# The Interaction of Memory Imperfections

Johannes K. Maier* Marcel R. Quint**

November 15, 2025

[Click here for the latest version]

**Abstract:** We investigate how two pervasive imperfections of human memory – motivated recall (remembering "good" rather than "bad" information) and similarity-based recall (remembering information similar to current information) – interact. We propose a theoretical model of recall in which these biases can reinforce or offset each other, depending on the cognitive effort invested in recall. With lower effort exerted, the two biases are more complementary, as people rely more on intuitive reasoning and thereby become more able to exploit similarity to self-servingly bias their recall. We test the model's predictions in a laboratory experiment that varies both motivation and contextual similarity. Consistent with the model, we find that the biases are complements, in particular when recall effort is low. The interaction shapes beliefs and behavior (such as trading and policy responses), so that simply "adding up" the biases' isolated effects would lead to misguided inferences under their coexistence. We further show that imperfect memory is equally important in explaining distorted beliefs as non-Bayesian updating.

**JEL Classification Numbers:** D90, D91, D83, D84, G11, G40, G41.

*University of Munich (LMU) and CESifo. Address: LMU Munich, INRIVER/MRIC, Schackstr. 4, D-80539 Munich, Germany. Phone: +49 (0)89 21803552. E-mail: johannes.maier@lmu.de.

**University of Munich (LMU). Address: LMU Munich, Seminar for Economic Theory, Ludwigstr. 28 Rgb., D-80539 Munich, Germany. Phone: +49 (0)89 21802238. E-mail: marcel.quint@econ.lmu.de.

Financial support from the *Deutsche Forschungsgemeinschaft* through CRC TRR 190 is gratefully acknowledged. The authors are grateful to Valeria Burdea, Keith M. Ericson, Katrin Gödker, Taisuke Imai, Klaus M. Schmidt, Andrew Schotter, and Matthew Rabin for helpful comments. We also thank seminar participants at New York University, Brown University, LMU Munich, MRICWW 2023 in St. Johann, NCBEE 2023 in Lund, BMD 2023 in Ingolstadt, BIW 2024 in Munich, EF 2024 in Stavanger, TBE 2024 at Caltech, BEEMA 2024 in Delaware, RBF 2024 in Amsterdam, LPEW 2024 in Lancaster, ESA 2025 in Osaka, FBEE 2025 in Tokyo, CRC 2025 in Berlin, TBE 2025 at Caltech, ESWC 2025 & ASSET 2025 in Seoul, BEEMA 2025 in Bethlehem, VfS 2025 in Cologne, NCBEE 2025 in Stavanger.

# 1 Introduction

Economists increasingly recognize that expectations (and subsequent behavior) are not only shaped by violations of Bayes' rule when processing information, but also by incorrectly recalling that information. Recent literature on memory emphasizes the role of both *similarity-based* and *motivated* recall biases. On the one hand, similarity-based (or associative) recall describes people's tendency to recall past information more accurately if it is similar to current information, where similarity is determined by contextual features which embed the information, such as stories and images. Thus, the context in which current information is transmitted strongly affects people's recall.[1] On the other hand, motivated recall describes people's tendency to recall past information in a self-serving way, so that more information in support of a desired view of oneself or the world is recalled than contradicting information.[2] The study of either one of these biases has uncovered important insights. Yet, such isolated treatment can mislead inference about real-world behavior when coexisting biases interact, and coexistence of similarity-based and motivated recall seems omnipresent: if people try to recall the past in order to form beliefs and act upon them, events from the past are usually "good" or "bad" for their desired view *and* have occurred in specific contexts that may be similar or dissimilar to the current context.

In this paper, we theoretically and experimentally investigate whether and how these two recall biases interact. We find that similarity-based and motivated recall complement each other. Thus, our analysis shows that simply "adding up" the biases' isolated effects would lead to misguided inferences under their coexistence. For instance, consider an investor forming expectations about her asset's performance. Motivated recall makes her more likely to remember good than bad news. Similarity-based recall makes her more likely to remember good news if the current context is good and bad news if the current context is bad, and this dependency on the current context increases with the number of "cues," i.e., with similarity (e.g., repeatedly published earnings reports). Without knowing how the two biases interact, we cannot answer whether more cues actually rise or lower the investor's expectation. Consider a situation, in which the current context is equally likely to be good or bad. With the two biases being complements, beliefs are expected to be more favorable with more cues, because more cues not only increase dependency on the current context, but also the investor's motivation to remember more good than bad news. If the two biases were substitutes, beliefs would instead be less favorable with more cues, as the investor's motivation to remember more good than bad news would shrink. Without any interaction, more cues would have no effect on the fa-

[1]See, e.g., Schacter, 2008; Kahana, 2012; Mullainathan, 2002; Gennaioli and Shleifer, 2010; Bordalo et al., 2020; Enke et al., 2024; Wachter and Kahana, 2024; Jiang et al., 2022; Charles, 2022; Charles, 2024; Charles and Sui, 2024

[2]See, e.g., Benabou and Tirole, 2002; Saucet and Villeval, 2019; Chew et al., 2020; Zimmermann, 2020; Gödker et al., 2025; Jiang et al., 2022; Sprengholz et al., 2023; Amelio and Zimmermann, 2023

vorability of beliefs. Thus, more cues induce more asymmetry in belief updating under complements and would induce less asymmetry under substitutes.

Furthermore, considering the aggregate stock market, since good news for asset holders are essentially bad news for non-holders (and vice versa),[3] disagreement between holders and non-holders increases with more cues under complements and would decrease with more cues under substitutes. Under complements, more cues induce more asymmetry in belief updating, so that the beliefs of holders and non-holders diverge with more cues. Under substitutes, more cues would have the opposite effect. However, on the aggregate stock market, interaction effects are not only important for similarity or cue changes, but also for swings in the current context. Consider, for instance, shifts in the ambient market state without substantive news. An earnings report by another, otherwise informationally irrelevant firm in the same stock portfolio can flip the context from good to bad, or vice versa, because overlapping earnings create cross-stock memory links that cue recall for the focal stock (see Charles, 2024, for empirical evidence). Under complements, since a current good context triggers more such signals among holders than non-holders, holders will increase their expectations more than non-holders. A current bad context triggers more such signals among non-holders than holders, so that non-holders will decrease their expectations more than holders. Thus, for both swings in the current context, the complementary effect rises disagreement between holders and non-holders. Substitutes have the opposite effect and lead to less disagreement between holders and non-holders for swings in the current context. Since disagreement between holders and non-holders affects prices and trade volume, interaction affects have substantial consequences for market outcomes.

More generally, this example illustrates that the interaction of memory imperfections affects belief polarization and convergence and may shed new light on the puzzling phenomenon that people's disagreement sometimes becomes more and sometimes less extreme with the different parties facing the same evidence on the issue at hand. For instance, consider a government trying to increase the vaccination rate by using a marketing campaign that affects people's mindset (rather than providing new information), such as using dramatic narratives or displaying visuals to make disease risks more salient and accessible. In the absence of an interaction, the marketing campaign increases vaccination rates as both pro- and anti-vaxers recall more supporting information. In case of substitutes, the marketing campaign is effective for anti-vaxers, but not necessarily for pro-vaxers because motivations are weakened by the associative recall. Here, the campaign is most effective in the sense of getting hold of non-vaxers. Under complements, the marketing campaign is effective for pro-vaxers, but not necessarily for anti-vaxers. Pro-vaxers recall more supporting information, but since they would have taken the vaccine anyway, vaccination rates are basically unaffected. For anti-vaxers the

[3]Good news for holders are bad news for non-holders, if non-holders have decided against purchasing the asset beforehand and regret their decision. Our empirical findings show that this is indeed the case.

campaign may even backfire, since by enhancing their motivations they could end up with even lower vaccination rates than without the campaign (see Nyhan et al., 2014; Nyhan and Reifler, 2015, for empirical evidence). This example illustrates that the interaction of memory imperfections may uncover potentially adverse policy effects that would otherwise remain unconsidered (see DellaVigna and Linos, 2022, for empirical evidence).

Finally, the way both memory imperfections interact may shed light on a more general relation between motivated and mechanical biases. While motivated recall falls into the former class of biases, associative recall falls into the latter. Their interaction determines whether individuals are able to exploit (complements) or whether they are hindered by (substitutes) similarity to self-servingly bias their recall. Our empirical findings show that individuals are indeed able to exploit similarity in a self-serving manner, in particular with low recall effort. Thus, our finding may well be interpreted as another manifestation of "motivated errors" (Exley and Kessler, 2024) in a different domain, namely memory, and may suggest a foundation for motivated reasoning, in which motivated reasoning is unconsciously enabled by the opportunity to make errors rather than constituting a conscious distortion of the truth.

We use the following stylized setting of an asset market for both our theory and experiment: An investor makes an informed investment decision in period 0 between a safe and a risky asset and then *subsequently* observes several price movements of the risky asset until period $\tau \in \mathbb{N}_+$. These price movements are informative for the unknown quality of the asset and determine investment performance. Between periods $\tau$ and $\tau + 1$, there is a time lag, inducing the investor's memory on previous signals to be incomplete. The investor observes a final price movement in period $\tau + 1$, has to recall all previous movements, and makes a new proportional investment into the risky asset. We are interested in how the investor recalls previous signals in period $\tau + 1$ as well as how this recall maps into her belief on the quality of the asset and subsequent investment decisions.

Our theoretical model of memory formation is based on Kahneman's (2011) "System 1-System 2" thinking. There are two stages. In the first stage, the agent's thinking – executed by her fast "System 1" – is instantaneous and unconscious. This stage captures the *incompleteness* of memory, i.e., an agent's tendency to either forget or confabulate signals. In the second stage, the agent's thinking – executed by her slow "System 2" – is conscious and requires cognitive resources. In our model, the agent's System 1 response acts as an anchor from which, by exerting costly effort, she can deliberately move away with her System 2. System 2 captures the *imperfection* of memory by allowing for both recall biases. Associative recall scales the cost of effort, so that increasing similarity reduces costs to recall more precisely. Motivated recall determines when and how much the agent weighs precision over cost, so that an increasing motivation enhances her recalled investment performance. Our model's main result is that associative and motivated memory are complements if recall effort is sufficiently low and they are substitutes for high recall effort. Intuitively, with low effort, the agent relies more on her instantaneous System 1 thinking. In this case, motivated agents are able to exploit similarity to self-servingly bias recall. With high effort, the agent relies more on her conscious System 2 thinking. Here, motivated agents are hindered by similarity to self-servingly bias recall. In contrast to other models generating interaction effects, complements and substitutes in our model are not the result of altering assumptions on functional forms. In fact, the size of the complementary effect is rather a *result* of our model and depends on exerted effort levels and, thereby, on the induced relative importance of System 1 vs. System 2. This feature of our model generates the testable prediction that the complementary effect of both recall biases becomes stronger with lower exerted effort levels.

We test this and other predictions of our model in a laboratory experiment that builds on the experimental design of Enke et al. (2024), albeit adjusts it in a way to simultaneously accommodate and vary both motivated and non-motivated memory imperfections. We use a laboratory setting because only in the lab can we orthogonally vary motivation and contextual similarity, directly measure recall after a controlled delay, and hold exposure to information and timing fixed – conditions that are not attainable in observational data. A subject in our experiment faces ten asset markets, where each market considers a specific company's stock, whose price path follows a high or low binomial process. Similarity is varied *within* subjects by splitting the ten markets into five *cue* and five *non-cue* markets. In a cue market, all up movements of the price occur in the same context and all down movements occur in the same context. A context is specified by a story causing the price movement and a story-fitting picture. In a non-cue market, all price movements occur in different contexts. Motivations are varied *across* subjects by assigning them to one of two possible trading schemes. Subjects in the *long position* profit from price increases, while subjects in the *short position* profit from price decreases in all ten markets. These diametrically opposing motivations assure that the stochastic environment of all subjects is exactly the same. After subjects have observed all ten markets, there is a time lag of three days before they receive the final price movement and then have to recall all earlier price movements in each market. While this recall is our main variable of interest, we also examine how recall influences subjects' beliefs and actions.

Our experimental findings show that both recall biases are indeed present in our experiment, which constitutes a pre-requisite for analyzing their interaction. Concerning their interaction, we find that associative and motivated memory are complements. Consistent with our theoretical model, we observe this complementary interaction effect for low, but not for high effort levels. We further find that the interaction effect that we document in subjects' recall transmits to their beliefs and subsequent investment decisions. Hence, the interaction of memory imperfections turns out to be belief- as well as action-relevant. When decomposing the difference between subjects' stated and Bayesian beliefs, incorrect recall turns out to be of equal importance in explaining distorted beliefs as non-Bayesian updating, and jointly they fully explain stated beliefs. Accounting for individual-specific risk aversion, we also show that subjects' belief distortion explains their investment mistakes, which demonstrates the economic relevance of our results.

**Related Literature.** Our paper relates to two strands of literature. First and foremost, our paper relates to a growing literature on memory imperfections. While the psychology literature has long recognized that memory is systematically distorted by contextual similarity (Schacter, 2008; Kahana, 2012; Tulving and Thomson, 1973; DuBrow and Davachi, 2013) as well as by self-serving motivations (Sedikides and Green, 2005; Anderson et al., 2004; Kensinger and Schacter, 2006; Carlson et al., 2020; Walker et al., 2003), the economics literature has only more recently investigated how important these recall biases are for belief formation and economic decision making. Both theoretical and empirical studies have convincingly demonstrated the existence and economic implications of similarity-based (Mullainathan, 2002; Bordalo et al., 2020; Enke et al., 2024; Wachter and Kahana, 2024; Charles, 2022, 2024; Charles and Sui, 2024; Gennaioli and Shleifer, 2010) and motivated recall (Benabou and Tirole, 2002; Saucet and Villeval, 2019; Chew et al., 2020; Gödker et al., 2025; Sprengholz et al., 2023; Amelio and Zimmermann, 2023; Jiang et al., 2022; Zimmermann, 2020). While all these papers analyze either one of those recall biases in isolation, our paper extends this literature by providing the first attempt to analyze their interaction. This is important, because the coexistence of these recall biases is omnipresent in the real world and their interaction can have significant effects on behavior.

Second, our paper relates more broadly to a large literature on belief formation (see, e.g., Benjamin, 2019, for an overview). While the standard model of belief formation builds on Bayes' Rule *and* perfect memory,[4] the literature on distorted beliefs has focused on the former, showing that the process of belief formation is typically non-Bayesian. The latter assumption, namely perfect memory, is retained. We contribute to this literature, firstly, by showing how the interaction of recall biases affects belief formation and subsequent decision making and, secondly, by showing that imperfect memory is equally important in explaining distorted beliefs as non-Bayesian updating.

**Outline.** The remainder of this paper is structured as follows. Section 2 introduces the setup that accommodates both our theory and experiment. In Section 3, we present our model and its theoretical results. Section 4 describes the experimental design and in Section 5 we derive hypotheses for the experiment based on our theoretical insights. Section 6 presents our empirical findings. We conclude in Section 7. All proofs and supplementary information can be found in the appendix.

[4] In the standard model of belief formation, pooled updating equals acceptive updating, which implies that agents recall perfectly.

## 2 Setup

The following setup accommodates both our theory and experiment. In Section 4, we further specify the exact parameterization of the experiment.

**Stochastic Environment.** We model a stylized asset market with a risky and a safe asset for a finite horizon in discrete time,

$$t \in \{k, \ldots, 0, \ldots, \tau, \tau+1, \ldots, T\},$$

from the viewpoint of an individual investor, where $k \leq 0$ and $T \geq 3$. The risky asset's price $P_t$ is exogenously determined by one of two possible stochastic processes, $F_H$ and $F_L$, with known distributions, but it is unknown to the investor whether the risky asset adheres to the high process $F_H$ or the low process $F_L$. Specifically, $F_H$ and $F_L$ are binomial processes with corresponding appreciation probabilities $p_H, p_L \in (0,1)$, where $p_H > p_L$. For simplicity, we further assume that $p_H = 1 - p_L$, so that $F_H$ and $F_L$ are symmetric around $1/2$. This allows us to implement the same processes in the two trading schemes that we investigate (the "long" and "short" position, see below). Each process appreciates by an absolute amount $U > 0$ and depreciates by $D = -U$.[5] We assume that $|D| < \frac{P_k}{T+|k|+1}$ is satisfied, so that the price of the risky asset stays always positive. While the risky asset's price evolves along one of two possible binomial trees, the rate of return of the safe asset is $r \in \mathbb{R}$ per period.

**Learning.** The investor is assumed to have an uninformative objective prior $q_k = 1/2$ about the likelihood that the asset adheres to the high process $F_H$. As the two price processes differ in their appreciation probabilities, observing the asset's price path is informative about whether it follows $F_H$ or $F_L$, i.e., price changes are noisy signals of asset quality. Hence, there is scope for learning. Let $u_t$ and $d_t$ respectively denote the number of price appreciations and depreciation of the risky asset between periods $k$ and $t$. Then, the agent's posterior belief in period $t$, $q_t$, is fully determined by the difference $\Delta_t := u_t - d_t = \frac{P_t - P_k}{U}$, i.e.,

$$q_t = \frac{p_H^{\Delta_t \beta_{\text{sgn}\Delta_t}}}{p_H^{\Delta_t \beta_{\text{sgn}\Delta_t}} + p_L^{\Delta_t \beta_{\text{sgn}\Delta_t}}}.$$

Given that the prior is uninformative, $\Delta_t = 0$ induces an uninformative posterior in period $t$, i.e., $q_t = 1/2$. While Bayesian updating requires $\beta_{\text{sgn}\Delta_t} = 1$, we also allow for non-Bayesian updating, i.e., $\beta_{\text{sgn}\Delta_t} \neq 1$. In particular, we allow the agent to overreact ($\beta_{\text{sgn}\Delta_t} > 1$) or to underreact ($\beta_{\text{sgn}\Delta_t} < 1$), i.e, to respectively perceive the overall signal as being more or less precise than it actually is, and differently so for positive ($\Delta_t > 0$)

[5] The amounts $U$ and $D$ are necessarily identical for $F_H$ and $F_L$, because otherwise observing price paths for only one period would suffice to deterministically infer whether the asset follows $F_H$ or $F_L$. Further, the same *absolute* amounts $|U| = |D|$ circumvent the complication that different magnitudes may be remembered differently.

or negative ($\Delta_t < 0$) overall signals. We only require the agent to update in the right direction, i.e., not to misinterpret the sign of the signal, so that $\beta_{\text{sgn}\Delta_t} > 0$ holds. Note that observing more (less) price appreciations than depreciations, i.e., $\Delta_t > (<)\ 0$, implies a posterior belief of $q_t > (<)\ 1/2$. A larger absolute value of $\Delta_t \in \mathbb{Z}$ induces a more extreme posterior belief.

**Trading.** While asset prices are observed in every period, trading is restricted to periods 0 and $\tau + 1$. For our theoretical model, there is no need to restrict investment decisions and we could allow the investor to invest any fraction of her current wealth into the risky asset in both trading periods. However, in order to compare incentives across subjects in our experiment, we need to restrict their investment decisions: Subjects' first investment decision in period 0 is a binary choice of investing their endowment $W_0 > 0$ in either the risky or safe asset. Their asset holdings are automatically liquidated in period $\tau + 1$ (in the background without feedback) and they take a second investment decision, which allows them to invest any fraction of a newly received endowment $W_{\tau+1} > 0$ into the risky asset. The second investment is automatically liquidated in period $T$.

In the experiment, we further need to exogenously vary the investor's motivation and therefore investigate two different trading regimes, the "long" and "short" position. In the long position, the investor benefits from a price increase and in the short position she benefits from a price decrease. More specifically, in the long position, the investor buys the risky asset for the price in the trading periods, i.e., $P_0$ for the first investment decision and $P_{\tau+1}$ for the second investment decision, and automatically sells the risky asset for the price in the liquidation periods, i.e., $P_{\tau+1}$ for the first and $P_T$ for the second investment decision. By contrast, in the short position, the investor buys the risky asset for the price in the liquidation periods, i.e., $P_{\tau+1}$ for the first investment decision and $P_T$ for the second investment decision, and automatically sells the risky asset for the price in the trading periods, i.e., $P_0$ for the first and $P_{\tau+1}$ for the second investment decision. Note that the long and short position induce mirror-image incentives for experimental subjects, so that motivations are varied without changing anything else. Such a desired isolated experimental variation could not be achieved by comparing, e.g., asset holders to non-holders, real to hypothetical investments, or informed to uninformed investments – all variations that arguably change more than just subjects' motivations.

**Timing.** Before the investor takes her first investment decision in period 0, she receives a statistical summary of all $|k| + 1$ pre-movements of the price. These pre-movements make the investment decision informed, so that the investor becomes "responsible" for her investment performance, which seems to be a necessity for self-serving motivations (see Gödker et al., 2025). Then, after her first investment decision, the agent observes subsequent price movements from period 1 to period $\tau$, and no further actions are taken. Between periods $\tau$ and $\tau + 1$, there is a time lag and subjects in the experiment are distracted in order to make their memory incomplete. In period $\tau + 1$, the investor observes another price movement and then has to recall all previous price movements since her first investment decision in order to make her second investment decision. Afterwards, the agent observes the remaining price movements without further actions until the final period $T$. We are especially interested in how the agent recalls price movements since her first investment decision, i.e., between periods 1 and $\tau$, when this recall becomes relevant for her second investment decision in period $\tau + 1$.

# 3 Model

In this section, we propose a simple model of memory formation that is based on the concept of "System 1 – System 2" thinking by Kahneman (2011). Our model has two stages. In the first stage, the agent's thinking – executed by her fast "System 1" – is instantaneous and unconscious. In the second stage, the agent's thinking – executed by her slow "System 2" – is conscious and requires cognitive resources. In our model, the agent's System 1 response acts as an anchor from which she can deliberately move away by activating System 2. We model the agent's memory retrieval for periods 1 to $\tau$ from the perspective of period $\tau + 1$.

**System 1.** The agent's response in the model's first stage uses an instantaneous estimate of the total number of signals $\bar{\tau} > 0$ and extrapolates the true proportions of ups and downs to that number. Thus, the agent's System 1 estimates of the respective number of ups and downs between periods 1 and $\tau$ are

$$\bar{u} = \frac{\bar{\tau}}{\tau} u \quad \text{and} \quad \bar{d} = \frac{\bar{\tau}}{\tau} d,$$

where $u$ and $d$ are the true respective numbers of ups and downs between periods 1 and $\tau$. Put differently, the instantaneous response induces recall to be incomplete but unbiased in the sense that the proportions of the estimated and the true number ups and downs are identical.

The agent's System 1 response identifies whether the agent tends to forget signals, so that $\bar{\tau} < \tau$, or whether she tends to confabulate signals, so that $\bar{\tau} > \tau$. The tendency to forget or confabulate signals is likely to be driven by environmental features, e.g., the number of signals or the way signals are embedded in narratives or statistics.[6] We make no restrictions on this tendency and allow the agent to either forget or confabulate signals. If the agent does neither of the two, so that $\bar{\tau} = \tau$, System 1 already produces correct estimates $\bar{u} = u$ and $\bar{d} = d$. However, with incomplete memory, $\bar{u}$ and $\bar{d}$ act as an initial anchor, from which the agent makes adjustments in the second stage of the model.

[6]In principle, this tendency could also be motivated because forgetting (confabulating) signals decreases (increases) the estimated absolute difference between ups and downs. However, as we show in Section 6, this is not the case in our experiment. We find that the tendency to forget or confabulate is mechanically based on a simple "averaging" rule that is independent of the agent's motivations.

**System 2.** The second stage of the model is a deliberate System 2 process that takes the agent's System 1 response as given. Here, the agent can exert costly effort to become more precise as compared to the response of System 1. The resulting estimates of the respective number of ups and downs between periods 1 and $\tau$ are

$$\hat{u}(e_u) = \left[\frac{\bar{\tau}}{\tau} + \left(1 - \frac{\bar{\tau}}{\tau}\right) e_u\right] u \quad \text{and} \quad \hat{d}(e_d) = \left[\frac{\bar{\tau}}{\tau} + \left(1 - \frac{\bar{\tau}}{\tau}\right) e_d\right] d, \tag{1}$$

where $e_u \in [0,1]$ denotes the effort level exerted on ups and $e_d \in [0,1]$ refers to the effort level exerted on downs. Thus, if the agent exerts no effort, she will recall $\hat{u}(0) = \bar{u}$, $\hat{d}(0) = \bar{d}$ and be as imprecise as her System 1 response. On the other hand, if she exerts maximal effort, she will recall $\hat{u}(1) = u$, $\hat{d}(1) = d$ and become perfectly precise.[7] If some effort in between is exerted, the agent will recall ups in between $\bar{u}$ and $u$ and downs in between $\bar{d}$ and $d$. In particular, since

$$\frac{d\hat{u}(e_u)}{de_u} > (<)\ 0 \iff \bar{\tau} < (>)\ \tau \quad \text{and} \quad \frac{d\hat{d}(e_d)}{de_d} > (<)\ 0 \iff \bar{\tau} < (>)\ \tau, \tag{2}$$

exerting more effort moves the agent's recall $\hat{u}$ and $\hat{d}$ closer to $u$ and $d$, respectively, either from below when the agent tends to forget ($\bar{\tau} < \tau$) or from above when the agent tends to confabulate signals ($\bar{\tau} > \tau$).

For any given effort level, how far $\hat{u}$ deviates from $u$ and $\hat{d}$ from $d$ depends on the distance between $\bar{\tau}$ and $\tau$, which is determined in System 1. System 2, on the other hand, determines how much effort the agent actually exerts for any given $\bar{\tau}$, which depends on how much she values recall precision over effort cost as well as on how costly effort actually is. The agent chooses the effort level that represents the best compromise in the trade-off of becoming more precise (captured by $\Pi$) and reducing cost of effort (captured by $C$), i.e.,

$$\max_{e_u, e_d \in [0,1]} \Pi\left(|u - \hat{u}(e_u)|, |d - \hat{d}(e_d)|\right) - C(e_u, e_d). \tag{3}$$

$\Pi$ is a decreasing and concave function in $|u - \hat{u}(e_u)|$ and $|d - \hat{d}(e_d)|$ and $C$ is an increasing and convex function in $e_u$ and $e_d$. Both functions reach their maximum at $(e_u, e_d) = (1,1)$ and their minimum at $(e_u, e_d) = (0,0)$. For simplicity, we impose some

[7] We could add noise terms in (1) to capture the fact that even maximal effort may not necessarily induce $\hat{u}(1) = u$ and $\hat{d}(1) = d$. However, for simplicity we abstract away from this feature.

more structure and assume that $\Pi$ is a quadratic loss function and $C$ is a quadratic cost function, i.e.,

$$\begin{aligned}\Pi\left(|u-\hat{u}(e_u)|,|d-\hat{d}(e_d)|\right) &= -\mu_u\left(u-\hat{u}(e_u)\right)^2-\mu_d\left(d-\hat{d}(e_d)\right)^2,\\ C(e_u,e_d) &= \frac{1-\mu_u}{\alpha_u}e_u^2+\frac{1-\mu_d}{\alpha_d}e_d^2.\end{aligned}$$

**Motivated Memory.** $(\mu_u,\mu_d)\in(0,1)^2$ measure how much the agent values precision over cost, separately for ups and downs. For instance, $(\mu_u,\mu_d)\to(0,0)$ means the agent cares only about avoiding effort cost and $(\mu_u,\mu_d)\to(1,1)$ means the agent only cares about becoming more precise in her recall. Motivated memory is captured by $\mu_u-\mu_d=:\mu\in(-1,1)$ with sgn $\mu(\text{sgn}\,\bar{\tau}-\tau,\text{asset position})$. If $\mu=0$, motivated recall is absent and the agent values precision over cost equally for ups and downs. The larger its distance to zero, i.e., the larger $|\mu|$, the stronger is motivated recall, because a larger $|\mu|$ increases the weight differential of precision over cost between ups and downs. $\mu>0$ means that the agent values precision over cost more for ups than downs. If the agent tends to forget (confabulate) signals, i.e., $\bar{\tau}<(>)\ \tau$, and is motivated by price increases (decreases), this induces her to recall a too rosy investment performance, because by (1) it increases (decreases) $\hat{u}(e_u)-\hat{d}(e_d)$. In contrast, $\mu<0$ means that the agent values precision over cost more for downs than ups. If the agent tends to confabulate (forget) signals, i.e., $\bar{\tau}>(<)\ \tau$, and is motivated by price increases (decreases), this again induces her to recall a too rosy investment performance, because by (1) it increases (decreases) $\hat{u}(e_u)-\hat{d}(e_d)$.

**Definition 1** *Motivated memory is captured by $\mu_u-\mu_d=:\mu\in(-1,1)$, where $|\mu|$ measures the strength and $\mu=0$ refers to the absence of motivated recall. Moreover, motivated recall induces $\mu_u^l=\mu_u\gtrless\mu_d=\mu_d^l\iff\bar{\tau}\lessgtr\tau$ in the long position and $\mu_u^s=\mu_d\wedge\mu_d^s=\mu_u$ in the short position.*

Motivated memory makes the agent prone to self-servingly bias her recall. In the long position, the agent is motivated to recall a larger price increase, and in the short position, she is motivated to recall a larger price decrease. Definition 1 assures that the long and short position are mirror images of each other, so that the overall weight of precision over cost remains stable when changing the asset position.

**Associative Memory.** $(\alpha_u,\alpha_d)\in\mathbb{R}_+^2$ measure how costly effort is, separately for ups and downs. For instance, $(\alpha_u,\alpha_d)\to(0,0)$ means that exerting effort is infinitely costly and $(\alpha_u,\alpha_d)\to(\infty,\infty)$ means that exerting effort is costless. Associative memory is captured by $\alpha_u-\alpha_d=:\alpha\in(-\infty,\infty)$ with sgn $\alpha(\text{context})$. If $\alpha=0$, associative recall is absent, so that exerting effort is equally costly for ups and downs. The larger its distance to zero, i.e., the larger $|\alpha|$, the stronger is associative recall so that the current context makes it increasingly less costly to recall same-context signals. $\alpha>0$ means that it is less costly for the agent to recall ups than downs, which is the case when the agent is currently in the up context, i.e., when the signal in period $\tau + 1$ is an up price movement. $\alpha < 0$ means that it is less costly for the agent to recall downs than ups, which is the case when the agent is currently in the down context, i.e., when the signal in period $\tau + 1$ is a down price movement.

**Definition 2** *Associative memory is captured by $\alpha_u - \alpha_d =: \alpha \in (-\infty, \infty)$ where $|\alpha|$ measures the strength and $\alpha = 0$ refers to the absence of associative recall. Moreover, associative recall induces $\alpha_u^u = \alpha_u > \alpha_d = \alpha_d^u$ in the up context and $\alpha_u^d = \alpha_d \wedge \alpha_d^u = \alpha_u$ in the down context.*

Similar to before, Definition 2 assures that the up and down context are mirror images of each other, so that overall effort costs remain stable when changing the context for a given level of similarity. Our experiment manipulates the level of similarity by having cued as well as non-cued contexts for each subject. The following definition shows how such a change in similarity is captured by our model.

**Definition 3** *Increasing similarity induces $\alpha_u$ and $\alpha_d$ to increase in the up and down context, respectively.*

An increase in similarity makes it less costly for the agent to exert effort on ups in the up context and on downs in the down context. In our model, these changes in effort cost are captured by changing $\alpha_u$ and $\alpha_d$ as stated in Definition 3.

**Optimal Effort.** The agent's maximization problem (3) induces optimal effort levels $(e_u^*, e_d^*)$ to satisfy the first-order conditions

$$\mu_u\big(u - \hat{u}(e_u^*)\big)(1 - \tfrac{\bar{\tau}}{\tau})u = \tfrac{1-\mu_u}{\alpha_u} e_u^*, \quad (4)$$
$$\mu_d\big(d - \hat{u}(e_d^*)\big)(1 - \tfrac{\bar{\tau}}{\tau})d = \tfrac{1-\mu_d}{\alpha_d} e_d^*, \quad (5)$$

which equate the agent's marginal benefit (left-hand side) to her marginal cost (right-hand side) of exerting effort. The effort levels that solve these optimality conditions[8] are

$$(e_u^*, e_d^*) = \left( \frac{\mu_u (1 - \frac{\bar{\tau}}{\tau})^2}{\mu_u (1 - \frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{u^2}}, \frac{\mu_d (1 - \frac{\bar{\tau}}{\tau})^2}{\mu_d (1 - \frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_d}{\alpha_d}\frac{1}{d^2}} \right). \quad (6)$$

We now use (6), (2), and (1) to derive our results. When explaining these results, we sometimes refer to (4) and (5).

**Results.** The first proposition shows that if System 1 already produces correct estimates, there is no reason to activate System 2.

**Proposition 1 (Complete Memory)** *If memory is complete, i.e., $\bar{\tau} = \tau$, the agent exerts no effort, i.e., $(e_u^*, e_d^*) = (0, 0)$, and recalls correctly, i.e, $\hat{u}(e_u^*) = u$ and $\hat{d}(e_d^*) = d$.*

[8] In Appendix B, we show that the Hessian is negative definite, so $e_u^*$ and $e_d^*$ represent the unique global maxima of the target function $\Pi\left(|u - \hat{u}(e_u)|, |d - \hat{d}(e_d)|\right) - C(e_u, e_d)$.

Proposition 1 states that complete memory implies perfect memory. Since recall biases affect System 2 and complete memory leaves System 2 unactivated, in the following we assume that memory is incomplete, i.e., $\bar{\tau} \neq \tau$.

**Proposition 2 (No Recall Biases)** *Let $\bar{\tau} \neq \tau$. If both motivated and associative recall are absent, i.e., $\mu = \alpha = 0$, the agent exerts effort $e_u^* \gtreqless e_d^*$ if and only if $u \gtreqless d$.*

Proposition 2 shows that, in the absence of recall biases, effort is equally exerted on ups and downs whenever $u = d$. If $u = d$, without exerting any effort the distance between the true and recalled number of ups and downs is identical, i.e., $|u - \hat{u}(0)| = |d - \hat{d}(0)| > 0$. Since the agent values precision over cost equally ($\mu_u = \mu_d$) and effort is equally costly ($\alpha_u = \alpha_d$) for ups and downs, the effort levels that equate marginal benefit to marginal cost in (4) and (5) are identical. If $u > d$, without exerting any effort the distance between the true and recalled number of ups is larger than that of downs, i.e., $|u - \hat{u}(0)| > |d - \hat{d}(0)| > 0$, so that the agent needs to exert more effort on ups than downs to equate marginal benefit to marginal cost, given that she values precision over cost equally ($\mu_u = \mu_d$) and effort is equally costly ($\alpha_u = \alpha_d$) for ups and downs. The same argument holds vice versa if $u < d$. Intuitively, if the true signal distribution is uneven, the fact that the agent's System 1 proportionally extrapolates her recalled total number of signals to that distribution makes the absolute difference between System 1's recall of the number of signals and the true number of signals larger for the signal that has occurred more often. The agent therefore has a higher incentive to become more precise for that signal.

An implication of Proposition 2 is that $u = d$ nicely functions as a baseline case when investigating recall biases, because then any differences of exerted effort levels on ups and downs and on the recalled number of ups and downs are due to those memory imperfections. The reason is that, with $u = d$ and without recall biases, exerted effort levels are identical, i.e., $e_u^* = e_d^*$, and the recalled number of ups and downs is identical, i.e., $\hat{u}(e_u^*) = \hat{d}(e_d^*)$. We exploit these properties in Corollaries 1 and 2 below.

The next two propositions concern the interaction of the agent's System 1 and System 2 thinking.

**Proposition 3 (Initial Response)** *Fix any $u$, $d$, $(\mu_u, \mu_d)$, and $(\alpha_u, \alpha_d)$. The more imprecise the initial response of System 1 is, i.e., the larger $|\bar{\tau} - \tau|$ is, the more effort the agent exerts in System 2, i.e., the higher are $e_u^*$ and $e_d^*$.*

Proposition 3 shows that the more imprecise System 1 is, the more the agent needs to compensate for that imprecision in System 2 by exerting more effort.

**Proposition 4 (Effort Cost)** *Fix any $u$, $d$, and $\bar{\tau} \neq \tau$ and let $\mu = \alpha = 0$. The more weight the agent puts on precision over cost and the less costly effort is, i.e., the larger $(\mu_u, \mu_d)$ and $(\alpha_u, \alpha_d)$ are, the more effort the agent exerts in System 2, i.e., the higher are $e_u^*$ and $e_d^*$.*

Proposition 4 shows that $(\mu_u, \mu_d)$ and $(\alpha_u, \alpha_d)$ measure the relative importance of System 2 over System 1: the larger $(\mu_u, \mu_d)$ and $(\alpha_u, \alpha_d)$ are, the more weight the agent puts on System 2 relative to System 1, so that more effort is exerted. Thus, in our model $(\mu_u, \mu_d)$ and $(\alpha_u, \alpha_d)$ not only capture the agent's memory imperfections, but also the relative importance of System 1 over System 2.

We now examine how changing an agent's motivated memory influences her recall of ups and downs.

**Proposition 5 (Motivated Memory)** *Let $\bar{\tau} \neq \tau$ and $\hat{\Delta} := \hat{u}(e_u^*) - \hat{d}(e_d^*)$. An increase in motivated recall, i.e., a larger $|\mu|$, increases (decreases) $\hat{\Delta}$ in the long (short) position.*

Proposition 5 shows that motivated memory induces the agent to recall more (less) ups relative to downs in the long (short) position. This tendency has direct implications for her estimation of asset performance, which we explore in the following corollary.

**Corollary 1 (Motivated Memory)** *Let $\bar{\tau} \neq \tau$ and $u = d$. Suppose motivated recall is present, i.e., $\mu \neq 0$, and associative recall is absent, i.e., $\alpha = 0$. For $\bar{\tau} \lesseqgtr \tau$, the agent exerts effort $e_u^* \gtreqless e_d^*$ in the long position and $e_u^* \lesseqgtr e_d^*$ in the short position. As a result, the agent overestimates her asset's performance, i.e., $\hat{u}(e_u^*) > \hat{d}(e_d^*)$ in the long position and $\hat{u}(e_u^*) < \hat{d}(e_d^*)$ in the short position.*

Given that $\alpha_u = \alpha_d$ and $u = d$, any difference between $e_u^*$ and $e_d^*$ is due to motivated recall. Since $\frac{\partial e_u^*}{\partial \mu_u} > 0$ and $\frac{\partial e_d^*}{\partial \mu_d} > 0$, the presence of motivated recall induces differences in effort levels as stated in Corollary 1. Intuitively, the agent puts more weight on precision over cost if, in the long position, more precision increases the recalled number of ups and decreases the recalled number of downs. In the short position, the agent puts more weight on precision over cost, if more precision decreases the recalled number of ups and increases the recalled number of downs. In both positions, motivated recall induces the agent to overestimate asset performance, so that she becomes overoptimistic in her belief to hold the good asset.

Next, we examine how changing the agent's associative memory influences her recall of ups and downs.

**Proposition 6 (Associative Memory)** *Let $\bar{\tau} \neq \tau$ and $\hat{\Delta} := \hat{u}(e_u^*) - \hat{d}(e_d^*)$. For $\bar{\tau} < \tau$, an increase in associative recall, i.e., a larger $|\alpha|$, increases (decreases) $\hat{\Delta}$ in the up (down) context. For $\bar{\tau} > \tau$, an increase in associative recall decreases (increases) $\hat{\Delta}$ in the up (down) context.*

Proposition 6 shows that, if the agent tends to forget signals, i.e., $\bar{\tau} < \tau$, associative memory induces the agent to recall more (less) ups relative to downs in the up (down) context. If the agent tends to confabulate signals, i.e., $\bar{\tau} > \tau$, associative memory induces the agent to recall less (more) ups relative to downs in the up (down) context. This tendency has direct implications for how the agent's recall reacts to the current context, which is established in the following corollary.

**Corollary 2 (Associative Memory)** *Let $\bar{\tau} \neq \tau$ and $u = d$. Suppose motivated recall is absent, i.e., $\mu = 0$, and associative recall is present, i.e., $\alpha \neq 0$. The agent exerts effort $e_u^* > (<)\ e_d^*$ in the up (down) context. As a result, for $\bar{\tau} < \tau$, the agent's recall overreacts to the current context, i.e., $\hat{u}(e_u^*) > (<)\ \hat{d}(e_d^*)$ in the up (down) context. For $\bar{\tau} > \tau$, the agent's recall underreacts to the current context, i.e., $\hat{u}(e_u^*) < (>)\ \hat{d}(e_d^*)$ in the up (down) context*

Given that $\mu_u = \mu_d$ and $u = d$, any difference between $e_u^*$ and $e_d^*$ is due to associative recall. Since $\frac{\partial e_u^*}{\partial \alpha_u} > 0$ and $\frac{\partial e_d^*}{\partial \alpha_d} > 0$, the presence of associative recall induces differences in effort levels as stated in Corollary 2. Intuitively, the up (down) context makes it less costly to exert effort on ups (downs). If the agent tends to forget signals ($\bar{\tau} < \tau$), the higher exerted effort on ups in the up context makes the agent to recall more ups and the higher exerted effort on downs in the down context makes the agent to recall more downs, so that her beliefs overreact to the current context. By contrast, if the agent tends to confabulate signals ($\bar{\tau} > \tau$), the higher exerted effort on ups in the up context makes the agent to recall less ups and the higher exerted effort on downs in the down context makes the agent to recall less downs, so that her beliefs underreact to the current context. Since recall precision of the current context increases, associative recall induces over- or underreaction in beliefs depending on the agent respectively forgetting or confabulating signals.[9]

Propositions 5 and 6 show that the optimally exerted effort on ups and down is different depending on the environment, and that the distinguishing environmental features are different for motivated and associative memory. The two propositions further identify for each memory imperfection *how* the agent's recall of ups and downs is biased, which directly relates to her belief bias. We next identify in which environments both memory imperfections impact the agent's recall in the same or opposite direction and the following definition is useful in this regard.

**Definition 4 (Alignment)** *Let $\bar{\tau} \neq \tau$ and $\hat{\Delta} := \hat{u}(e_u^*) - \hat{d}(e_d^*)$. Associative and motivated memory are ...*

*(i) aligned upwards (downwards), if an increase in each recall bias increases (decreases) $\hat{\Delta}$;*

*(ii) misaligned, if an increase in one recall bias increases $\hat{\Delta}$, while an increase in the other recall bias decreases $\hat{\Delta}$.*

Whether motivated and associative memory are aligned or misaligned depends on the context, the asset position, and whether the agent tends to forget or confabulate signals. Concerning the latter, note that the agent's System 1 response $\bar{\tau}$ is unobservable, whereas

[9] Note that the general overreaction implication of associative recall in Enke et al. (2024) is due to the fact that they only considers agents who tend to forget signals.

her System 2 response $\hat{\tau} := \hat{u}(e_u^*) + \hat{d}(e_d^*)$ is observable. Because the agent optimally exerts effort $(e_u^*, e_d^*) \in (0,1)^2$ if $\bar{\tau} \neq \tau$ and $(e_u^*, e_d^*) = (0,0)$ if $\bar{\tau} = \tau$, (1) implies that

$$\hat{\tau} \gtrless \tau \iff \bar{\tau} \gtrless \tau.$$

Since we need to identify alignment and misalignment both theoretically and empirically, we base the following lemma on the observable System 2 response $\hat{\tau}$ rather than the unobservable System 1 response $\bar{\tau}$.

**Lemma 1 (Alignment)** *Let $\bar{\tau} \neq \tau$ and $\hat{\tau} := \hat{u}(e_u^*) + \hat{d}(e_d^*)$. Associative and motivated memory are . . .*

*(i) aligned upwards, if*
$\{up\ context, \hat{\tau} < \tau, long\ position\} \cup \{down\ context, \hat{\tau} > \tau, long\ position\} =: ++$

*(ii) aligned downwards, if*
$\{up\ context, \hat{\tau} > \tau, short\ position\} \cup \{down\ context, \hat{\tau} < \tau, short\ position\} =: --$

*(iii) misaligned, if*
$\{up\ context, \hat{\tau} < \tau, short\ position\} \cup \{down\ context, \hat{\tau} > \tau, short\ position\} =: +-$
$\{up\ context, \hat{\tau} > \tau, long\ position\} \cup \{down\ context, \hat{\tau} < \tau, long\ position\} =: -+$

Lemma 1 identifies environments that align or misalign both memory imperfections. $++$ and $--$ are the aligned environments, in which an increase in each recall bias respectively increases and decreases $\hat{\Delta}$. $+-$ and $-+$ are the misaligned environments, in which an increase in one recall bias increases $\hat{\Delta}$, while an increase in the other recall bias decreases $\hat{\Delta}$.

We now turn to our main investigation, namely the interaction between associative and motivated memory. Both in the theory and the experiment, we explore this interaction by changing the level of similarity.

**Definition 5 (Interaction)** *Let $\bar{\tau} \neq \tau$ and $\hat{\Delta} := \hat{u}(e_u^*) - \hat{d}(e_d^*)$. Associative and motivated memory are complements (substitutes, unrelated), if the impact that an increase in similarity has on $\hat{\Delta}$ is larger (smaller, equal) in aligned than misaligned environments, i.e., in $++$ than $+-$ and in $--$ than $-+$.*

Definition 5 makes clear what complements and substitutes are in our setting: Associative and motivated memory are complements (substitutes, unrelated), if the impact of increasing similarity on $\hat{\Delta}$ is larger (smaller, equal) when both memory imperfections are aligned than misaligned.

The following proposition constitutes our main result.

**Proposition 7 (Interaction)** *Associative and motivated memory are complements (substitutes, unrelated) if*

$$\mu_u < (>,=) \frac{1}{1+\alpha_u u^2\left(1-\frac{\bar{\tau}}{\tau}\right)^2} \iff \alpha_u < (>,=) \frac{\frac{1}{\mu_u}-1}{u^2\left(1-\frac{\bar{\tau}}{\tau}\right)^2}$$

$$\textit{and} \quad \mu_d < (>,=) \frac{1}{1+\alpha_d d^2\left(1-\frac{\bar{\tau}}{\tau}\right)^2} \iff \alpha_d < (>,=) \frac{\frac{1}{\mu_d}-1}{d^2\left(1-\frac{\bar{\tau}}{\tau}\right)^2}.$$

Proposition 7 shows that associative and motivated memory are complements if $(\mu_u, \mu_d)$ and $(\alpha_u, \alpha_d)$ are sufficiently small, and they are substitutes if $(\mu_u, \mu_d)$ and $(\alpha_u, \alpha_d)$ are sufficiently large. Together with Proposition 4, this result implies the following:

**Corollary 3 (Interaction and Effort)** *Associative and motivated memory are complements (substitutes) if System 1 (2) is sufficiently important relative to System 2 (1), so that exerted effort is sufficiently low (high).*

Thus, if the agent's recall relies more on System 1 and is instantaneous and instinctive, she can well exploit similarity to self-servingly bias her recall. In this case, associative and motivated memory are complements. On the other hand, if the agent's recall relies more on System 2 and is deliberate and conscious, she is rather hindered by similarity to self-servingly bias her recall. In this case, associative and motivated memory are substitutes.

There are two effects that work against each other. Since the explanation is similar across environments, for the sake of brevity we focus on the up context for agents that tend to forget signals. Recall that $e_u^*$ is determined by equating the marginal benefit to the marginal cost of effort, so that (4) holds. Now, increasing similarity, i.e., increasing $\alpha_u$ decreases marginal cost, i.e., the right-hand side (RHS) of (4). To restore equality, $e_u^*$ has to increase as increasing $e_u^*$ increases the RHS of (4) and decreases its left-hand side (LHS) because $\hat{u}(e_u^*)$ moves closer to $u$ with increasing $e_u^*$. Moving from the short to long position for agents that tend to forget signals increases $\mu_u$ (and decreases $\mu_d$). A larger $\mu_u$ has two effects on this impact of increasing $\alpha_u$ (a smaller $\mu_d$ has no effect). First, it makes the decrease in marginal cost smaller because $\frac{\partial(1-\mu_u/\alpha_u)}{\partial\alpha_u} = \frac{-(1-\mu_u)}{\alpha_u^2}$ is increasing in $\mu_u$, so that $e_u^*$ has to increase by less. This is the substitutive effect. Second, a larger $\mu_u$ increases marginal benefit, i.e., the LHS of (4), so that $e_u^*$ has to increase by more to restore equality. This is the complementary effect. Whether the model predicts complements or substitutes depends on which effect outweighs the other. If $\mu_u$ and $\alpha_u$ are rather small to start with, $e_u^*$ is relatively low and $|u - \hat{u}(e_u^*)|$ is relatively large. In this case, the complementary effect outweighs the substitutive effect. On the other hand, when $\mu_u$ and $\alpha_u$ are rather large, $e_u^*$ is relatively high and $|u - \hat{u}(e_u^*)|$ is relatively small. In this case the substitutive effect dominates the complementary effect. In the knife-edge case where the two effects exactly balance each other, associative and motivated memory are unrelated.

**Implications for Beliefs.** Our final investigation concerns implications of interaction effects for agents' beliefs.

**Corollary 4 (Beliefs)** *If associative and motivated memory are complements (substitutes, unrelated), an increase in similarity induces more (less, equally) favorable beliefs regarding asset performance in aligned than misaligned environments, i.e., in* $++$ *than* $+-$ *and in* $--$ *than* $-+$.

As an example, consider first an agent who tends to forget signals ($\hat{\tau} < \tau$) and is in the up context. If the agent is in the long position, she will be in the aligned $++$ environment, and if the agent is in the short position, she will be in the misaligned $+-$ environment. Under complements (substitutes), an increase in similarity implies that $\hat{\Delta}$ is larger (smaller) in $++$ than $+-$, so that the agent's belief is more (less) favorable in the aligned than misaligned environment. Next, consider the same scenario but an agent who tends to confabulate signals ($\hat{\tau} > \tau$). Then, the agent is in the misaligned $-+$ environment in the long position and in the aligned $--$ environment in the short position. Under complements (substitutes), an increase in similarity now implies that $\hat{\Delta}$ is smaller (larger) in $--$ than $-+$, so that again the agent's belief becomes more (less) favorable in the aligned than misaligned environment. A similar reasoning applies to an agent in the down context.

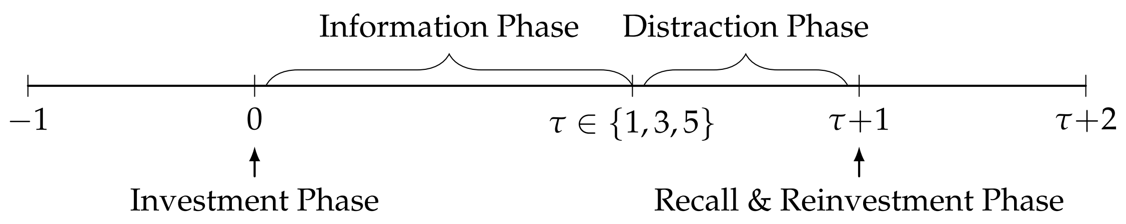
Corollary 4 can also be applied to a scenario in which there are two agents that are identical, except that one agent is in the long while the other is in the short position. If we suppose that the aligned agent has more favorable beliefs to start with than the misaligned agent, a marginal increase in similarity causes the two agents' beliefs to polarize (converge) under complements (substitutes). We can further interpret Corollary 4 against the backdrop of Corollary 3: Suppose that the importance of System 1 over System 2 varies across decision contexts. For instance, concerning investment decisions it may be reasonable to assume that in highly volatile spot markets recall is mainly driven by unconscious System 1, while trading on fundamentals is rather driven by conscious System 2. Then, we would expect a larger complementary effect being present in the former spot market, so that we would expect a larger disagreement between asset holders and non-holders than under fundamental trading.

# 4 Experimental Design

The experiment was pre-registered and conducted at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA).[10] The currency in our experiment was

[10]We pre-registered the experiment via AsPredicted #116752. The (translated) experimental instructions, that were given to subjects before they entered the oTree program interface, can be found on *https://drive.google.com/file/d/1lEWYT58AZcyNHHcrV3Pl7gIz46YWjuxe/view*. A link to the oTree program can be obtained from the authors upon request.

Experimental Monetary Units (EMU) and the exchange rate was 2000 EMU = €1. In our main treatment (MAIN), 291 subjects participated and earned €15.65 on average with a maximum payment of €26.87 and a minimum payment of €7.02. Our experiment used a specific parameterization of the general setup introduced in Section 2 and separated the asset market into different phases:




**Investment Phase.** First, a subject saw a statistical summary of two pre-movements of the risky asset's price ($k = -1$) to make an informed choice of putting her market endowment $W_0 = 20000$ EMU in either the risky or safe asset. In order to increase subjects' incentives to be invested in the risky asset, we imposed safe asset account fees of 10000 EMU for the entire investment horizon.[11] It was randomly drawn whether the risky asset followed the high or low binomial process (both were equally likely). The high process appreciated with probability $p_H = 0.6$ and the low process appreciated with probability $p_L = 0.4$. Both processes appreciated and depreciated by the same absolute amount $U = 10$ EMU and the starting price of the risky asset was set to $P_k = 100$ EMU.

**Information Phase.** After their investment decision subjects subsequently received one, three, or five signals (i.e., up or down price movements), so that $\tau \sim \mathcal{U}\{1, 3, 5\}$. Every signal screen appeared for 20 seconds and any two signal screens were divided by a waiting screen that appeared for 8 seconds (to help making signals more differentiable in recall). After observing the signals, in period $\tau$ subjects were reminded of the price $P_0$ and had to state the current price $P_\tau$. If their answer was correct, they received 2000 EMU. We used these answers as an attention check to make sure that later deviations are in fact caused by memory distortions rather than earlier inattention. After answering the attention check, subjects also had to state their belief that the risky asset follows the good process. These elicited beliefs are used to control for individual-specific updating biases independent from memory distortions.

**Distraction Phase.** After the investment and information phases had been completed in a market, a new market began. In total, subjects faced ten of these asset markets. After completion of all ten markets, subjects had to do an unrelated real effort task to blur their memory. This real-effort task lasted for 10 minutes. In each of these tasks, subjects received a random combination of 9 (lower- and upper-case) letters and a random combination of numbers from 1 to 9. They then had to reorder the given letters in the given

[11] We conjectured that for motivated memory it is essential that the investor is invested in the risky asset. Such high account fees should induce the investor to invest in the risky asset, while allowing to hold $F_h$ and $F_l$ entirely symmetric and constant between the two trading schemes (the long and short position).

order. Per correctly solved task they received 200 EMU. After the real effort task, subjects left the laboratory and had to return to the experiment online three days later. The payment for the experiment was electronically transferred to their bank accounts after finishing the online experiment at the second date and they received no payments when they failed to come back to the online part of the experiment.

**Recall Phase.** After three days the experiment continued with the recall phase. In each of the ten asset markets (presented in a different random order as on the first date), subjects first received a final signal (i.e., up or down price movement), were then reminded of the price $P_0$ and whether they were invested in the stock, and then had to recall all signals (i.e., the number of ups and the number of downs) since their initial investment decision, i.e., from $t = 1$ to $t = \tau + 1$.[12] Furthermore, subjects again had to state their belief that the risky asset follows the good process.

**Reinvestment Phase.** The reinvestment phase came as a surprise to subjects to avoid interference with motivated memory. From the beginning of the experiment, subjects knew that only one of the ten markets would be randomly selected for payment. In the beginning of the reinvestment phase they were informed which market was selected and that their previous investment in this market is liquidated according to the actual price path, but they received no feedback on the outcome. They were also informed about their previously recalled number of ups and downs, and their belief that the risky asset follows the good process in this market. Subjects then received a new market endowment $W_{\tau+1} = 4000$ and could distribute this new endowment freely between the risky and safe asset, knowing that investments would be automatically liquidated after another price movement in $t = T = \tau + 2$. The goal of this reinvestment phase was to identify whether memory distortions are action relevant.

**Feedback.** At the end of the experiment, after subjects answered a short questionnaire, they received feedback about their total earnings from the experiment. We avoided giving them more detailed information about payments to allow for unconstrained motivated cognition (Drobner, 2022).

**Manipulating Associativeness.** We manipulated associativeness by having five cue (C) and five non-cue (NC) markets.[13] Each market considered a specific company (e.g., a car producer) and the risky asset referred to the company's stock. Each signal in a market was embedded in a specific context, consisting of a story that described the reason for the price movement and a story-fitting picture.[14] We told subjects that these contexts may be helpful to remember price movements at a later stage in the experiment. In a cue

[12] We incentivized this recall in the following way: If subjects recalled the number of ups and downs correctly, they received 8000 EMU with certainty. If they made one deviation mistake, they received 8000 EMU with probability 0.8. If they made two mistakes, they received 8000 EMU with probability 0.2. And if they made three or more mistakes, they received nothing.

[13] Specifically, for each draw of $\tau \sim \mathcal{U}\{1, 3, 5\}$, we had a cue and a non-cue market. Thus, for each subject $\tau$ was drawn five times.

[14] Figures 9 and 10 in Appendix A show an example of a down and up signal, respectively.

market, the contexts which embedded the ups and downs were kept constant, so that each up occurred in the same context and each down occurred in the same context. In a non-cue market, the context for each price movement was different. Thus, the level of signals' similarity in cue markets should be higher than in non-cue markets and this manipulation consisted a within-subject variation.

**Diametrically Opposing Motivations.** If subjects were invested in stocks, price movements constituted good or bad news for their investment performance. In order to further boost their motivated cognition, subjects whose actual investment performance was in the top half of their experimental session additionally participated in a coin-flip lottery to win a prize of 8000 EMU. Whether ups or downs are good or bad news depends on the incentive structure. It was randomly assigned whether a subject was in a short or long position. A subject in the long position profited from price increases while a subject in the short position profited from price decreases, and since the stochastic environment was entirely symmetric, both positions induced diametrically opposing motivations. The asset position of each subject was kept constant over the entire experiment, so this manipulation consisted an across-subject variation.

**Robustness.** In addition to our main treatment (MAIN) described above, we also conducted the immediate treatment (IMM). IMM differed from MAIN by adding another recall phase at the end of the first date of the experiment. The idea of this design feature was to manipulate the level of both motivated and associative memory. We expected that motivations play a smaller role when subjects had less time to build up these motivations, and we also expected that similarity plays a smaller role when recalled signals were more recent. Hence, manipulating motivated and associative memory through this treatment variation should result in a substantially decreased or even absent interaction effect. 94 subjects participated in IMM, of which 7 subjects did not come back three days later and hence received no payment. The remaining 87 subjects earned €19.00 on average with a maximum payment of €34.89 and a minimum payment of €7.93.

# 5 Hypotheses

In this section, we present empirical hypotheses that are derived from our theoretical model. These hypotheses cover the presence of motivated and associative memory, their interaction, the link between memory and beliefs, and the link between beliefs and actions.

**Presence of Motivated Memory.** In our model, we assumed that the initial response of System 1 is exogenous and not motivated, so that any presence of motivated memory is driven by System 2. In other words, an agent's tendency to forget or confabulate signals should be independent of her motivations. Thus, we should firstly observe that the probabilities of forgetting and confabulating signals are not different for subjects in the long and short position, conditional on having seen more ups than downs ($u > d$) and more downs than ups ($u < d$).[15] Motivated memory in System 2, however, distorts the true difference between ups and downs ($\Delta$) upwards in the long position and downwards in the short position. Thus, the presence of motivated memory in System 2 implies that the stated difference between ups and downs ($\hat{\Delta}$) is larger than the true difference in the long position and smaller in the short position. Thus, we should secondly observe that $\hat{\Delta} - \Delta$ is larger in the long than short position. In order to isolate the presence of motivated memory from potential interaction effects with associative memory, we restrict attention to non-cue markets.

**Hypothesis 1 (Presence Motivated Memory)** *Motivated memory is present if (i) and (ii) hold:*

*(i) Conditional on $u > d$ and $u < d$, $\Pr(\hat{\tau} > \tau)$ and $\Pr(\hat{\tau} < \tau)$ are not different between the long and short position.*

*(ii) $\hat{\Delta} - \Delta$ is larger for the long than short position in non-cue markets.*

**Presence of Associative Memory.** Similar to motivated memory, our model assumed that the initial response of System 1 is exogenous and not influenced by associative memory, so that any presence of associative memory is driven by System 2. We may again validate this assumption by looking at an agent's tendency to forget or confabulate signals, which should be independent of a change in similarity. Thus, we should firstly observe that the probabilities of forgetting and confabulating signals are not different for subjects in cue and non-cue markets. The presence of associative memory in System 2, however, makes recall more precise in cue than non-cue markets. Thus, we should secondly observe that $|\hat{u} - u|$ in the up context and $|\hat{d} - d|$ in the down context are smaller in cue than non-cue markets. Finally, we can ask what the effects of increased precision in cue markets for the stated difference between ups and downs ($\hat{\Delta}$) compared to the true difference ($\Delta$) are. Note that increased precision may increase or decrease $\hat{\Delta}$ depending on the context *and* the tendency to forget or confabulate signals. According to our model (see Lemma 1), associative memory increases $\hat{\Delta}$ in $++$ and $+-$ environments, while motivated memory is balanced across these environments. Likewise, associative memory decreases $\hat{\Delta}$ in $--$ and $-+$ environments, while motivated memory is balanced across these environments. Thus, the presence of associative memory in System 2 thirdly implies in cue markets that the stated difference between ups and downs ($\hat{\Delta}$) is larger than the true difference ($\Delta$) when conditioning on $+ + \cup + -$ and smaller when conditioning on $- - \cup - +$.

[15]The case $u = d$ does not occur in our experiment because $\tau \in \{1, 3, 5\}$. Further, note that this is not an obvious prediction: If $u > d$, motivated subjects in the long position would benefit (lose) from confabulating (forgetting) signals, because $\bar{u} - \bar{d}$ would increase (decrease). Likewise, if $u < d$, motivated subjects in the long position would lose (benefit) from confabulating (forgetting) signals, because $\bar{u} - \bar{d}$ would decrease (increase).

**Hypothesis 2 (Presence Associative Memory)** *Associative memory is present if (i), (ii), and (iii) hold:*

*(i)* $\Pr(\hat{\tau} > \tau)$ *and* $\Pr(\hat{\tau} < \tau)$ *are not different between cue and non-cue markets.*

*(ii)* $\{|\hat{u} - u| \mid$ *up context*$\}$ *and* $\{|\hat{d} - d| \mid$ *down context*$\}$ *are smaller in cue than non-cue markets.*

*(iii)* $\{\hat{\Delta} - \Delta| ++ \cup +-\} > \{\hat{\Delta} - \Delta| -- \cup -+\}$ *in cue markets.*

**Interaction of Memory Imperfections.** Our main investigation concerns the interaction of memory imperfections. In order to control for the incompleteness of memory and to separate incomplete from imperfect memory, we construct a benchmark difference between ups and downs:

$$\bar{\Delta} := \frac{\hat{\tau}}{\tau}\Delta$$

where $\hat{\tau} = \hat{u} + \hat{d}$, $\tau = u + d$, and $\Delta := u - d$. This benchmark difference adjusts the true difference between ups and downs by the factor with which subjects tend to forget or confabulate signals. For instance, if the true difference between ups and downs is $\Delta = 4$ and the agent forgets half of the signals, her benchmark difference will be $\bar{\Delta} = 2$. So far, there was no need to control for memory incompleteness, because we did not investigate $\hat{\Delta}$ *across* cue and non-cue markets. However, investigating a change in similarity by comparing cue to non-cue markets requires such control, because $\hat{\Delta}$ mechanically changes if increasing similarity makes memory more complete (which is the case in our model).

Since we identify the interaction of memory imperfections by investigating the effect of changes in similarity, we need to abstract away from mechanically induced memory incompleteness and focus on the distortion of subjects' stated difference between ups and downs, i.e., $\hat{\Delta} := \hat{u} - \hat{d}$, from the benchmark difference $\bar{\Delta}$, depending on the alignment of both memory imperfections. Our investigation uses the following difference-in-difference measures:

$$\begin{aligned}
D^{++} &:= \{\hat{\Delta} - \bar{\Delta} \mid ++\}, \\
D^{-+} &:= \{\hat{\Delta} - \bar{\Delta} \mid -+\}, \\
D^{+-} &:= \{\hat{\Delta} - \bar{\Delta} \mid +-\}, \\
D^{--} &:= \{\hat{\Delta} - \bar{\Delta} \mid --\}.
\end{aligned}$$

According to Lemma 1 and Definitions 4 and 5, we can identify whether associative and motivated memory complement or substitute each other by two independent interaction tests that use these distortion measures, separately for cue (C) and non-cue (NC) markets:

**Hypothesis 3 (Interaction)** *Associative and motivated memory are complements (substitutes, unrelated) if (i) and (ii) hold:*

(i) *Interaction Test I:* $D_C^{++} - D_{NC}^{++} > (<, =)\ D_C^{+-} - D_{NC}^{+-}$.

(ii) *Interaction Test II:* $D_C^{--} - D_{NC}^{--} < (>, =)\ D_C^{-+} - D_{NC}^{-+}$.

An implication of Proposition 7 and Corollary 3 is that the complementary effect becomes stronger for lower effort levels. Consistent with our theory, we can proxy the effort level in the experiment with the absolute distance between $\hat{\tau}$ and $\tau$ and split the sample into high and low effort.[16]

**Hypothesis 4 (Interaction and Effort)** *Associative and motivated memory are more complementary for low than high effort if (i) and (ii) hold:*

(i) *The difference between the left- and right-hand side of Interaction Test I is larger for low than high effort.*

(ii) *The difference between the left- and right-hand side of Interaction Test II is smaller for low than high effort.*

**Mapping Memory to Beliefs.** Theoretically, recall biases matter because they influence agents' expectation formation. In order to empirically evaluate how recall biases indeed influence subjects' belief formation, we use four different beliefs in our empirical analysis. First, we use subjects' empirically stated belief in period $t$ that the asset follows the high process, i.e., $\hat{q}_t$. Second, we use the Bayesian belief in period $t$,

$$q_t^B = \frac{0.6^{\Delta_t}}{0.6^{\Delta_t} + 0.4^{\Delta_t}},$$

which is calculated based on the true difference between the number of ups and downs until period $t$, i.e, $\Delta_t$. Third, the "true subjective" belief in period $t$, $q_t^{TS}$, is also computed based on $\Delta_t$, but allows for individual-specific non-Bayesian updating, captured by the parameter $\beta$:

$$q_t^{TS} = \frac{0.6^{\Delta_t \beta}}{0.6^{\Delta_t \beta} + 0.4^{\Delta_t \beta}}.$$

Finally, the "recalled subjective" belief in period $t$, $q_t^{RS}$, is computed based on the recalled difference between the number of ups and downs until period $t$, namely $\hat{\Delta}_t$, and allows for non-Bayesian updating:

$$q_t^{RS} = \frac{0.6^{\hat{\Delta}_t \beta}}{0.6^{\hat{\Delta}_t \beta} + 0.4^{\hat{\Delta}_t \beta}}.$$

For each subject in each market, we compute the non-Bayesian updating parameter in period $\tau$ (when subjects still perfectly recall all signals) by letting $\hat{q}_\tau = q_\tau^{TS}$ and solving for $\beta$. In any given market, we seek to explain subjects' stated beliefs in period $\tau + 1$, i.e.,

[16] In our theoretical model, for a given $\bar{\tau}$, higher effort implies a smaller absolute distance between $\hat{\tau}$ and $\tau$.

$\hat{q}_{\tau+1}$. In our model, $\hat{q}_{\tau+1}$ may deviate from the Bayesian belief $q^B_{\tau+1}$ for two reasons, non-Bayesian updating and imperfect recall. Accounting only for non-Bayesian updating by plugging the calculated $\beta$ from period $\tau$ in $q^{TS}_{\tau+1}$ should not fully explain $\hat{q}_{\tau+1}$ if recall is incorrect. However, when additionally accounting for incorrect recall by plugging the calculated $\beta$ from period $\tau$ in $q^{RS}_{\tau+1}$, we should be able to fully explain $\hat{q}_{\tau+1}$ according to our model.

**Hypothesis 5 (Explaining Beliefs)** *We have*

$$|\hat{q}_{\tau+1} - q^B_{\tau+1}| \geq |\hat{q}_{\tau+1} - q^{TS}_{\tau+1}| \geq |\hat{q}_{\tau+1} - q^{RS}_{\tau+1}| = 0.$$

*With non-Bayesian updating, the first inequality is strict and with incorrect recall, the second inequality is strict.*

Given that subjects' stated beliefs can be (partly) explained by their recall, we can further test Corollary 4 against the backdrop of Corollary 3. Together, the two corollaries imply that aligned subjects should have more favorable beliefs than misaligned subjects in cue markets, and more so for low than high effort levels. There are two ways to identify the favorability of beliefs when comparing situations in which subjects do not observe identical information. First, we can measure favorability by the difference between subjects' stated belief $\hat{q}_{\tau+1}$ and the Bayesian belief $q^B_{\tau+1}$. A second measure of favorability additionally controls for the incompleteness of memory and, instead of the Bayesian belief, uses the "benchmark Bayesian belief" $q^{BB}_{\tau+1}$ as comparison, which is based on the benchmark difference between the number of ups and downs $\bar{\Delta}_{\tau+1} := \Delta_0 + \frac{\hat{\tau}}{\tau}\Delta + \Delta_{\tau+1} - \Delta_\tau$, so that

$$q^{BB}_{\tau+1} = \frac{0.6^{\bar{\Delta}_{\tau+1}}}{0.6^{\bar{\Delta}_{\tau+1}} + 0.4^{\bar{\Delta}_{\tau+1}}}.$$

While the first favorability measure captures a common understanding of belief distortion, the second favorability measure additionally identifies that belief distortion is due to imperfection rather than incompleteness of memory. The following hypothesis takes both measures into account and builds on our theoretical insight that associative and motivated memory are more complementary for low than high effort.

**Hypothesis 6 (Beliefs and Effort)** *There is a larger difference in the favorability of beliefs between aligned and misaligned environments for low than high effort in cue markets if (i) and (ii) hold:*

*(i)* $\{\hat{q} - q^B | ++ \cup --\} - \{\hat{q} - q^B | -+ \cup +-\}$ *is larger for low than high effort in cue markets.*

*(ii)* $\{\hat{q} - q^{BB} | ++ \cup --\} - \{\hat{q} - q^{BB} | -+ \cup +-\}$ *is larger for low than high effort in cue markets.*

Cases $(i)$ and $(ii)$ of Hypothesis 6 are identical except that favorability of subjects' stated belief is measured against the Bayesian belief in the former and against the benchmark Bayesian belief in the latter case. Note that in order to measure favorability of beliefs independent of the asset position, Hypotheses 6, 7, and 8 make use of the following transformation: $\hat{q}$ is the agent's stated belief in period $\tau+1$ that the asset follows the *good* process, so $\hat{q}$ equals $\hat{q}_{\tau+1}$ in the long position and $1-\hat{q}_{\tau+1}$ in the short position. Likewise, $q^B$ ($q^{BB}$) is the (benchmark) Bayesian belief in period $\tau+1$ that the asset follows the *good* process, so $q^B$ ($q^{BB}$) equals $q^B_{\tau+1}$ ($q^{BB}_{\tau+1}$) in the long position and $1-q^B_{\tau+1}$ ($1-q^{BB}_{\tau+1}$) in the short position.

**Mapping Beliefs to Actions.** Once we have established that recall biases determine subjects' beliefs, we can investigate the final link, namely whether these beliefs translate into actual behavior. In our experiment, subjects received a new endowment in period $\tau+1$, which they could freely distribute between the risky and safe asset. This reinvestment phase came as a surprise and hence does not interfere with previous recall or expectation formation.

Theoretically, we assume that the agent is a risk-averse expected utility maximizer with Bernoulli utility function $u(x)$ (where $u'(x)>0$, $u''(x)<0$), who can invest $X_{\tau+1}\in[0,W_{\tau+1}]$ into the risky asset. While not investing (i.e., the safe asset) yields a gross return of 1, the risky asset yields a random gross return of $\tilde{z}=[z^+,q^+\,;\,z^-,q^-\,]$, where $z^+>1$, $z^-<1$ and $q^+:=\hat{q}p_G+(1-\hat{q})p_B$, $q^-:=\hat{q}(1-p_G)+(1-\hat{q})(1-p_B)$. $\hat{q}$ is the agent's belief in period $\tau+1$ that the asset follows the good process (see above). $p_G$ ($p_B$) is the probability that the asset moves in the favorable direction if it is the good (bad) asset, so $p_G$ ($p_B$) equals $p_H$ ($p_L$) in the long position and $1-p_L$ ($1-p_H$) in the short position. The symmetric nature of our stochastic processes implies $z^+-1=1-z^-$ and $p_H=1-p_L$, so that $\mathbb{E}[\tilde{z}]\gtrless 1 \iff \hat{q}\gtrless 1/2$. Given her beliefs in period $\tau+1$, the agent's optimal investment $X^*_{\tau+1}$ solves

$$\max_{X_{\tau+1}\in[0,W_{\tau+1}]} q^+\,u\left(W_{\tau+1}+X_{\tau+1}(z^+-1)\right)+q^-\,u\left(W_{\tau+1}+X_{\tau+1}(z^--1)\right) \tag{7}$$

and thus exhibits $X^*_{\tau+1}>(=)\;0$ if $\hat{q}>(\leq)\;1/2$. In case of the corner solution, there may be no effect of increasing beliefs $\hat{q}$ on the optimal investment $X^*_{\tau+1}$. In case of the interior solution, we can use the implicit function theorem to show that $\frac{dX^*_{\tau+1}}{d\hat{q}}>0$, so that increasing beliefs increase optimal investments into the risky asset.

**Hypothesis 7 (Actions)** *Suppose $\hat{q}>1/2$. There is a positive correlation between the belief to own the good asset $\hat{q}$ and the share invested into the risky asset $\frac{\hat{X}_{\tau+1}}{W_{\tau+1}}$.*

In the experiment, we can test Hypothesis 7 across subjects since each subject states her belief $\hat{q}$ as well as her investment $\hat{X}_{\tau+1}$ in one randomly drawn market.

While Hypothesis 7 establishes that beliefs are indeed action-relevant, we can further investigate whether the favorability of beliefs translates into investment mistakes. We define investment mistakes individually for each subject by accounting for individual-specific risk aversion in the following way: Restricting attention to the interior solution (i.e., $\hat{q} > 1/2$), we assume that the agent exhibits constant relative risk aversion, i.e.,

$$u(x) = \frac{x^{1-\gamma} - 1}{1-\gamma},$$

and then solve the first-order condition of (7) for $\gamma$ by letting $X^*_{\tau+1} = \hat{X}_{\tau+1}$, so that each subject reveals her individual-specific risk aversion parameter with her investment choice.[17] We then use this risk aversion parameter to determine the agent's optimal investment given the Bayesian and benchmark Bayesian belief. Specifically, $X^{*B}_{\tau+1}$ solves

$$\max_{X^B_{\tau+1} \in [0, W_{\tau+1}]} q^{+B}\, u\left(W_{\tau+1} + X^B_{\tau+1}(z^+ - 1)\right) + q^{-B}\, u\left(W_{\tau+1} + X^B_{\tau+1}(z^- - 1)\right),$$

when plugging in the agent's revealed $\gamma$ and using $q^{+B} := q^B p_G + (1 - q^B) p_B$ and $q^{-B} := q^B(1 - p_G) + (1 - q^B)(1 - p_B)$, where $q^B > 1/2$ is the Bayesian belief in period $\tau + 1$ that the asset follows the good process (see above). Likewise, $X^{*BB}_{\tau+1}$ solves

$$\max_{X^{BB}_{\tau+1} \in [0, W_{\tau+1}]} q^{+BB}\, u\left(W_{\tau+1} + X^{BB}_{\tau+1}(z^+ - 1)\right) + q^{-BB}\, u\left(W_{\tau+1} + X^{BB}_{\tau+1}(z^- - 1)\right),$$

when plugging in the same revealed $\gamma$, but using $q^{+BB} := q^{BB} p_G + (1 - q^{BB}) p_B$ and $q^{-BB} := q^{BB}(1 - p_G) + (1 - q^{BB})(1 - p_B)$, where $q^{BB} > 1/2$ is the benchmark Bayesian belief in period $\tau + 1$ that the asset follows the good process (see above). Investment mistakes, that control for subjects' risk aversion, can then be captured by comparing subjects' actual investment $\hat{X}_{\tau+1}$ to the Bayesian investment $X^{*B}_{\tau+1}$ and to the benchmark Bayesian investment $X^{*BB}_{\tau+1}$.

**Hypothesis 8 (Investment Mistakes)** *Suppose $\hat{q} > 1/2$, $q^B > 1/2$, and $q^{BB} > 1/2$. There is a positive correlation between the favorability of beliefs and over-investments if (i) and (ii) hold:*

*(i) $\hat{q} - q^B$ positively correlates with $\frac{\hat{X}_{\tau+1} - X^{*B}_{\tau+1}}{W_{\tau+1}}$.*

*(ii) $\hat{q} - q^{BB}$ positively correlates with $\frac{\hat{X}_{\tau+1} - X^{*BB}_{\tau+1}}{W_{\tau+1}}$.*

Hypothesis 8 identifies whether belief distortions translate into behavioral distortions when accounting for individual-specific risk attitudes.

[17]Empirically, we assume that nothing invested amounts to $X^*_{\tau+1} = 0.00001 \times W_{\tau+1}$ and everything invested to $X^*_{\tau+1} = 0.99999 \times W_{\tau+1}$ to get closed-form solutions for $\gamma$.

# 6 Empirical Findings

The presence of both recall biases is a pre-requisite for being able to measure their interaction. Sections 6.1 and 6.2 establish that both recall biases are actually present in our decision environment and we investigate how they interact in Section 6.3. To identify whether subjects' recall biases actually translate into their beliefs beyond non-Bayesian updating, we analyze the mapping from memory to beliefs in Section 6.4. The mapping from beliefs to actions is analyzed in Section 6.5, which establishes that our documented recall biases and induced beliefs further translate into subjects' investment behavior and corresponding investment mistakes. All these sections use the data of treatment MAIN. The data of treatment IMM is used in Section 6.6, where we investigate the robustness of our findings.

Throughout, we use various empirical tests: In addition to Spearman correlations, OLS and logit regressions, we use non-parametric Mann-Whitney U tests (MWU) for across-subject comparisons and Wilcoxon signed-rank tests (WSR) for within-subject comparisons. For these tests, we report the $p$-values of a two-sided ($p_2$) or one-sided ($p_1$) test (depending on the tested hypothesis) and the number of observations $N$. Standard errors are clustered at the subject level. As specified in the pre-registration, our empirical analysis excludes markets for which subjects failed the attention check in the information phase and, if not stated otherwise, markets in which subjects invested into the safe asset in the investment phase.[18] The reason for mostly focusing on markets in which subjects invested into the risky asset is that in these markets recall directly translates into perceived investment performance, whereas investing into the safe asset prevents recall from affecting perceived investment performance. We conjectured that perceived investment performance functions as a vehicle for motivated memory. In some instances, we additionally need to exclude "neutral" markets, i.e., markets with $\hat{u} = u \wedge \hat{d} = d$ or $\hat{u} \gtrless u \wedge \hat{d} \lessgtr d$, because the former means that memory is complete and we assumed incomplete memory in the theory, and the latter two combinations are inconsistent with an agent's tendency to either forget or confabulate signals. Thus, whenever we need to distinguish markets with respect to environments ($++$, $--$, $+-$, $-+$), we exclude neutral markets to adhere to the definitions of these environments.

## 6.1 Presence of Motivated Memory

Motivated memory is present if both cases of Hypothesis 1 hold. We start with case $(i)$, which postulates that an agent's tendency to forget or confabulate signals is independent of her motivations. Table 1 shows the results of logit regressions, in which the dependent

[18] In total, we had 2910 markets. Subjects failed the attention check in 138 markets and invested into the safe asset in 442 markets. By excluding both, we exclude 546 markets. If not stated otherwise, we calculate our variables of interest per market rather than per subject.

variable is $\Pr(\hat{\tau} > \tau) = 1 - \Pr(\hat{\tau} < \tau)$. Models (1) - (4) concern the presence of motivated memory and relate to case $(i)$ of Hypothesis 1. As can be seen in Table 1, conditional on $u > d$ (Up Market) and $u < d$ (Down Market), $\Pr(\hat{\tau} > \tau)$ and $\Pr(\hat{\tau} < \tau)$ are not different between the long and short position. Neither the asset position (long or short) nor the overall signal (up or down market), and especially not their interaction, influences $\Pr(\hat{\tau} > \tau)$. The only factor that influences an agent's tendency to forget or confabulate signals is the number of signals in a market. A plausible reason for this effect is that the number of signals is varied across markets between 1, 3 and 5, and agents seem to use the average number of signals across markets and apply this number to each market. As a result, the tendency to confabulate decreases in the number of signals as can be seen in Model (1). In markets with only one signal, the tendency to confabulate is largest because the average number of signals is clearly larger than 1, and in markets with five signals, this tendency is smallest because the average number of signals is clearly smaller than 5. This mechanical effect on the tendency to forget or confabulate signals can be seen by comparing Models (2), (3), and (4), but it does not affect our empirical investigation, because the number of signals was randomly drawn (from a uniform distribution) for each market and subject.

Table 1: Recall Biases and Completeness

| Logit, Dependent Variable: $\Pr(\hat{\tau} > \tau)$ | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Long Position | 0.112 | 0.097 | 0.018 | 0.059 | | | | |
| | (0.204) | (0.186) | (0.160) | (0.187) | | | | |
| Up Market | 0.245 | 0.250 | 0.199 | 0.204 | | | | |
| | (0.165) | (0.154) | (0.138) | (0.156) | | | | |
| Long Position × Up Market | 0.342 | 0.290 | 0.195 | 0.297 | | | | |
| | (0.239) | (0.228) | (0.204) | (0.225) | | | | |
| Cue Market | | | | | 0.081 | 0.077 | 0.082 | 0.084 |
| | | | | | (0.098) | (0.088) | (0.080) | (0.094) |
| # Signals ($\tau$) | -0.802*** | | | | -0.784*** | | | |
| | (0.052) | | | | (0.052) | | | |
| $\tau = 1$ | | 2.090*** | | | | 2.038*** | | |
| | | (0.155) | | | | (0.157) | | |
| $\tau = 3$ | | | 0.412*** | | | | 0.421*** | |
| | | | (0.134) | | | | (0.134) | |
| $\tau = 5$ | | | | -2.347*** | | | | -2.322*** |
| | | | | (0.176) | | | | (0.175) |
| Constant | 1.711*** | -1.303*** | -0.859*** | -0.028 | 1.898*** | -1.069*** | -0.738*** | 0.139 |
| | (0.209) | (0.130) | (0.114) | (0.132) | (0.189) | (0.094) | (0.087) | (0.097) |
| $Pseudo-R^2$ | 0.215 | 0.133 | 0.012 | 0.175 | 0.205 | 0.124 | 0.008 | 0.168 |
| $N$ | 1811 | 1811 | 1811 | 1811 | 1811 | 1811 | 1811 | 1811 |

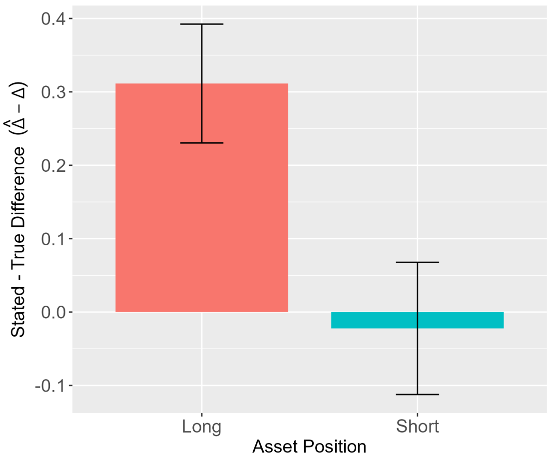
*Notes: Clustered standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. Markets with no cued signal are excluded (to differentiate cue from non-cue markets).*

Regarding case $(ii)$ of Hypothesis 1, Figure 1 displays $\hat{\Delta} - \Delta$ in the long and short position in non-cue markets. As hypothesized by case $(ii)$, we find that $\hat{\Delta} - \Delta$ is significantly larger in the long than short position ($p_1 = 0.0111$, $N = 1191$, MWU). Note that the effect is not symmetric around zero between the long and short position, which may be due to the fact that subjects are less experienced with short than long trading.[19] For

[19] Although in the experiment we reminded subjects repeatedly of their asset position and whether they benefit from price increases or decreases, the short position may still have appeared less intuitive than the

our purposes, however, it is only important that there is a difference in the displayed direction between the long and short position.

Figure 1: $\hat{\Delta} - \Delta$ in Non-Cue Markets per Asset Position




*Notes: Results show stated minus true differences between ups und downs in the long vs. short position. Data are presented as mean values. Error bars indicate standard errors.*

Interestingly, Figure 11 in Appendix A further shows that subjects, who invested into the safe asset, show the reverse pattern of Figure 1, i.e., their recall is downward biased in the long and upward biased in the short position. Here, we find that $\hat{\Delta} - \Delta$ is significantly smaller in the long than short position ($p_2 = 0.0103$, $N = 201$, MWU). Thus, somewhat surprisingly, motivated memory is not only present for subjects invested into the risky asset, but also for those invested into the safe asset, so that another vehicle for motivated recall could be to justify the investment decision. While including safe asset holders into the empirical analysis makes our results even stronger, we continue to exclude them to conform to the pre-registration of our experiment. Note that this result is in line with our motivating example in Section 1, in which non-holders are also not neutral: their recall is motivated in the opposite direction, serving to justify the decision not to hold the asset.

**Result 1 (Presence Motivated Memory)** *Both cases of Hypothesis 1 can be confirmed. Thus, motivated memory is present.*

## 6.2 Presence of Associative Memory

Associative memory is present if all three cases of Hypothesis 2 hold. We start with case $(i)$, which, similar to motivated memory, states that an agent's tendency to forget or confabulate signals should not be influenced by associative memory. Case $(i)$ of Hypothesis 2 is addressed by Models (5) - (8) in Table 1. As can be seen in the table, $\Pr(\hat{\tau} > \tau)$ and

long position, so that subjects had to think more about investment returns and such deliberate thinking may have partly eroded motivated cognition.

$\Pr(\hat{\tau} < \tau)$ are not different between cue and non-cue markets. Thus, a change in similarity does not affect agents' tendency to forget or confabulate signals. Again, the only factor that influences an agent's tendency to forget or confabulate signals is the number of signals in a market, which, most likely, is a mechanical effect caused by the "averaging" rule described above.

Case $(ii)$ of Hypothesis 2 states that increasing similarity should make recall more precise. This case is addressed by Table 2, which displays OLS regressions with the dependent variable being the collection of $|\hat{u} - u|$ in up contexts and $|\hat{d} - d|$ in down contexts. Models (1) - (3) all show that less similarity, i.e., non-cue as opposed to cue markets, increases imprecision. Model (1) shows this effect without additional controls. Model (2) additionally controls for the number of signals and shows that more signals also increase imprecision. Model (3) is more tailored to our dependent variable and shows that the effect of the number of signals is driven by the number of cued signals, i.e., the number of ups in the up context and the number of downs in the down context. Thus, when being in the up context, observing more ups makes recalling ups less precise, and when being in the down context, observing more downs makes recalling downs less precise. Nevertheless, in line with associative memory, similarity remains a significant driver of recall precision in all models.

Table 2: Determinants of Recall Precision

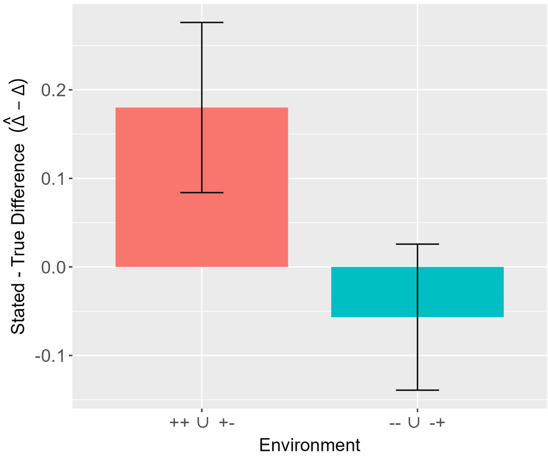
| OLS, Dependent Variable: $\{\|\hat{u} - u\| \| \textit{up context}\} \cup \{\|\hat{d} - d\| \| \textit{down context}\}$ | (1) | (2) | (3) |
|---|---|---|---|
| Non-Cue Market | 0.127** | 0.129** | 0.124** |
| | (0.063) | (0.062) | (0.061) |
| # Signals ($\tau$) | | 0.208*** | 0.042 |
| | | (0.033) | (0.032) |
| # Cued Signals | | | 0.458*** |
| | | | (0.048) |
| Constant | 1.036*** | 0.162 | -0.383*** |
| | (0.050) | (0.134) | (0.147) |
| $R^2$ | 0.004 | 0.045 | 0.164 |
| $N$ | 1022 | 1022 | 1022 |

*Notes: Clustered standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. Markets with no or one cued signal are excluded (to make the change in similarity meaningful).*

Case $(iii)$ of Hypothesis 2 states that in cue markets associative memory distorts the recalled difference between ups and downs upwards in $++$ and $+-$ environments and downwards in $--$ and $-+$ environments, so that $\{\hat{\Delta} - \Delta | ++ \cup +-\} > \{\hat{\Delta} - \Delta | -- \cup -+\}$ should hold. This is addressed by Figure 2. Indeed, we find that $\{\hat{\Delta} - \Delta | ++ \cup +-\}$ is significantly larger than $\{\hat{\Delta} - \Delta | -- \cup -+\}$ ($p_1 = 0.0524$, $N = 532$, MWU).[20]

[20] Here, we exclude neutral markets (to follow definitions of environments) and markets with no cued signal (to differentiate cue from non-cue markets).

Figure 2: $\hat{\Delta} - \Delta$ in Cue Markets per Environment




*Notes: Results show stated minus true differences between ups und downs in the* $++$ *and* $+-$ *vs.* $--$ *and* $-+$ *environments. Data are presented as mean values. Error bars indicate standard errors.*

**Result 2 (Presence Associative Memory)** *All three cases of Hypothesis 2 can be confirmed. Thus, associative memory is present.*

## 6.3 Interaction of Memory Imperfections

Having established that both recall biases are indeed present, we can now analyze their interaction. The two cases of Hypothesis 3 correspond to two independent interaction tests, that investigate whether an increase in similarity, i.e., moving from non-cue (NC) to cue (C) markets, has a larger distortion effect on recall in aligned than misaligned environments. In contrast to Sections 6.1 and 6.2, we now measure distortion not against the true difference $\Delta$, but rather against the benchmark difference $\bar{\Delta} = \frac{\hat{\tau}}{\tau}\Delta$. The reason is that $\hat{\Delta}$ mechanically changes if increasing similarity makes memory more complete and we need to control for this effect.

Table 3: Determinants of Completeness

| OLS, Dependent Variable: $\frac{\hat{\tau}}{\tau}$ | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Long Position | -0.017 | 0.058 | | | | |
| | (0.088) | (0.087) | | | | |
| $++\cup+-$ | | | -0.075 | -0.131 | | |
| | | | (0.133) | (0.088) | | |
| Cue Market | | | | | 0.068*** | 0.070*** |
| | | | | | (0.024) | (0.023) |
| # Signals ($\tau$) | | -0.550*** | | -0.697*** | | -0.180*** |
| | | (0.024) | | (0.037) | | (0.016) |
| Constant | 1.699*** | 3.286*** | 1.857*** | 4.031*** | 0.956*** | 1.709*** |
| | (0.064) | (0.122) | (0.101) | (0.173) | (0.023) | (0.081) |
| $R^2$ | 0.000 | 0.406 | 0.001 | 0.512 | 0.006 | 0.166 |
| $N$ | 1191 | 1191 | 532 | 532 | 1022 | 1022 |

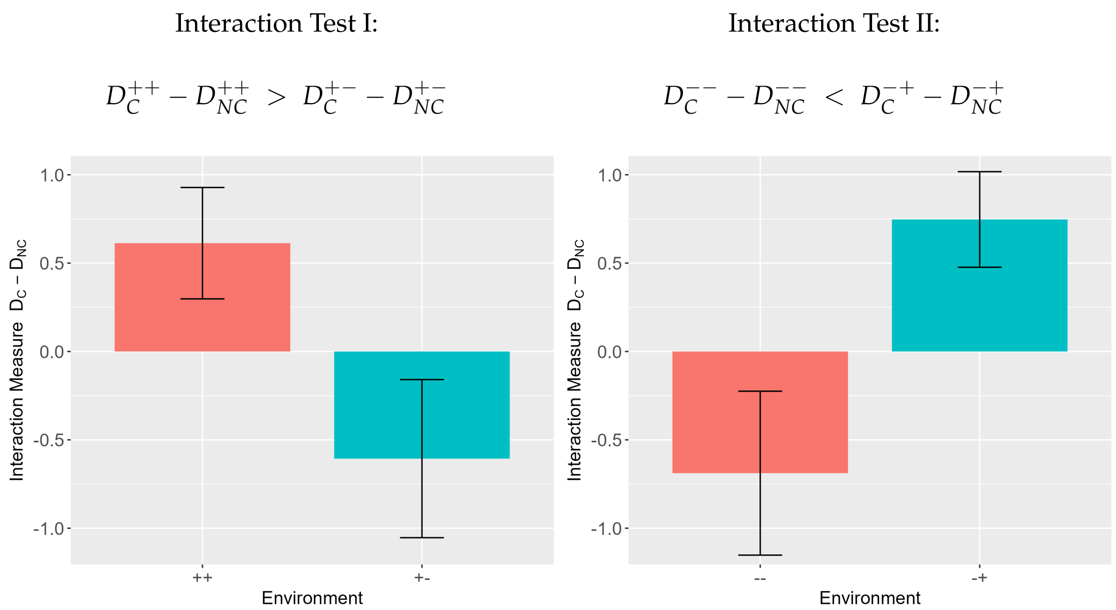
*Notes: Clustered standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. Models (1) and (2) use non-cue markets. Models (3) and (4) use cue markets and exclude neutral markets (to follow definitions of environments) and markets with no cued signal (to differentiate cue from non-cue markets). Models (5) and (6) use cue and non-cue markets and exclude markets with no or one cued signal (to make the change in similarity meaningful).*

Table 3 validates that such control is indeed necessary by showing OLS regressions with the dependent variable being $\frac{\hat{\tau}}{\tau}$. Models (1) and (2) show that in non-cue markets the asset position (long vs. short) does not affect $\frac{\hat{\tau}}{\tau}$, so there is no need to control for completeness in case $(ii)$ of Hypothesis 1. Similarly, Models (3) and (4) show that there is no need to control for completeness in case $(iii)$ of Hypothesis 2 as in cue markets the environment $(++\cup+-$ vs. $--\cup-+)$ does also not affect $\frac{\hat{\tau}}{\tau}$. However, Models (5) and (6) show that an increase in similarity (moving from non-cue to cue markets) affects $\frac{\hat{\tau}}{\tau}$, so we need to control for completeness when investigating changes in similarity, and we do so by measuring distortion against the benchmark difference $\bar{\Delta} = \frac{\hat{\tau}}{\tau}\Delta$ rather than against the true difference $\Delta$ (as we did in case $(ii)$ of Hypothesis 1 and case $(iii)$ of Hypothesis 2). Note that in all three cases, $\frac{\hat{\tau}}{\tau}$ is negatively affected by the number of signals, which is consistent with the "averaging" rule described in Section 6.1.

The results for both independent interaction tests, expressed as cases $(i)$ and $(ii)$ of Hypothesis 3 are displayed in Figure 3. The left graph displays Interaction Test I and the right graph displays Interaction Test II. In both tests, we observe that the two recall biases are complements. In Interaction Test I, $D_C^{++} - D_{NC}^{++}$ is significantly larger than $D_C^{+-} - D_{NC}^{+-}$ ($p_2 = 0.0299$, $N = 125$, MWU), and in Interaction Test II, $D_C^{--} - D_{NC}^{--}$ is significantly smaller than $D_C^{-+} - D_{NC}^{-+}$ ($p_2 = 0.0186$, $N = 120$, MWU).[21] Applying this complementarity result to our holders/non-holders example of Section 1 suggests that more similarity not only increases dependence on the current context, but also the investor's motivation to recall more good than bad news.

[21] Each observation $D_C - D_{NC}$ per environment represents a different subject. If a subject had multiple markets for $D_C$ or $D_{NC}$ in some environment, we used the average value. We excluded subjects, who did not have both $D_C$ and $D_{NC}$ in a certain environment. Per interaction test, a subject is either on the right or left hand side, but can occur in both interaction tests. We exclude neutral markets (to follow definitions of environments) and markets with no cued signal (to differentiate cue from non-cue markets).

Figure 3: Interaction of Recall Biases



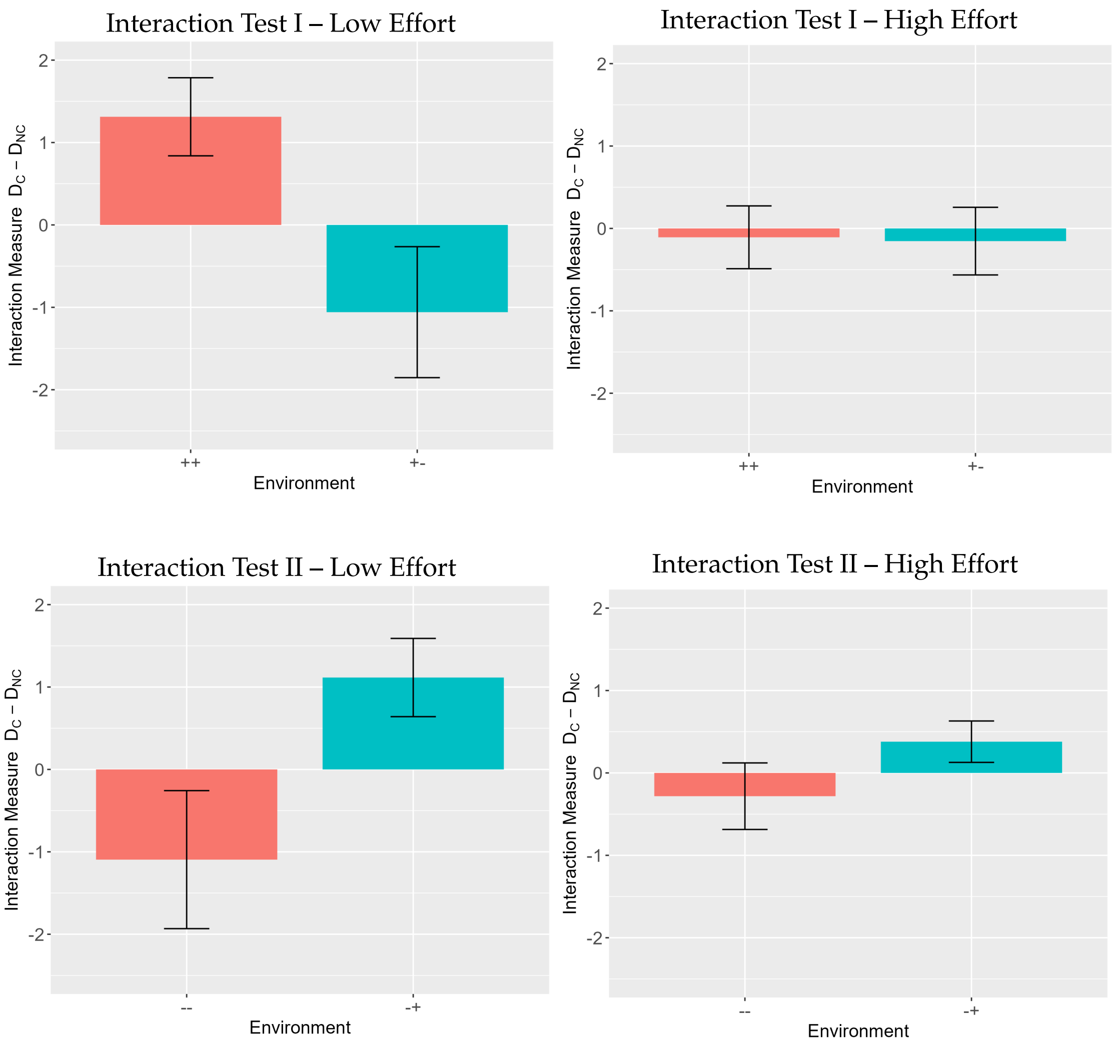

*Notes: Results show the interaction measure conditional on each environment. The left panel compares environments relevant for Interaction Test I and the right panel compares environments relevant for Interaction Test II. Data are presented as mean values. Error bars indicate standard errors.*

**Result 3 (Interaction)** *Both interaction tests of Hypothesis 3 independently reveal that the two recall biases are complements.*

Interestingly, Figure 3 not only shows that associative and motivated memory are complements, it also shows that this interaction makes similarity-induced recall effects to go even in the opposite direction when environments are misaligned. The blue bars in Figure 3 have a different sign than the red bars, so that increases in similarity lead to *opposing* effects in aligned vs. misaligned environments. Relating to our example in Section 1, this observation is consistent with the backfiring effect of vaccination campaigns that Nyhan et al. (2014) find for anti-vaxers. It demonstrates that the interaction of recall biases not only has quantitative effects, but is strong enough to change the direction of an intended effect.

We now investigate how effort affects the way both recall biases interact. Our theory predicts that associative and motivated memory are more complementary for low than high effort. Hypothesis 4 expresses this prediction in empirical terms, namely that in Interaction Test I the difference between $D_C^{++} - D_{NC}^{++}$ and $D_C^{+-} - D_{NC}^{+-}$ is larger for low than high effort, and in Interaction Test II the difference between $D_C^{--} - D_{NC}^{--}$ and $D_C^{-+} - D_{NC}^{-+}$ is smaller for low than high effort. This can be observed in Figure 4. Interaction Test I is particularly pronounced towards complements when subjects exert low effort as predicted by our model. When high effort is exerted, the two recall biases seem to be rather unrelated. We find complements under low effort ($p_2 = 0.0105$, $N = 63$, MWU) but not under high effort ($p_2 = 0.9550$, $N = 62$, MWU). Similarly, Interaction Test II is particularly pronounced towards complements when subjects exert low effort. Here, we again find complements under low effort ($p_2 = 0.0498$, $N = 60$, MWU) but not under high effort ($p_2 = 0.1934$, $N = 60$, MWU).[22]

Figure 4: Interaction Tests per Effort Level




*Notes: Results show the interaction measure conditional on each environment. The left panels show these results for low effort and the right panels show results for high effort. The top panels compare environments relevant for Interaction Test I and the bottom panels compare environments relevant for Interaction Test II. Data are presented as mean values. Error bars indicate standard errors*

We further investigate the effect of effort with an OLS regression, in which the dependent variable is $D_C - D_{NC}$. Table 4 in Appendix A shows similar results as above. Models (1) and (2) relate to Interaction Tests I and II, respectively. Model (1) reveals that $D_C - D_{NC}$ is significantly larger in the $++$ than $+-$ environment only when effort is low. Likewise, Model (2) reveals that $D_C - D_{NC}$ is significantly smaller in the $--$ than $-+$ environment only when effort is low.

[22] As before, each observation per environment represents a different subject. For each $D_C - D_{NC}$ per subject, we calculate her average effort value $|\hat{\tau} - \tau|$ (by first averaging separately over $D_C$ and $D_{NC}$, and then averaging the two values for $D_C$ and $D_{NC}$). We split the subjects in half to separate low (larger values of $|\hat{\tau} - \tau|$) from high effort (smaller values of $|\hat{\tau} - \tau|$) subjects. We exclude neutral markets (to follow definitions of environments) and markets with no cued signal (to differentiate cue from non-cue markets).

**Result 4 (Interaction and Effort)** *Both cases of Hypothesis 4 can be confirmed and independently reveal that the two recall biases are more complementary for low than high effort.*

While Result 3 shows that subjects are able to exploit similarity to self-servingly bias their recall, Result 4 identifies an interesting comparative static: subjects are more able to exploit similarity to self-servingly bias their recall when they spend less cognitive resources on it.

## 6.4 Mapping Memory to Beliefs

A fundamental question is whether distorted memory translates into distorted beliefs and actions. In this subsection, we investigate the link between memory and beliefs and the next subsection investigates the link between beliefs and action.

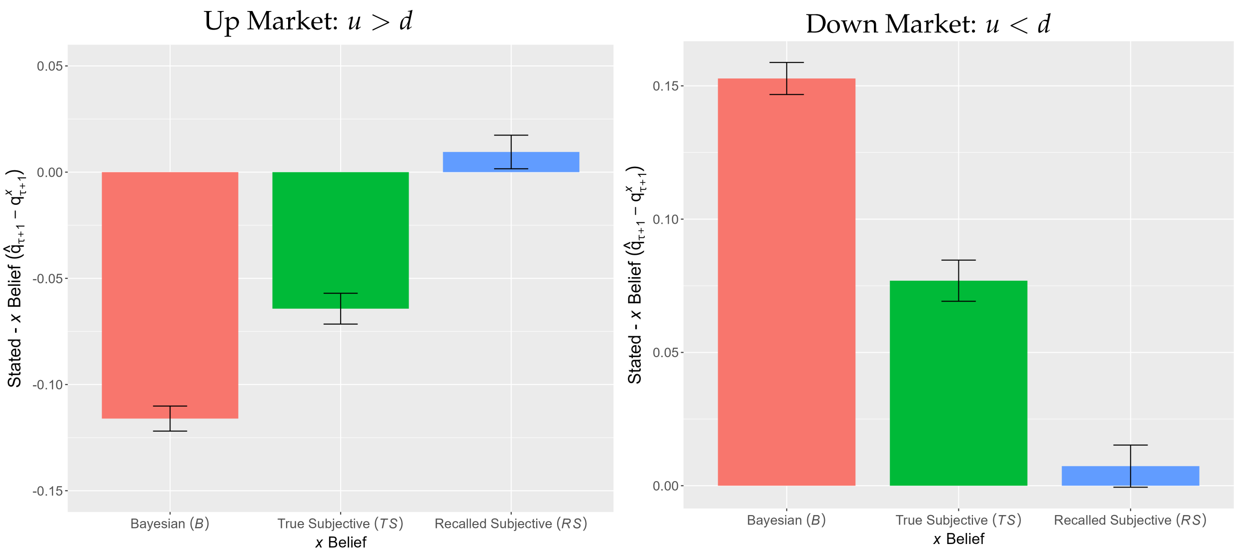
Hypothesis 5 seeks to explain distorted beliefs, i.e., the difference between stated and Bayesian beliefs, $|\hat{q}_{\tau+1} - q^{B}_{\tau+1}|$, by non-Bayesian updating and/or incorrect recall. First, if non-Bayesian updating explains distorted beliefs, we should observe that $|\hat{q}_{\tau+1} - q^{B}_{\tau+1}|$ is larger than the difference between stated and true subjective beliefs (which account for non-Bayesian updating but not for incorrect recall), $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$. Second, if incorrect recall further explains distorted beliefs, we should observe that $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ is larger than the difference between stated and recalled subjective beliefs (which account for both non-Bayesian updating and incorrect recall), $|\hat{q}_{\tau+1} - q^{RS}_{\tau+1}|$. Third, if non-Bayesian updating and incorrect recall fully explain stated beliefs, we should observe that $|\hat{q}_{\tau+1} - q^{RS}_{\tau+1}|$ is zero. All three observations are visible in Figure 5. $|\hat{q}_{\tau+1} - q^{B}_{\tau+1}|$ is represented by the red bars, $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ by the green bars, and $|\hat{q}_{\tau+1} - q^{RS}_{\tau+1}|$ by the blue bars. The left graph represents an overall up signal, i.e., $u > d$, and the right graph represents an overall down signal, i.e., $u < d$.[23] Looking at the red bars, we observe a common result in the belief updating literature, namely conservatism. When accounting for individual-specific non-Bayesian updating, we can explain a significant amount of these belief distortions since $|\hat{q}_{\tau+1} - q^{B}_{\tau+1}|$ is significantly larger than $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ both when $u > d$ ($p_1 = 0.0000$, $N = 1403$, MWU) and when $u < d$ ($p_1 = 0.0000$, $N = 1369$, MWU). However, incorrect recall also explains a significant amount of belief distortions since $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ is significantly larger than $|\hat{q}_{\tau+1} - q^{RS}_{\tau+1}|$ both when $u > d$ ($p_1 = 0.0000$, $N = 1403$, MWU) and when $u < d$ ($p_1 = 0.0000$, $N = 1369$, MWU). In fact, we can explain stated beliefs almost entirely by non-Bayesian updating *and* incorrect recall.[24] We further perform a within-subject analysis and observe similar results: $|\hat{q}_{\tau+1} - q^{B}_{\tau+1}|$ is

[23] Here, we only exclude subjects who failed the attention check, but include subjects who invested into the safe asset in the investment phase, because their memory should similarly map into their beliefs as for subjects who invested into the risky asset.

[24] $|\hat{q}_{\tau+1} - q^{B}_{\tau+1}|$ is significantly larger than zero both when $u > d$ ($p_1 = 0.0000$, $N = 1403$, MWU) and when $u < d$ ($p_1 = 0.0000$, $N = 1369$, MWU). Also, $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ is significantly larger than zero both when $u > d$ ($p_1 = 0.0000$, $N = 1403$, MWU) and when $u < d$ ($p_1 = 0.0000$, $N = 1369$, MWU). However, $|\hat{q}_{\tau+1} - q^{RS}_{\tau+1}|$ is not significantly different from zero when $u > d$ ($p_2 = 0.5319$, $N = 1403$, MWU) and only marginally so when $u < d$ ($p_2 = 0.0372$, $N = 1369$, MWU).

significantly larger than $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ both when $u > d$ ($p_1 = 0.0000$, $N = 1403$, WSR) and when $u < d$ ($p_1 = 0.0000$, $N = 1369$, WSR). Likewise, $|\hat{q}_{\tau+1} - q^{TS}_{\tau+1}|$ is significantly larger than $|\hat{q}_{\tau+1} - q^{RS}_{\tau+1}|$ both when $u > d$ ($p_1 = 0.0000$, $N = 1403$, WSR) and when $u < d$ ($p_1 = 0.0000$, $N = 1369$, WSR).

Figure 5: Explaining Stated Beliefs




*Notes: Results show the stated belief minus the Bayesian (red), true subjective (green), and recalled subjective (blue) belief. The left panel shows markets with more ups than downs (i.e., overall up signal) and the right panel shows markets with more downs than ups (i.e., overall down signal). Data are presented as mean values. Error bars indicate standard errors.*

**Result 5 (Beliefs)** *We observe that*

$$|\hat{q}_{\tau+1} - q^{B}_{\tau+1}| > |\hat{q}_{\tau+1} - q^{TS}_{\tau+1}| > |\hat{q}_{\tau+1} - q^{RS}_{\tau+1}| \approx 0.$$

*Thus, incorrect recall explains subjects' beliefs well beyond non-Bayesian updating.*

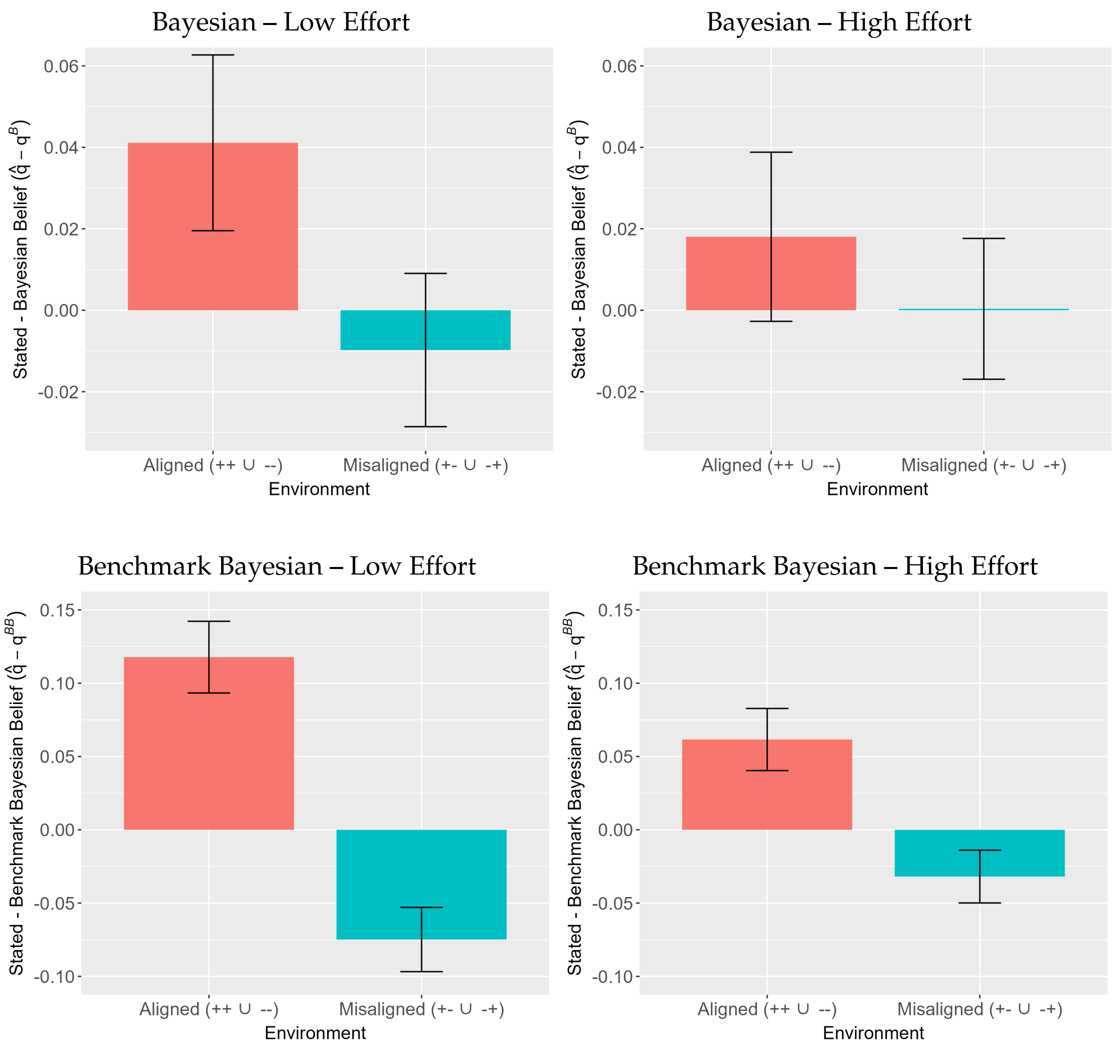
Interestingly, when we compare the quantitative effects of moving from the red to the green and from the green to the blue bars in Figure 5, we observe that non-Bayesian updating and incorrect recall are *equally* important in explaining distorted beliefs.

We now investigate whether the interaction of both recall biases translates into distorted beliefs. Hypothesis 6 specifies that the difference in the favorability of beliefs between aligned and misaligned environments is larger for low than high effort in cue markets. Figure 6 shows the favorability of beliefs for aligned ($+ + \cup - -$) and misaligned ($- + \cup + -$) environments. The left graphs display low effort and the right graphs high effort. The top graphs represent the favorability measure $\hat{q} - q^B$, adressing case $(i)$ of Hypothesis 6, and the bottom graphs represent the favorability measure $\hat{q} - q^{BB}$, adressing case $(ii)$ of Hypothesis 6.[25] As predicted, both the top and bottom graphs show that the

[25]To control for the number of signals ($\tau$) in our effort variable, we separated low from high effort conditional on $\tau$. Hence, the left graphs in Figure 6 have the same distribution of $\tau$ as the right graphs. We exclude neutral markets (to follow definitions of environments) and markets with no cued signal (to differentiate cue from non-cue markets).

differences between red and blue bars are larger under low effort (left graphs) than under high effort (right graphs). Concerning the top graphs, we find that $\{\hat{q} - q^B | ++ \cup --\}$ is larger than $\{\hat{q} - q^B | -+ \cup +-\}$ under low ($p_1 = 0.0461$, $N = 266$, MWU), but not under high effort ($p_1 = 0.3954$, $N = 266$, MWU), which implies that the difference in the favorability of beliefs between aligned and misaligned environments is larger for low than high effort. Concerning the bottom graphs, we find that $\{\hat{q} - q^B | ++ \cup --\}$ is larger under low than high effort ($p_1 = 0.0308$, $N = 255$, MWU) and $\{\hat{q} - q^B | -+ \cup +-\}$ is smaller under low than high effort ($p_1 = 0.0314$, $N = 277$, MWU), which also implies that the difference in the favorability of beliefs between aligned and misaligned environments is larger for low than high effort.

Figure 6: Favorabilty of Beliefs per Effort Level in Cue Markets



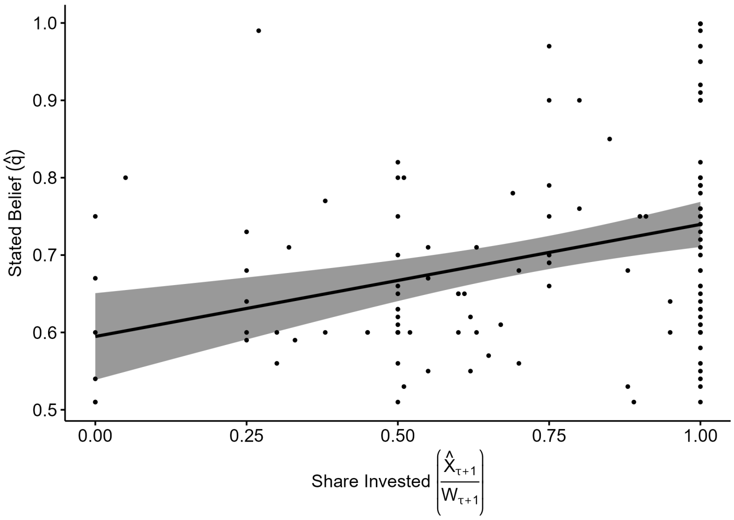

*Notes: Results show the stated belief minus the Bayesian (top panels) or benchmark Bayesian (bottom panels) belief in aligned vs. misaligned environments. The left panels show these results for low effort and the right panels show results for high effort. Data are presented as mean values. Error bars indicate standard errors.*

**Result 6 (Beliefs and Effort)** *Both cases of Hypothesis 6 can be confirmed. Thus, the difference in the favorability of beliefs between aligned and misaligned environments is larger for low than high effort.*

## 6.5 Mapping Beliefs to Actions

Our final investigation concerns the link between beliefs and actions.[26] Figure 7 displays subjects' belief to own the good asset (on the $y$-axis) and the share invested into the risky asset (on the $x$-axis).[27] As stated by Hypothesis 7, we find a positive correlation between the two ($\rho = 0.34$, $p_1 = 0.0000$, $N = 131$, Spearman).

Figure 7: Beliefs and Investments




*Notes: Results show a positive correlation between the stated belief (y-axis) and the share invested into the risky asset (x-axis). Data are presented in a scatter plot.*

**Result 7 (Actions)** *There is a positive correlation between the belief to own the good asset and the share invested into the risky asset.*

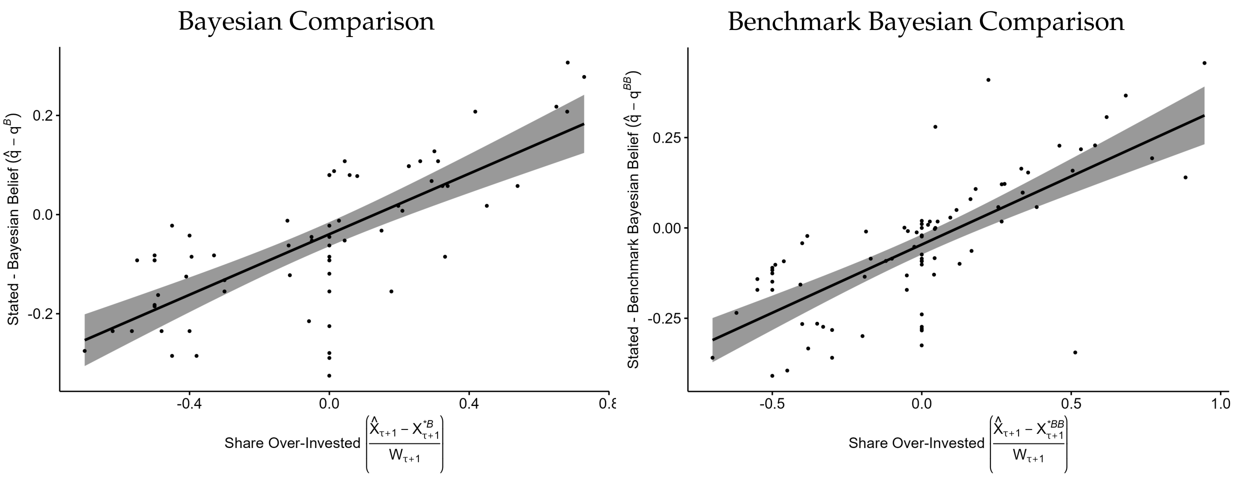
While Result 7 shows that being more confident to own the good asset translates into more investments, we do not know yet whether being more *over*-confident also translates into more *over*-investments. In other words, it remains to be show that *distorted* beliefs translate into *distorted* actions. Our final Hypothesis 8 states that there is a positive correlation between the favorablity of beliefs and over-investments, case $(i)$ compares stated beliefs and actual actions to Bayesian and case $(ii)$ to benchmark Bayesian beliefs and actions. Our benchmark investments account for individual-specific risk aversion, revealed from subjects' actual investments and stated beliefs when assuming expected utility and constant relative risk aversion. Figure 12 in Appendix A shows the parameters of relative risk aversion that subjects reveal in our experiment. We find reasonable and typical empirical values of relative risk aversion with our method and quite some variation across subjects that warrants the need to control for individual-specific risk aversion. Figure 8

[26] Here, we only exclude subjects who failed the attention check, but include subjects who invested into the safe asset in the investment phase, because their beliefs should similarly map into reinvestment actions as for subjects who invested into the risky asset.

[27] Since only one reinvestment market was randomly chosen per subject, each observation represents a different subject. This applies to all results in this section.

displays subjects' over-confidence to own the good asset (on the $y$-axis) and their over-investments (on the $x$-axis). As stated by Hypothesis 8, we find a positive correlation between the two, both when comparing stated beliefs and actions against Bayesian beliefs and actions ($\rho = 0.74$, $p_1 = 0.0000$, $N = 67$, Spearman) and when comparing against benchmark Bayesian beliefs and actions ($\rho = 0.77$, $p_1 = 0.0000$, $N = 84$, Spearman).

Figure 8: Distorted Beliefs and Investment Mistakes




*Notes: Results show a positive correlation between belief distortion (y-axis), i.e., stated belief minus Bayesian (left panel) and benchmark Bayesian (right panel) belief, and investment mistakes (x-axis), i.e., the share over-invested into the risky asset. Data are presented in a scatter plot.*

We further investigate the link between over-confidence and over-investments with a logit regression, in which the dependent variable is the probability to invest more than prescribed by the benchmark, either using the Bayesian comparison, i.e., $\Pr(\hat{X}_{\tau+1} > X^{*B}_{\tau+1})$ or the benchmark Bayesian comparison, i.e., $\Pr(\hat{X}_{\tau+1} > X^{*BB}_{\tau+1})$. Table 5 in Appendix A shows for both comparisons that being over-confident makes over-investing substantially more likely and that over-confidence explains roughly half of the variation in over-investments.

**Result 8 (Investment Mistakes)** *Both cases of Hypothesis 8 can be confirmed. Thus, there is a positive correlation between the favorability of beliefs and over-investments.*

An empirical implication of Results 6 and 8 is that disagreement of asset holders and non-holders should affect trade volume on financial markets and we would expect more disagreement and less trade volume in low effort markets (e.g., markets with more sentimental traders) than in high effort markets (e.g., markets with more fundamental traders).

## 6.6 Robustness

The main insight of this paper is that the interaction of recall biases matters for expectation formation as well as economic decision making. We find that motivated and associative memory are complements and, in particular, for low effort levels. To assure that the interaction effect, which we measure through two independent interaction tests, is not just an artefact in our experiment, we conducted the robustness treatment IMM (described in Section 4). IMM was designed to manipulate the importance of both associative and motivated memory. We expected that the elimination of the three-days time lag between periods $\tau$ and $\tau+1$, on the one hand, made it more difficult for subjects to build up motivations, and, on the other hand, limited the importance of associative memory by relying less on contextual features when recalling, so that the same increase in similarity should have a smaller effect. Both limitations should result in a weakened or even eliminated interaction effect.[28]

Figure 13 in Appendix A shows that motivated memory is absent in treatment IMM. While there is a significant positive difference of $\hat{\Delta}-\Delta$ between the long and short position in non-cue markets in treatment MAIN (see also Figure 1), this difference completely vanishes in treatment IMM ($p_2 = 0.6487$, $N = 384$, MWU). Concerning associative memory, we firstly analyze the effect of increasing similarity on recall precision and, secondly, $\{\hat{\Delta}-\Delta|++\cup+-\}$ vs.$\{\hat{\Delta}-\Delta|--\cup-+\}$ in cue markets. Table 7 in Appendix A replicates Table 2 for the IMM instead of the MAIN treatment. While Table 2 showed that less similarity, i.e., non-cue as opposed to cue markets, significantly increases imprecision, Table 7 in Appendix A shows that less similarity still has a positive but now insignificant effect on imprecision. Figure 14 in Appendix A shows that in cue markets $\{\hat{\Delta}-\Delta|++\cup+-\} > \{\hat{\Delta}-\Delta|--\cup-+\}$ in both treatments. However, while this inequality is large and significant in treatment MAIN (see also Figure 2), it is rather small and insignificant in treatment IMM ($p_2 = 0.7053$, $N = 163$, MWU).

Since both motivated and associative memory are basically absent in treatment IMM, we also expect that any interaction between the two is absent. Figure 15 in Appendix A shows that this is indeed the case. The top graphs represent Interaction Test I and the bottom graphs represent Interaction Test II. The left graphs display treatment MAIN (see also Figure 3) and the right graphs display treatment IMM. Comparing the left to right graphs shows that in both independent interaction tests, there is a large and significant interaction effect in MAIN, but only a small and insignificant interaction effect in IMM (Interaction Test I: $p_2 = 0.9367$, $N = 41$, MWU; Interaction Test II: $p_2 = 0.4040$, $N = 39$, MWU). Thus, as expected, eliminating the influence of motivated and associative memory through a treatment variation makes the interaction effect disappear as well.

[28] Although subjects had to recall twice, immediately and three days later, we cannot perform a within-subject analysis within treatment IMM, because subjects tend to recall their answers at the first date rather than the observed signals. Table 6 in Appendix A shows that subjects' stated belief at the first date is the best predictor for their stated belief at the second date. We therefore compare treatments IMM and MAIN across subjects in this section.

# 7 Conclusion

When behavioral biases interact, simply “adding up” their isolated effects can mislead inference about real-world behavior under their coexistence. In this paper, we investigate how similarity-based and motivated recall biases interact, which is an important endeavor given their steady coexistence in reality: for people trying to recall the past in order to form beliefs and act upon them, events from the past are usually “good” or “bad” for their desired view and have occurred in specific contexts that may be similar or dissimilar to the current context.

We propose a model of recall, in which the interaction of similarity-based and motivated recall biases switches between complementarity and substitutability as a function of recall effort. In the data, the biases are complements, and especially for low effort levels, which is consistent with the model. The interaction matters for individual beliefs and actions. With both biases being complements, similarity amplifies motivated distortion, producing asymmetric patterns in belief updating, which in turn affect investment behavior. We also show that imperfect memory is equally important in explaining distorted beliefs as non-Bayesian updating.

Moreover, the interaction matters in the aggregate. As the environment shifts, disagreement can rise predictably, and this can translate into systematic patterns in trading, volatility, and the sensitivity of prices to news. This provides a micro-foundation for when beliefs polarize in response to the same information. From a policy perspective, complementarity implies a real risk of backfiring: changes in context or salience can strengthen motivated recall even when no new information is provided, so the same message may entrench opposing views across audiences with different motivational stakes. In practice, this means that attempts to persuade highly motivated groups can make beliefs diverge rather than converge, and neutral disclosures can polarize assessments of the very same evidence. Effective information design must therefore anticipate heterogeneity in motivational stakes and recognize that persuasion is not symmetric: what convinces one audience may harden convictions in another.

Finally, our finding that complementarity depends on low cognitive effort suggests systematic heterogeneity across markets and market participants. Settings with higher cognitive load, tighter time pressure, or weaker incentives to exert effort should display stronger complementarity. Environments with more monitoring, higher stakes, or greater scope for deliberation should display weaker complementary interaction. Tracing how this variation maps into prices, volumes, and disagreement is an important task for future work. More broadly, applying the same interaction lens to other behavioral biases may establish where and when biases complement each other and by how much. The present paper is a first, domain-specific step in that direction.

# Appendix A: Additional Tables and Figures

Figure 9: Example Down Signal

**Stock C**

Company C produces Smartphones. A new mobile communications standard is being introduced, which is not yet supported by most of the company's smartphones. Customers are dissatisfied and therefore switch to competitor products which causes the company's revenue to fall.

The stock **decreases** by 10 EG.

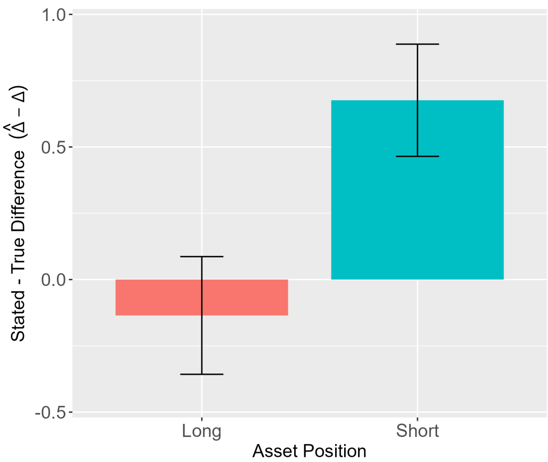
Figure 10: Example Up Signal

**Stock C**

Company C produces Smartphones. The company invented a new production technique through which new smartphones are predominantly produced with recycled materials. This not only reduces production costs, but also increases demand of environmentally conscious customers.

The stock **increases** by 10 EG.

Figure 11: $\hat{\Delta} - \Delta$ in Non-Cue Markets when Holding Safe Asset



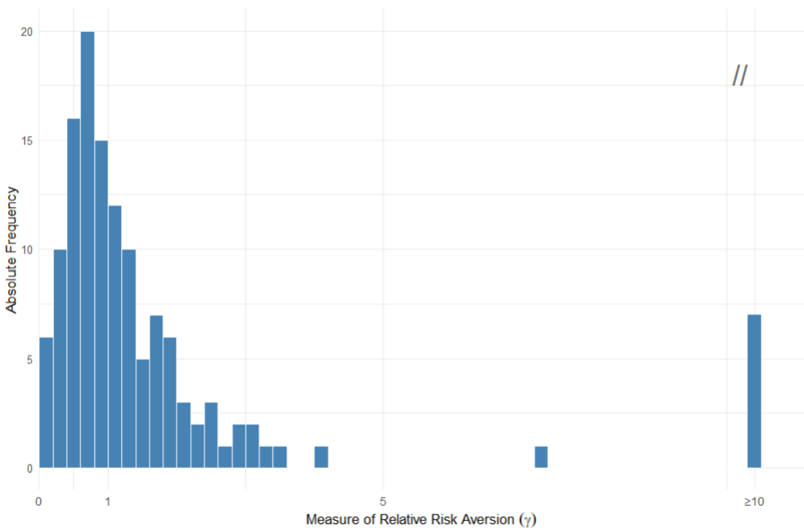

*Notes: Results show stated minus true differences between ups und downs in the long vs. short position for safe asset holders. Data are presented as mean values. Error bars indicate standard errors.*

Table 4: Interaction and Effort

| OLS, Dependent Variable: $D_C - D_{NC}$ | (1) | (2) |
|---|---|---|
| Low Effort | -0.905 | 0.736 |
| | (0.775) | (0.690) |
| Aligned | 0.046 | -0.661 |
| | (0.752) | (0.727) |
| Low Effort × Aligned | 2.325** | -1.548* |
| | (1.059) | (1.029) |
| Constant | -0.154 | 0.379 |
| | (0.548) | (0.488) |
| $R^2$ | 0.080 | 0.080 |
| $N$ | 125 | 120 |

*Notes: Standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. Neutral markets are excluded (to follow definitions of environments) and markets with no cued signal (to differentiate cue from non-cue markets).*

Figure 12: Distribution of Relative Risk Aversion




*Notes: Results show the distribution of parameters of relative risk aversion elicited in the experiment. Data are presented in a histogram with frequency on the y-axis and the measure of relative risk aversion on the x-axis.*

## Table 5: Over-Confidence and Over-Investments

| Logit, Dependent Variable: $\Pr(\hat{X}_{\tau+1} > X^{*B}_{\tau+1})$ in (1)-(2) and $\Pr(\hat{X}_{\tau+1} > X^{*BB}_{\tau+1})$ in (3)-(4) | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Over-Confident ($\hat{q} > q^B$) | 4.406*** | 4.402*** | | |
| | (0.879) | (0.882) | | |
| Over-Confident ($\hat{q} > q^{BB}$) | | | 4.102*** | 4.940*** |
| | | | (0.731) | (1.012) |
| # Signals ($\tau$) | | 0.013 | | -0.448 |
| | | (0.263) | | (0.283) |
| Constant | -2.054*** | -2.095** | -1.904*** | -0.853 |
| | (0.475) | (0.958) | (0.405) | (0.705) |
| Pseudo-$R^2$ | 0.500 | 0.500 | 0.461 | 0.487 |
| $N$ | 67 | 67 | 84 | 84 |

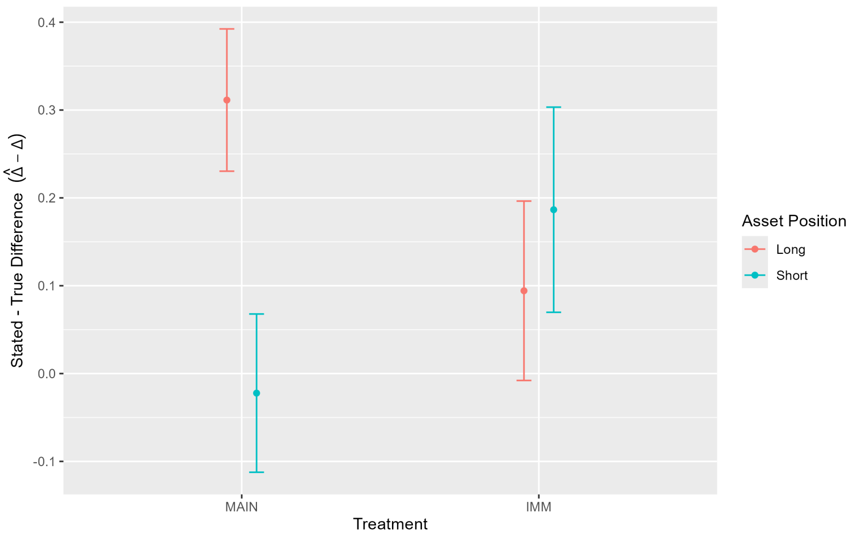
*Notes: Standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. All models exclude subjects with stated beliefs $\hat{q} \leq 1/2$, Models (1)-(2) additionally exclude subjects with Bayesian beliefs $q^B \leq 1/2$, and Models (3)-(4) additionally exclude subjects with benchmark Bayesian beliefs $q^{BB} \leq 1/2$.*

## Table 6: Explaining Stated Beliefs at the Second Date in Treatment IMM

| OLS, Dependent Variable: Stated Belief $2^{nd}$ Date | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Stated Belief $1^{st}$ Date | 0.475*** | | | |
| | (0.0445) | | | |
| True Subjective Belief | | 0.287*** | | |
| | | (0.0326) | | |
| Bayesian Belief | | | 0.418*** | |
| | | | (0.0389) | |
| Recalled Subjective Belief | | | | 0.364*** |
| | | | | (0.0428) |
| Constant | 0.261*** | 0.348*** | 0.281*** | 0.307*** |
| | (0.0248) | (0.0195) | (0.0230) | (0.0246) |
| $R^2$ | 0.222 | 0.122 | 0.165 | 0.172 |
| $N$ | 848 | 848 | 848 | 848 |

*Notes: Clustered standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. We only exclude subjects who failed the attention check, but include subjects who invested into the safe asset in the investment phase.*

## Figure 13: $\hat{\Delta} - \Delta$ in Non-Cue Markets per Treatment and Asset Position



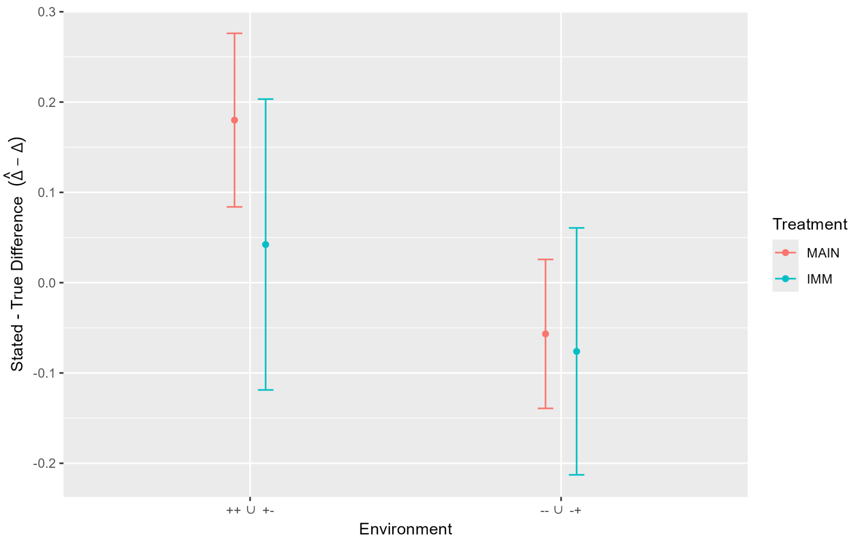

*Notes: Results show stated minus true differences between ups und downs in the MAIN vs. IMM treatment, conditional on the asset position (long and short). Data are presented as mean values. Error bars indicate standard errors.*

Table 7: Determinants of Recall Precision in IMM

| OLS, Dependent Variable: $\{|\hat{u} - u||up\ context\} \cup \{|\hat{d} - d||down\ context\}$ | (1) | (2) | (3) |
|---|---|---|---|
| Non-Cue Market | 0.099 | 0.092 | 0.123 |
| | (0.099) | (0.101) | (0.096) |
| # Signals ($\tau$) | | 0.087* | 0.008 |
| | | (0.052) | (0.055) |
| # Cued Signals | | | 0.195** |
| | | | (0.077) |
| Constant | 0.841*** | 0.480** | 0.262 |
| | (0.094) | (0.213) | (0.229) |
| $R^2$ | 0.003 | 0.012 | 0.042 |
| $N$ | 337 | 337 | 337 |

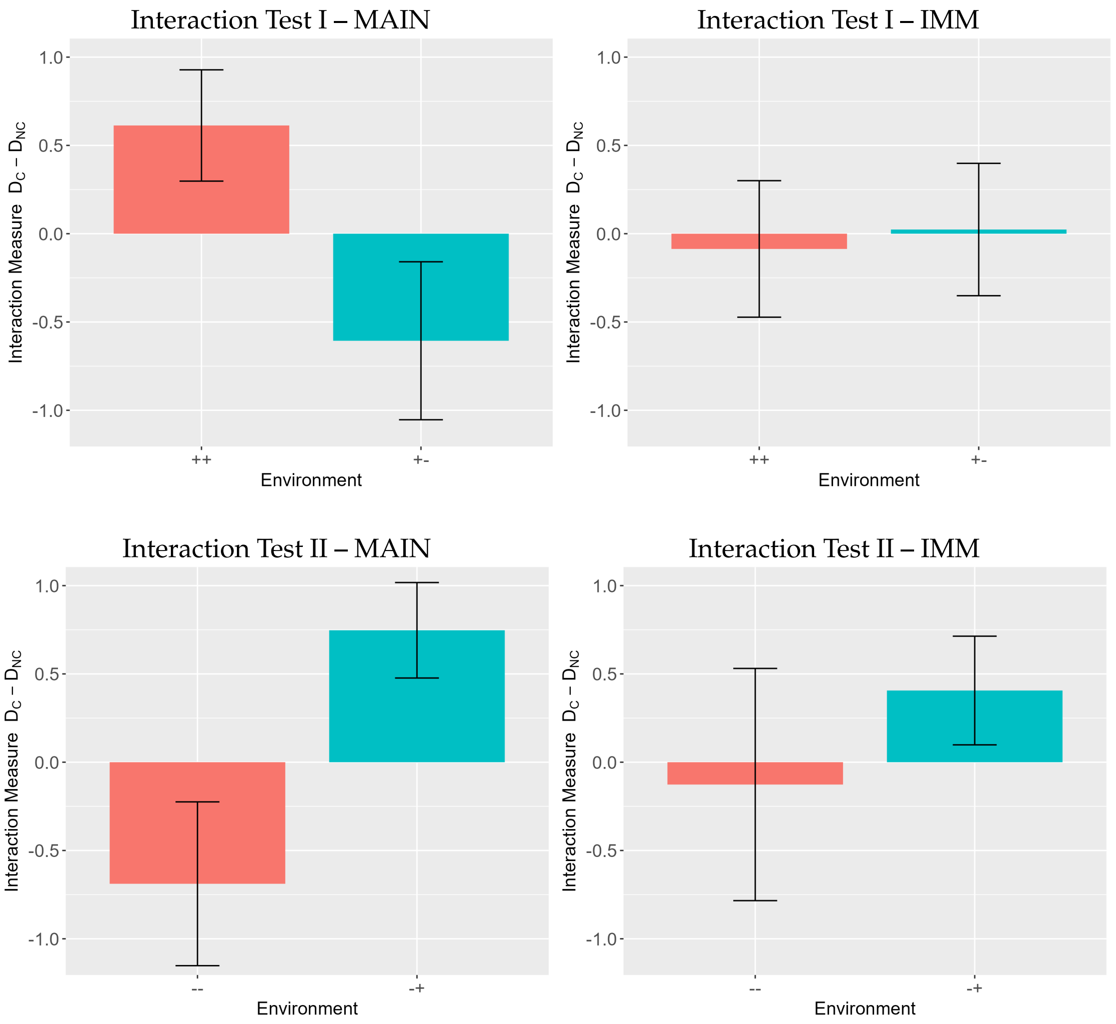
*Notes: Clustered standard errors are in parentheses. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively. Markets with no or one cued signal are excluded (to make the change in similarity meaningful).*

Figure 14: $\hat{\Delta} - \Delta$ in Cue Markets per Treatment and Environment




*Notes: Results show stated minus true differences between ups und downs in the ++ and +− vs. −− and −+ environments, conditional on the treatment (MAIN and IMM). Data are presented as mean values. Error bars indicate standard errors*

Figure 15: Interaction Tests per Treatment




*Notes: Results show the interaction measure conditional on each environment. The left panels show these results for the MAIN treatment and the right panels show results for the IMM treatment. The top panels compare environments relevant for Interaction Test I and the bottom panels compare environments relevant for Interaction Test II. Data are presented as mean values. Error bars indicate standard errors*

## Appendix B: Proofs

**Proof of Strict Concavity of Target Function.** The second-order partial derivatives of the target function $F := \Pi\left(|u - \hat{u}(e_u)|, |d - \hat{d}(e_d)|\right) - C(e_u, e_d)$ are as follows:

$$\begin{aligned}
\frac{\partial^2 F}{\partial e_u \partial e_u} &= -\mu_u \frac{d\hat{u}(e_u^*)}{de_u^*}(1 - \tfrac{\bar{\tau}}{\tau})u - \frac{1-\mu_u}{\alpha_u}, \\
\frac{\partial^2 F}{\partial e_d \partial e_d} &= -\mu_d \frac{d\hat{d}(e_d^*)}{de_d^*}(1 - \tfrac{\bar{\tau}}{\tau})d - \frac{1-\mu_d}{\alpha_d}, \\
\frac{\partial^2 F}{\partial e_u \partial e_d} &= \frac{\partial^2 F}{\partial e_d \partial e_u} = 0.
\end{aligned}$$

It holds that $\frac{\partial^2 F}{\partial e_u \partial e_u} < 0$ since $\frac{1-\mu_u}{\alpha_u} > 0$, $\mu_u > 0$, $u \geq 0$, and, by (2), $\frac{d\hat{u}(e_u)}{de_u} > (<)\, 0 \iff \bar{\tau} < (>)\, \tau$. Likewise, it holds that $\frac{\partial^2 F}{\partial e_d \partial e_d} < 0$ since $\frac{1-\mu_d}{\alpha_d} > 0$, $\mu_d > 0$, $d \geq 0$, and, by (2), $\frac{d\hat{d}(e_d)}{de_d} > (<)\, 0 \iff \bar{\tau} < (>)\, \tau$. Therefore, the Hessian is negative definite. ■

**Proof of Proposition 1.** Plugging $\bar{\tau} = \tau$ into (6) yields $(e_u^*, e_d^*) = (0, 0)$. Then, plugging $\bar{\tau} = \tau$ and $(e_u^*, e_d^*) = (0, 0)$ into (1) yields $\hat{u}(e_u^*) = u$ and $\hat{d}(e_d^*) = d$. ■

**Proof of Proposition 2.** Let $\bar{\tau} \neq \tau$. Since $\mu = \alpha = 0$ implies $\mu_u = \mu_d$ and $\alpha_u = \alpha_d$, we can write (6) as

$$(e_u^*, e_d^*) = \left( \frac{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2}{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{u^2}}, \frac{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2}{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{d^2}} \right),$$

or $(e_u^*, e_d^*) = (\frac{x}{x + y\frac{1}{u^2}}, \frac{x}{x + y\frac{1}{d^2}})$ with $x := \mu_u(1 - \frac{\bar{\tau}}{\tau})^2 > 0$ and $y := \frac{1-\mu_u}{\alpha_u} > 0$. Then, $e_u^* \gtreqless e_d^* \iff x(x + y\frac{1}{d^2}) \gtreqless x(x + y\frac{1}{u^2}) \iff \frac{1}{d^2} \gtreqless \frac{1}{u^2} \iff u \gtreqless d$. ■

**Proof of Proposition 3.** Fix any $u$, $d$, $(\mu_u, \mu_d)$, and $(\alpha_u, \alpha_d)$. Then, $(1 - \frac{\bar{\tau}}{\tau})^2$ is increasing in $|\bar{\tau} - \tau|$, and both $e_u^* = \frac{\mu_u(1-\frac{\bar{\tau}}{\tau})^2}{\mu_u(1-\frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{u^2}}$ and $e_d^* = \frac{\mu_u(1-\frac{\bar{\tau}}{\tau})^2}{\mu_u(1-\frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{d^2}}$ are increasing in $(1 - \frac{\bar{\tau}}{\tau})^2$. ■

**Proof of Proposition 4.** Fix any $u$, $d$, and $\bar{\tau} \neq \tau$ and let $\mu = \alpha = 0$. Since $\mu = \alpha = 0$ implies $\mu_u = \mu_d$ and $\alpha_u = \alpha_d$, we can write (6) as

$$(e_u^*, e_d^*) = \left( \frac{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2}{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{u^2}}, \frac{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2}{\mu_u(1 - \frac{\bar{\tau}}{\tau})^2 + \frac{1-\mu_u}{\alpha_u}\frac{1}{d^2}} \right).$$

Now, $e_u^*$ and $e_d^*$ are increasing in both $\mu_u$ and $\alpha_u$. ■

**Proof of Proposition 5.** Let $\bar{\tau} \neq \tau$ and $\hat{\Delta} := \hat{u}(e_u^*) - \hat{d}(e_d^*)$. We start with the long position. Since $\mu = \mu_u - \mu_d \gtrless 0 \iff \bar{\tau} \lesseqgtr \tau$, a larger $|\mu|$ implies that $\mu_u$ ($\mu_d$) increases (decreases) when $\bar{\tau} < \tau$ and that $\mu_u$ ($\mu_d$) decreases (increases) when $\bar{\tau} > \tau$. From (6), we can derive that

$$\begin{aligned}\frac{\partial e_u^*}{\partial \mu_u} &= \frac{\alpha_u(1-\frac{\bar{\tau}}{\tau})^2 u^2}{\left(\mu_u(\alpha_u(1-\frac{\bar{\tau}}{\tau})^2 u^2 - 1) + 1\right)^2} > 0,\\ \frac{\partial e_d^*}{\partial \mu_d} &= \frac{\alpha_d(1-\frac{\bar{\tau}}{\tau})^2 d^2}{\left(\mu_d(\alpha_d(1-\frac{\bar{\tau}}{\tau})^2 d^2 - 1) + 1\right)^2} > 0.\end{aligned}$$

Suppose first that $\bar{\tau} < \tau$, so that a larger $|\mu|$ implies that $\mu_u$ ($\mu_d$) increases (decreases). Thus, $e_u^*$ increases and $e_d^*$ decreases. By (2), we now that $\frac{d\hat{u}(e_u)}{de_u} > 0$ and $\frac{d\hat{d}(e_d)}{de_d} > 0$ if $\bar{\tau} < \tau$, so $\hat{u}(e_u^*)$ increases and $\hat{d}(e_d^*)$ decreases, and hence $\hat{\Delta}$ increases through a larger $|\mu|$. Suppose next that $\bar{\tau} > \tau$, so that a larger $|\mu|$ implies that $\mu_u$ ($\mu_d$) decreases (increases). Thus, $e_u^*$ decreases and $e_d^*$ increases. By (2), we now that $\frac{d\hat{u}(e_u)}{de_u} < 0$ and $\frac{d\hat{d}(e_d)}{de_d} < 0$ if $\bar{\tau} > \tau$, so, again, $\hat{u}(e_u^*)$ increases and $\hat{d}(e_d^*)$ decreases, and hence $\hat{\Delta}$ increases through a larger $|\mu|$.

Next, we consider the short position. Since now $\mu = \mu_u - \mu_d \lesseqgtr 0 \iff \bar{\tau} \lesseqgtr \tau$, a larger $|\mu|$ implies that $\mu_u$ ($\mu_d$) decreases (increases) when $\bar{\tau} < \tau$ and that $\mu_u$ ($\mu_d$) increases (decreases) when $\bar{\tau} > \tau$. From above, we know that $\frac{\partial e_u^*}{\partial \mu_u} > 0$ and $\frac{\partial e_d^*}{\partial \mu_d} > 0$. Suppose first that $\bar{\tau} < \tau$, so that a larger $|\mu|$ implies that $\mu_u$ ($\mu_d$) decreases (increases). Thus, $e_u^*$ decreases and $e_d^*$ increases. By (2), we now that $\frac{d\hat{u}(e_u)}{de_u} > 0$ and $\frac{d\hat{d}(e_d)}{de_d} > 0$ if $\bar{\tau} < \tau$, so $\hat{u}(e_u^*)$ decreases and $\hat{d}(e_d^*)$ increases, and hence $\hat{\Delta}$ decreases through a larger $|\mu|$. Suppose next that $\bar{\tau} > \tau$, so that a larger $|\mu|$ implies that $\mu_u$ ($\mu_d$) increases (decreases). Thus, $e_u^*$ increases and $e_d^*$ decreases. By (2), we now that $\frac{d\hat{u}(e_u)}{de_u} < 0$ and $\frac{d\hat{d}(e_d)}{de_d} < 0$ if $\bar{\tau} > \tau$, so, again, $\hat{u}(e_u^*)$ decreases and $\hat{d}(e_d^*)$ increases, and hence $\hat{\Delta}$ decreases through a larger $|\mu|$. ■

**Proof of Proposition 6.** Let $\bar{\tau} \neq \tau$ and $\hat{\Delta} := \hat{u}(e_u^*) - \hat{d}(e_d^*)$. Since $\alpha = \alpha_u - \alpha_d > (<)\ 0$ in the up (down) context, a larger $|\alpha|$ implies that $\alpha_u$ ($\alpha_d$) increases (decreases) in the up context and that $\mu_u$ ($\mu_d$) decreases (increases) in the down context. From (6), we can derive that

$$\begin{aligned}\frac{\partial e_u^*}{\partial \alpha_u} &= \frac{(1-\mu_u)\mu_u(1-\frac{\bar{\tau}}{\tau})^2 u^2}{\left(\mu_u(\alpha_u(1-\frac{\bar{\tau}}{\tau})^2 u^2 - 1) + 1\right)^2} > 0,\\ \frac{\partial e_d^*}{\partial \alpha_d} &= \frac{(1-\mu_d)\mu_d(1-\frac{\bar{\tau}}{\tau})^2 d^2}{\left(\mu_d(\alpha_d(1-\frac{\bar{\tau}}{\tau})^2 d^2 - 1) + 1\right)^2} > 0.\end{aligned}$$

Consider first the up context, so that a larger $|\alpha|$ implies that $\alpha_u$ ($\alpha_d$) increases (decreases). Thus, $e_u^*$ increases and $e_d^*$ decreases. By (2), we now that $\frac{d\hat{u}(e_u)}{de_u} > 0$ and $\frac{d\hat{d}(e_d)}{de_d} > 0$ if $\bar{\tau} < \tau$, so $\hat{u}(e_u^*)$ increases and $\hat{d}(e_d^*)$ decreases, and hence $\hat{\Delta}$ increases through a larger $|\mu|$ when $\bar{\tau} < \tau$. By (2), we also now that $\frac{d\hat{u}(e_u)}{de_u} < 0$ and $\frac{d\hat{d}(e_d)}{de_d} < 0$ if $\bar{\tau} > \tau$, so $\hat{u}(e_u^*)$ decreases and $\hat{d}(e_d^*)$ increases, and hence $\hat{\Delta}$ decreases through a larger $|\mu|$ when $\bar{\tau} < \tau$.

Next, we consider the down context, where a larger $|\alpha|$ implies that $\alpha_u$ ($\alpha_d$) decreases (increases).Thus, $e_u^*$ decreases and $e_d^*$ increases. By (2), we now that $\frac{d\hat{u}(e_u)}{de_u} > 0$ and $\frac{d\hat{d}(e_d)}{de_d} > 0$ if $\bar{\tau} < \tau$, so $\hat{u}(e_u^*)$ decreases and $\hat{d}(e_d^*)$ increases, and hence $\hat{\Delta}$ decreases through a larger $|\mu|$ when $\bar{\tau} < \tau$. By (2), we also now that $\frac{d\hat{u}(e_u)}{de_u} < 0$ and $\frac{d\hat{d}(e_d)}{de_d} < 0$ if $\bar{\tau} > \tau$, so $\hat{u}(e_u^*)$ increases and $\hat{d}(e_d^*)$ decreases, and hence $\hat{\Delta}$ increases through a larger $|\mu|$ when $\bar{\tau} < \tau$. ■

**Proof of Lemma 1.** Propositions 5 and 6 state when increases in each recall bias increase or decrease $\hat{\Delta}$. Lemma 1 uses these statements to identify the environments – consisting of up/down context, forgetting/confabulating signals, and long/short position – in which both recall biases are upward aligned (++), downward aligned (−−), or misaligned (+− or −+). The lemma follows directly from Propositions 5 and 6. ■

**Proof of Proposition 7.** We investigate the effect of increasing similarity, i.e., increasing $\alpha_u$ ($\alpha_d$) in the up (down) context. According to Definition 5, associative and motivated memory are complements (substitutes, unrelated), if increasing similarity has a larger impact on $\hat{\Delta}$ in ++ than +− and in −− than −+ environments. ++ and +− only differ in that the former entails the long whereas the latter entails the short position. Likewise, −− and −+ only differ in that the latter entails the long whereas the former entails the short position. An increase in similarity has an effect on these two comparison if and only if associative memory interacts with motivated memory.

Consider $\hat{\Delta}$ in ++ vs. +− first. In both environments, we consider the up context when forgetting signals and the down context when confabulating signals. Thus, increasing similarity increases $\alpha_u$ in the former and $\alpha_d$ in the latter case. Suppose that $\frac{\partial^2 e_u^*}{\partial\mu_u\partial\alpha_u} > 0$, $\frac{\partial^2 e_d^*}{\partial\mu_d\partial\alpha_d} > 0$ and $\frac{\partial^2 e_u^*}{\partial\mu_d\partial\alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial\mu_u\partial\alpha_d} = 0$. Then, increasing similarity makes the positive effects that $\mu_u$ and $\mu_d$ respectively have on $e_u^*$ and $e_d^*$ (see Proof of Proposition 5) stronger. According to Definition 1, moving from the short to long position induces $\mu_u$ ($\mu_d$) to increase (decrease) if $\bar{\tau} < \tau$ and $\mu_u$ ($\mu_d$) to decrease (increase) if $\bar{\tau} > \tau$. Since increasing similarity increases $\alpha_u$ for $\bar{\tau} < \tau$, moving from the short to long position increases $e_u^*$ more when similarity is increased. By (2), this leads to a larger increase in $\hat{u}(e_u)$ and hence in $\hat{\Delta}$. For $\bar{\tau} > \tau$, increasing similarity increases $\alpha_d$, so moving from the short to long position increases $e_d^*$ more when similarity is increased. By (2), this leads to a larger decrease in $\hat{d}(e_d)$, and hence a larger increase in $\hat{\Delta}$. If $\frac{\partial^2 e_u^*}{\partial\mu_u\partial\alpha_u} < 0$, $\frac{\partial^2 e_d^*}{\partial\mu_d\partial\alpha_d} < 0$ and $\frac{\partial^2 e_u^*}{\partial\mu_d\partial\alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial\mu_u\partial\alpha_d} = 0$, the reverse happens. Then, increasing similarity leads to a smaller increase in $\hat{u}(e_u)$ and a smaller decrease in $\hat{d}(e_d)$, and hence a smaller increase in $\hat{\Delta}$. Finally, if $\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d} = 0$ and $\frac{\partial^2 e_u^*}{\partial \mu_d \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_u \partial \alpha_d} = 0$, increasing similarity leads to the same increase in $\hat{u}(e_u)$ and $\hat{d}(e_d)$, and thus also to the same increase in $\hat{\Delta}$.

Next, consider $\hat{\Delta}$ in $--$ vs. $-+$. In both environments, we consider the down context when forgetting signals and the up context when confabulating signals. Thus, increasing similarity increases $\alpha_d$ in the former and $\alpha_u$ in the latter case. Suppose that $\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u} > 0$, $\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d} > 0$ and $\frac{\partial^2 e_u^*}{\partial \mu_d \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_u \partial \alpha_d} = 0$. Then, increasing similarity makes the positive effects that $\mu_u$ and $\mu_d$ respectively have on $e_u^*$ and $e_d^*$ (see Proof of Proposition 5) stronger. According to Definition 1, moving from the long to short position induces $\mu_u$ ($\mu_d$) to decrease (increase) if $\bar{\tau} < \tau$ and $\mu_u$ ($\mu_d$) to increase (decrease) if $\bar{\tau} > \tau$. Since increasing similarity increases $\alpha_d$ for $\bar{\tau} < \tau$, moving from the long to short position increases $e_d^*$ more when similarity is increased. By (2), this leads to a larger increase in $\hat{d}(e_d)$ and hence a larger decrease in $\hat{\Delta}$. For $\bar{\tau} > \tau$, increasing similarity increases $\alpha_u$, so moving from the long to short position increases $e_u^*$ more when similarity is increased. By (2), this leads to a larger decrease in $\hat{u}(e_u)$, and hence a larger decrease in $\hat{\Delta}$. If $\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u} < 0$, $\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d} < 0$ and $\frac{\partial^2 e_u^*}{\partial \mu_d \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_u \partial \alpha_d} = 0$, the reverse happens. Then, increasing similarity leads to a smaller increase in $\hat{d}(e_d)$ and a smaller decrease in $\hat{u}(e_u)$, and hence a smaller decrease in $\hat{\Delta}$. Finally, if $\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d} = 0$ and $\frac{\partial^2 e_u^*}{\partial \mu_d \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_u \partial \alpha_d} = 0$, increasing similarity leads to the same increase in $\hat{d}(e_d)$ and $\hat{u}(e_u)$, and thus also to the same increase in $\hat{\Delta}$.

To sum up, the impact that an increase in similarity has on $\hat{\Delta}$ is larger (smaller, equal) in aligned than misaligned environments, i.e., in $++$ than $+-$ and in $--$ vs. $-+$, if $\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u} > (<,=)\ 0$, $\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d} > (<,=)\ 0$ and $\frac{\partial^2 e_u^*}{\partial \mu_d \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_u \partial \alpha_d} = 0$. From (6), we can derive that $\frac{\partial^2 e_u^*}{\partial \mu_d \partial \alpha_u} = 0$, $\frac{\partial^2 e_d^*}{\partial \mu_u \partial \alpha_d} = 0$ and

$$\begin{aligned}
\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u} &= -\frac{u^2(1-\frac{\bar{\tau}}{\tau})^2(\alpha_u \mu_u (1-\frac{\bar{\tau}}{\tau})^2 u^2 + \mu_u - 1)}{\left(\mu_u(\alpha_u(1-\frac{\bar{\tau}}{\tau})^2 u^2 - 1) + 1\right)^3}, \\
\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d} &= -\frac{d^2(1-\frac{\bar{\tau}}{\tau})^2(\alpha_d \mu_d (1-\frac{\bar{\tau}}{\tau})^2 d^2 + \mu_d - 1)}{\left(\mu_d(\alpha_d(1-\frac{\bar{\tau}}{\tau})^2 d^2 - 1) + 1\right)^3}.
\end{aligned}$$

The denominators are positive since $\mu_u \alpha_u (1-\frac{\bar{\tau}}{\tau})^2 u^2 + 1 - \mu_u > 0$ and $\mu_d \alpha_d (1-\frac{\bar{\tau}}{\tau})^2 d^2 + 1 - \mu_d > 0$. The numerators, and hence $\frac{\partial^2 e_u^*}{\partial \mu_u \partial \alpha_u}$ and $\frac{\partial^2 e_d^*}{\partial \mu_d \partial \alpha_d}$, are positive (negative, zero) whenever

$$
\begin{aligned}
& 1-\mu_u-\alpha_u \mu_u (1-\tfrac{\bar{\tau}}{\tau})^2 u^2 > (<,=)\ 0 \\
\iff\quad & \mu_u (1+\alpha_u (1-\tfrac{\bar{\tau}}{\tau})^2 u^2) < (>,=)\ 1 \\
\iff\quad & \mu_u < (>,=)\ \frac{1}{1+\alpha_u u^2 \left(1-\frac{\bar{\tau}}{\tau}\right)^2} \\
\iff\quad & \alpha_u < (>,=)\ \frac{\frac{1}{\mu_u}-1}{u^2 \left(1-\frac{\bar{\tau}}{\tau}\right)^2}
\end{aligned}
$$

and

$$
\begin{aligned}
& 1-\mu_d-\alpha_d \mu_d (1-\tfrac{\bar{\tau}}{\tau})^2 d^2 > (<,=)\ 0 \\
\iff\quad & \mu_d (1+\alpha_d (1-\tfrac{\bar{\tau}}{\tau})^2 d^2) < (>,=)\ 1 \\
\iff\quad & \mu_d < (>,=)\ \frac{1}{1+\alpha_d d^2 \left(1-\frac{\bar{\tau}}{\tau}\right)^2} \\
\iff\quad & \alpha_d < (>,=)\ \frac{\frac{1}{\mu_d}-1}{d^2 \left(1-\frac{\bar{\tau}}{\tau}\right)^2},
\end{aligned}
$$

respectively. ■

## References


Amelio, A. and F. Zimmermann (2023). Motivated memory in economics — a review. *Games 14*(1), 15.

Anderson, M. C., K. N. Ochsner, B. Kuhl, J. Cooper, E. Robertson, and J. D. E. Gabrieli (2004). Neural systems underlying the suppression of unwanted memories. *Science 303*(5655), 232–235.

Benabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics 117*(3), 871–915.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1 2*, 69–186.

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

Carlson, R. W., M. A. Maréchal, B. Oud, E. Fehr, and M. J. Crockett (2020). Motivated misremembering of generosity. *Nature Communications 11*(1), 2100.

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

Charles, C. (2024). Memory moves markets. *The Review of Financial Studies 38*(6), 1641–1686.

Charles, C. and P. Sui (2024). Marketwide memory. *Working Paper*.

Chew, S. H., W. Huang, and X. Zhao (2020). Motivated false memory. *Journal of Political Economy 128*(10), 3913–3939.

DellaVigna, S. and E. Linos (2022). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica 90*(1), 81–116.

Drobner, C. (2022). Motivated beliefs and anticipation of uncertainty resolution. *American Economic Review: Insights 4*(1), 89–105.

DuBrow, S. and L. Davachi (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General 142*(4), 1277–1286.

Enke, B., F. Schwerter, and F. Zimmermann (2024). Associative memory, beliefs and market interactions. *Journal of Financial Economics 157*, 103853.

Exley, C. L. and J. B. Kessler (2024). Motivated errors. *American Economic Review 114*(4), 961–987.

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

Gödker, K., P. Jiao, and P. Smeets (2025). Investor memory. *The Review of Financial Studies 38*(6), 1595–1640.

Jiang, Z., H. Liu, C. Peng, and H. Yan (2022). Investor memory and biased beliefs: Evidence from the field. *Working Paper*.

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

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kensinger, E. A. and D. L. Schacter (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience 6*(2), 110–126.

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

Nyhan, B. and J. Reifler (2015). Does correcting myths about the flu vaccine work? an experimental evaluation of the effects of corrective information. *Vaccine 33*(3), 459–464.

Nyhan, B., J. Reifler, S. Richey, and G. L. Freed (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics 133*(4), e835–e842.

Saucet, C. and M. C. Villeval (2019). Motivated memory in dictator games. *Games and Economic Behavior 117*, 250–275.

Schacter, D. L. (2008). *Searching for Memory: The Brain, the Mind, and the Past*. Basic Books.

Sedikides, C. and J. D. Green (2005). Mnemic neglect: Selective amnesia of one's faults. *European Review of Social Psychology 16*(1), 107–143.

Sprengholz, P., L. Henkel, R. Böhm, and C. Betsch (2023). Historical narratives about the covid-19 pandemic are motivationally biased. *Nature 623*(7987), 588–593.

Tulving, E. and D. M. Thomson (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review 80*(5), 352–373.

Wachter, J. A. and M. J. Kahana (2024). A retrieved-context theory of financial decisions. *The Quarterly Journal of Economics 139*(2), 1095–1147.

Walker, W. R., J. J. Skowronski, and C. P. Thompson (2003). Life is pleasant – and memory helps to keep it that way! *Review of General Psychology 7*(2), 203–210.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review 110*(2), 337–361.