$. For the second case, assume that e is identical to the cue, while all other experiences in E have the lowest similarity level with the cue. In this alternative scenario, the recall probability $r(e, \kappa_t)$ is (very close to) one. Notice that the similarity between e and κ_t is the same across the two cases, showing that interference from other memories is driving the difference in the recall probability across the two cases. In sum, subjective probability distributions derived from memory vary with the cue and with the exact composition of the memory database.

3 An Empirical Measure of Memory Associations in Financial Markets

In this section, we apply the insights from the conceptual framework to construct an empirical measure of memory associations in financial markets. In our setting, experiences e and cues κ_t are vectors in M dimensions. We calculate the similarity between an experience e and a cue κ_t, $S(e, \kappa_t)$, using the cosine similarity measure:[2]

$$S(e, \kappa_t) = 0.5 \times \frac{e \cdot \kappa_t}{\|e\| \cdot \|\kappa_t\|} + 0.5 \tag{2}$$

Here, $e \cdot \kappa_t$ represents the dot product of vectors e and κ_t, $\|e\|$ represents the magnitude (length) of vector e, and $\|\kappa_t\|$ represents the magnitude (length) of vector κ_t. The cosine similarity value ranges from -1 to 1. A value closer to 1 indicates a higher similarity between the vectors, a value closer to -1 indicates a lower similarity. We normalize the cosine similarity into the range $[0, 1]$ by scaling it with a factor of 0.5 and then shifting it by 0.5. For the normalized measure, the similarity is equal to 1 if an experience e is identical to the cue. Conversely, the similarity value is 0 if e is the opposite of κ_t.[3]

As in our conceptual framework, the vectors e and κ_t consist of stock- and firm-level features as well as broader, contextual features of the economic state. While the potential list of features is near endless, we focus on features to which investors are likely paying attention, and which are therefore plausibly encoded in memory.[4] These features can be sorted into three groups: (i) stock-level variables, (ii) firm-level financial ratios, and (iii) broad macroeconomic variables.

At the stock-level, we include stock returns, stock prices, dollar volume, and trading volume. At the firm-level, we follow Van Binsbergen et al. (2023) and focus on a large set of financial variables, such as the book-to-market ratio and the dividend yield. At the macro-level, we consider consumption growth, GDP growth, growth of industrial production, the unemployment rate, and the inflation rate. These variables can affect firms' earnings and are thus potentially relevant to investors. All variables are at the monthly level.

It is worth noting that the above variables exhibit variation at different scales, which can lead to undesired weighting effects when using the cosine similarity method. To see this, consider the following example with only two features (M = 2): volatility and dollar trading volume. Suppose that currently both volatility and

[2] The cosine similarity of two vectors, A and B, is calculated by taking the dot product of the vectors and dividing it by the product of their magnitudes: $\frac{A \cdot B}{\|A\| \cdot \|B\|}$

[3] Due to the normalization, two randomly chosen vectors have an expected similarity of 0.5. While at first blush this may appear high, this is not an issue for our empirical approach. The reason is that when we apply equation (1) to construct our measure of the recall probability, $r(e, \kappa_t)$, it is relative similarity that matters, not absolute similarity.

[4] In Section 4.3, we analyze which types of features are most important for our results.

dollar trading volume are equal to 1, i.e., $\kappa_t = (1, 1)$. Suppose further that the standard deviation of volatility is 1, while the standard deviation of dollar trading volume is 100. Now, consider two distinct experiences, $e1 = (2, 1)$ and $e2 = (1, 101)$, which both differ from the cue κ_t by one standard deviation in only one feature. Their similarity levels with the cue κ_t should be identical, but the cosine similarity method yields $S(e2, \kappa_t)$ = 0.714, which is significantly lower than $(e1, \kappa_t)$ = 0.949. This difference is due to the fact that the similarity measure is overly sensitive to dollar trading volume because dollar trading volume fluctuates with a larger magnitude. We address this issue by normalizing each variable to have a mean of zero and a standard deviation of one.

Using the normalized variables, we calculate similarity over a rolling 15-year window.[5] We make one important simplification: we calculate similarity stock-by-stock. This approach assumes that investors narrowly frame each stock, an assumption that is supported by previous work showing that investors' thinking is quite "local" (Barberis (2013)). We make the assumption of narrow framing at the stock-level throughout the paper.

Finally, we use the estimated similarity to construct an empirical measure of the recall probability $r(e, \kappa_t)$ by applying equation (1). The resulting measure captures the probability that a representative investor recalls experience e when cued with κ_t. This recall probability represents the relative frequency with which experience e comes to the investor's mind when faced with cue κ_t. Therefore, this recall probability can also be interpreted as a measure of endogenous attention driven by memory retrieval.

3.1 Data

Our sample consists of U.S. firms listed on the NYSE, Amex, and Nasdaq that have been listed for at least 15 years.[6] We source stock-level variables from CRSP and firm-level variables from the Financial Ratios Suite offered by Wharton Research Data Services. The macroeconomic variables are obtained from

[5]The 15-year horizon enables us to have a long enough look-back window when we analyze the recall of past crises during the Covid-19 pandemic.

[6]In Figure A1 in the Appendix, we focus on even longer memories by constructing our measure for firms that have been listed for at least 50 years.

the real-time dataset provided by the Federal Reserve Bank of Philadelphia. Our monthly sample starts in January 1966, as the firm-level variables are only available after 1951 and we require a rolling window of 15 years. Our sample ends in December 2021. Detailed variable descriptions are presented in Table A1 in the Appendix.

We normalize each variable stock-by-stock, using data from the past 15 years to calculate the mean and standard deviation. Specifically, we subtract the mean from each data point and then divide it by the standard deviation. If firm-level variables are missing, we follow Van Binsbergen et al. (2023) and fill them with the monthly industry-level median of the firm's Fama-French 49 industry. After filling, approximately 3% of our observations still have at least one variable with missing data. The majority of these missing values are due to missing trading volume. We set these values to zero when calculating the cosine similarity.

4 Validation of the Measure

The goal of this section is to validate our empirical measure by showing that it naturally generates signature patterns of cued recall. We also show that *actual* recall data, extracted from transcripts of earnings conference calls, mirrors the patterns generated by our measure.[7]

4.1 Recency

We first analyze the distribution of the recall probability implied by our measure and show that it displays a strong recency effect. To construct this distribution, we proceed in three steps. In the first step, we apply equation (2) to calculate for each stock-month the relative similarity between the cue κ_t (comprising the current stock-month's features) and each experience e (comprising each historical stock-month's features) over the past 180 months. In the second step, we apply equation (1) to calculate for each stock-month the probability that a representative investor recalls a historical stock-month from the previous 15 years. Finally, in the third step, we average the recall probability across stocks for each month to arrive at an average recall

[7] In Appendix B, we provide further evidence, showing that our measure generates patterns typically observed in subjective beliefs data (Da et al. (2021)).

probability for each month over the past 15 years.[8]

We present this average recall probability for years 1967 to 2021 in Figure 1. For ease of reading, the recall probability in this figure is aggregated to the quarterly level. The x-axis represents the quarter in which the cued recall occurs, while the y-axis represents the probability weights on past quarters. A darker shade of red indicates a higher probability weight.

The figure displays a striking recency effect: the probability weight on recent quarters is much larger than the probability weight on distant quarters, as evidenced by the dark red area near the lower diagonal line. The recency effect is considered a fundamental law of memory (Kahana (2012)) and its role has also been studied in the finance literature (Malmendier and Nagel (2011); Greenwood and Shleifer (2014); Barberis et al. (2015); Nagel and Xu (2022); Jin and Sui (2022)). However, previous studies often resort to ad-hoc configurations to generate a recency effect. For example, Barberis et al. (2015) and Jin and Sui (2022) directly assume that investors use exponentially-decaying weights. Similarly, Nagel and Xu (2022) generate the recency effect by assuming that investors use a constant-gain learning rule. In contrast to these studies, we do not impose any direct assumptions related to recency. Instead, the principles of similarity and interference naturally generate the recency effect.

4.2 Recall of Extreme Episodes

The recency effect is a strong regularity of recall, but there are important empirical exceptions to it. For instance, Malmendier and Nagel (2011) show that disastrous experiences, like the Great Depression, can have long-lasting effects on financial risk-taking and return expectations. In a more direct test of the memory channel, Jiang et al. (2023) show that distant dramatic events are more likely to be recalled when investors are cued with drastic economic dynamics. Thus, an empirical approach that only focuses on recency (e.g., by imposing exponentially-decaying weights) misses these important recall patterns.

Here, we show that our measure naturally generates high recall probabilities for past crises if recall occurs during extreme episodes. In Figure 1, we highlight three dramatic episodes during which recall

[8] When calculating this average, we only include stocks for which we have data on the previous 180 months, to ensure that we have a valid recall probability for each of the previous 180 months.

occurs: the 2020 Covid-19 Pandemic, the 2008 Financial Crisis, and the 1979 Oil Crisis. For each of these episodes, our measure assigns high recall probabilities to past dramatic episodes. For instance, during the 2020 Covid-19 Pandemic, our measure assigns a high recall probability to the 2008 Financial Crisis. Similarly, during the 2008 Financial Crisis, the Tech Bubble Burst receives a high weight. Finally, during the 1979 Oil Crisis, past experiences of the 1973 Oil Crisis are weighted more heavily.[9]

These high probability weights on distant but similar events are a key distinguishing feature of our measure, one that a model of simple extrapolation would miss. To make this point visually salient, in Figure 2, we show the recall probabilities implied by the extrapolative model of Greenwood and Shleifer (2014), using the average decay parameter $\lambda = 0.56$ from their study. As Figure 2 shows, this model generates a strong recency effect with a fast decay, but it cannot generate the high recall probabilities on past crises generated by our measure.

In Figures 3 and 4, we dive into the dynamics of recall generated by our measure, by providing snapshots of recall probabilities immediately before and then during a dramatic event. The x-axis in these figures indicates the number of months between the cueing month and the recalled month, while the y-axis indicates the recall probability assigned to a historical month by our measure.

Take Figure 3 as an example. This figure illustrates the dynamics of memory retrieval before and during the 2008 Financial Crisis. The upper panel shows (in red) the recall probability if recall occurs in June 2007, just before the outbreak of the crisis. Memory recall gradually decays, consistent with the average recall probability in our sample (displayed in green). In contrast, the lower panel shows that if recall occurs in December 2008, in the middle of the Financial Crisis, our measure assigns a larger weight to the months around the 2001 Tech Bubble Burst.

The patterns in in Figure 4 are perhaps even more striking. In this figure, we present memory recall just before and then during the 2020 Covid-19 Pandemic. In the upper panel, we show that if recall occurs in December 2019, immediately before the outbreak of the pandemic, the recall probability gradually decays,

[9] In Figure A1 in the Appendix, we recreate Figure 1 using firms that existed for at least 50 years. Focusing on these firms allows us to extend the look-back window from 15 to 50 years. The resulting figure shows that past crises–going all the way back to the 1973 Oil Crisis–receive high weights if recall occurs during the 2008 Financial Crisis or the 2020 Covid-19 Crisis.

aligning with the average recall probability in our sample. In the lower panel, we show that six months later, when the Covid-19 Pandemic is in full swing, the probability weights on the 2008 Financial Crisis are dramatically higher.

Our measure naturally generates these recall patterns through the principles of similarity and interference. Intuitively, as a dramatic event like the 2020 Covid-19 Pandemic unfolds, our rich set of stock-level, firm-level, and macroeconomic features accurately captures the economic state of the period and, through the principle of similarity, places a higher probability weight on past periods with similar features. Further, since crises occur relatively infrequently, these experiences are subject to less interference during recall, which also leads to a higher probability weight.

4.3 Which Features are Driving the Recall Patterns?

Our baseline measure comprises three types of features: (i) stock-level variables, (ii) firm-level financial ratios, and (iii) broad macroeconomic variables. Here, we analyze which features are driving the recall patterns discussed in the previous two sections. To do so, we create three different flavors of our measure, where each flavor is constructed using only data from one type of feature. We then redo Figure 1 for each flavor.

The first flavor is constructed using only stock-level variables, specifically, monthly stock returns, stock prices, dollar volume, and trading volume. Figure A2 in the Appendix shows that this flavor displays a strong recency effect but that the strength of the recency effect fluctuates over time. For instance, the recency effect was very strong in the early 1980s and early 2000s. Figure A2 also hints at recall of extreme periods, but overall, this pattern is much less pronounced for this flavor.

The second flavor is constructed using only firm-level financial ratios, such as the book-to-market ratio and the dividend yield. As Figure A3 in the Appendix shows, this flavor also produces a strong recency effect, but one that is much more stable over time. Further, the recall of extreme periods is very salient for this flavor.

The third flavor is constructed using only broad macroeconomic variables, specifically, consumption growth, GDP growth, growth of industrial production, the unemployment rate, and the inflation rate. Figure

A4 in the Appendix shows that this flavor produces a certain degree of recency, but that many distant episodes also receive high weights. The weights on these distant periods somewhat mask the high weights that past crises receive during extreme periods.

Another visually striking feature of Figure A4 is that it displays a checkered pattern, especially for recent decades. One interpretation of this pattern is that the economy moves through two types of macroeconomic states, which are picked up by our measure. For instance, during the 2008 Financial Crisis, economic growth was low and unemployment was high, while in the boom period that followed, growth was high and unemployment was low. These changes in the economic state might plausibly create the checkered pattern. In contrast, there is no checkered pattern for the other two flavors. A possible explanation is that there are many more combinations of stock-level and firm-level features – many more states – which result in smoother patterns.[10]

Overall, the signature patterns of cued recall appear for all three flavors. The recency effect is more pronounced for stock-level and firm-level features, while the recall of extreme periods is more pronounced for firm-level and broad macroeconomic features. In our tests going forward, we focus on our baseline measure that combines all features.

4.4 Actual Recall Patterns Extracted from Transcripts of Corporate Events

Perhaps the most direct way of validating our measure is to show that it generates recall patterns that directly match *actual* recall patterns observed in financial markets. Therefore, to further validate our measure, we show that it generates recall patterns that closely align with actual recall patterns observed during corporate events, like earnings calls. We measure actual recall in these events by extracting which past episodes are mentioned during the event.

We collect transcripts of corporate events from Refinitiv StreetEvents for January 2001 to December 2021. These transcripts are verbatim representations of Earnings Calls, M&A Calls, Sales Calls, Analyst

[10] It is also worth mentioning that for the flavors using stock-level and firm-level features, the recall probability for each month is averaged across many firms, which may also lead to smoother patterns. In contrast, the macroeconomic features are the same for all firms in a given month.

Meetings as well as Corporate Conference Presentations. Q&A sessions of these calls are also included in the transcripts. In Section A in the Appendix, we describe in detail how we construct the sample and how we process the data to extract past episodes mentioned during each event.

To get a feel for the extracted recall patterns, we begin by aggregating the data to the monthly level. This aggregation allows us to calculate the relative frequency with which different historical months are mentioned across all events taking place in a given month. As a concrete example, assume that 1,000 firms each had one event (e.g., an earnings call) in January 2015. Further assume that in 200 of these calls, July 2014 was mentioned at least once. In this case, we would assign a 20% recall probability to July 2014 if recall occurs in January 2015.[11]

Figure 5 shows a heatmap of the extracted recall probabilities. To ease the comparison with Figure 1, we aggregate the recall probabilities to the quarterly level. The x-axis represents the quarter in which corporate events take place, while the y-axis represents the relative frequency with which past quarters are mentioned during these events. A darker shade of red indicates a higher probability weight. The figure displays a very strong recency effect: the past 8-10 quarters are recalled with a very high probability. There is also strong recall of past crises during extreme episodes. For example, during the 2020 Covid-19 Pandemic, the 2008 Financial Crisis receives a very high weight. Similarly, during the 2008 Financial Crisis, the 2001 Tech Bubble Burst receives a high weight. Both of these patterns mirror the recall patterns generated by our measure (see Figure 1).

Visually, the strong overlap between recall patterns implied by our measure (displayed in Figure 1) and the recall patterns extracted from corporate events (displayed in Figure 5) is apparent. We can also quantify it: the correlation between the two is 0.88 (rank correlation is 0.79). This strong correlation shows that our measure does a good job at capturing *actual* recall patterns, at least for the aggregate market.

In a further step, we show that our measure also captures actual recall patterns for individual firms. Our tests are simple. For a corporate event of stock l in month t, we ask if our measure can predict whether a past month was mentioned during the event. We construct a dummy that is equal to one if during a corporate

[11] If a firm has multiple events in a month, we consider a historical month as being recalled if it was mentioned during at least one of the events.

event of stock l in month t the past month $t - h$ was mentioned, where $h \in \{1, 2, 3, \ldots, 180\}$. We regress this dummy on our measure $r(e, \kappa_t)$, which is the probability that a representative investor recalls a past experience e when cued in month t, where past experiences correspond to past months $t - h$. Thus, our regression is at the stock-by-current-month-by-past-month level. Accordingly, we cluster standard errors by stock, current month, and past month.

We present summary statistics for the sample in Table 1 and results from estimating the above regression in Table 2. We find that our measure strongly predicts actual recall. In all columns, we include quarter fixed effects to account for seasonality. In column (1), we also include stock-by-past-month fixed effects, which capture the possibility that some past stock-months are more/less likely to be recalled. Such effects could be due to stock-specific shocks that occurred in the past month and that impacted the encoding of a memory trace for all market participants.

The coefficient on $r(e, \kappa_t)$ in column (1) implies that a one percentage point (pp) increase in the probability of recalling month $t - h$ translates into a 67 pp increase in the probability that month $t - h$ is mentioned during a corporate event in month t. At first blush, this effect size may seem unreasonably large. However, it is important to realize that while our measure captures a probability distribution for which the sum of all probabilities must be equal to one, there is no restriction on how many past months can be mentioned during a corporate event. Thus, this result simply says that if, according to our measure, the recall probability for a past month is reasonably high, this month will almost surely be mentioned at some point during the corporate event. Consistent with this intuition, the deep red lower-left diagonal Figure 5 shows that recent months have a very high probability of being mentioned. This result is intuitive: communicating recent firm performance is an important part of corporate events like earnings calls.

Given the strong degree of recency in the extracted recall patterns shown in Figure 5, some readers may wonder whether a simple model of extrapolation or recency performs just as well as our measure. Therefore, in column (2), we replace our measure with the probability weights generated by the exponentially-decaying weighting approach of Greenwood and Shleifer (2014), using their estimated decay parameter λ of 0.56.[12] While the coefficient on these exponentially-decaying weights is large and positive, it is not significant.

[12]This decay parameter is the average across a range of different investor surveys.

In column (3), we include both our measure as well as the exponentially-decaying weights as explanatory variables and find that the coefficient on our measure remains virtually unchanged, both in magnitude and significance. In column (4), we further augment the regression with stock-by-current-month fixed effects, which capture stock-specific shocks in the current month that affect the likelihood of recalling (any) past month. Including these fixed effects does not change the results.

One potential concern with the results in columns (1) through (4) is that they may pick up a spurious correlation. Specifically, one may worry that managers almost always talk about recent firm performance, and that this just happens to coincide with the strong degree of recency in our measure. The fact that an exponentially-decaying model performs worse than our measure should already alleviate some of these worries. Nevertheless, in the following tests, we show that the overlap between our measure and extracted recall is not only due to the recency effect. Specifically, we focus on recall of historical months that are at least five years in the past. Visually, once can think of this as slicing off the deep red lower-left diagonal in Figure 5.[13]

In columns (5) through (8), we re-estimate the specifications from the first four columns for these distant months. The coefficient estimates on $r(e, \kappa_t)$ are much lower, but still highly significant. In terms of magnitude, these estimates imply that a one percentage point (pp) increase in the recall probability translates to a 3-4 pp increase in the probability that the distant month is mentioned during a corporate event in month t. These lower magnitudes are due to the fact that managers are much less likely to discuss the distant past. The results in columns (6) through (8) also show that the coefficients on the exponentially-decaying weights are zero, indicating that a model based purely on exponential decay does a poor job at predicting recall of the distant past. Overall, the takeaways from Table 2 are that recency is an important reason for the strong overlap between our measure and extracted recall patterns, that our measure outperforms a simple model of

[13] As an alternative approach to addressing the concern that managers are prone to mentioning recent firm performance, we distinguish actual recall patterns for different types of participants in the corporate event. The transcript of each event identifies the speaker of each sentence, such as an analyst or a manager, during an earnings call. This distinction allows us to analyze recall patterns separately for different types of speakers. In Tables A2 and A3 in the Appendix, we focus on analysts and managers, respectively, and show that the effects are very similar across these two types of speakers. Beyond helping to address the aforementioned concern, these tables show that our measure can explain the recall patterns of different types of market participants.

exponential decay, and that our measure also predicts extracted recall of the distant past.

5 Memory and Beliefs

In this section, we use our measure to construct memory-based return expectations for the aggregate stock market. We then link these memory-based return expectations with survey-based return expectations. We also construct memory-based perceived volatility and link it with return volatility implied by stock options.

5.1 Constructing Memory-Based Return Expectations

We begin by describing how we use the measure introduced in Section 3 to construct memory-based return expectations for the aggregate stock market. For each stock-month, our measure yields the probability of recalling each of the past 180 months. That is, for each stock in each month, our measure yields a distinct recall probability distribution over the past 180 months. To arrive at an aggregate recall probability distribution in each month, we take the equally-weighted average of the stock-level recall probability distributions in each month. We then use the resulting aggregate recall probability distribution in each month, along with historical monthly returns of the S&P 500 Index, to construct the memory-based return expectation for the aggregate market in each month. Intuitively, we are weighting each historical return with its associated recall probability. More formally, we can express the memory-based probability distribution of monthly returns as:

$$F^S(ret) = \frac{\sum_{ret=ret_e,\text{where } e\in E} S(e,\kappa_t)}{\sum_{e'\in E} S(e',\kappa_t)}. \tag{3}$$

Here, $F^S(ret)$ represents the memory-based probability for a return with a value of ret, and ret_e denotes the specific value of the return in experience e. We think of each past month as an experience. This formulation can accommodate scenarios in which multiple experiences (i.e., multiple past months) contain the same return. If the historical returns in all experiences are distinct from one another, the memory-based probability for a historical return in experience e is simply the recall probability $r(e,\kappa_t)$. However, if returns in multiple experiences are identical to each other, then the memory-based probability for that particular return is the sum of the recall probabilities across the corresponding experiences. Using these memory-based probabilities, we construct the subjective monthly return expectation as follows:

$$\mathbb{E}^S(ret) = \sum_{ret} F^S(ret) \times ret. \tag{4}$$

5.2 Linking Memory-Based and Survey-Based Return Expectations

We next test whether the memory-based return expectations constructed using our measure can explain actual investor expectations from surveys. The survey expectations in our tests are from the Gallup investor survey, as it provides a large sample size and follows a consistent methodology (Greenwood and Shleifer (2014)). The Gallup survey does not ask investors for the percentage return they expect to earn in the market, but rather asks investors whether they are bullish or bearish about the market over the next 12 months. We follow Greenwood and Shleifer (2014) and define survey expectations as the percentage point difference in bullish and bearish investors. We source data from Greenwood and Shleifer (2014) for the period from October 1996 to December 2011, and extend the sample period to May 2020 with data sourced directly from Gallup.[14]

Figure 6 gives a visual impression of the relationship between memory-based and survey-based return expectations. The dashed blue line shows survey-based expectations in month $t+1$.[15] The solid black line shows annualized memory-based return expectations as of the end of month t constructed using our measure. Clearly, the two time series comove heavily, especially from 2002 onward. The correlation between the two time series is 0.67 (the rank correlation is 0.68). Overall, the figure suggests that memory-based and survey-based return expectations are tightly linked. To establish this relationship more rigorously, we estimate the following regression:

$$\text{Survey Expectations}_{t+1} = a + b\text{Memory-Based Expectation}_t + cX_t + e_{j,t},$$

[14] The Gallup survey elicited survey expectations every three months from 1996-1998, then switched to monthly elicitations, and reverted back to elicitations every three months from 2011 onward. As in Greenwood and Shleifer (2014), some months are missing in the period from 1998-2011.

[15] We use survey expectations from month $t+1$, since survey responses are typically collected throughout a month. This approach ensures that all the information used to construct memory-based expectations at the end of month t is available to survey respondents throughout the month $t+1$.

where Memory-Based Expectation$_t$ is the annualized memory-based return expectation as of the end of month t, and Survey Expectations$_{t+1}$ are annual return expectations in month $t+1$ from the Gallup investor survey. X_t is a set of control variables that captures factors which potentially influence investors' belief formation, including the price level (proxied for by Log(P/D)), the risk-free rate, earnings growth, and the unemployment rate. This set of controls follows Greenwood and Shleifer (2014). We show summary statistics of the sample in Panel A of Table 3, and present regression results in Table 4.

In column (1) of Table 4, we estimate the above regression and find that memory-based return expectations significantly predict survey-based return expectations with a positive sign. In terms of magnitude, the coefficient implies that a one standard deviation increase in memory-based return expectations (corresponding to a 3.9 percentage point higher annualized return) is associated with a 14.5 percentage point increase in the number of investors who are bullish about the stock market over the next 12 months.

One potential concern with this result is that our memory-based measure might just be a proxy for recent returns, which are known to predict survey expectations (Greenwood and Shleifer (2014)). To address this concern, in column (2), we decompose the memory-based return expectation into a component capturing only expectations from recent memory (defined as the most recent 12 months) and expectations from distant memory (defined as more than 12 months in the past). The sum of these two components equals our baseline memory-based expectation from column (1). We find that both components independently explain survey expectations. In terms of magnitude, a one standard deviation increase in expectations from recent (distant) memory is associated with a 12.2 (8.2) percentage point increase in the number of investors who are bullish about the stock market over the next 12 months. This result highlights that both recent and distant memory help explain survey expectations.

In column (3), we further augment this regression with the cumulative return over the past 12 months, since Greenwood and Shleifer (2014) show that recent returns are an important determinant of investors' return expectations. This control variable is heavily correlated with our variable capturing expectations from recent memory, and it washes out the coefficient on expectations from recent memory. However, expectations from distant memory continue to explain survey expectations.

In column (4), we directly benchmark our measure against a simple model of extrapolation by controlling for the exponentially-weighted average of returns over the past 5 years. To construct this exponentially-weighted return, we first calculate quarterly returns by compounding 3-month returns on a monthly basis. We then use the weighting approach of Greenwood and Shleifer (2014) and their estimated quarterly λ of 0.77 to calculate an exponentially-weighted average return over the past five years. The λ of 0.77 from Greenwood and Shleifer (2014) is estimated using the Gallup investor survey, which is the same survey data that we use in our tests. Thus, if the concern is that our measure is just a fancy way to generate extrapolation, controlling for this exponentially-weighted average return is arguably a high bar to clear for our measure. As the results in column (4) show, expectations from distant memories continue to explain investors' expectations, highlighting that our measure captures more than a simple model of extrapolation. In columns (5) to (7), we replicate the regressions from columns (2) to (4), using a five year cutoff to distinguish between recent and distant memory. Our results remain very similar.

Finally, in column (8) we ask whether expectations from distant memory are particularly useful in explaining survey expectations during bad times. These tests are motivated by the results presented in 4.2, which show that our measure generates high recall probabilities for distant crises if recall occurs during turbulent economic times. We implement this test by interacting expectations from recent and from distant memory with a dummy variable that is equal to one during NBER recessions. Indeed, expectations from distant memory have more explanatory power, and expectations from recent memory have less, during recessions. This is intuitive, as distant memories include, e.g., memories of the 2008 Financial Crisis for survey expectations that are elicited during the 2020 Covid Pandemic, and memories of the Tech Bubble Burst for survey expectations that are elicited during the 2008 Financial Crisis.

Overall, the results in Table 4 show that memory-based expectations predict survey expectations well. Importantly, the results show that memory-based expectations capture more than mere extrapolation of past returns. The principle of similarity, which is the underlying driver of our measure, allows our measure to draw on distant experiences precisely when the current economic environment resembles these distant experiences.

5.3 Memory-Based Perceived Volatility

The tests in the previous section focus on the time series relationship between memory-based and survey-based return expectations. This focus on the time series is due to the fact that surveys expectations are typically not elicited for each stock individually, but rather for the stock market as a whole. However, narrowly focusing on the time series masks a key strength of our measure: its rich cross-sectional variation. Indeed, our measure yields a full recall probability distribution for each stock-month. In this section, we leverage this variation by constructing memory-based perceived volatility for each stock in each month, and link it with option-implied volatility.

We define memory-based volatility as the standard deviation of monthly returns over the past 180 months, where each historical return is weighted with its associated recall probability. We then link the resulting memory-based volatility for each stock with implied volatility from stock options.[16] We calculate implied volatility for each stock in each month using data from OptionMetrics, following the methodology described in An et al. (2014). We provide summary statistics of option-implied and memory-based volatility in Panel B of Table 3.

In our tests, we regress option-implied volatility in month $t + 1$ on memory-based volatility in month t. The results in column (1) of Table 5 imply that a one standard deviation increase in memory-based volatility is associated with an increase in option-implied volatility of 0.13 units, which corresponds to about 59% of the standard deviation of option-implied volatility. In column (2), we control for perceived volatility from an exponentially-decaying model using the decay parameter $\lambda = 0.56$ from Greenwood and Shleifer (2014).[17] We construct exponential-decay-based volatility as the standard deviation of monthly returns over past months, where each historical return is weighted with exponentially-decaying weights. While

[16]We focus on implied volatility in our cross-sectional tests for two reasons. First, implied volatility is readily available at the stock-level, as it can be easily extracted from stock options. In contrast, return expectations at the stock-level are much harder to come by. Second, by focusing on implied volatility, we can test whether memory-based beliefs from our measure also explain higher-moment beliefs.

[17]We use $\lambda = 0.56$, since this value is the average across a range of different surveys in Greenwood and Shleifer (2014). Our results are similar if we use $\lambda = 0.77$, which is the value of λ that is calibrated to the Gallup survey data.

the coefficient on memory-based volatility in column (2) shrinks by about 40% compared to column (1), it remains statistically significant and economically meaningful, emphasizing that our measure captures more than mere extrapolation of recent volatility.

In column (3), we augment the regression with stock and month fixed effects. These fixed effects leave the effect of memory-based volatility largely unchanged. In column (4) we further control for lagged implied volatility, as implied volatility is known to be persistent. The coefficient on memory-based volatility shrinks substantially once we control for lagged implied volatility, highlighting the importance of this control, but memory-based volatility continues to predict option-implied volatility.[18] Finally, in column (5), we add a host of control variables that capture firm fundamentals and which are know to predict returns in the cross-section, namely (i) a firm's size, which we measure as the logarithm of a firm's market capitalization (in million $), (ii) idiosyncratic volatility in %, which we construct following Ang et al. (2006), (iii) asset growth, which we construct as the book asset growth rate (= current book assets – lagged book assets)/lagged book assets), (iv) operating profit following Fama and French (2006), and (v) the logarithm of the book-to-market ratio, where the book-to-market ratio is constructed following Fama and French (1992). In terms of magnitude, the estimate in column (5) implies that a one-standard-deviation increase in memory-based volatility is associated with an increase in option-implied volatility of 0.016 units, which corresponds to about 7% of the standard deviation of option-implied volatility.

Summarizing, our results in this section show that memory-based volatility derived from our measure can explain cross-sectional variation in implied volatility, establishing a link between memory and higher-order beliefs in financial markets.[19] The explanatory power of memory-based volatility survives even when we control for a large set of alternative predictors of option-implied volatility. From a methodological perspective, the tests in this section also highlight how our measure can be used to easily construct higher-moment beliefs for a range of financial and economic variables.

---

[18] In Table A5 in the Appendix, we use an alternative approach to handle the persistence in option-implied volatility. We estimate Fama-MacBeth regressions, which only exploit cross-sectional variation, and find similar results.

[19] In Appendix C, we construct memory-based volatility for the aggregate market and show that it can explain variation in the Volatility Index (VIX) over time.

6 Memory and Individual Trading Decisions

In this section, we use our measure to revisit previously documented evidence that past trading experiences affect investors' future repurchasing decisions (Kaustia and Knüpfer (2008); Choi et al. (2009); Strahilevitz et al. (2011)). We show that the probability of recalling a past trading experience strongly modulates the likelihood of repurchasing a stock. Our tests in this section establish that memory and trading are linked, validating our measure in a different setting than our previous tests. Further, as we discuss in more detail below, the trading setting allows us to construct memory associations for each investor individually, thus allowing us to test the role of memory not only in the cross-section of stocks, but also in the cross-section of investors.

6.1 Memory Retrieval and Repurchasing Decisions

Consider an investor who debates the question *"Should I repurchase stock l?"*. Strahilevitz et al. (2011) find that the answer to this question depends on how well stock l previously performed for the investor. In particular, if the stock was previously sold for a gain, the investor is more likely to repurchase the stock. We propose that memory plays a key role in modulating this effect. Specifically, we argue that when an investor debates the above question, it brings to mind current stock-level and contextual features, which cue the recall of past experiences. We therefore propose that past trading experiences *do* affect the likelihood of repurchase, but that the probability of recalling a past trading experience strongly modulates this effect. Crucially, this implies that identical past realized returns differentially affect the probability of repurchase, depending on how likely it is that they are recalled by the investor.

We test these ideas using data from a large US discount brokerage. These are the same data as in Barber and Odean (2000). The data include the holdings and trades of approximately 78,000 households between January 1991 and November 1996. We retain only common stocks, drop trades with negative commissions, and match the data to CRSP for information on stock prices. Since our measure of memory associations is constructed monthly, we match this frequency and aggregate each investor's holdings and trades to the monthly level. This means that we pool all trades (buys, sells, liquidations, and repurchases) that the investor executes in a given month. Since we are interested in analyzing investors' repurchasing decisions, our sample consists of all stocks that an investor can repurchase in a given month. These are stocks that an investor once held and subsequently liquidated, but that the investor has not (yet) repurchased since the latest liquidation. Once an investor repurchases a stock, we exclude that stock from the sample going forward, until the investor potentially liquidates the position again. For each position of stock l in account i, we use the past trading history to calculate the weighted average purchase price (WAPP). When an investor fully liquidates a position in a month, we use the WAPP and the average transacted price in the liquidation month to calculate the realized return $Ret_{i,l}$ (winsorized at the 1st and 99th percentiles). We only consider positions for which we know all historical purchase prices, so that we can accurately calculate the WAPP. To this end, we exclude positions that investors held in the first month of the position files, since we do not know at which price the investor acquired these positions.

In our tests, we examine whether the probability that investor i repurchases stock l in month t is modulated by (i) the past realized return, (ii) the probability that the investor recalls the stock-month in which that return was realized, and (iii) the interaction between the two. Specifically, we estimate the following regression:

$$\text{Buy}_{i,l,t} = a\,\text{Ret}_{i,l} + b\,r(e,\kappa_t) + c\,\text{Ret}_{i,l} \times r(e,\kappa_t) + \text{Controls} + \text{FEs} + \epsilon_{i,l,t} \qquad (5)$$

The dependent variable in this regression is a dummy variable, $\text{Buy}_{i,l,t}$, which is equal to one if investor i repurchases stock l in month t. The variable $Ret_{i,l}$ is the previously realized return of stock l for investor i. Finally, the variable $r(e,\kappa_t)$ is the probability that a representative investor recalls the stock-month in which the past return of stock l was realized. This recall probability is constructed using the measure that we introduced in Section 3. Importantly, this recall probability does not vary across investors, since it is the probability that a representative investor recalls the stock-month in which investor i realized the return $Ret_{i,l}$.[20] Standard errors in this regression are clustered by account, liquidation month, and current month.

The coefficients in the above regression have intuitive interpretations. Coefficient a captures the effect

[20]In Sections 6.2 and 6.3 below, we implement tests in which we estimate memory associations for each investor individually. This allows us to construct recall probabilities that are unique to each investor. Our results are very similar, suggesting that, on average, the recall probability of a representative investor does a good job at capturing memory associations of individual investors.

of past realized returns on the probability of repurchase if the recall probability implied by our measure is zero. Coefficient b captures the effect of the recall probability on the likelihood of repurchase if the past realized return is zero. Finally, coefficient c captures the additional effect of the recall probability on the likelihood of repurchase for a given (non-zero) past realized return.

Across specifications, we include different sets of fixed effects, which we discuss in more detail below. We also control for the return of stock l between the liquidation month and the end of month $t-1$, since Strahilevitz et al. (2011) show that this return, which the investor could have hypothetically realized had she not sold the stock, affects the probability of repurchase. Moreover, following Ben-David and Hirshleifer (2012), we control for the logarithm of the initial purchase price (WAPP), the square root of the number of days between initial purchase and liquidation, and the stock volatility calculated using daily returns over the 250 days preceding the initial purchase.

Before showing the results from estimating equation (5), we briefly discuss the summary statistics of our sample, presented in Table 6. On average, there are about 6 stocks (median 4) that an investor can repurchase in a given month.[21] These are stocks that the investor previously held and subsequently liquidated. Investors realized an average gain of 6.90% when liquidating previously-held positions. The unconditional probability of repurchasing a previously-held stock is low at only 0.5%, but as we show next, the probability of recalling previous gains and losses strongly affects the likelihood of repurchase.

Table 7 presents our regression results. We first replicate the finding of Strahilevitz et al. (2011) using a simple linear probability model.[22] We regress the repurchase dummy $\text{Buy}_{i,l,t}$ on the past realized return, $Ret_{i,l}$ and the above control variables. We also include account-month and stock-month fixed effects. The first set of fixed effects controls for characteristics of investor i in month t that might be driving the propensity to repurchase, such as investor wealth or sophistication. The second set of fixed effects controls for stock-level information revealed in month t that might be driving repurchasing behavior across all investors.

In column (1), the coefficient on past realized returns is positive and significant. In terms of magnitude, this coefficient implies that a realized return of 6.90% (the average in our sample) increases the likelihood

[21] For comparison, investors hold about 3 stocks on average (median 1) in their portfolio in a given month.

[22] Strahilevitz et al. (2011) use both ratio analysis and a hazard model to document their finding.

of repurchase by 0.02 percentage points (pp). This effect, while statistically significant, is minuscule even compared to the relatively low unconditional probability of repurchase in our sample of 0.5%. However, as we show in the remaining columns of Table 7, this average effect masks strong heterogeneity with respect to the recall probability.

In column (2), we augment the regression with the recall probability, $r(e, \kappa_t)$, and its interaction with the past realized return, $Ret_{i,l}$. The coefficients on the recall probability as well as on the interaction term are positive, economically meaningful, and highly significant. Consider first the coefficient on the recall probability. This coefficient implies that a 10% probability of recalling the month in which the investor liquidated the stock is associated with a 2.49 pp increase in the likelihood of repurchasing the stock (for a past realized return of zero). This effect size is roughly five times the unconditional probability of repurchase. Further, the coefficient on the interaction term implies that for the same recall probability of 10%, if the investor previously realized a return of 6.90%, the repurchase probability increases by an additional 0.14 pp. Compared to the unconditional probability of repurchase (0.5%), this additional effect represents about a 28% increase.

In column (3), we further tighten up the regression by including account-by-month-by-liquidation-month fixed effects. These fixed effects rule out that we are merely picking up a horizon effect, where investors simply repurchase recently-liquidated stocks. These fixed effects also fully soak up (potentially investor-specific) extrapolative beliefs. Intuitively, including these fixed effects allows us to compare an investor's repurchasing decisions for stocks that were previously liquidated in the same month. While these stocks have been out of the investor's portfolio for the same amount of time, they generally differ in their previously realized returns, $Ret_{i,l}$, as well as in their similarity with the cue κ_t, resulting in different recall probabilities, $r(e, \kappa_t)$.

Notice first that in column (3) the direct effect of the recall probability on the likelihood of repurchase becomes much weaker. This is intuitive: the strongest characteristic of the recall probability is the recency effect (see Figure 1), which is soaked up by the fixed effects. Importantly, however, the coefficient on the interaction term between past realized returns and the recall probability remains strong and significant. The robustness of this coefficient highlights the key result of this table, namely that the probability of recalling a past realized return strongly modulates the likelihood of repurchase.

6.2 The Role of Similarity and Interference

So far, we approximate the memory associations of individual investors with the memory associations of a representative investor. We now turn to estimating memory associations for each investor individually. This allows for a more granular test of memory effects at the investor-level.

When estimating memory associations at the investor-level, we assume that a stock enters an investor's memory database E_i when the investor i first acquires the stock (the "entering month"). We also retain the assumption of narrow framing at the stock-level. We use equation (2) to calculate similarity $S_i(e, \kappa_t)$, which is the cosine similarity between the vector of features from the current stock-month and the vector of features from the stock-month in which the investor realized the return $Ret_{i,l}$. We also construct interference as $I_i(e, \kappa_t) = \sum_{e' \in E_i} S_i(e', \kappa_t)$, that is, the sum of all $S_i(e', \kappa_t)$ across months in which stock l is in investor i's memory database E_i, starting with the entering month and ending with month $t-1$.

Intuitively, the ratio of $S_i(e, \kappa_t)$ and $I_i(e, \kappa_t)$ yields $r_i(e, \kappa_t)$, which is the probability that investor i recalls the stock-month in which $Ret_{i,l}$ was realized when cued with κ_t. This investor-level recall probability can be interpreted analogously to the recall probability of a representative investor, which we use in previous tests. The key difference is that the investor-level recall probability is based on the memory database E_i, which is unique to investor i.

In column (1) of Table 8, we begin by replicating column (2) of Table 7 using investor-specific recall probabilities. We find that the coefficient on the interaction term is about half as large compared to column (2) of Table 7. At first glance, this might suggest that the effect is weaker when using investor-specific recall probabilities rather than the recall probability of a representative investor, which we have used in our tests so far. However, it is important to note that the mean and standard deviation of investor-specific recall probabilities are about four times larger than those of the representative investor, as shown in Table 6. This difference in scale indicates that the effect size is not necessarily weaker, but rather reflects the increased variability in investor-specific recall probabilities.

The difference in scale is intuitive: the memory database of a given individual investor is sparser than that of the representative investor, since we assume that a stock enters an individual investor's memory database when the investor first acquires the stock. Thus, if we compare the effect size for a one standard deviation increase, the magnitude is about double when using investor-specific memory associations instead of memory associations of a representative investor. Further, this effect is estimated with more precision, since the t-statistic is larger even though the coefficient is smaller.

In column (2), we add $S_i(e, \kappa_t)$ and $I_i(e, \kappa_t)$ as separate regressors, and interact each with the previously realized return, $Ret_{i,l}$. Including similarity and interference separately as independent variables allows for a targeted test of associative memory theory. Specifically, since these two forces have opposing effects on the recall probability – similarity increases recall, while interference reduces recall – the interaction of $Ret_{i,l}$ with similarity should have a positive coefficient and the interaction of $Ret_{i,l}$ with interference should have a negative coefficient. This is precisely what we find. In terms of magnitude, for a previously realized return of 6.90$\%$, a one standard deviation increase of similarity (one std. dev. = 0.193) implies a 0.01 pp increase in the probability of repurchase. Conversely, a one standard deviation decrease in interference (one std. dev. = 10.301) implies a 0.014 pp decrease in the probability of repurchase.

In columns (3) and (4) of Table 8, we further augment the regressions from columns (1) and (2) with account-by-month-by-liquidation-month fixed effects. Including these very tight fixed effects does not change the coefficients on the interaction terms. Overall, the results in Tables 7 and 8 show that the probability of recalling past experiences strongly modulates the effect that these experiences have on future repurchasing decisions.

6.3 The Role of Encoding Strength

In this next set of tests, we examine whether the encoding strength of an experience matters for the effects documented in Tables 7 and 8. We hypothesize that, conditional on all else being equal, a strongly encoded experience is more likely to be recalled than a weakly encoded one.

To estimate the encoding strength of a past trading experience, we use the attention that the investor devoted to her portfolio in the month of the experience. We proxy for this attention using the total number of transactions that the investor executed in the month of the experience. Specifically, we construct a dummy variable that is equal to one if the investor executed at least two transactions in the month of the experience, and zero otherwise.[23]

In our tests, we augment equation (5) with a triple interaction term between the attention dummy, the past realized return, $Ret_{i,l}$, and the investor-level recall probability, $r_i(e, \kappa_t)$. If the intensity of encoding matters, the coefficient on this triple interaction term should be positive. Intuitively, a positive coefficient on the triple interaction implies that experiences that were encoded in months with high attention are more likely to be recalled, and therefore have a stronger effect on the likelihood of repurchase. Notably, the direction of this effect depends on the sign of the previously realized return: a positive realized return increases the likelihood of repurchase, while a negative realized return decreases it.

Table 9 presents the results. In both columns, we find that the coefficient on the triple interaction term is positive and significant. Consider the estimates in column (1), which includes account-month and stock-month fixed effects. The coefficient on the triple interaction term implies that for a realized return of 6.90% and a recall probability of 10%, the likelihood of repurchase is about 0.035 pp higher if the past return was realized in a month in which the investor paid a lot of attention to her portfolio. In column (2), in which we include extremely tight account-by-month-by-liquidation-month fixed effects, the effect size is, if anything, somewhat stronger. Overall, the results in this table show that even for past experiences that have identical realized returns and that are equally likely to be recalled, those that are encoded more strongly are more likely to be recalled, resulting in a higher likelihood of repurchase.

7 Conclusion

Models of human memory have the potential to microfound a wide range of empirical phenomena in financial markets, from long-lasting experience effects to short-term return extrapolation. While recent surveys and experimental studies support many predictions of these models, they often prioritize internal validity at the expense of external validity. Our study contributes to this literature by examining memory at scale using data from the field.

[23] The median number of transactions per month is one.

We base our empirical approach on two well-established regularities of associative recall – similarity and interference – that have been robustly documented in the experimental laboratory. We estimate memory associations that a representative investor *would* have if these regularities also operate in the field. We then show that the resulting memory associations predict (i) actual recall patterns extracted from transcripts of corporate events, (ii) return expectations from investor surveys, (iii) higher moment beliefs implied by the VIX, and (iv) trading decisions of individual investors. Our results provide strong evidence that models of human memory can be broadly applied in financial markets.

On the methodological front, we provide a straightforward, theory-based approach for retro-actively constructing memory-based beliefs for a wide range of economic and financial variables. For instance, using the recall probability distributions generated by our measure, it is possible to construct memory-based expectations of macro-variables such GDP growth or the inflation rate, as well as firm-level variables such as earnings or cash flow growth. Being able to construct memory-based expectations of different variables may be particularly valuable when survey data are not available. A further advantage of our approach is that it is very flexible. By extending or reducing the set of features used to calculate similarity and interference, our measure can be tailored to different applications. Thus, our approach may help researchers study the role of memory in other domains of financial markets and in the economy more generally.

References

Adam, K. and Nagel, S. (2023). Expectations data in asset pricing. In *Handbook of Economic Expectations*. Elsevier.

An, B.-J., Ang, A., Bali, T. G., and Cakici, N. (2014). The joint cross section of stocks and options. *The Journal of Finance*, 69(5):2279–2337.

Andre, P., Pizzinelli, C., Roth, C., and Wohlfart, J. (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. *Review of Economic Studies*, 89(6):2958–2991.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, 61(1):259–299.

Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, 55(2):773–806.

Barber, B. M. and Odean, T. (2013). The behavior of individual investors. In *Handbook of the Economics of Finance*, volume 2, pages 1533–1570. Elsevier.

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

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

Ben-David, I. and Hirshleifer, D. (2012). Are investors really reluctant to realize their losses? trading responses to past returns and the disposition effect. *Review of Financial Studies*, 25(8):2485–2532.

Bodoh-Creed, A. L. (2020). Mood, memory, and the evaluation of asset prices. *Review of Finance*, 24(1):227–262.

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., and Shleifer, A. (2024). Imagining the future: Memory, simulation, and beliefs. *Review of Economic Studies*, page rdae070.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., and Shleifer, A. (2023). Memory and probability. *Quarterly Journal of Economics*, 138(1):265–311.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, attention, and choice. *Quarterly Journal of Economics*, 135(3):1399–1442.

Brunnermeier, M., Farhi, E., Koijen, R. S., Krishnamurthy, A., Ludvigson, S. C., Lustig, H., Nagel, S., and Piazzesi, M. (2021). Perspectives on the future of asset pricing. *Review of Financial Studies*, 34(4):2126–2160.

Cassella, S. and Gulen, H. (2018). Extrapolation bias and the predictability of stock returns by price-scaled variables. *Review of Financial Studies*, 31(11):4345–4397.

Charles, C. (2022). Memory and trading. *Working Paper*.

Charles, C. (2024). Memory moves markets. *Working Paper*.

Chen, Z. and Huang, J. (2023). Memory and beliefs in financial markets: A machine learning approach. *Working Paper*.

Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. (2009). Reinforcement learning and savings behavior. *Journal of Finance*, 64(6):2515–2534.

Colonnelli, E., Gormsen, N. J., and McQuade, T. (2023). Selfish corporations. *Review of Economic Studies*, forthcoming.

Da, Z., Huang, X., and Jin, L. J. (2021). Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics*, 140(1):175–196.

Da Silveira, R. A., Sung, Y., and Woodford, M. (2020). Optimally imprecise memory and biased forecasts. *Working paper*.

Enke, B., Schwerter, F., and Zimmermann, F. (2024). Associative memory, beliefs and market interactions. *Journal of Financial Economics*, forthcoming.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, 82(3):491–518.

Gennaioli, N. and Shleifer, A. (2010). What comes to mind. *Quarterly Journal of Economics*, 125(4):1399–1433.

Giglio, S., Maggiori, M., Stroebel, J., and Utkus, S. (2021). Five facts about beliefs and portfolios. *American Economic Review*, 111(5):1481–1522.

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

Gödker, K., Jiao, P., and Smeets, P. (2024). Investor memory. *Review of Financial Studies*, forthcoming.

Goetzmann, W. N., Watanabe, A., and Watanabe, M. (2022). Evidence on retrieved context: How history matters. *Working Paper*.

Graeber, T., Zimmermann, F., and Roth, C. (2024). Stories, statistics, and memory. *Quarterly Journal of Economics*, forthcoming.

Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *Review of Financial Studies*, 27(3):714–746.

Hirshleifer, D. and Welch, I. (2002). An economic approach to the psychology of change: Amnesia, inertia, and impulsiveness. *Journal of Economics & Management Strategy*, 11(3):379–421.

Jiang, Z., Liu, H., Peng, C., and Yan, H. (2023). Investor memory and biased beliefs: Evidence from the field. *Working Paper*.

Jin, L. J. and Sui, P. (2022). Asset pricing with return extrapolation. *Journal of Financial Economics*, 145(2):273–295.

Kahana, M. J. (2012). *Foundations of human memory*. OUP USA.

Kaustia, M. and Knüpfer, S. (2008). Do investors overweight personal experience? evidence from ipo subscriptions. *Journal of Finance*, 63(6):2679–2702.

Li, K., Mai, F., Shen, R., and Yan, X. (2021). Measuring corporate culture using machine learning. *Review of Financial Studies*, 34(7):3265–3315.

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

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

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

Strahilevitz, M. A., Odean, T., and Barber, B. M. (2011). Once burned, twice shy: How naive learning, counterfactuals, and regret affect the repurchase of stocks previously sold. *Journal of Marketing Research*, 48(SPL):S102–S120.

Van Binsbergen, J. H., Han, X., and Lopez-Lira, A. (2023). Man versus machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies*, 36(6):2361–2396.

Voigt, M. (2023). Investor beliefs and asset prices under selective memory. *Working Paper*.

Wachter, J. A. and Kahana, M. J. (2023). A retrieved-context theory of financial decisions. *Quarterly Journal of Economics*, forthcoming.

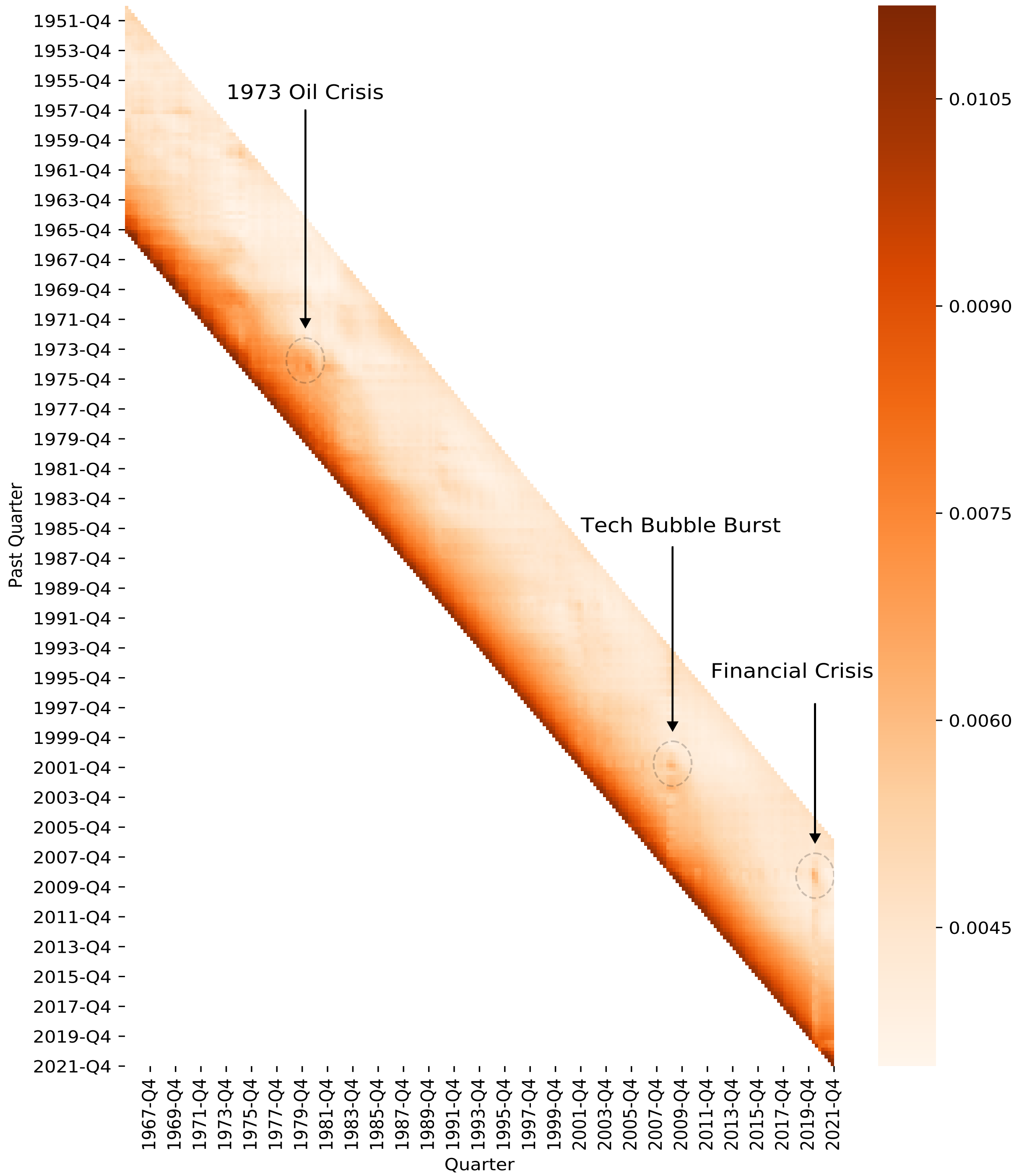

Figure 1: Recall Probabilities from our Measure

This figure displays the recall probability that our measure assigns to the past 60 quarters (= 15 years) for each quarter from 1966 to 2021. The x-axis indicates the quarter in which recall occurs, while the y-axis indicates the probability weight assigned to past quarters. A darker shade of red indicates a higher probability weight. We highlight three episodes with exceptionally high probability weights: the recall of the 1973 Oil Crisis during the 1979 Oil Crisis, the recall of the Tech Bubble Burst during the 2008 Financial Crisis, and the recall of the 2008 Financial Crisis during the 2020 Covid-19 Pandemic. This figure is also repeated at the beginning of the manuscript.

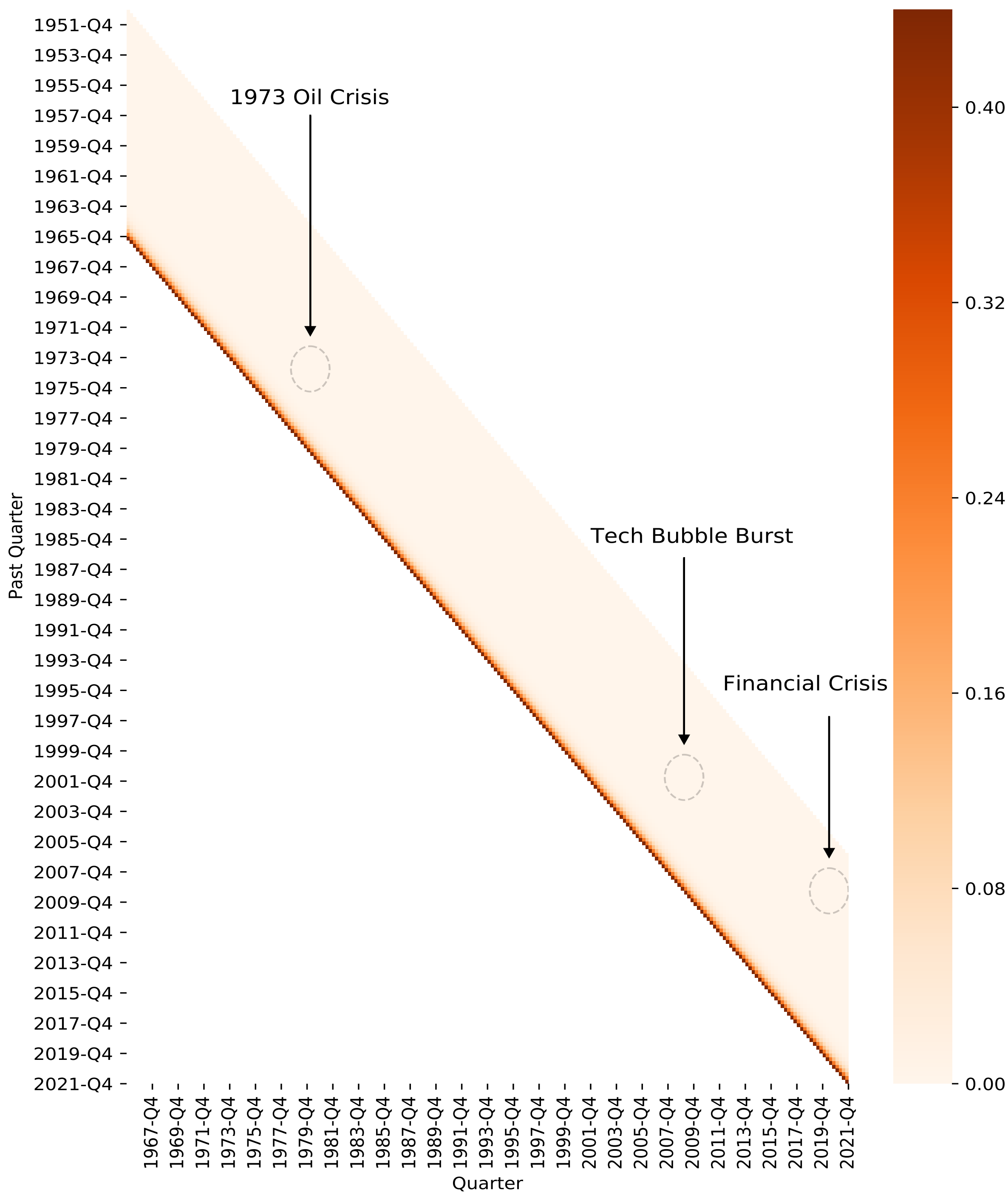

Figure 2: Recall Probabilities from a Simple Model of Exponential Decay

This figure displays the recall probability that a simple model of exponential decay assigns to the past 60 quarters (= 15 years) for each quarter from 1966 to 2021. The probability weight assigned to the j-lagged quarter is given by:

$$\omega_j = \frac{\lambda^j}{\Sigma_{k=0}^{59} \lambda^k},$$

where $\lambda = 0.56$ following Greenwood and Shleifer (2014).

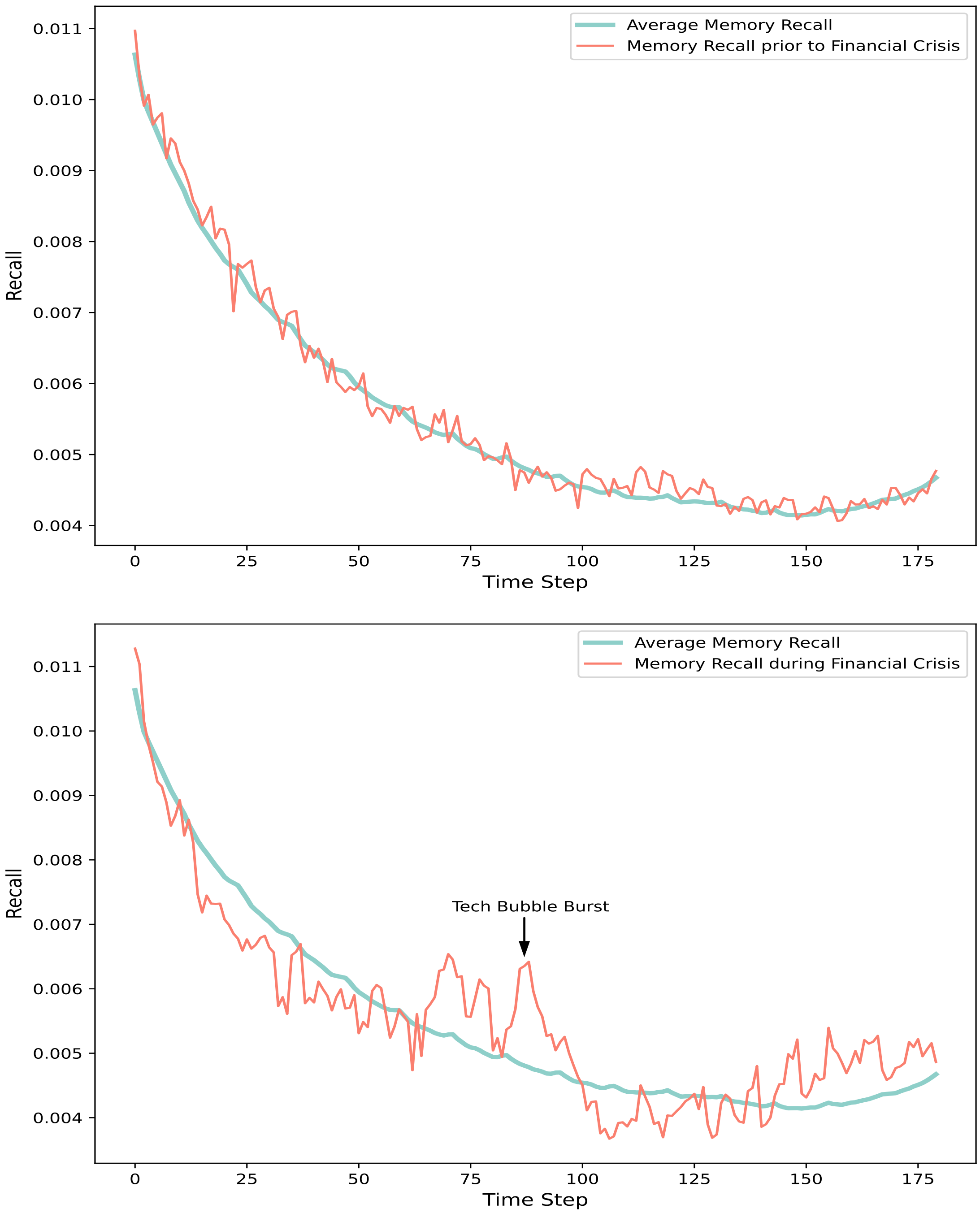

Figure 3: Recall Probabilities from our Measure: Before and During the 2008 Financial Crisis

This figure displays recall probabilities implied by our measure immediately before and during the 2008 Financial Crisis. The x-axis shows the number of months between the cueing month and the recalled month, while the y-axis represents the recall probability assigned to a historical month by our measure. In the upper panel, we present recall probabilities of past months if recall occurs in June 2007 (in red), which is a date that is immediately before the outbreak of the 2008 Financial Crisis. In the lower panel, we present recall probabilities of past months if recall occurs in December 2008 (in red), which is a date that is during the 2008 Financial Crisis. In both panels, we also display the average recall probabilities of past months in green.

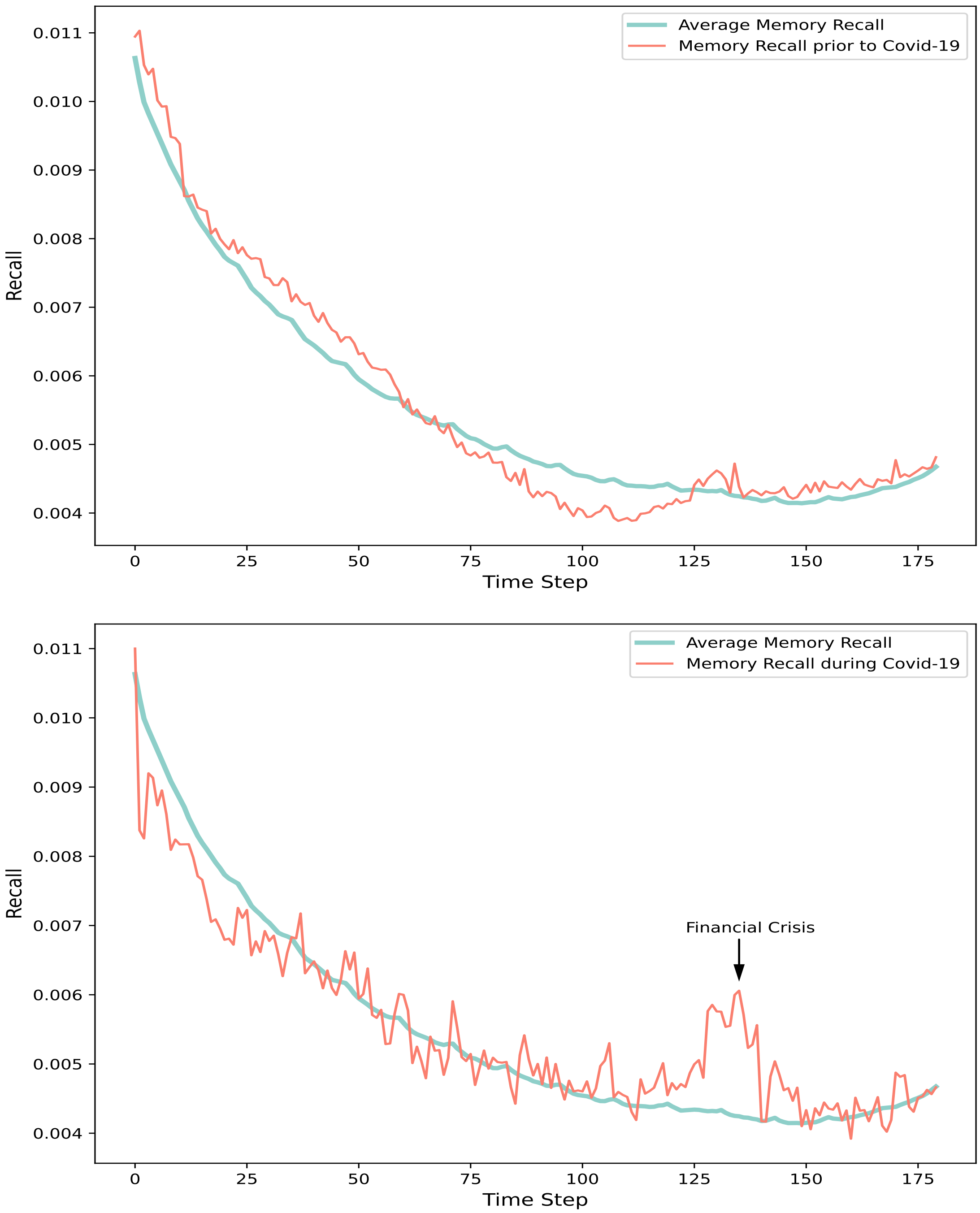

Figure 4: Recall Probabilities from our Measure: Before and During the Covid-19 Pandemic

This figure displays recall probabilities implied by our measure immediately before and during the 2008 Financial Crisis. The x-axis shows the number of months between the cueing month and the recalled month, while the y-axis represents the recall probability assigned to a historical month by our measure. In the upper panel, we present recall probabilities of past months if recall occurs in December 2019 (in red), which is a date that is immediately before the outbreak of the 2020 Covid-19 Pandemic. In the lower panel, we present recall probabilities of past months if recall occurs in June 2020 (in red), which is a date that is during the 2020 Covid-19 Pandemic. In both panels, we also display the average recall probabilities of past months in green.

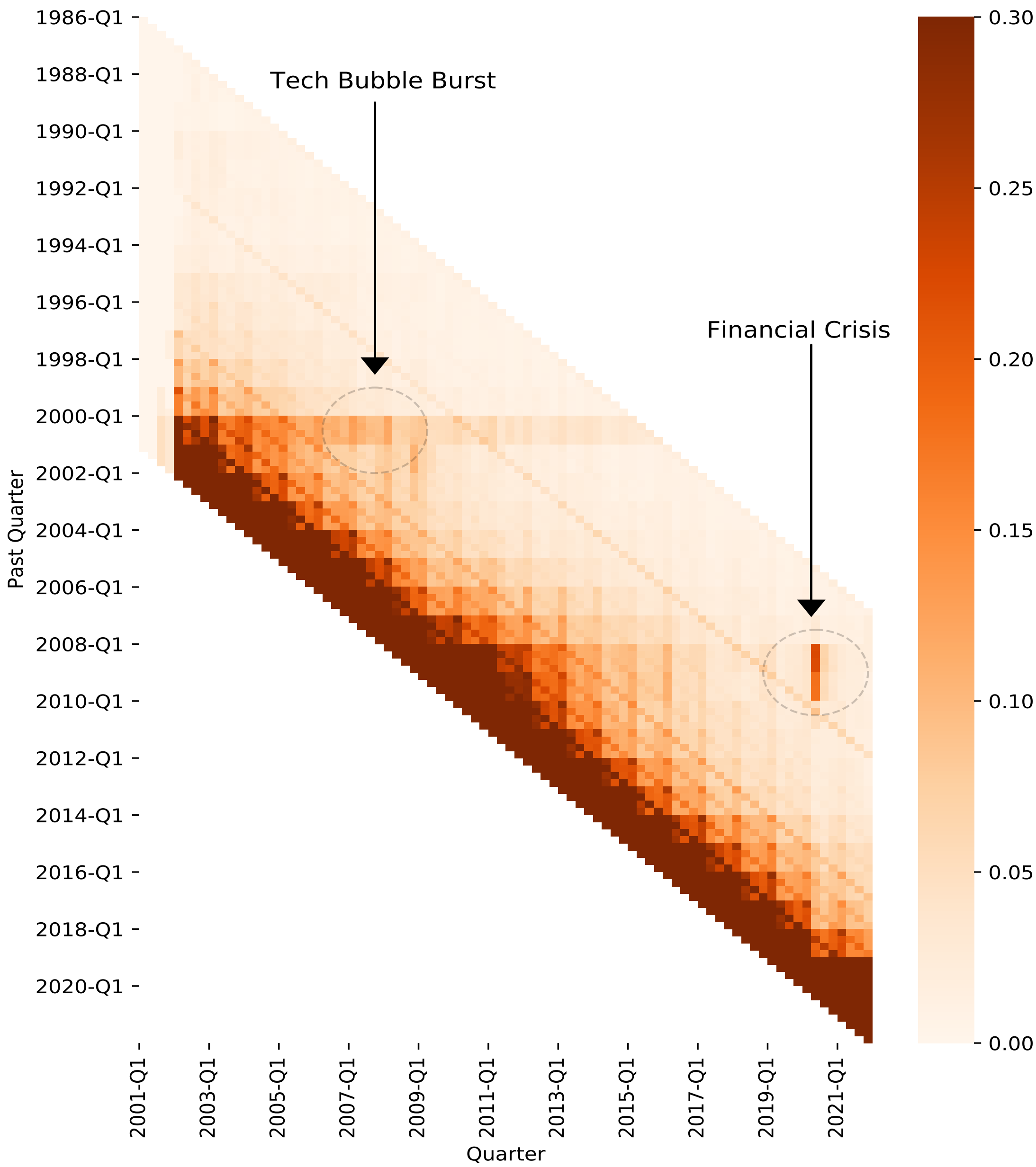

Figure 5: Actual Recall Patterns Extracted from Transcripts of Corporate Events

This figure displays the probability that a previous quarter is mentioned during a corporate event. The sample includes corporate events taking place between 2001 and 2021. The x-axis represents the quarter in which the corporate event occurs, while the y-axis represents the probability with which each of the past 60 quarters (= 15 years) is mentioned during the event. A darker shade of red indicates a higher probability weight, and we require the darkest color to be capped at a probability of 0.3.

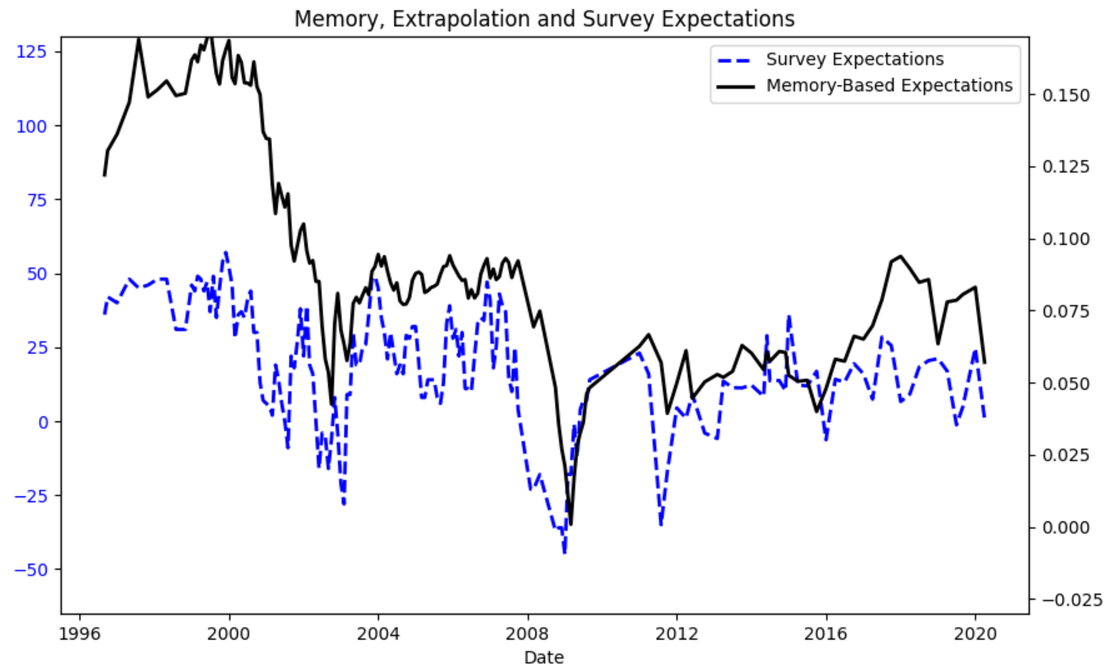

Figure 6: Memory-Based Survey-Based Return Expectations

This figure plots memory-based return expectations (as of the end of month t) and survey-based return expectations (elicited over the course of month $t+1$). Survey-based return expectations are the percentage point difference in bullish and bearish investors from the Gallup investor survey. We source survey data from Greenwood and Shleifer (2014) for the period from October 1996 to December 2011, and extend the sample period to May 2020 with data sourced directly from Gallup. Memory-based return expectations are annualized return expectations of the S&P 500 index derived from our measure.

Table 1: Summary Statistics for Tests of Actual Recall Patterns Extracted from Corporate Events

This table presents summary statistics for the sample used in the tests displayed in Table 2. The sample covers corporate events taking place between January 2001 and December 2021. Actual Recall is an indicator that equals one if, during a firm's corporate event in month t, any participant mentions the historical month $t-h$ at least once, and zero otherwise. We focus on the previous 180 months, i.e., $h \in \{1, 2, 3, \ldots, 180\}$. The recall probability $r(e, \kappa_t)$ captures the probability that month $t-h$ is recalled by a representative investor in month t. Exponential weight is the probability weight on month $t-h$ derived from a simple model of exponential decay following Greenwood and Shleifer (2014) using $\lambda = 0.56$. Size is the logarithm of a firm's market capitalization (in million \$). Turnover is the monthly dollar trading volume over market capitalization at the end of the month. The firm's book-to-market ratio (BM) is constructed following Fama and French (1992), using book equity from (at least) six months ago and market capitalization from the most recent December. Ivol is idiosyncratic volatility (in %) from CAPM regressions and is constructed following Ang et al. (2006), using daily data from the past month. Price is the firm's stock price (in \$) at the end of the month.

	N	Mean	Median	Std.Dev	P25	P75	Min	Max
Actual Recall	15,237,765	0.163	0.000	0.369	0.000	0.000	0.000	1.000
$r(e, \kappa_t)$	15,237,765	0.006	0.005	0.002	0.004	0.007	0.001	0.014
Exponential Weight	15,237,765	0.024	0.000	0.102	0.000	0.000	0.000	1.000
Size	15,237,765	7.569	7.580	1.986	6.277	8.906	0.591	14.659
Turnover	15,237,765	0.176	0.133	0.148	0.084	0.216	0.001	0.836
BM	15,237,765	0.601	0.505	0.693	0.293	0.791	-26.782	27.284
Ivol	15,237,765	0.018	0.014	0.015	0.009	0.021	0.001	0.655
Price	15,237,765	44.912	29.590	77.195	14.220	52.950	0.054	3440.160

Table 2: Actual Recall Patterns Extracted from Corporate Events

This table shows that our measure predicts actual recall patterns extracted from transcripts of corporate events. The dependent variable is an indicator that equals one if, during a firm's corporate event in month t, any participant mentions the historical month $t-h$ at least once, and zero otherwise. The recall probability $r(e, \kappa_t)$ captures the probability that month $t-h$ is recalled by a representative investor in month t. Exponential weight captures the probability weight on month $t-h$ derived from a simple model of exponential decay following Greenwood and Shleifer (2014) using λ = 0.56. The first four columns focus on the previous 180 months, i.e., $h \in \{1, 2, 3, \ldots, 180\}$. The last four columns focus on months that are at least 5 years in the past, i.e., $h \in \{61, 62, 63, \ldots, 180\}$. Columns (1) to (3) and (4) to (6) include stock-by-past-month fixed effects as well as control variables. The control variables are as of the end of month $t-1$, given that recall occurs in month t. Columns (4) and (8) further include stock-by-current-month fixed effects. These fixed effects soak up the control variables. In all columns, we include quarter fixed effects. Standard errors are clustered by stock, current month, and past month, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Actual Recall of Month *t-h* in Month *t*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	Past 15 Years				At Least 5 Years in the Past			
$r(e, \kappa_t)$	66.909***		65.052***	67.267***	2.952***		2.908***	3.860***
	(41.47)		(35.30)	(35.89)	(7.67)		(7.52)	(8.86)
Exponential Weight		448.540	301.707	294.847		0.000	0.000	0.000
		(1.58)	(1.52)	(1.52)		(1.52)	(1.50)	(1.52)
Size	0.011***	0.011***	0.011***		0.004***	0.004***	0.004***	
	(6.63)	(7.87)	(6.59)		(4.25)	(4.19)	(4.24)	
Turnover	0.002	0.005	0.002		0.012***	0.012***	0.012***	
	(0.46)	(1.05)	(0.48)		(3.06)	(3.14)	(3.06)	
BM	0.001	0.001	0.001		-0.000	-0.000	-0.000	
	(1.15)	(0.91)	(1.12)		(-0.41)	(-0.45)	(-0.41)	
Ivol	-0.081**	-0.131***	-0.082**		-0.049*	-0.049*	-0.049*	
	(-2.23)	(-3.58)	(-2.26)		(-1.66)	(-1.67)	(-1.66)	
Price	0.000*	-0.000	0.000*		0.000	0.000	0.000	
	(1.91)	(-1.26)	(1.87)		(0.63)	(0.39)	(0.63)	
Adjusted R-Squared	0.405	0.361	0.408	0.457	0.083	0.083	0.084	0.212
N	15,237,765	15,237,765	15,237,765	15,262,267	10,152,541	10,152,541	10,152,541	10,168,898
Stock x current month FE	NO	NO	NO	YES	NO	NO	NO	YES
Stock x past month FE	YES	YES	YES	YES	YES	YES	YES	YES
Quarter FE	YES	YES	YES	YES	YES	YES	YES	YES

Table 3: Summary Statistics for Tests of Memory-Based Beliefs

This table presents summary statistics for the samples used in the tests displayed in Tables 4 and 5. Panel A displays summary statistics for the sample used in Table 4. The sample period for these tests ranges from October 1996 to May 2020. We source Gallup Survey Expectations from Greenwood and Shleifer (2014) until December 2011, and extend the sample until May 2020 with data sourced directly from Gallup. Gallup Survey Expectations are defined as the percentage point difference in bullish and bearish investors from the Gallup investor survey. Memory-Based Expectation is the annualized return expectation for the S&P 500 index derived from our measure. We decompose the memory-based expectation into expectations from recent and distant memory, where recent refers either to the most recent 12 months or the most recent five years, and distant refers either to more than 12 months or more than five years in the past. The sum of recent and distant expectations equals the baseline memory-based expectation. The cumulative return over the previous 12 months, the exponentially-weighted past return, the logarithm of the price-dividend ratio (Log(P/D)), the risk-free rate, the earnings growth rate, and the unemployment rate are constructed following Greenwood and Shleifer (2014). US Rec is a dummy variable that is equal to one during NBER recessions. Panel B displays summary statistics for the sample used in Table 5. The sample period for these tests ranges from January 1996 to December 2021. Option-implied volatility is constructed following An et al. (2014). Memory-based volatility is the standard deviation of monthly returns over the past 180 months, where each historical return is weighted with its associated recall probability. Exponential-decay-based volatility is the standard deviation of monthly returns over past months, where each historical return is weighted with exponentially-decaying weights, using the decay parameter $\lambda = 0.56$ from Greenwood and Shleifer (2014). Size is the logarithm of a firm's market capitalization (in million $), idiosyncratic volatility (in %) is constructed following Ang et al. (2006), asset growth is the book asset growth rate (= current book assets – lagged book assets)/lagged book assets), operating profit is constructed following Fama and French (2006), and the book-to-market ratio is constructed following Fama and French (1992).

Panel A: Return Expectations								
	N	Mean	Median	Std.Dev	P25	P75	Min	Max
Gallup Survey Expectations	171	18.079	18.303	20.636	7.000	34.000	-45.000	57.000
Memory-Based Expectation	171	0.091	0.085	0.039	0.063	0.105	0.001	0.174
Exp from Recent Mem (<=12 months)	171	0.005	0.011	0.021	-0.007	0.019	-0.065	0.048
Exp from Distant Mem (>12 months)	171	0.086	0.076	0.033	0.066	0.117	0.021	0.150
Exp from Recent Mem (<=5 years)	171	0.033	0.030	0.038	-0.000	0.060	-0.055	0.110
Exp from Distant Mem (>5 years)	171	0.058	0.063	0.026	0.045	0.076	-0.018	0.092
Cumulative Return (Past 12 Months)	171	0.066	0.109	0.178	-0.035	0.185	-0.425	0.409
Exponentially-Weighted Past Return	171	0.019	0.026	0.034	0.007	0.039	-0.097	0.077
US Rec	171	0.123	0.000	0.329	0.000	0.000	0.000	1.000
Log(P/D)	171	4.059	4.041	0.235	3.941	4.202	3.324	4.502
Risk-free Rate	171	1.005	1.007	0.019	0.994	1.016	0.950	1.101
Earnings Growth	171	0.042	0.134	0.330	-0.055	0.195	-0.886	0.767
Unemployment	171	5.455	5.000	1.576	4.400	5.800	3.500	14.700

Panel B: Volatility Perceptions								
	N	Mean	Median	Std.Dev	P25	P75	Min	Max
Option-Implied Volatility	410,569	0.454	0.402	0.221	0.299	0.553	0.100	1.473
Memory-Based Volatility	410,569	0.480	0.437	0.209	0.328	0.584	0.159	1.447
Exponential-Decay-Based Volatility	410,569	0.357	0.291	0.248	0.194	0.440	0.022	1.609
Size	410,569	7.537	7.342	1.566	6.385	8.473	3.704	14.813
Idiosyncratic Volatility (3F)	410,569	0.020	0.016	0.014	0.011	0.025	0.000	1.383
Asset Growth	410,569	0.216	0.078	0.901	-0.004	0.221	-0.937	233.36
Operating Profit	410,569	0.255	0.254	5.082	0.144	0.381	-553.75	268.35
log(Book-to-Market Ratio)	410,569	-0.976	-0.903	0.851	-1.431	-0.432	-9.872	3.175

Table 4: Memory-Based and Survey-Based Return Expectations of the S&P 500 Index

This table shows that memory-based return expectations explain survey-based return expectations. In all columns, the dependent variable is the difference in the percentage of bullish and bearish investors from the Gallup investor survey (elicited over the course of month $t + 1$). In column (1), the main independent variable is the memory-based return expectation of the S&P 500 Index derived from our measure (as of the end of month t). In column (2), we decompose this memory-based expectation into expectations from recent memory (most recent 12 months) and expectations from distant memory (more than 12 months in the past). The sum of these two variables equals our baseline memory-based expectation from column (1). In column (3), we add the cumulative return over the past 12 month as an additional control, following Greenwood and Shleifer (2014). In column (4), we control for the exponentially-weighted average return over the past five years. To construct this exponentially-weighted return, we first calculate quarterly returns by compounding 3-month returns on a monthly basis. We then use the weighting approach of Greenwood and Shleifer (2014) and their estimated quarterly λ of 0.77 to calculate an exponentially-weighted average return over the past five years. Columns (5) - (7) mirror columns (2) - (4), except that the cutoff between recent and distant memory is five years instead of 12 months. In column (8), we interact expectations from recent and distant memory with a dummy variable equal to one during NBER recessions. All columns include the same control variables as Panel B of Table 3 in Greenwood and Shleifer (2014). The sample period ranges from October 1996 to May 2020. t-statistics, in parentheses, are Newey-West adjusted with twelve lags. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Gallup Survey Expectations							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Memory-Based Expectation	371.970***							
	(4.97)							
Exp from Recent Mem (<=12 months)		581.296***	-87.026	-43.262				
		(7.33)	(-0.69)	(-0.36)				
Exp from Distant Mem (>12 months)		249.131***	203.420***	168.358**				
		(3.76)	(3.84)	(2.55)				
Exp from Recent Mem (<=5 years)					382.190***	114.281	34.740	19.974
					(5.23)	(1.57)	(0.42)	(0.20)
Exp from Distant Mem (>5 years)					333.652***	244.758***	220.195***	178.967***
					(3.90)	(3.76)	(3.29)	(2.69)
Exp from Recent Mem (<=5 years) X US Rec								-294.003***
								(-3.57)
Exp from Distant Mem (>5 years) X US Rec								817.134***
								(3.22)
Cumulative Return (Past 12 Months)			84.441***			70.444***		17.069
			(5.20)			(5.10)		(1.20)
Exponentially-Weighted Past Return				368.441***			389.757***	331.594***
				(5.39)			(6.60)	(4.13)
Log(P/D)	-8.713	3.333	11.488	9.409	-8.046	15.132	16.026	23.272*
	(-0.54)	(0.25)	(1.00)	(0.71)	(-0.50)	(1.46)	(1.52)	(1.70)
Risk-free Rate	-189.556***	-169.016***	-116.927**	-110.157**	-189.184***	-114.985**	-93.143*	-61.654
	(-3.38)	(-2.86)	(-2.53)	(-2.35)	(-3.37)	(-2.37)	(-1.94)	(-1.32)
Earnings Growth	19.826***	5.441	-0.344	8.646	20.440***	-9.144	-4.041	-4.852
	(3.00)	(0.76)	(-0.05)	(1.10)	(2.89)	(-1.13)	(-0.59)	(-0.58)
Unemployment	0.543	0.369	-0.131	-0.242	0.704	-0.578	-1.018	-0.935
	(0.47)	(0.31)	(-0.12)	(-0.19)	(0.59)	(-0.53)	(-0.79)	(-0.57)
Adjusted R-Squared	0.553	0.594	0.640	0.650	0.554	0.648	0.673	0.707
N	171	171	171	171	171	171	171	171

Table 5: Memory-Based Volatility and Option-Implied Volatility

This table shows that memory-based volatility explains option-implied volatility in the cross-section of stocks. In all columns, the dependent variable is option-implied volatility in month $t + 1$, constructed following An et al. (2014). All independent variables are as of month t. Memory-based volatility is the standard deviation of monthly returns over the past 180 months, where each historical return is weighted with its associated recall probability. Similarly, exponential-decay-based volatility is the standard deviation of monthly returns over past months, where each historical return is weighted with exponentially-decaying weights, using the decay parameter $\lambda = 0.56$ from Greenwood and Shleifer (2014). In column (1), we only include memory-based volatility as an independent variable. In column (2), we control for exponential-decay-based volatility. In column (3), we add stock and month fixed effects. In column (4) we control for lagged option-implied volatility, and in column (5) we control for size, idiosyncratic volatility (following Ang et al. (2006)), asset growth, operating profit (following Fama and French (2006)), and the logarithm of the book-to-market ratio (following Fama and French (1992)). We multiply all coefficients by 100. Standard errors are clustered by stock and month, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Option-Implied Volatility				
	(1)	(2)	(3)	(4)	(5)
Memory-Based Volatility	61.706***	37.351***	34.761***	11.169***	7.626***
	(38.94)	(25.22)	(21.57)	(14.73)	(10.90)
Exponential-Decay-Based Volatility		33.089***	15.222***	4.866***	4.442***
		(24.93)	(26.70)	(17.46)	(17.32)
Option-Implied Volatility (lagged)				63.153***	57.800***
				(69.59)	(59.85)
Size					-2.231***
					(-15.00)
Idiosyncratic Volatility (3F)					102.611***
					(20.19)
Asset Growth					0.175***
					(2.85)
Operating Profit					-0.006
					(-1.25)
log(Book-to-Market Ratio)					-0.306***
					(-4.06)
Adjusted R-Squared	0.342	0.427	0.710	0.819	0.824
N	410,569	410,569	410,569	410,569	410,569
Month FE	NO	NO	YES	YES	YES
Stock FE	NO	NO	YES	YES	YES

Table 6: Summary Statistics for Tests of Memory-Based Trading Decisions

This table presents summary statistics for the samples used in the tests displayed in Tables 7, 8, and 9. The data in these tests are the same as in Barber and Odean (2000) and the sample period ranges from January 1991 to November 1996. Number of Stocks for Repurchase is the number of distinct stocks that the investor once owned but does not own in month t. Number of Stocks in Portfolio is the number of distinct stocks that the investor owns in month t. These two variables are at the account-month level. Repurchase Dummy (Buy) is a dummy variable that is equal to one if the investor repurchases a previously-held stock in month t. Ret is the return that the investor realized when liquidating a previously-held position. The recall probability $r(e, \kappa_t)$ is the probability that a representative investor in month t recalls the month in which the investor liquidated a previously-held position. In contrast, the recall probability $r_i(e, \kappa_t)$ is investor-specific and estimated for each investor individually using the investor's historical holdings. Return between Sell and Repurchase is the return that a stock realized between the previous liquidation and a (potential) repurchase. The logarithm of the initial purchase price (ln(WAPP)), the square root of the number of days between initial purchase and liquidation ($\sqrt{\text{Time Owned}}$), and the volatility calculated using daily returns over the 250 days preceding the initial purchase (Return Volatility) are constructed following Ben-David and Hirshleifer (2012).

	N	Mean	Median	Std.Dev	P25	P75	Min	Max
Number of Stocks for Repurchase	1,096,908	5.840	4.000	7.902	2.000	6.000	1.000	440.000
Number of Stocks in Portfolio	1,096,908	2.744	1.000	5.673	0.000	4.000	0.000	645.000
Repurchase Dummy (Buy)	6,413,970	0.005	0.000	0.070	0.000	0.000	0.000	1.000
Ret	6,413,970	0.069	0.050	0.321	-0.102	0.196	-0.706	1.429
$r(e, \kappa_t)$	6,126,108	0.013	0.010	0.012	0.008	0.013	0.001	0.582
$r_i(e, \kappa_t)$	6,116,040	0.046	0.030	0.045	0.020	0.053	0.003	0.687
Return between Sell and Repurchase	6,413,970	0.323	0.093	1.084	-0.134	0.483	-0.999	138.342
$\ln$(WAPP)	6,413,970	2.389	2.464	1.184	1.764	3.029	-3.814	15.023
$\sqrt{\text{Time Owned}}$	6,413,970	2.388	2.000	1.239	1.414	3.162	1.000	8.185
Return Volatility	6,413,970	0.140	0.123	0.098	0.082	0.171	0.012	6.953

Table 7: Memory Retrieval and Repurchasing Decisions

This table shows that the probability of recalling a past trading experience strongly modulates the likelihood of repurchasing a stock. The dependent variable in all columns is a dummy that is equal to one if the investor repurchases a previously-held stock. The main independent variables are (i) the return that the investor realized when liquidating a previously-held position, (ii) the probability that a representative investor in month t recalls the month in which the investor liquidated a previously-held position, and (iii) the interaction of (i) and (ii). All columns include a set of control variables as well as stock-by-current-month fixed effects. Columns (1) and (2) further include account-by-current-month fixed effects, and column (3) includes account-by-liquidation-month-by-current-month fixed effects instead. All coefficients are multiplied by 100. Standard errors are clustered by account, liquidation month, and current month, and the t-statistics are reported in parentheses. *, **, and ***, indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Repurchase Dummy (Buy)		
	(1)	(2)	(3)
Ret	0.342***	-0.006	0.018
	(8.93)	(-0.18)	(0.37)
$r(e, \kappa_t)$		24.872***	5.208***
		(21.16)	(2.85)
$r(e, \kappa_t)$ x Ret		20.210***	16.327***
		(10.77)	(6.42)
Return between Sell and Repurchase	-0.145***	-0.098***	-0.093***
	(-7.41)	(-6.14)	(-5.16)
$\ln$(WAPP)	0.259***	0.152***	0.103**
	(7.51)	(5.53)	(2.60)
$\sqrt{\text{Time Owned}}$	0.010	-0.004	-0.009
	(1.55)	(-0.72)	(-0.97)
Return Volatility	-0.563***	-0.331***	0.020
	(-4.68)	(-3.36)	(0.13)
Adjusted R-Squared	0.036	0.027	0.038
N	6,413,970	6,126,108	2,842,767
Stock x Current Month FE	YES	YES	YES
Account x Current Month FE	YES	YES	NO
Account x Past Month x Current Month FE	NO	NO	YES

Table 8: The Role of Similarity and Interference

This table shows the separate effect of investor-specific recall probabilities, similarity, and interference on the likelihood of repurchasing a stock. The dependent variable in all columns is a dummy that is equal to one if the investor repurchases a previously-held stock. The main independent variables in columns (1) and (3) are (i) the return that the investor realized when liquidating a previously-held position, (ii) the probability that the investor in month t recalls the month in which the previously-held position was liquidated, and (iii) the interaction of (i) and (ii). Columns (2) and (4) break out the investor-specific recall probability separately by similarity and interference. All columns include a set of control variables as well as stock-by-current-month fixed effects. Columns (1) and (2) further include account-by-current-month fixed effects, and columns (3) and (4) include account-by-liquidation-month-by-current-month fixed effects instead. All coefficients are multiplied by 100. Standard errors are clustered by account, liquidation month, and current month, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Repurchase Dummy (Buy)			
	(1)	(2)	(3)	(4)
Ret	-0.173***	0.188***	-0.144***	0.242**
	(-5.77)	(3.18)	(-3.20)	(2.62)
$r_i(e, \kappa_t)$	3.676***		-5.305***	
	(18.44)		(-8.14)	
$r_i(e, \kappa_t)$ x Ret	10.593***		8.824***	
	(14.29)		(7.38)	
Similarity		1.296***		-0.067
		(17.04)		(-1.14)
Interference		0.008***		0.038***
		(5.32)		(11.46)
Similarity x Ret		0.768***		0.629***
		(9.70)		(5.52)
Interference x Ret		-0.019***		-0.019***
		(-11.13)		(-6.79)
Return between Sell and Repurchase	-0.101***	-0.076***	-0.101***	-0.091***
	(-6.32)	(-5.60)	(-5.42)	(-5.05)
$\ln(\text{WAPP})$	0.088***	0.108***	0.080**	0.094**
	(3.64)	(4.36)	(2.16)	(2.51)
$\sqrt{\text{Time Owned}}$	0.031***	-0.041***	-0.067***	-0.126***
	(6.57)	(-6.13)	(-5.32)	(-8.45)
Return Volatility	-0.278***	-0.151	0.088	0.040
	(-2.84)	(-1.57)	(0.53)	(0.25)
Adjusted R-Squared	0.027	0.027	0.037	0.037
N	6,116,040	6,116,040	2,836,957	2,836,957
Stock x Current Month FE	YES	YES	YES	YES
Account x Current Month FE	YES	YES	NO	NO
Account x Past Month x Current Month FE	NO	NO	YES	YES

Table 9: The Role of Encoding Strength

This table shows that the strength with which a trading experience is encoded affects the probability of recalling the experience, and ultimately the likelihood of repurchasing the associated stock. Columns (1) and (2) augment columns (1) and (3) of Table 8, respectively, with an additional interaction with the dummy variable Attention. This dummy is equal to one if the investor executed at least two transactions in the month of the experience, and zero otherwise. All coefficients are multiplied by 100. Standard errors are clustered by account, liquidation month, and current month, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Repurchase Dummy (Buy)	
	(1)	(2)
Attention x $r_i(e, \kappa_t)$ x Ret	5.109***	5.517***
	(4.33)	(3.27)
$r_i(e, \kappa_t)$ x Ret	8.093***	5.472***
	(12.43)	(5.02)
Attention x $r_i(e, \kappa_t)$	0.299	-3.139***
	(1.00)	(-3.00)
Attention x Ret	-0.047	-0.012
	(-1.05)	(-0.17)
Attention	-0.014	
	(-1.17)	
Ret	-0.134***	-0.103**
	(-5.14)	(-2.08)
$r_i(e, \kappa_t)$	3.544***	-3.323***
	(14.28)	(-4.23)
Return between Sell and Repurchase	-0.100***	-0.100***
	(-6.30)	(-5.39)
$\ln(\text{WAPP})$	0.091***	0.085**
	(3.75)	(2.28)
$\sqrt{\text{Time Owned}}$	0.031***	-0.063***
	(6.55)	(-5.12)
Return Volatility	-0.279***	0.103
	(-2.86)	(0.62)
Adjusted R-Squared	0.027	0.038
N	6,116,040	2,836,957
Stock x Current Month FE	YES	YES
Account x Current Month FE	YES	NO
Account x Past Month x Current Month FE	NO	YES

APPENDIX

A Construction of the Corporate Event Sample

In this section, we describe how we construct the sample that we use in our tests in Section 4.4 of the paper. We collect transcripts of corporate events from Refinitiv StreetEvents for January 2001 to December 2021. The set of corporate events covered by these transcripts includes Earnings Calls, M&A Calls, Sales Calls, Analyst Meetings as well as Corporate Conference Presentations. Q&A sessions of these calls are also included in the transcripts. Each transcript of an event provides a verbatim representation of what was spoken and by whom during the event, along with metadata that allows us to match the company to the Compustat database. This metadata includes the ticker symbol header, company name, event title, and date of the event. We match company names in the transcripts to the corresponding 'GVKEY' (the unique Compustat identifier) following Li et al. (2021). We also match the sample to CRSP and retain only stocks with a share code of 10 or 11 and an exchange code of 1, 2, or 3.

We parse the text of each transcript and segment it into sentences. For each sentence, we identify the dates mentioned in it using *sutime*, a Python wrapper for Stanford CoreNLP's SUTime Java library.[24] *sutime* allows us to identify date formats in both absolute and relative forms. Absolute date formats are phrases that directly represent a specific date, such as "2011-03-31" or "2011-Q3". When dealing with absolute dates, *sutime* can interpret date strings in various formats; it recognizes that strings like "2011-03-31" and "March 31, 2011" are equivalent. In contrast, relative date formats involve phrases that indirectly convey a date, such as "two weeks ago". For instance, if a conference call occurs on December 31, 2011, and the phrase "two weeks ago" is mentioned, *sutime* codes the corresponding date as December 17, 2011. *sutime* labels both absolute and relative date formats as belonging to the category *date*.

There are other forms of date-related terms that *sutime* can detect. For instance, in a sentence containing the phrase "century long commitment", *sutime* would classify the term "century long" as belonging to the category *duration*. We disregard sentences with terms falling into this category. In our sample, we focus only on dates from sentences with phrases that are grouped into the *date* category.

[24]Link: https://pypi.org/project/sutime/

The identified dates mainly fall into three different frequencies: days (e.g., 2011-03-31), quarters (e.g., 2011-Q3), and years (e.g., 2011). To ensure consistency with the monthly frequency of our measure of memory associations, we aggregate dates to the monthly level. Specifically, we assume that when referring to a date of lower frequency, all corresponding months within that date are included. For instance, if a conference participant mentions the year 2011, we treat it as if every month in 2011 was mentioned. Moreover, since our focus is on memory recall, we only focus on dates that occurred before the date of the corporate event, and ignore references to the future.

B Variation in the Speed of Memory Decay

Our goal in this section is to show that our measure can generate patterns that are typically observed in subjective beliefs data. Specifically, our measure generates recall patterns that are largely consistent with empirical patterns on return extrapolation documented in Da et al. (2021). In this study, the authors show that investors extrapolate from past returns, placing higher weights on recent returns compared to distant returns. Moreover, these extrapolative weights decay faster for small firms and value stocks. Here, we show that our measure generates variation in the speed of memory decay in the cross-section of stocks that is very consistent with the findings in Da et al. (2021).

To construct our measure of decay speed, we first calculate the recall probability of the most recent month if recall occurs in the current month. Then, we calculate the recall probability of the same (most recent) month assuming that recall occurs one year from now. We construct the decay speed as the ratio of the two recall probabilities, with the first recall probability in the numerator and the second recall probability in the denominator. A higher ratio indicates a faster decay speed.

We show how decay speed varies along four firm-level characteristics, which are each measured as of the end of each month: the logarithm of a firm's market capitalization (in million \$), the book-to-market ratio following Fama and French (1992), the idiosyncratic volatility (in %) from CAPM regressions following Ang et al. (2006), and the stock price (in \$). At the end of each month, we sort stocks into deciles based on these characteristics and calculate both the equal-weighted and value-weighted average of decay speed.

Table A4 presents our findings. The first row of Panel A shows that, when weighting stocks equally in each portfolio, the decay speed of the smallest firms is 10.7% (= 1.34/1.21 - 1) faster than that of the largest firms. We find differences of a similar magnitude when we compare stocks in the extreme deciles of idiosyncratic volatility and stock price, as shown in the third and fourth rows, with memory decaying faster for volatile and low-priced stocks. We also find that decay speed is lower for growth stocks, but the magnitude of this effect is smaller. Panel B shows similar results when we value-weight stocks in each portfolio.

Overall, the consistency between our findings and those of Da et al. (2021) suggests a strong link between memory and return extrapolation. We view the results of these tests as suggestive evidence that our measure is linked with subjective beliefs data from the field. In section 5.2 of the paper, we empirically establish the connection between memory-based and survey-based return expectations.

C Linking Memory-Based Perceived Volatility and the VIX

In this section, we use our measure to construct memory-based perceived volatility for the aggregate market and link it to the actual perceived volatility captured by the Volatility Index (VIX). Since our measure yields a full probability distribution of monthly returns for each stock, we can use these distributions to construct a measure of memory-based return variance for each stock in each month. Using the resulting stock-level variances, we construct monthly memory-based variance at the market-level as follows:

$$\mathrm{Var}^S(\mathrm{Ret}_m) = \underbrace{\sum_{l=1}^{L} \omega_l^2 \mathrm{Var}^S(\mathrm{Ret}_l)}_{\text{sum of variances}} + \underbrace{\sum_{l \neq k} \omega_l \omega_k \mathrm{Cov}^S(\mathrm{Ret}_l, \mathrm{Ret}_k)}_{\text{sum of covariances}}, \tag{A1}$$

where ω_l is the weight of stock l, which is simply its value-weight in the S&P 500 index, and $\mathrm{Var}^S(\mathrm{Ret}_l)$ is the memory-based monthly return variance of stock l, constructed from the memory-based return distribution of stock l as of the end of month t. Further, $\mathrm{Cov}^S(\mathrm{Ret}_l, \mathrm{Ret}_k)$ is the memory-based perceived return covariance between stocks l and k. As discussed in Section 3, we adopt the concept of narrow framing at the stock-level throughout the paper, and therefore assume that investors perceive $\mathrm{Cov}^S(\mathrm{Ret}_l, \mathrm{Ret}_k)$ to be zero. Finally, the perceived monthly volatility at the market-level is simply the square root of $\mathrm{Var}^S(\mathrm{Ret}_m)$. We annualize this volatility by multiplying it with the square root of twelve.

To connect this memory-based perceived volatility with actual perceived volatility, we need a proxy for the latter. We use the VIX. This is a natural choice, as the VIX is designed to capture the market's annualized expectation of volatility over the next 30 days. We download the VIX from the website of the Federal Reserve Bank of St. Louis, calculate the average of daily VIX over the course of each month, and normalize it by dividing it by 100. We provide summary statistics of the VIX and memory-based volatility in Panel B of Table A6.

We begin by visualizing the relationship between memory-based volatility and the VIX in Figure A5. The dashed blue line represents the average daily VIX over the course of month $t + 1$, while the solid black line represents memory-based perceived volatility as of the end of month t. The two time series track each other well, with the VIX being more volatile over the sample period. The correlation between the two time series is 0.37 (the rank correlation is 0.50).

Next, we regress the VIX (from month $t + 1$) on memory-based perceived volatility (as of the end of month t) and present the results in Table A7. Column (1) shows that memory-based volatility predicts the VIX significantly with a positive sign. In terms of magnitude, a one standard deviation increase in memory-based volatility is associated with an increase in the VIX of 0.026 units, which corresponds to about 35% of the standard deviation of the VIX. In column (2), we add the same control variables as in column (1) of Table 4. Adding these controls approximately doubles the magnitude of the coefficient, implying that a one standard deviation increase in memory-based volatility is associated with an increase in the VIX by about 65% of its standard deviation. In column (3), we also control for the realized volatility over the past 12 months, to account for the possibility that investors extrapolate from past volatility in the same way that they do from past returns. While the coefficient on memory-based volatility loses some of its statistical significance, the economic magnitude of the effect remains very similar to the estimate from column (1). Overall, our results in Table A7 show that memory can not only explain return expectations, but also higher-moment beliefs, such as volatility perceptions.

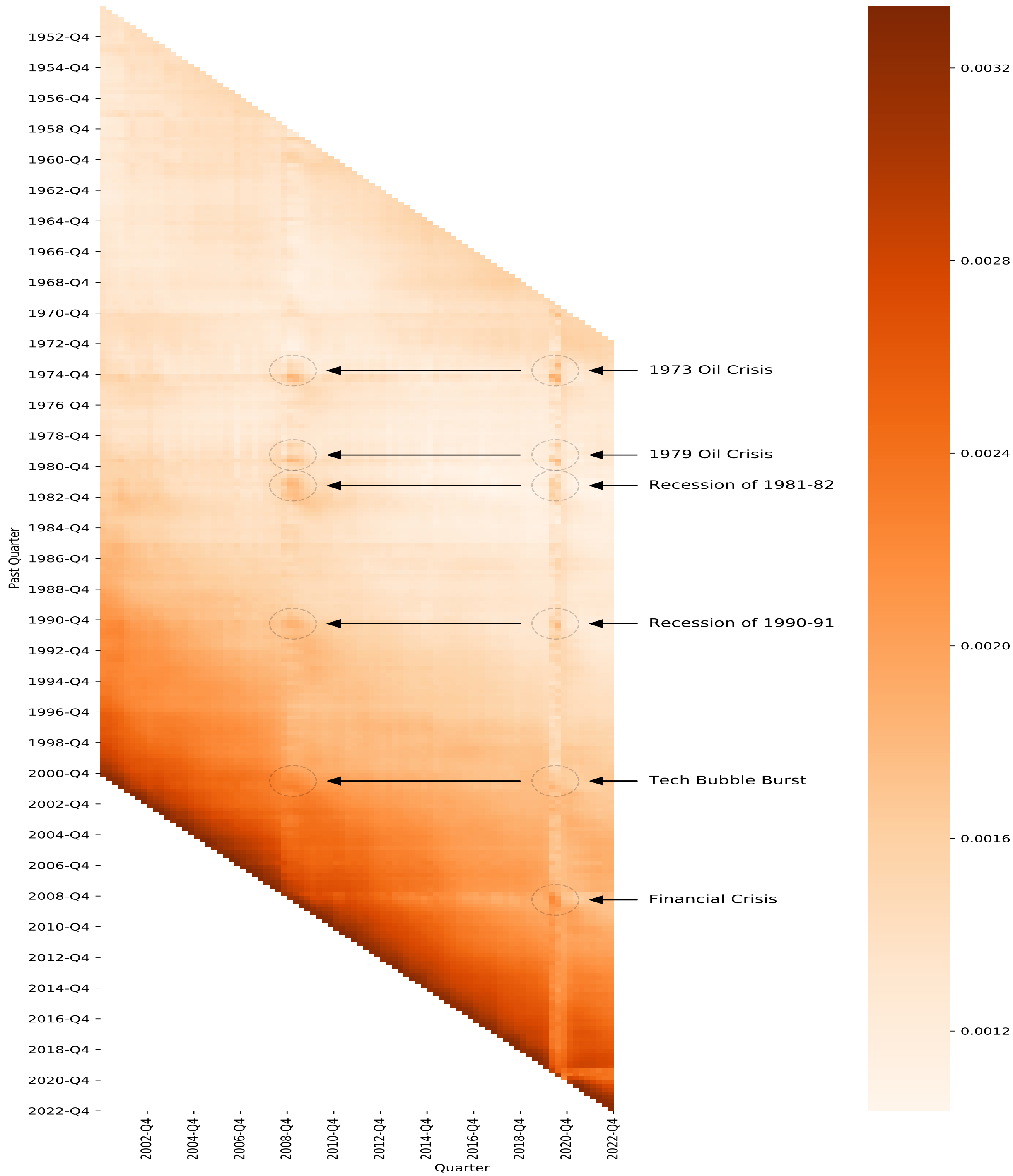

Figure A1: Recall Probabilities from our Measure: Going Back 50 Years

This figure displays the probability weights on the past 200 quarters (= 50 years) implied by our measure for each quarter from 2001 to 2021. The x-axis indicates the quarter in which recall occurs, while the y-axis indicates the recall probability that our measure assigns to past quarters. A darker shade of red indicates a higher probability weight. We highlight several episodes with exceptionally high probability weights during the 2020 Covid-19 Pandemic and the 2008 Financial Crisis: the 1973 Oil Crisis, the 1979 Oil Crisis, the Recession of 1981-82, the Recession of 1990-91, the Tech Bubble Burst, and the 2008 Financial Crisis.

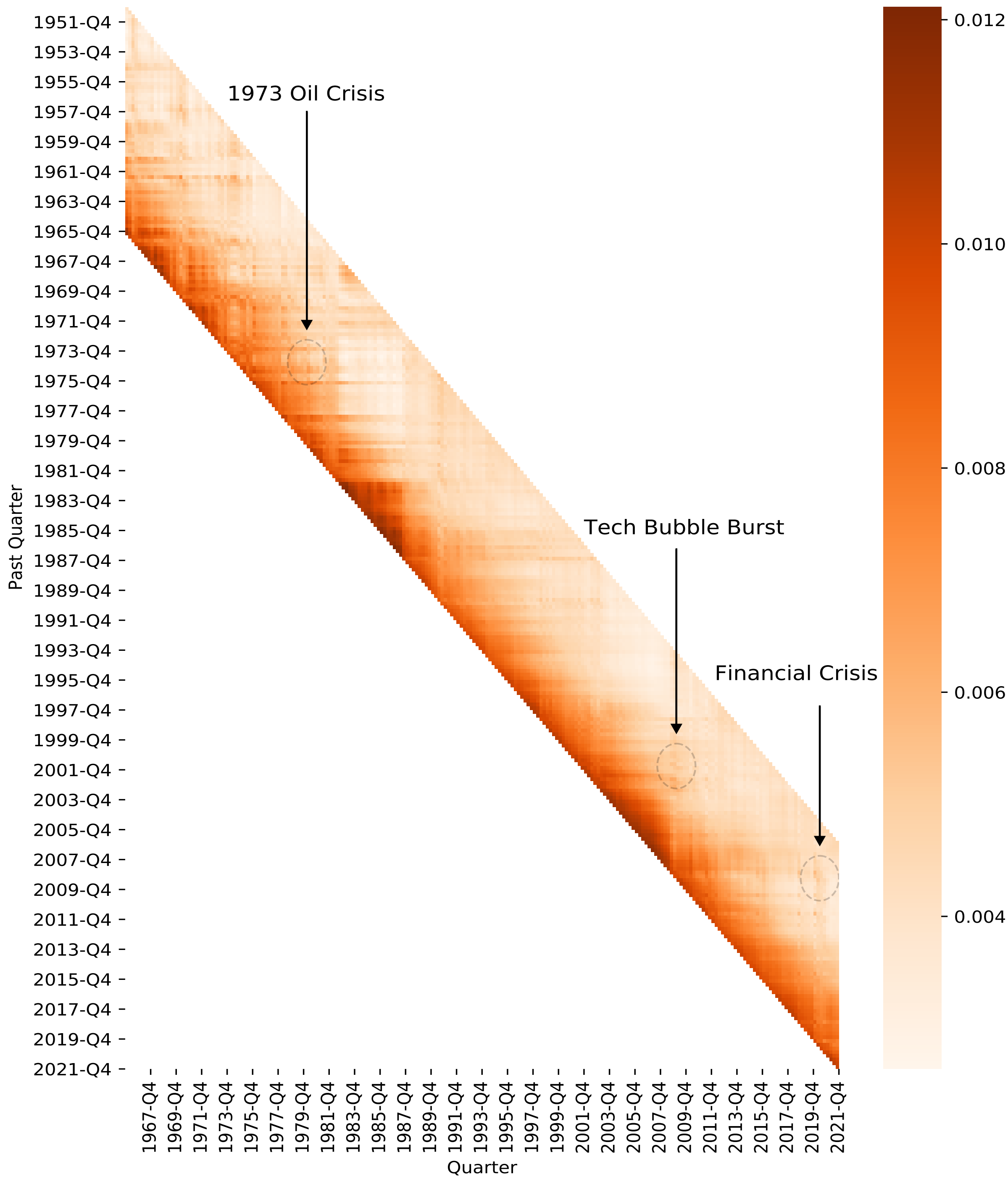

Figure A2: Recall Probabilities from our Measure: Only Stock-Level Features

This figure replicates Figure 1 using only stock-level features (see Table A1 for details on these features). The x-axis indicates the quarter in which recall occurs, while the y-axis indicates the probability weight that our measure assigns to past quarters. A darker shade of red indicates a higher probability weight. We highlight three episodes with exceptionally high probability weights: the recall of the 1973 Oil Crisis during the 1979 Oil Crisis, the recall of the Tech Bubble Burst during the 2008 Financial Crisis, and the recall of the 2008 Financial Crisis during the 2020 Covid-19 Pandemic.

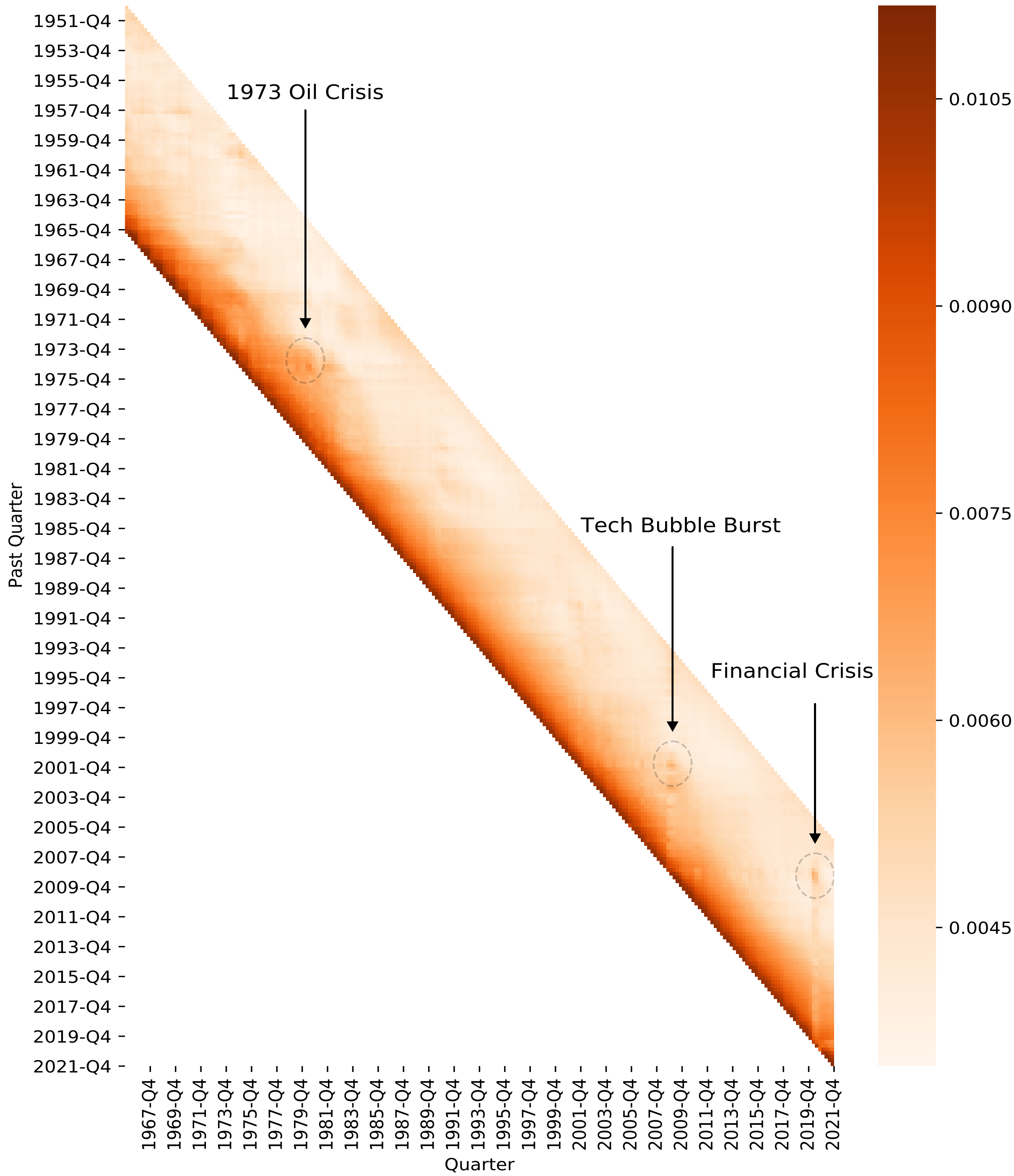

Figure A3: Recall Probabilities from our Measure: Only Firm-Level Financial Ratios

This figure replicates Figure 1 using only firm-level financial ratios (see Table A1 for details on these features). The x-axis indicates the quarter in which recall occurs, while the y-axis indicates the probability weight that our measure assigns to past quarters. A darker shade of red indicates a higher probability weight. We highlight three episodes with exceptionally high probability weights: the recall of the 1973 Oil Crisis during the 1979 Oil Crisis, the recall of the Tech Bubble Burst during the 2008 Financial Crisis, and the recall of the 2008 Financial Crisis during the 2020 Covid-19 Pandemic.

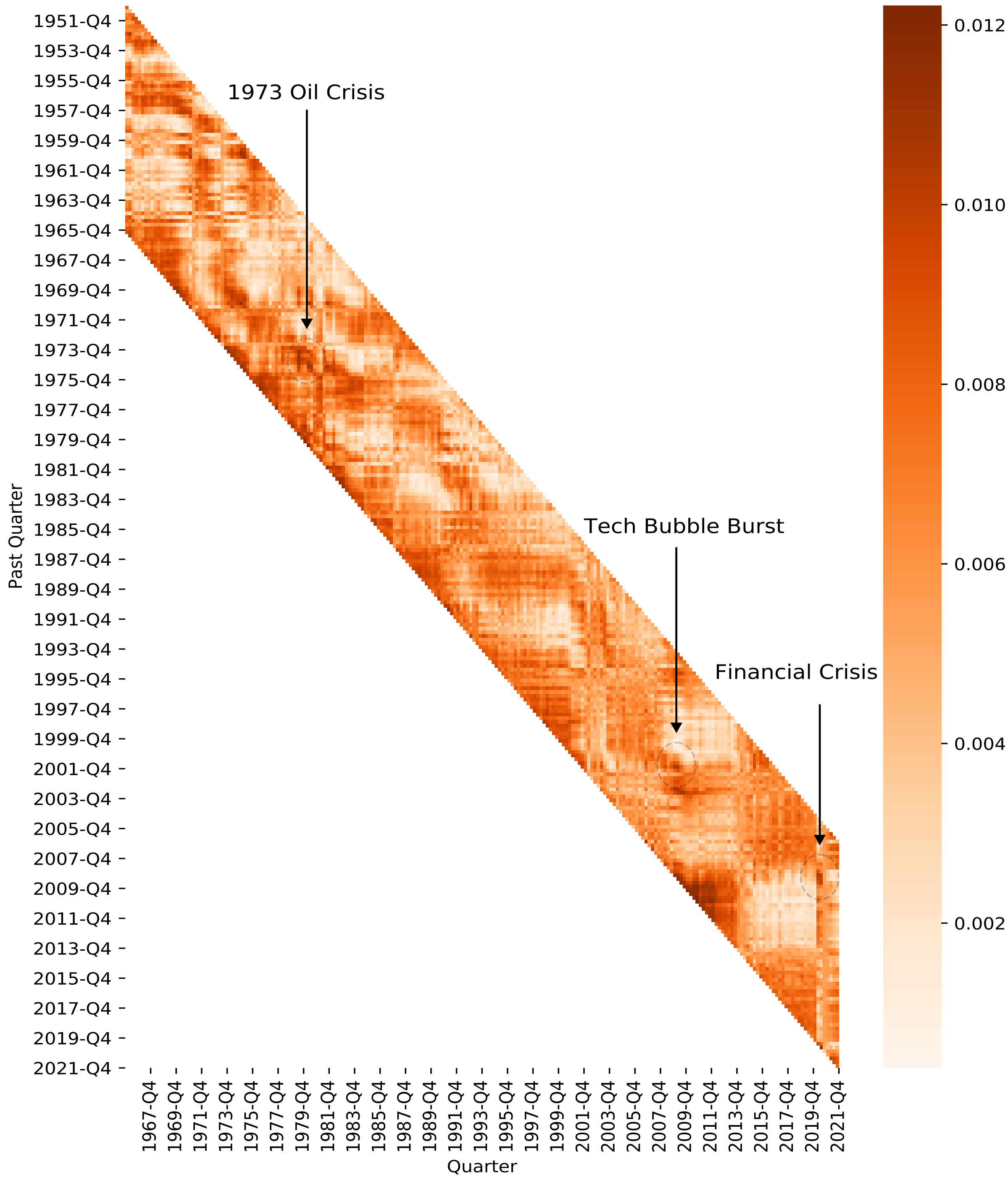

Figure A4: Recall Probabilities from our Measure: Only Broad Macroeconomic Features

This figure replicates Figure 1 using only broad macroeconomic variables (see Table A1 for details on these features). The x-axis indicates the quarter in which recall occurs, while the y-axis indicates the probability weight that our measure assigns to past quarters. A darker shade of red indicates a higher probability weight. We highlight several episodes with exceptionally high probability weights: the recall of the 1973 Oil Crisis during the 1979 Oil Crisis, the recall of the Tech Bubble Burst during the 2008 Financial Crisis, and the recall of the 2008 Financial Crisis during the 2020 Covid-19 Pandemic.

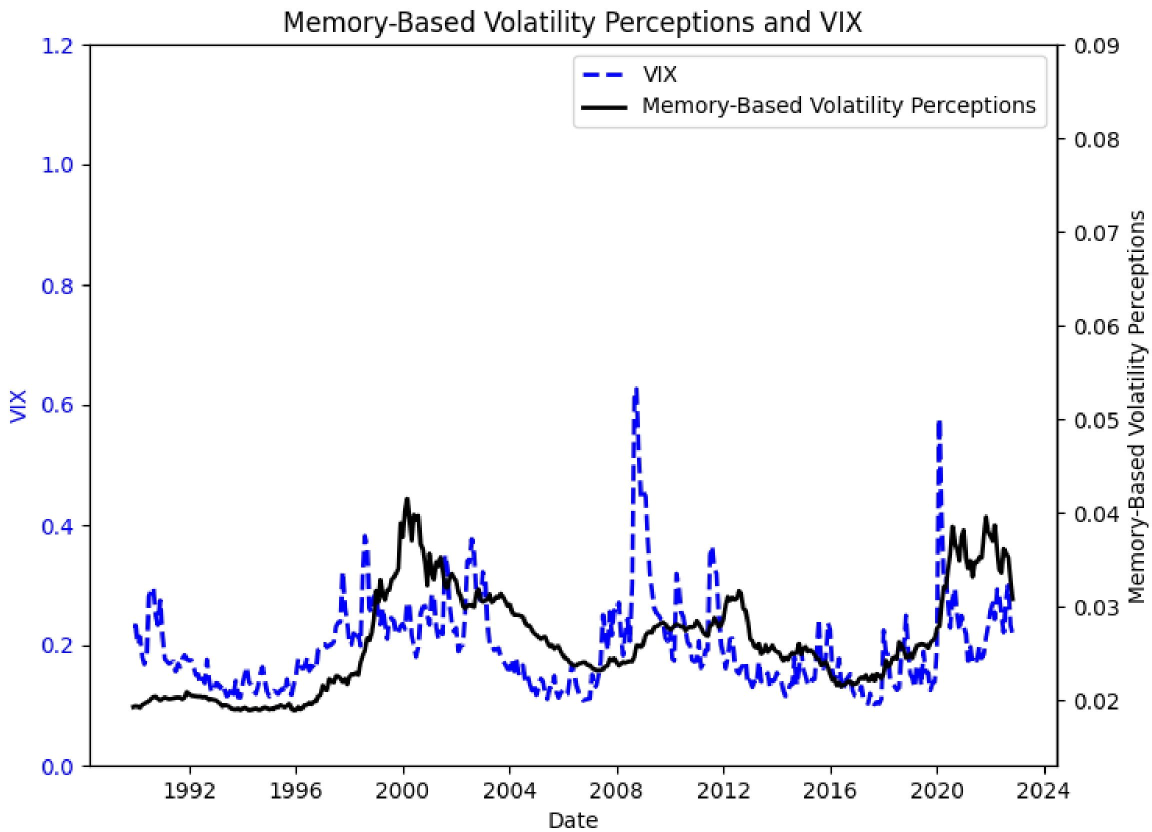

Figure A5: Memory-Based Volatility Perceptions and the VIX

This figure plots memory-based perceived volatility (as of the end of month t) and the average daily VIX over the course of month $t+1$. VIX is divided by 100. Memory-Based Volatility is the annualized perceived volatility of the S&P 500 index derived from our measure.

Table A1: Set of Features

Part 1. Macroeconomic variables			
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
INFLATION	Inflation Rate		
Part 2. Stock-level variables			
return	Monthly Stock Return	PRICE	Stock Price
DOLLARVOL	Dollar Volume	VOL	Trading Volume
Part 3. Firm-level variables			
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
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		

Table A2: Analysts' Recall Patterns from Corporate Events

This replicates Table 2 using only recall patterns extracted from sentences spoken by analysts during corporate events. The dependent variable is an indicator that equals one if, during a firm's corporate event in month t, an analyst mentions the historical month $t-h$ at least once, and zero otherwise. The recall probability $r(e, \kappa_t)$ captures the probability that month $t-h$ is recalled by a representative investor in month t. The first two columns focus on the previous 180 months, i.e., $h \in \{1, 2, 3, \ldots, 180\}$. The last two columns focus on months that are at least 5 years in the past, i.e., $h \in \{61, 62, 63, \ldots, 180\}$. Columns (1) and (3) include stock-by-past-month fixed effects as well as control variables. Columns (2) and (4) further include stock-by-current-month fixed effects. These fixed effects soak up the control variables. In all columns, we include quarter fixed effects. Standard errors are clustered by stock, current month, and past month, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Actual Recall of Month *t-h* in Month *t*			
	(1)	(2)	(3)	(4)
Sample:	Past 15 Years		At Least 5 Years in the Past	
$r(e, \kappa_t)$	45.878***	46.997***	0.743***	0.883***
	(32.22)	(32.76)	(5.15)	(5.26)
Size	0.006***		0.001**	
	(5.49)		(2.26)	
Turnover	0.016***		0.006***	
	(4.55)		(4.56)	
BM	0.001		-0.000**	
	(1.46)		(-2.48)	
Ivol	-0.090***		-0.021**	
	(-3.59)		(-2.45)	
Price	-0.000*		-0.000	
	(-1.85)		(-1.09)	
Adjusted R-Squared	0.275	0.319	0.011	0.147
N	15,237,765	15,237,765	10,152,541	10,152,541
Stock x Current Month FE	NO	YES	NO	YES
Stock x Past Month FE	YES	YES	YES	YES
Quarter FE	YES	YES	YES	YES

Table A3: Managers' Recall Patterns from Corporate Events

This replicates Table 2 using only recall patterns extracted from sentences spoken by managers during corporate events. The dependent variable is an indicator that equals one if, during a firm's corporate event in month t, a manager mentions the historical month $t-h$ at least once, and zero otherwise. The recall probability $r(e, \kappa_t)$ captures the probability that month $t-h$ is recalled by a representative investor in month t. The first two columns focus on the previous 180 months, i.e., $h \in \{1, 2, 3, \ldots, 180\}$. The last two columns focus on months that are at least 5 years in the past, i.e., $h \in \{61, 62, 63, \ldots, 180\}$. Columns (1) and (3) include stock-by-past-month fixed effects as well as control variables. Columns (2) and (4) further include stock-by-current-month fixed effects. These fixed effects soak up the control variables. In all columns, we include quarter fixed effects. Standard errors are clustered by stock, current month, and past month, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Actual Recall of Month *t-h* in Month *t*			
	(1)	(2)	(3)	(4)
Sample:	Past 15 Years		At Least 5 Years in the Past	
$r(e, \kappa_t)$	65.330***	67.385***	2.505***	3.428***
	(39.89)	(40.64)	(7.15)	(8.71)
Size	0.012***		0.004***	
	(6.98)		(4.52)	
Turnover	0.006		0.009**	
	(1.24)		(2.46)	
BM	0.001		-0.000	
	(0.88)		(-0.81)	
Ivol	-0.112***		-0.038	
	(-2.88)		(-1.37)	
Price	0.000		0.000	
	(1.60)		(0.54)	
Adjusted R-Squared	0.397	0.448	0.081	0.208
N	15,237,765	15,237,765	10,152,541	10,152,541
Stock x Current Month FE	NO	YES	NO	YES
Stock x Past Month FE	YES	YES	YES	YES
Quarter FE	YES	YES	YES	YES

Table A4: Cross-Sectional Variation in the Speed of Memory Decay

This table presents the relation between stock-level characteristics and the speed of memory decay. The speed of memory decay speed is defined as the recall probability of the most recent month if recall occurs in the current month divided by the recall probability of the same (most recent) month assuming that recall occurs one year from now. A higher ratio indicates a faster memory decay speed. The table shows how decay speed varies along four firm-level characteristics, which are each measured as of the end of each month: the logarithm of a firm's market capitalization (in million \$), the book-to-market ratio following Fama and French (1992), the idiosyncratic volatility (in %) from CAPM regressions following Ang et al. (2006), and the stock price (in \$). The sample runs from January 1966 to December 2021. The t-statistics for tests that the difference between the highest and lowest decile is equal to zero are displayed in parentheses.

Panel A: Equal-Weighted Memory Decay Speed											
Model	Low	P2	P3	P4	P5	P6	P7	P8	P9	High	High-Low
Size	1.34	1.31	1.30	1.28	1.27	1.25	1.24	1.24	1.23	1.21	–0.13
											(–29.33)
BM	1.26	1.25	1.25	1.26	1.27	1.26	1.27	1.27	1.28	1.30	0.04
											(8.56)
Ivol	1.22	1.23	1.23	1.24	1.25	1.27	1.28	1.30	1.32	1.34	0.12
											(29.37)
Price	1.34	1.32	1.29	1.27	1.26	1.25	1.24	1.24	1.23	1.22	–0.12
											(–21.17)

Panel B: Value-Weighted Memory Decay Speed											
Model	Low	P2	P3	P4	P5	P6	P7	P8	P9	High	High-Low
Size	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.24	1.23	1.19	–0.14
											(–32.72)
BM	1.19	1.20	1.22	1.22	1.22	1.22	1.23	1.23	1.23	1.27	0.09
											(9.55)
Ivol	1.18	1.20	1.21	1.22	1.22	1.23	1.25	1.26	1.28	1.29	0.11
											(20.33)
Price	1.32	1.29	1.27	1.25	1.23	1.23	1.22	1.22	1.21	1.19	–0.13
											(–18.20)

Table A5: Memory-Based Volatility and Option-Implied Volatility: Fama-MacBeth Regressions

This table shows that memory-based volatility explains option-implied volatility in the cross-section of stocks. In all columns, the dependent variable is option-implied volatility in month $t+1$, constructed following An et al. (2014). All independent variables are as of month t. Memory-based volatility is the standard deviation of monthly returns over the past 180 months, where each historical return is weighted with its associated recall probability. Similarly, exponential-decay-based volatility is the standard deviation of monthly returns over past months, where each historical return is weighted with exponentially-decaying weights, using the decay parameter λ = 0.56 from Greenwood and Shleifer (2014). In all columns, we run Fama-MacBeth regressions. In column (1), we only include memory-based volatility as an independent variable. In column (2), we control for exponential-decay-based volatility. In column (3), we control for lagged option-implied volatility, and in column (4) we control for size, idiosyncratic volatility (following Ang et al. (2006)), asset growth, operating profit (following Fama and French (2006)), and the logarithm of the book-to-market ratio (following Fama and French (1992)). We multiply all coefficients by 100. Standard errors are Newey-West adjusted with twelve lags, and the t-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	Option-Implied Volatility			
	(1)	(2)	(3)	(4)
Memory-Based Volatility	61.703***	46.227***	11.681***	8.876***
	(20.87)	(15.59)	(13.60)	(12.80)
Exponential-Decay-Based Volatility		24.550***	5.872***	4.473***
		(13.67)	(7.66)	(8.31)
Option-Implied Volatility (lagged)			74.352***	68.346***
			(56.11)	(43.48)
Size				-0.969***
				(-11.68)
Idiosyncratic Volatility (3F)				110.009***
				(11.89)
Asset Growth				0.154***
				(2.85)
Operating Profit				-0.086***
				(-3.60)
log(Book-to-Market Ratio)				-0.090
				(-1.53)
Adjusted R-Squared	0.427	0.479	0.747	0.758
N	410,569	410,569	410,569	410,569

Table A6: Summary Statistics for Table A7

This table presents summary statistics for the sample used in Table A7. The sample period for these tests ranges from January 1990 to December 2021. VIX is the average daily Volatility Index over the course of a month. We normalize VIX by dividing it by 100. Memory-Based Volatility is the annualized perceived volatility of the S&P 500 index derived from our measure. The logarithm of the price-dividend ratio (Log(P/D)), the risk-free rate, the earnings growth rate, and the unemployment rate are constructed following Greenwood and Shleifer (2014). Realized Volatility is the the annualized realized volatility of monthly S&P 500 index returns over the past 12 months.

	N	Mean	Median	Std.Dev	P25	P75	Min	Max
VIX Index	396	0.197	0.178	0.076	0.141	0.234	0.101	0.627
Memory-Based Volatility	396	0.026	0.025	0.005	0.022	0.030	0.019	0.041
Log(P/D)	396	3.933	3.946	0.271	3.825	4.086	3.248	4.502
Risk-free Rate	396	1.004	1.005	0.018	0.993	1.016	0.948	1.101
Earnings Growth	396	0.221	0.116	1.014	-0.060	0.202	-0.886	7.935
Unemployment	396	5.827	5.500	1.730	4.600	6.700	3.500	14.700
Realized Volatility (Past 12 Months)	396	0.135	0.136	0.054	0.087	0.172	0.039	0.300

Table A7: Memory-Based Volatility Perceptions of the S&P 500 Index

This table shows that memory-based perceived volatility predicts the VIX. In all columns, the dependent variable is the average daily VIX over the course of month $t + 1$. The main independent variable is memory-based perceived volatility of the S&P 500 Index derived from our measure (as of the end of month t). Columns (2) and (3) add control variables. The sample period ranges from January 1990 to December 2021. t-statistics, in parentheses, are Newey-West adjusted with twelve lags. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	VIX		
	(1)	(2)	(3)
Memory-Based Volatility	5.284***	9.884***	5.502*
	(5.19)	(3.11)	(1.72)
Realized Volatility (Past 12 Months)			0.541***
			(3.54)
Log(P/D)		-0.113	-0.075
		(-1.44)	(-1.08)
Risk-free Rate		0.875***	0.810***
		(3.00)	(2.69)
Earnings Growth		-0.004	-0.000
		(-0.58)	(-0.07)
Unemployment		-0.001	-0.004
		(-0.17)	(-1.15)
Adjusted R-Squared	0.136	0.245	0.327
N	396	396	396