Trading Narratives*

Katrin Gödker Christine Laudenbach Cameron Peng

September 2026

Abstract

We study how investors cognitively represent their investment experience. In a survey of retail investors, we elicit open-text narratives about their own stock trades and link them to administrative transaction records. We find that investors retrieve stock purchases from memory far more often than sales, a pattern largely organized by which positions they still hold. Retrieved purchases and sales are, in turn, represented through systematically different narratives: purchase narratives emphasize company fundamentals and are forward-looking, whereas sale narratives emphasize prices and trading rules and are backward-looking. These representations are not static: as the consequences of their trades unfold, investors selectively reconstruct their narratives in ways that justify the original decision. Finally, these representations are linked to beliefs and behavior: the recalled trade type is correlated with overconfidence, and narrative types with trading styles and performance. Overall, we show that how experience shapes beliefs and behavior depends not only on what investors experience, but on how they represent it mentally.

(*JEL codes*: D14, D91, G11, G41.)

*Katrin Gödker: Bocconi University, CESifo, IGIER (katrin.goedker@unibocconi.it); Christine Laudenbach: Goethe University, Leibniz Institute for Financial Research SAFE, CEPR (laudenbach@safe-frankfurt.de); Cameron Peng: London School of Economics, CEPR (c.peng9@lse.ac.uk). We thank participants at HBS Junior Finance Conference, Workshop on Beliefs, Narratives, and Memory, and Yale Junior Finance Conference as well as seminar participants at Aalto University School of Business and Hanken School of Economics Helsinki, FGV-EESP and INSPER, and the VCCE seminar at the University of Vienna for very helpful comments and suggestions. We thank Vivien Bonten and Alireza Moslemihaghighi for great research assistance.

1 Introduction

Experience matters. What investors have lived through shapes their preferences, their beliefs, and their behavior.[1] And there is growing evidence that experience affects choice through a memory channel: what investors believe depends on which experiences they retrieve and how.[2] But what constitutes an investment experience to begin with? Existing research approaches this question quantitatively: experience is a number—a realized return reconstructed from prices or transaction records (Strahilevitz et al., 2011; Malmendier and Nagel, 2011), or a recalled outcome elicited in surveys and experiments (Gödker et al., 2025; Jiang, Liu, Peng, and Yan, 2025).

Common sense and introspection, however, suggest that our investment experience is much more than a number. It also involves the actions we took—bought, sold—and the rationales we attached to those actions: "I bought Nvidia because AI will rule the world." Remembering a trade is therefore not just recalling a number; it is retelling a story: recalling, or reconstructing, why and in what context the decision was made. And stories are precisely what memory holds on to: narrative material is remembered far more durably than statistical information (Graeber, Roth, and Zimmermann, forthcoming). The result of such remembering is what we call a trading narrative: an investor's own account, in their own words, of the context and reasoning behind a particular trading decision. These narratives are an object of interest in their own right. Sitting between investment experiences and investor beliefs and behavior, they reveal how past experiences are encoded in the investor's memory, and the reasoning they contain may itself shape subsequent beliefs and trades.

This paper measures trading narratives directly. In a survey of retail clients of a German online broker, we elicit open-text narratives about individual stock trades and ask three

[1] Macroeconomic and personal experiences have been shown to shape risk taking and portfolio choice (Malmendier and Nagel, 2011; Knüpfer, Rantapuska, and Sarvimäki, 2017; Andersen, Hanspal, and Nielsen, 2019), beliefs and expectations (Malmendier and Nagel, 2016; Anagol, Balasubramaniam, and Ramadorai, 2021), and trading and savings behavior (Kaustia and Knüpfer, 2008; Choi, Laibson, Madrian, and Metrick, 2009; Strahilevitz, Odean, and Barber, 2011). See Malmendier (2021) and Malmendier and Wachter (2024) for surveys of this literature.

[2] Retrieval is cue-dependent and similarity-based (Bordalo, Gennaioli, and Shleifer, 2020; Enke, Schwerter, and Zimmermann, 2024; Wachter and Kahana, 2024), and it can be selective and self-serving (Zimmermann, 2020; Gödker, Jiao, and Smeets, 2025).

questions.[3] Which trades are remembered? How are they represented, and rewritten? Do they matter for beliefs and behavior? We document that investors do not remember their trading history as a neutral record of past decisions. Instead, purchases are recalled more readily than sales; purchases and sales are represented through different narratives; and those narratives are selectively reconstructed as the consequences of the trades unfold.

Our survey design combines three features that allow us to study trading narratives. First, the direction of the recall prompt is randomized: each respondent is randomly assigned to recall and describe either a recent stock purchase (*PURCHASE* condition), a recent stock sale (*SALE* condition), or a recent trade of either direction (*FREE* condition). In addition, each respondent provides a pair of general purchase and sale narratives. This allows us to compare purchase and sale narratives both across randomly assigned investors and within the same investor. Second, we elicit narratives about the same recalled trade twice. Approximately ten months later, respondents describe that trade again, allowing us to study how narratives evolve over time as subsequent investment outcomes are revealed. Third, we link the survey responses to administrative transaction records at the individual level, benchmarking each investor's reports against their own transactions.

To quantify the elicited narratives, we develop a transparent dictionary-based annotation scheme and validate the results with independent human coders and a hypothesis-blind large language model (LLM). The resulting measures provide the first systematic evidence on how investors interpret their past trading experiences and how these interpretations relate to subsequent belief formation, trading behavior, and ultimately trading performance.

Four findings emerge. First, trade experiences are recalled selectively. In the *FREE* condition, 80% of recalled trades are purchases, compared with a purchase share of 70% in investors' own transaction histories ($p < .001$). This asymmetry is largely explained by current ownership: purchases that investors still hold are recalled several times more often than any other type of trade. The determinants of recall also differ by trade type.

[3] Methodologically, the open-text elicitation follows a growing literature that uses free-form survey responses to study economic reasoning and decision-making (Stantcheva, 2021; Andre, Haaland, Roth, and Wohlfart, 2024a; Andre, Schirmer, and Wohlfart, 2024b; Haaland, Roth, Stantcheva, and Wohlfart, 2024, 2025).

Purchase recall has many drivers: current ownership above all, but also trade size, recency, and subsequent returns. Sale recall depends almost entirely on size and recency. In this sense, purchases are remembered as ongoing investments and sales as past events. Memory errors show the same asymmetry: respondents who made no recent trade, when prompted to recall one, are almost twice as likely to report an illusory purchase (36%) as an illusory sale (19%).

Second, purchase and sale experiences are represented in systematically different ways: purchase narratives emphasize company fundamentals and are forward-looking, whereas sale narratives emphasize prices and trading rules and are backward-looking. Purchase narratives are more likely to emphasize company fundamentals (38% versus 17%) and recommendations from others, whereas sale narratives instead emphasize prices and trading rules (57% versus 28%) and external circumstances.[4] Beyond content, purchase narratives are more forward-looking and more likely to embed the trade in its broader context, while sale narratives are more backward-looking, episodic, and trade-specific, emphasizing transaction mechanics and quantitative details. These patterns reappear with remarkably similar magnitudes when the same investor describes why they generally buy and why they generally sell stocks. The two are similarly elaborate; they differ in structure. Sale narratives read like records of a completed transaction, whereas purchase narratives read like the case for an ongoing investment. Purchase and sale narratives are not mirror images of the same decision; they are different mental accounts of it.

Third, trading narratives evolve over time. Approximately ten months later, narratives are 56% shorter and contain 30% fewer content features, yet they remain recognizably related to the original account. More importantly, the content that survives depends on subsequent investment outcomes. When purchases are confirmed by rising prices, investors disproportionately retain fundamental reasoning (61% versus 41%); when purchases are disconfirmed, they instead retain external recommendations (50% versus 19%). Conversely, when sales are disconfirmed by rising prices, investors disproportionately

[4]Part of the circumstance gap is definitional, as some motives (liquidity needs, profit-taking, taxes) structurally exist only on the sell side. Dropping narratives that invoke such direction-specific motives eliminates the circumstance gap but leaves the fundamentals and recommendation asymmetries essentially unchanged and most of the price/rule asymmetry intact (Appendix J).

retain trading rules—coincidentally by the same 61%-versus-41% margin—and external circumstances, preserving justifications that remain defensible despite the outcome. The pattern survives a battery of design checks. Trading narratives therefore evolve through selective reconstruction that justifies the original decision, not through uniform forgetting.

Finally, trading narratives are associated with beliefs and behavior. Which trades come to mind predicts overconfidence: investors who spontaneously recall a purchase rather than a sale are about 18 percentile points more overconfident about their relative performance, controlling for their actual trading. Narrative types predict trading styles on both sides of the trade. Investors who articulate no reason for their trades are the most extrapolative, whereas fundamentals narrators are the most contrarian buyers. On the selling side, investors whose narratives use profit-taking vocabulary realize gains at nearly double the rate of others, price/rule narrators disproportionately sell after price run-ups, and episodic narrators hold positions twice as long before selling. Narrative types also predict the quality of trades: fundamentals narrators are the only group whose purchases outperform their sales, by about two percentage points over the following quarter. We present these patterns as associations, but they locate trading narratives at the point where beliefs and trading decisions form.

The paper contributes to a growing literature on narratives in economics and finance (Shiller, 2017; Andre et al., 2024b,a; Haaland et al., 2025), which so far has focused largely on aggregate narratives about macroeconomic events. Despite this growing interest, little is known about how narratives evolve at the individual level and over time, especially in high-stakes environments such as investing. We measure individual trading narratives longitudinally. We provide evidence that investor narratives are distinct for purchases and sales, selectively reconstructed as investment outcomes unfold, and predictive of subsequent investor beliefs and trading behavior. Further, our approach of combining open-text narratives with administrative transaction records extends the study of narratives from macroeconomic settings to the micro-behavior of individual retail investors.

Moreover, we contribute to the emerging economics literature on memory and beliefs (Bordalo et al., 2020; Zimmermann, 2020; Bordalo, Conlon, Gennaioli, Kwon, and Shleifer, 2023; Enke et al., 2024; Fudenberg, Lanzani, and Strack, 2024; Wachter and Kahana, 2024;

Gödker et al., 2025; Jiang et al., 2025) in several ways. First, we show that investors selectively retrieve trading experiences. Second, existing empirical studies typically measure memory in relatively narrow or controlled ways, for example by eliciting recall of specific outcomes, returns, or episodes (Zimmermann, 2020; Gödker et al., 2025; Jiang et al., 2025). Little is known about the richer structure of investor memory. We introduce a methodological approach that combines free recall, guided recall, and open-ended narrative elicitation, allowing us to capture what is cognitively "top of mind." Beyond this application, our annotation framework provides a transparent process for studying investor cognition from open-text data.

Third, most economics and finance work on memory treats memory primarily as a retrieval problem of which experiences are more accessible (Bordalo et al., 2020, 2023; Wachter and Kahana, 2024). Our evidence suggests something richer: investors do not merely retrieve trading experiences; they reconsolidate and reconstruct them: retrieval itself alters what is stored. This process gives rise to the selective retention of content, the addition of new material, and change over time, patterns that ultimately connect to beliefs and behavior. Moreover, we document two novel empirical facts that bear on theories of memory-based beliefs. First, memory is organized by ownership: a position still held in the portfolio is much more likely to be recalled or constructed. The experiences most readily retrieved are not a representative sample of past trading activity but a selective subset centered on ongoing investments. This provides field evidence in support of the cue-based view of memory retrieval advanced by recent theory (Bordalo et al., 2020; Wachter and Kahana, 2024): a still-held position acts as a standing retrieval cue, making the memory of its purchase accessible. Second, we show that narratives of the same trade experience change in outcome-contingent ways. This is consistent with motivated memory as a microfoundation for experience effects (Bénabou and Tirole, 2002, 2016; Gödker et al., 2025).

Our paper also extends the substantial body of work on purchase–sale asymmetries in retail trading. Traditional portfolio theory treats stock purchases and sales symmetrically, viewing both as integral parts of the portfolio rebalancing process. However, empirical evidence suggests pronounced asymmetries between these actions. For example, purchases

are often extrapolative, driven by expectations of continued price increases, whereas sales tend to be contrarian, motivated by the desire to realize gains or limit losses (Barber and Odean, 2013). The asymmetry extends even to experts: institutional portfolio managers add value when buying but destroy it when selling (Akepanidtaworn, Di Mascio, Imas, and Schmidt, 2023). Existing work primarily studies asymmetries in attention allocation, information acquisition, or trading outcomes. We complement this work by showing that purchases and sales already differ at the level of recall and narrative representation, even by the same investor in the same survey. These cognitive asymmetries are reflected in behavior: the narratives investors use to describe buying and selling are linked to differences in subsequent trading styles and, for purchases, the quality of trading decisions. A long literature documents that active trading often leads to suboptimal outcomes (Odean, 1998, 1999; Barber and Odean, 2008; Strahilevitz et al., 2011), yet relatively little is known about the cognitive representations underlying trading decisions; our evidence speaks directly to them.

The paper proceeds as follows. Section 2 describes the surveys and the transaction records. Section 3 studies which trades investors retrieve from memory. Section 4 documents how purchases and sales are represented in narratives, and Section 5 how those narratives are reconstructed as the consequences of the trades unfold. Section 6 relates trading narratives to investor confidence and trading behavior. Section 7 concludes.

2 Survey Design and Data

We conduct a two-wave survey among clients of a German online broker and link the responses to administrative transaction records. The broker is a large platform serving self-directed retail investors. Figure 1 summarizes the survey flow. The survey instructions are provided in Appendix O (main survey) and Appendix P (follow-up survey).

2.1 Main Survey

The main survey (wave 1) was sent to a random sample of 20,000 retail broker clients aged 18 years or older, excluding clients who had opted out of email communication from the broker. Administered in German, the survey took approximately 15 minutes to complete, and respondents received 5.00 EUR (via PayPal or an Amazon voucher). Data collection took place between October and December 2023, and 1,524 clients completed the survey, corresponding to a response rate of 7.6%. Respondents first recall a trade and describe it in their own words. We then elicit general purchase and sale motives, return expectations for the stock, portfolio, and market; confidence; decision-making style; and demographics.

Recall. To elicit respondents' recall of recent trading activity, the survey first asks them whether they have traded a stock through their brokerage account within the previous four weeks. We randomly assign respondents to one of three conditions that vary the wording of this question: it refers to a purchase (the *PURCHASE* condition), a sale (*SALE*), or a trade of either direction (*FREE*). Respondents who report such a trade are asked to think of one specific trade—the first that comes to mind if multiple trades qualify—and to report the name of the stock. Respondents who have not traded within the previous four weeks, or are unsure, are instead asked to recall a trade from further in the past. In the *FREE* condition, respondents can recall either a purchase or a sale and then indicate the direction of the trade. Comparing the *PURCHASE* and *SALE* conditions allows us to examine differences in the recall and narratives of purchases and sales. The *FREE* condition reveals which type of trade respondents recall when we do not prompt a direction.

Trading narratives. A trading narrative, as defined in the introduction, is an investor's own account, in their own words, of the context and reasoning behind a particular trading decision. We elicit these narratives through open-text questions. For the recalled trade, respondents explain why they made the trade. We also ask what generally motivates them to buy and to sell stocks, yielding a parallel pair of general trading narratives for each respondent. Open-text elicitation has become an established tool for measuring how

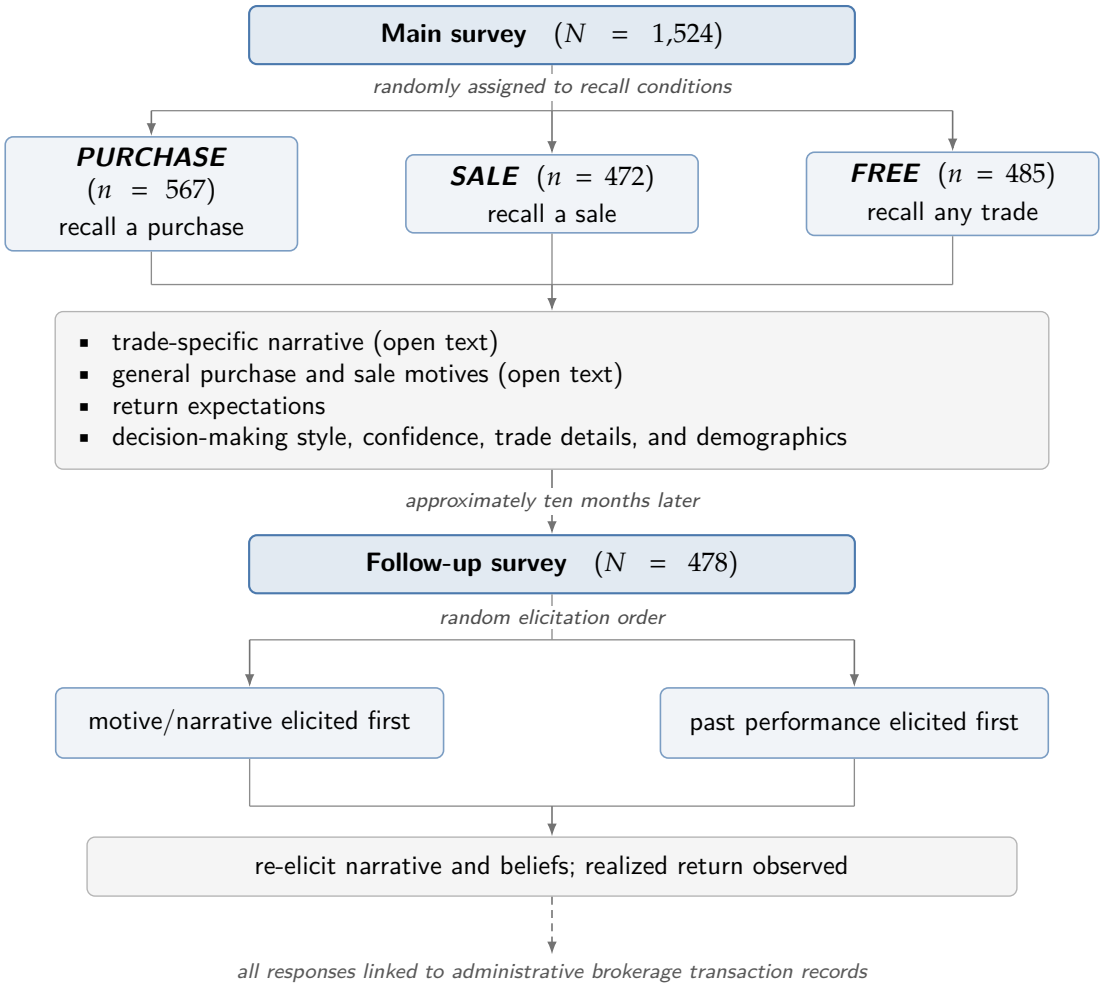


Figure 1: **Survey design.** In the main survey, respondents are randomly assigned to one of three recall conditions: *PURCHASE*, in which they recall a stock purchase; *SALE*, in which they recall a stock sale; or *FREE*, in which they recall a stock trade of either direction. All respondents subsequently provide an open-text narrative describing the recalled trade, open-text descriptions of their general motives for buying and selling stocks, return expectations, measures of decision-making style and confidence, information on the recalled trade, and demographic characteristics. Approximately ten months later, respondents are invited to complete a follow-up survey. In the follow-up survey, we randomize whether respondents first report the realized performance of the recalled stock or first provide the trade narrative and beliefs again. All survey responses are linked to administrative brokerage transaction records.

people reason (Ferrario and Stantcheva, 2022; Stantcheva, 2023) and has been applied in settings including taxation (Stantcheva, 2021), the macroeconomy (Andre et al., 2024a), and stock prices (Andre et al., 2024b). Rather than requiring economists to hypothesize why investors trade, we let respondents explain their motivations directly. Section 4.1 describes how we turn the open-text narratives into quantitative measures.

Return expectations. To elicit return expectations, we ask respondents to report their expected percentage change over the next twelve months for three objects: the recalled stock, their overall portfolio at the broker, and the aggregate market (the German stock index, DAX). Responses are elicited on a categorical scale ranging from "less than −20%" to "more than +20%" in five-percentage-point bins. This format lets respondents express directional beliefs and approximate magnitudes without the burden of precise point forecasts, a scale design common in the expectations literature (Manski, 2004; Giglio, Maggiori, Stroebel, and Utkus, 2021). After each forecast, we also ask how confident the respondent is in that estimate, on a scale from 1 (not confident at all) to 7 (extremely confident). For the empirical analysis, we map the categories onto an ordered numerical scale, preserving the ordinal information in respondents' answers.

Decision-making style, confidence, and demographics. In addition to the trading narratives and return expectations, we collect a set of structured measures. To characterize respondents' decision-making style, we ask how long they deliberated before making the recalled trade, how long they planned to hold the stock, how frequently they trade, how important they consider market timing, and about their self-reported financial literacy and risk tolerance. We also measure respondents' confidence in their own investment ability. Specifically, we ask them to estimate the percentage of the broker's clients who achieved a higher annual portfolio return than they did over the twelve months preceding the survey (between 0% and 100%). Subtracting this figure from 100 yields a self-assessed percentile rank, so higher values mean the respondent believes they outperformed more of the broker's clients. Finally, we collect standard demographic information, including year of birth, gender, education, and marital status.

2.2 Follow-up Survey

To examine how trading narratives evolve over time, we re-surveyed all respondents from the main survey who had agreed to be contacted again by email (914 clients). Data collection took place in October 2024. The follow-up survey (wave 2) took approximately five minutes to complete and compensated respondents with 5.00 EUR. A total of 478

respondents completed the survey, corresponding to a response rate of 52.3%. The survey begins with a reminder of the previously recalled stock (name and security identifier), followed by the open-text trading narrative for that trade (the re-elicitation) and questions on return expectations, decision-making style, and confidence.

The follow-up survey includes a second randomization: the order of the narrative re-elicitation and the performance question is randomly assigned across respondents. We use this order randomization in Appendix C to test whether making the realized outcome salient before re-elicitation changes how respondents reconstruct their narratives.

We compare respondents who completed the follow-up with the full main-survey sample (Appendix E). Selection into the follow-up survey is modest. Follow-up respondents are somewhat older, more likely to be male, more engaged (as reflected in longer trading narratives), and slightly more optimistic and confident. Importantly, they do not differ in their actual trading activity over the sample period.

2.3 Investor Transaction Data

We obtain transaction-level data from the broker for all 20,000 randomly sampled clients. For survey respondents, we link the survey responses to the transaction records at the individual level using a pseudonymized identifier. Of the 1,524 survey respondents, 1,518 executed at least one trade during the sample period. The records cover every transaction in respondents' investment accounts between October 2021 and January 2024, including the transaction date and time, asset identifier (ISIN), transaction price (in EUR), number of shares, and transaction type. Our analysis focuses on common-stock transactions. Multiple purchases or sales of the same stock in the same account on the same day are aggregated into a single transaction. We restrict the analysis to investment accounts held in the respondent's own name. If a respondent holds more than one such account, we merge the accounts.

Actual investment performance and overconfidence. The administrative transaction records allow us to construct each respondent's realized portfolio performance over the twelve months preceding the survey. We reconstruct the portfolio held on each trading day

by rolling the January 2024 holdings snapshot backward through the complete transaction history and compute the portfolio's daily value-weighted total return, compounded over the twelve-month period.[5] We then rank each respondent's realized return against the returns of all broker clients over the same window, yielding an *actual percentile rank*. Comparing this realized rank with the respondent's self-assessed percentile rank from the survey yields our measure of *overconfidence*, defined as the difference between the self-assessed and actual percentile ranks.[6]

Trading behavior. The transaction records further allow us to characterize respondents' realized trading behavior. We construct several investor-level measures capturing trading styles, including extrapolative buying (the share of purchases following recent stock-price increases), profit-taking (the propensity to realize gains, following Odean 1998), selling after price run-ups, holding periods, and other trading characteristics. Unless otherwise noted, these measures are computed over all common-stock transactions in the administrative records prior to the survey. Appendix N provides the precise construction of each measure.

Trading performance. Finally, we measure the quality of investors' trading decisions by comparing the subsequent performance of purchased and sold stocks. Following Odean (1999), we compute the market-adjusted return of each traded stock over the three months following the transaction and compare the average performance of purchased and sold stocks for each narrative group. Because market movements affect purchased and sold stocks roughly equally, this purchase–sale return differential isolates whether investors systematically buy stocks that subsequently outperform the stocks they sell. We view these comparisons as descriptive evidence linking trading narratives to realized trading performance rather than as causal estimates.

Throughout the paper, we study investors' single-stock trades. Stocks are the natural object for our analysis because they represent claims on individual firms, allowing in-

[5]The measure is computable for 843 of 1,524 respondents; construction, coverage, and a static buy-and-hold variant that yields the same conclusions are described in Appendix K.

[6]Because the survey asks respondents to compare themselves with other clients of the broker, the percentile ranking provides the natural benchmark.

vestors to form firm-specific narratives and return expectations. This focus also reflects investors' actual portfolios: for the median respondent, stocks account for 76% of portfolio value and the majority of both holdings and transactions (Appendix Table AI).

2.4 Summary Statistics

Table 1 reports summary statistics for the sample of 1,524 survey respondents, randomly assigned to the *PURCHASE* ($n = 567$), *SALE* ($n = 472$), or *FREE* ($n = 485$) condition. Overall, 46% of the survey respondents report having traded a stock in the previous four weeks. Return expectations are moderately optimistic. The mean expectation for the recalled stock is 6.50 on the 1–10 scale; the median of 7 corresponds to expected one-year returns between +5% and +10%. Respondents' confidence in their relative investment ability is centered around the population midpoint. The mean self-assessed percentile rank is 49.3 (median = 50), indicating that respondents, on average, view their performance as similar to that of the typical broker client. Trading activity varies widely across respondents. The mean number of stock transactions over the sample period (October 2021 to January 2024) is 57.5, compared with a median of only 6, reflecting a strongly right-skewed distribution. The sample is predominantly male: 15% of respondents are female and 2% report another gender category. Respondents are, on average, 44.8 years old (median = 43), and educational attainment is relatively high, with a mean of 6.22 on a 1–9 scale (median = 6, corresponding to a university of applied sciences degree). Further, 65% report this broker as their primary brokerage account. Appendix B reports detailed summary statistics for the narratives respondents provide in the surveys.

Table 1: Summary Statistics: Survey Respondents

	Mean	Median	St. Dev.
N = 1,524			
Traded in last 4 weeks (y/n)	0.46	0	0.50
Stock return expectations (1y, 1–10)	6.50	7	2.51
	(Median = +5% to +10%)		
Portfolio return expectations (1y, 1–10)	6.70	7	1.95
Market return expectations (1y, 1–10)	6.14	6	1.58
Confidence (percentile rank, 0–100)	49.27	50	23.29
	(Median = 50th percentile)		
Number of stock trades	57.46	6	218.89
Female (dummy)	0.15	0	0.36
Age (in years)	44.77	43	14.71
Educational level (1–9)	6.22	6	1.85
Main account at broker (dummy)	0.65	1	0.48

Notes: Stock, portfolio, and market return expectations are measured on a 10-point categorical scale (1 = <−20% to 10 = >+20%). Number of stock trades counts transactions in the broker database (Oct. 2021–Jan. 2024), aggregated to the same-day, same-stock, same-direction level. Confidence is measured as 100− the respondent's self-reported share of clients who earned a higher return. Main account at broker equals 1 if the respondent indicated this broker as their primary brokerage account. Statistics are computed over non-missing responses, so the effective N varies by row. Descriptive statistics for the constructed outcome measures (actual performance rank, overconfidence, and the trading-behavior measures) appear in Appendix Table AXIII.

3 Selective Recall of Trades

Our first question is which trading experiences are retrieved from memory. To answer it, we exploit two complementary experimental designs. First, we examine *free recall*, asking respondents to recall a stock trade without specifying whether it was a purchase or a sale (the *FREE* condition). Second, we examine *guided recall*, asking respondents to recall either a stock purchase or a stock sale (the *PURCHASE* and *SALE* conditions). Together, these designs reveal three purchase–sale asymmetries. First, purchases are more likely to be recalled than sales. Second, investors are far more likely to remember purchases that did not occur than sales that did not occur—the asymmetry extends beyond selective retrieval of genuine experiences. Third, the determinants of recall differ by trade type. Recall of purchases is primarily organized by whether the stock is still held, whereas recall of sales depends mainly on trade size and recency.

3.1 Purchases Are More Readily Retrieved Than Sales

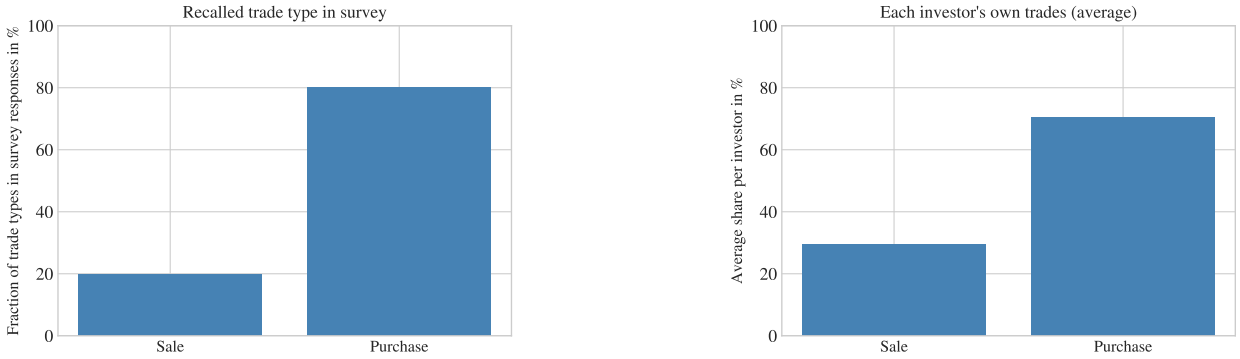
The *FREE* condition provides a direct test of whether purchases and sales are equally likely to be retrieved from memory: respondents recall a stock trade without being prompted toward either direction. We find that 80% of recalled trades are purchases (Figure 2(a)). This imbalance exceeds what investors' own trading histories would predict. Although these investors purchase stocks more frequently than they sell them, the average purchase share in their transaction records is 70%, well below the 80% observed in recall. We formalize this comparison using a per-investor random-recall benchmark that draws one trade at random from each respondent's individual transaction history. The resulting null distribution is centered at 70%, with the observed recall share lying far in its upper tail ($p < .001$; Figure 2(b)).

The over-representation of purchases in recall does not depend on what investors trade. Figure 2(c) groups respondents by the purchase share of their actual transactions. Across most of the distribution, recalled purchases exceed the random-recall benchmark. The difference is largest for investors with a balanced share of purchases and sales in their trading histories or histories tilted toward sales, where an over-representation of purchases is least mechanical. As investors' own purchase share approaches 100%, the gap naturally narrows. The documented pattern is stable across subsamples: the probability of recalling a purchase remains close to 80% regardless of whether the observed brokerage account is the respondent's main account and whether the respondent traded in the previous four weeks (Appendix Table AVI).[7]

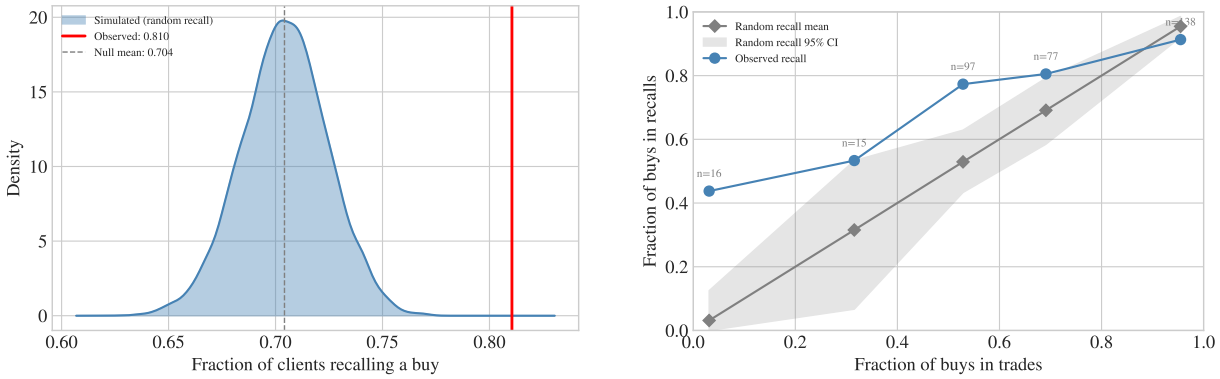
Why are purchases more readily retrieved than sales? Figure 3 assesses whether the asymmetry can be explained by mechanical differences in trade characteristics. It cannot. Recency, trade size, and return salience all predict purchase shares close to the random benchmark (which, unlike Figure 2, draws only from pre-survey trades), whereas current ownership almost fully reproduces the observed purchase advantage.[8]

[7] Nor can it be attributed to automated saving, since savings plans at this broker are limited to ETFs and mutual funds and do not involve individual stocks.

[8] This generalizes the random-recall benchmark of Section 3.1, which draws uniformly; here the draw is weighted by a trade's recall propensity within bins of the feature (age, size, salience, or ownership), estimated from the recalled trades we can match. The uniform benchmark is 72% here, against the 70% of Figure 2, because it draws only from each investor's pre-survey trades—the recalled trade must predate the



(a) Recalled trades (left) vs. each investor's own trades (right)



(b) Observed recall vs. random-recall null

(c) Recall by trade composition

Figure 2: **Purchases are disproportionately recalled.** The figure compares the direction of recalled trades with investors' realized trading histories. In the *FREE* condition, purchases account for 80% of recalled trades, exceeding the 70% purchase share in investors' own transaction histories. A random-recall benchmark confirms that this difference is unlikely to arise by chance and persists across investors with different trading compositions.

Weighting recall solely by whether the stock is still held increases the predicted purchase share to 79%, close to the observed 81% (Figure 3). Investors are therefore disproportionately likely to retrieve purchases that remain in their portfolios.

We next examine which of an investor's past trades is retrieved from memory. For each respondent in the *FREE* condition whose recalled trade can be matched to the transaction records, we stack the investor's complete pre-survey trading history and use a conditional logit model to estimate the probability that a given trade is recalled. Each investor contributes a single choice occasion, so identification comes entirely from comparisons among

survey—whereas Figure 2 draws from the full trade history. Appendix Table AVIII reports the recency case in detail.

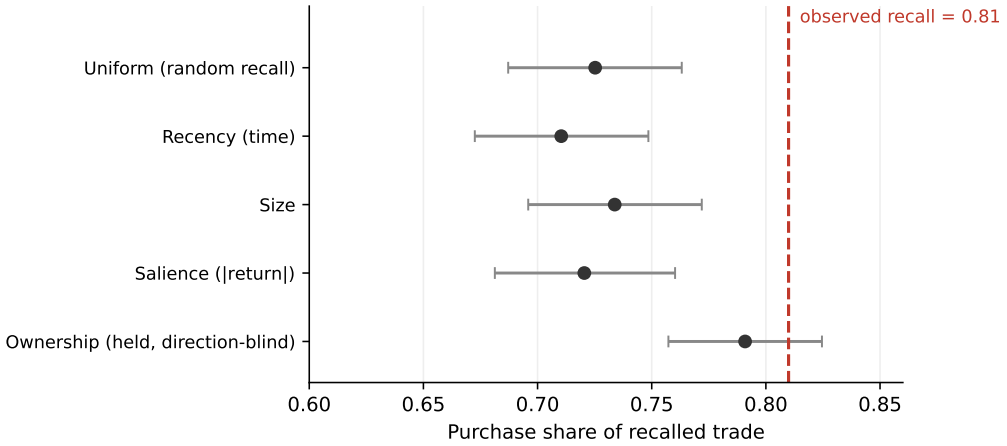


Figure 3: **No mechanical model reproduces the purchase tilt, but ownership does.** Purchase share of the recalled trade under single-factor recall models. Each model draws one trade per investor from that investor's own pre-survey trades, with draw probability proportional to the empirical recall rate along one dimension, and reports the mean purchase share (dot) with its 95% bootstrap interval (10,000 draws). The dashed line is the observed purchase-recall share (81%).

Table 2: Recall rate by trade direction and ownership

	Still held	Exited
Purchase	3.4%	0.3%
Sale	0.5%	0.4%

Notes: Share of trades in each cell that are the respondent's recalled trade, among the complete pre-survey trades of the 214 investors in the *FREE* condition whose recalled trade is located in their history under the fuzzy match (stock name and direction); this is the sample of columns 1–3 of Table 3. "Still held" indicates a positive position in the stock at the survey date.

that investor's own trades.[9] The raw data already reveal a striking pattern (Table 2). Purchases that investors still own are recalled far more often than any other type of trade. Exited purchases and both types of sales are recalled only rarely. Roughly three-quarters of all recalled trades are purchases that remain in the investor's portfolio.

Table 3 confirms this pattern at the level of the individual trade. Column 1 reproduces the recall asymmetry: purchases are significantly more likely than sales to be recalled. Once we account for whether the stock is still held, however, the purchase coefficient becomes small and statistically insignificant, while the ownership effect is large and precisely

[9] An investor who trades hundreds of times and one who trades a handful of times each contribute a single choice occasion, so the estimates weight every investor equally and are not dominated by the most active traders, as a trade-level linear model would be. We estimate the conditional (McFadden) logit for the case of exactly one recalled trade per investor.

Table 3: Which trade is recalled: the role of ownership (*FREE* condition)

	(1)	(2)	(3)	(4)	(5)
Purchase	+0.72***	+0.13	−0.42	−0.29	−0.52
	(0.22)	(0.23)	(0.36)	(0.39)	(0.55)
Still held		+1.80***	+1.09**	+1.06**	+0.66
		(0.26)	(0.44)	(0.49)	(0.61)
Purchase × still held			+1.08**	+1.19**	+1.99***
			(0.52)	(0.59)	(0.77)
Log trade size				+0.41***	+0.55***
				(0.13)	(0.16)
Return since × purchase				+0.93***	+0.83**
				(0.26)	(0.36)
Return since × sell				−0.03	−0.30
				(0.55)	(0.75)
Return before × purchase				+0.06	+0.59
				(0.40)	(0.51)
Return before × sell				−0.23	−1.12
				(0.87)	(1.20)
McFadden R^2	0.011	0.063	0.067	0.086	0.105
Investors	214	214	214	192	139

Notes: Investor-weighted conditional logit; the outcome equals one for the recalled trade, and each column stacks the complete pre-survey trades of the investors in the *FREE* condition whose recalled trade matches a transaction. Columns 1–4 use the fuzzy match (stock name and direction); column 5 repeats the full specification on the strict match (name, direction, and amount within 20%). Investor counts are reported at the foot of each column. *Still held*: a positive position in the stock at the survey date, rolled back from the January 2024 holdings snapshot. *Return since / before trade*: the stock's total return between the trade and the survey, and over the 91 days before the trade; both winsorized and interacted with direction. Trade recency is omitted (the elicitation prompts a recent trade). Standard errors in parentheses. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

estimated (column 2). Column 3 shows that this effect is concentrated among still-held purchases: both the ownership indicator and its interaction with purchases are significant. Adding trade size and stock returns leaves this pattern unchanged (column 4). Recalled purchases are somewhat more likely to be those whose stock subsequently appreciated,[10] whereas neither pre-sale returns nor subsequent stock returns predict which sales are recalled. The results are qualitatively similar using the stricter matching criterion (column 5).

Ownership appears to act as a persistent retrieval cue. Stocks that remain in the portfolio stay visible, are monitored, and are repeatedly encountered, making the original

[10]This tilt toward winners is modest and, because respondents choose which trade to recall, need not reflect memory retrieval alone.

purchase easier to retrieve. Yet retrieval is not simply a readout of the current portfolio: even among still-held positions, recall declines sharply with time since purchase, from 6.6% within three months to 1.3% beyond one year.[11] Ownership thus facilitates retrieval even as accessibility fades.[12]

3.2 Recalled Trades Are Real, but Illusions Favor Purchases

We now turn to how accurately investors remember past trades. Comparing recalled trades with the administrative transaction records serves two purposes. First, it establishes whether the recalled trades correspond to genuine transactions. Second, it reveals a second purchase–sale asymmetry, this time in memory errors. In the guided recall conditions, respondents are prompted to recall either a stock purchase (*PURCHASE*) or a stock sale (*SALE*), allowing us to compare recall accuracy across the two directions.[13]

We compare recalled trades with the administrative transaction records using a three-step matching procedure. First, we match the reported stock to the transaction records after cleaning and standardizing stock names by removing legal suffixes, share-class markers, and broker-specific abbreviations. Matching is based on token-level fuzzy string matching, requiring every token in the shorter name to match a token in the longer name either exactly or as a prefix.[14] Second, the trade direction must agree with the transaction record. In the *PURCHASE* and *SALE* conditions, direction is fixed by the recall prompt. Third, the reported transaction value must be within 20% of the recorded transaction value. If multiple transactions satisfy these criteria, we assign the recalled trade to the transaction with the smallest absolute difference in transaction value.

Panel A of Table 4 asks whether the recalled trade corresponds to a recorded transaction.

[11] Recall rates among still-held purchases in the estimation sample of Table 3 are 6.6% at one to three months, 4.5% at three to six months, 2.9% at six to twelve months, and 1.3% beyond one year.

[12] The pattern is not driven by the survey's recent-trade prompt: among respondents who reported no trade within the previous four weeks, and were therefore free to recall any past trade, recall still declines with time since trade and ownership continues to organize retrieval.

[13] The guided conditions are better suited for identifying recall determinants because they separate purchase and sale recall. In the *FREE* condition, current ownership is highly collinear with recalling a purchase, and the initial four-week recall window mechanically truncates the distribution of trade age. The guided conditions also provide more matched observations, particularly for sales.

[14] Prefix matches require at least five characters and at least 60% coverage of the longer token.

Table 4: Accuracy of recall in guided recall conditions

	PURCHASE cond.	*SALE* cond.
Answered survey	567	472
Panel A. Is the recalled trade real?		
Recalled a stock trade	472 (83.2%)	334 (70.8%)
Matched, fuzzy	295 (52.0%)	144 (30.5%)
Matched, strict	169 (29.8%)	81 (17.2%)
Panel B. Accuracy (traded in last four weeks)		
Traded, reported	126 (94.0%)	78 (91.8%)
Traded, not reported (forgetfulness)	8 (6.0%)	7 (8.2%)
Not traded, reported (illusion)	157 (36.3%)	75 (19.4%)
Not traded, not reported	276 (63.7%)	312 (80.6%)
Panel C. What are the illusions?		
Named a stock	109 (69.4%)	58 (77.3%)
Matched, fuzzy	54 (34.4%)	21 (28.0%)
Matched, strict	24 (15.3%)	6 (8.0%)

Notes: Recall outcomes in the guided recall conditions. *Panel A* assesses whether the recalled trade is real: "Recalled a stock trade" means the reported transaction was a stock trade; a "fuzzy" match requires the recalled stock name (via token-based fuzzy matching) and direction to agree with a transaction in the records, and a "strict" match additionally requires the self-reported value to be within 20% of the actual amount. *Panel B* compares each respondent's self-report of trading in the last four weeks ("Reported") against whether the records show a trade of the prompted direction in that window ("Traded"): forgetfulness is a Type II error (traded but not reported), illusion a Type I error (reported but not traded). *Panel C* restricts to the illusion cases and asks whether the falsely reported trade matches a real trade of the prompted direction on any date; percentages there are of the illusion count. Percentages are shown in parentheses.

Among respondents prompted to recall a purchase, 83% name a stock trade, 52% correctly identify a stock they purchased (a fuzzy match on stock name and direction), and 30% correctly identify the specific transaction (a strict match including transaction value). The corresponding rates for sales are 71%, 31%, and 17%. The decline from the fuzzy to the strict match primarily reflects imperfect recall of transaction values rather than incorrect recall of the trade itself, and the purchase advantage appears at every stage. Thus, where recall can be verified, respondents generally remember genuine trades, providing the foundation for treating recalled trades as meaningful observations throughout the paper.[15]

Panel B turns to the accuracy of recall by comparing respondents' reports of trading

[15]Recall in the *FREE* condition is comparably accurate: 52% of recalled trades match a recorded transaction on stock and direction, and 35% also match on transaction value, similar to the guided conditions. Thus, the recall asymmetry documented in Section 3.1 reflects genuine transactions rather than fabricated ones.

within the previous four weeks with the transaction records. One possible error is *forgetfulness* (a Type II error): respondents traded during the relevant period but report that they did not. Such errors are relatively rare. Among respondents who traded within the previous four weeks, only 6.0% in the *PURCHASE* condition and 8.2% in the *SALE* condition fail to report a recent trade.

The second, and more consequential, error is a *memory illusion* (a Type I error): respondents report a recent trade despite having made none during the relevant time period. Among respondents who did not trade within the previous four weeks, 36.3% in the *PURCHASE* condition and 19.4% in the *SALE* condition nevertheless report a recent trade. *Panel C* examines these memory illusions. Only about one-third correspond to a stock that respondents actually traded in the reported direction (a fuzzy match, typically to a transaction approximately three months earlier), and fewer than one-fifth match a specific recorded transaction. Most illusions therefore cannot be explained simply as misdated trades (i.e., a wrong sense of timing). Although an unmatched recall need not represent a false memory—some respondents also trade through brokers outside our data—the asymmetry is not driven by such transactions. It remains among respondents for whom this broker is the primary brokerage account (Appendix Table AVII).

3.3 Purchase and Sale Recall Run on Different Cues

What makes one trade more memorable than another? We again focus on the guided conditions, which allow us to study whether the determinants of recall differ between purchases and sales. On the purchase side (Table 5), ownership again emerges as the dominant predictor of recall (+0.75 to +1.49), alongside trade size and time since trade. Initial purchases that open a position are significantly more likely to be recalled than additional purchases into an existing position (+0.43), suggesting that investors particularly remember the decision to establish a new investment. The modest tilt toward subsequent winners also reappears (+0.84).

The determinants of sale recall look markedly different (Table 6). Ownership no longer matters: whether the sale leaves a position open or closes it has no detectable effect on

Table 5: Which purchase is recalled (*PURCHASE* condition)

	(1)	(2)	(3)	(4)	(5)
Still held	+0.94***	+0.95***	+0.75***	+1.00***	+1.49***
	(0.20)	(0.20)	(0.22)	(0.26)	(0.38)
Initial buy (opens position)		+0.21	+0.43**	+0.40**	+0.35
		(0.17)	(0.19)	(0.20)	(0.29)
Log days since trade			−0.92***	−0.88***	−0.98***
			(0.07)	(0.07)	(0.09)
Log trade size			+0.64***	+0.57***	+1.05***
			(0.11)	(0.12)	(0.19)
Return since trade				+0.84***	+0.76*
				(0.32)	(0.45)
Return before trade				−0.16	−0.15
				(0.37)	(0.52)
McFadden R^2	0.018	0.019	0.193	0.195	0.271
Investors	271	271	271	241	144

Notes: Investor-weighted conditional logit over the pre-survey purchases of *PURCHASE*-condition investors whose recalled trade matches. Columns 1–4 use the fuzzy match, column 5 the strict match; investor counts (choice occasions) are reported at the foot of each column. *Still held*: a positive position in the stock at the survey date. *Initial buy*: the purchase opens a position rather than adding to one. *Return since / before trade*: as in Table 3. Standard errors in parentheses. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

recall. Instead, recall depends primarily on trade size and recency; sales of longer-held positions are marginally more likely to be remembered (+0.20, $p < .10$), even conditional on when the sale occurred. By contrast, no return measure predicts which sale is recalled: neither the realized holding-period return, nor the run-up before the sale, nor subsequent price movements.

Taken together, purchase and sale recall run on different retrieval cues: purchases are remembered as ongoing investments, sales as past events. This asymmetry in what memory makes available foreshadows an asymmetry in how purchases and sales are represented, to which we turn next.

4 Asymmetric Representation of Trades in Narratives

Section 3 showed that investors selectively retrieve purchases and sales from memory. We now turn to how those retrieved experiences are mentally represented by investors. We ask whether purchase and sale narratives are simply mirror images of the same invest-

Table 6: Which sale is recalled (*SALE* condition)

	(1)	(2)	(3)	(4)	(5)
Still held	+0.27	+0.29	−0.16	−0.33	−0.25
	(0.30)	(0.33)	(0.44)	(0.48)	(0.57)
Log days since trade		−1.15***	−1.18***	−1.15***	−1.29***
		(0.09)	(0.11)	(0.12)	(0.15)
Log trade size		+0.32**	+0.50**	+0.51**	+0.72**
		(0.15)	(0.21)	(0.24)	(0.34)
Holding-period return				−0.49	+1.54
				(1.13)	(1.84)
Log holding period			+0.20*	+0.15	+0.18
			(0.11)	(0.12)	(0.15)
Return since trade				+0.53	+0.67
				(0.59)	(0.83)
Return before trade				−0.35	−0.42
				(0.80)	(1.03)
McFadden R^2	0.001	0.288	0.344	0.340	0.399
Investors	123	123	90	81	53

Notes: Investor-weighted conditional logit over the pre-survey sales of *SALE*-condition investors whose recalled trade matches. Columns 1–4 use the fuzzy match, column 5 the strict match; investor counts (choice occasions) are reported at the foot of each column. *Still held*: a position remains in the stock at the survey date (a partial sale) rather than a full liquidation. *Holding-period return*: the average-cost realized return from purchase to sale; *Log holding period*: days from first purchase to the sale. *Return since / before trade*: the stock's return after the sale and over the 91 days before it. Standard errors in parentheses. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

ment decision or whether they represent qualitatively different mental accounts of trading experiences. Using both randomized recall prompts and within-investor comparisons, we document systematic asymmetries in narrative content, form, and time orientation across purchases and sales. Purchase narratives represent trades as forward-looking investment theses centered on company fundamentals and future prospects. Sale narratives instead represent trades as completed decisions centered on prices, trading rules, and implementation.

4.1 Textual Analysis of Narratives

The narratives are open text, and three established approaches can be used to analyze such text (Grimmer and Stewart, 2013; Gentzkow, Kelly, and Taddy, 2019). Dictionary methods apply predefined word lists mechanically and are transparent and fully replica-

ble, but ignore linguistic context (Tetlock, 2007; Loughran and McDonald, 2011). Human coding incorporates context but is costly and subject to coder disagreement (Krippendorff, 2019). Large language models (LLMs) also incorporate context and are increasingly used for text annotation in economics (Gilardi, Alizadeh, and Kubli, 2023; Korinek, 2023), but their predictions are less transparent and require validation.

We use all three approaches. The deterministic dictionary provides our baseline measures precisely because of its transparency and replicability. Human coding and LLM annotation serve as validation and robustness checks and are provided in Appendices I and J. Agreement across the three approaches increases confidence that the findings do not depend on a particular annotation methodology. All analyses are conducted on the original German text.

Work on economic narratives has so far focused on content, classifying what open-ended accounts are about (Stantcheva, 2023; Andre et al., 2024b). We extend this approach and measure each narrative on three dimensions: its *content* (what the investor talks about), its *form* (what the account is made of, and how it is told), and its *temporal orientation* (whether the language looks forward or backward).[16]

Content features. Our main content variables come from the deterministic word-list coder. Each feature is defined by a fixed list of German tokens, stems, and short phrases, and a narrative receives a value of one for a feature if it contains at least one corresponding entry.[17] The coder produces four core content features. *Fundamentals* captures business-level reasons for the trade (earnings, valuation, dividends, the business model, growth prospects, company quality). *Price/Rule* captures price signals and trading rules (past price movements, gains and losses, stop-losses, profit-taking, target prices, rule-based exits). *Recommendation* captures external attribution (analysts, advisors, friends, family, financial media, or social-media tips). *Circumstance* captures non-return-based motives (liquidity needs, rebalancing, taxes, life-stage considerations, and retirement saving, as

[16]Word-list features are based solely on keyword occurrence and do not distinguish between affirmative and negated statements.

[17]The coder is implemented in Python: text is lower-cased, then matched by stem-prefix for inflected German nouns and verbs and by substring for multi-word phrases. The complete word lists are provided in the replication package.

well as identity, familiarity, and social motives).[18]

These four categories span the broad classes of reasons investors give for trading: beliefs about the company, reactions to prices and rules, advice from others, and needs unrelated to returns. They are also comprehensive in practice.[19] Overall, 70% of the trade-specific narratives and roughly three-quarters of the general narratives contain at least one of the four features, and the uncovered remainder is disproportionately made up of very short texts.[20]

Narrative form. Narrative form is measured using deterministic text statistics inspired by autobiographical-memory research, which distinguishes event-specific from incidental detail (Levine, Svoboda, Hay, Winocur, and Moscovitch, 2002). We adapt this framework to trading narratives. We measure five types of *event-specific detail*: transaction mechanics (order types, quantities), anchoring in time or to market events, market context, quantitative anchors (prices, percentages, amounts), and the investor's contemporaneous mental state. *Incidental detail* counts sentences that add contextual information that is not specific to the recalled trade. Separate word lists identify episodic narration (past-tense auxiliaries and demonstratives referring to a specific episode) and habitual narration (e.g., "immer," "in der Regel"). From these we build three indices, fixed ex ante rather than extracted from the data: *Elaboration*, the standardized log word count; *Specificity*, standardized event-specific detail minus standardized incidental detail; and *Episodicity*, standardized episodic minus habitual markers—whether the account narrates a specific episode or states a standing policy.[21]

[18]The features are not mutually exclusive: about a third of the narratives that mention any feature mention more than one. In the coder, *Circumstance* is implemented through nine sub-lists—liquidity, rebalancing, profit realization, taxes, long-term saving, emotion, identity, familiarity, and social motives; profit-realization vocabulary (e.g., "Gewinnmitnahme") typically also matches *Price/Rule* through its gain and loss stems, so such narratives register under both features.

[19]Three of the six dimensions that the LLM pipeline later induces from the text, without seeing our categories, map directly onto the four categories used (Appendix I).

[20]Among the trade-specific narratives, uncovered texts have a median length of 6 words against 11 for covered texts, and 44% of uncovered texts contain five or fewer words, against 27% of covered texts. The general narratives show the same pattern (median 8 vs. 14 words; 34–35% vs. 16–18% at five or fewer words).

[21]Standardization is within each design's estimation sample. The construction was set before any purchase–sale contrast was computed, and two independent derivations of the lists produce nearly identical results; a hypothesis-blind LLM instrument cross-validates the constructs (Appendix I).

Temporal orientation. Using the same word-list procedure, we code two temporal-orientation features. Each combines a grammatical component with a content component. *Future orientation* equals one if the narrative uses the future tense or mentions future developments (expectation and forecast stems, growth terms, long-horizon phrases); *Past orientation* equals one if it uses the past tense or refers to past periods or past performance (temporal adverbs, time-anchoring phrases).

4.2 Trading Narratives Under Randomized Recall

We first compare narratives about a specific recalled trade, having randomly assigned each respondent to describe either a stock purchase (*PURCHASE* condition) or a stock sale (*SALE* condition). Table 7 asks what purchase and sale narratives emphasize. Purchase narratives are more likely to discuss company fundamentals (38% versus 17%) and recommendations (10% versus 3%). Sale narratives instead emphasize prices and trading rules (57% versus 28%) and non-return-based circumstances (24% versus 15%). The differences are economically large, ranging from 8 to 29 percentage points, and all are statistically significant at the 1% level.

The narratives analyzed here describe the trades investors actually recalled, and Section 3 showed that recall is selective. Are the differences in Table 7 differences in representation, or differences in the trades that come to mind? Three findings suggest the former. First, controlling for the characteristics of the matched trade—current ownership, trade size, time since trade, and subsequent return—leaves the content differences essentially unchanged.[22] Second, we obtain the same qualitative patterns in respondents' general purchase and sale narratives (Section 4.3), which are not tied to any particular recalled trade. Third, the *Fundamentals*, *Price/Rule*, and *Circumstance* gaps keep their sign under conservative selection bounds following Lee (2009); the smallest gap, *Recommendation*, is not sign-identified under the worst case (Appendix H). At the same time, the selectivity

[22]Restricted to fuzzy-matched recalls (295 *PURCHASE*- and 144 *SALE*-condition respondents), the four content gaps are +19, −28, +8, and −18 percentage points (*Fundamentals*, *Price/Rule*, *Recommendation*, *Circumstance*; all $p < .01$), similar to the full-sample gaps. Pooling the matched recalls ($n = 369$) in a linear model, the direction coefficient moves little when the matched trade's characteristics are added: *Fundamentals* +20.6 to +19.2, *Price/Rule* −26.2 to −26.3, *Recommendation* +8.3 to +10.1, *Circumstance* −24.2 to −21.3 (all $p < .01$).

Table 7: Features of Purchase vs. Sale Narratives (Across Conditions)

Feature	*PURCHASE* (n = 472)	*SALE* (n = 334)	Δ	p
Panel A: Content features				
Fundamentals	38%	17%	+21	< .001
Price/Rule	28%	57%	−29	< .001
Recommendation	10%	3%	+8	< .001
Circumstance	15%	24%	−9	.001
Panel B: Narrative form				
Elaboration	+0.04	−0.06	+0.10	.155
word count	16.5	13.6	+2.9	.035
Specificity	−0.10	+0.14	−0.24	.004
event-specific detail	0.41	0.47	−0.06	.467
incidental detail	0.44	0.31	+0.13	.006
Episodicity	−0.20	+0.28	−0.48	< .001
episodic markers (per 100w)	2.27	4.32	−2.05	< .001
habitual markers (per 100w)	0.29	0.04	+0.25	.039
Panel C: Temporal orientation				
Future orientation	37%	22%	+15	< .001
Past orientation	16%	25%	−9	.002

Notes: Content and temporal features are binary indicators; values show the share of respondents whose narrative contains at least one word-list match. Δ is the percentage-point difference (*PURCHASE* cond. minus *SALE* cond.). In Panel B, Elaboration, Specificity, and Episodicity are pre-specified indices: the standardized log word count; standardized event-specific detail minus standardized incidental detail; and standardized episodic minus habitual narration markers (construction in Appendix I). The indented rows report each index's components in raw units. For index rows, Δ is the raw difference in index scores (not in percentage points). Future and Past orientation equal one if the narrative matches a fixed German word list combining grammatical markers (future- and past-tense auxiliaries) with content terms (expectation and forecast stems; references to past periods and past performance); the lists are provided in Appendix I. p-values from two-sample Welch's t-tests.

of recall is itself part of the phenomenon we study: investors form beliefs from the trades that come to mind, and those are precisely the trades described here.

The differences extend beyond content to how trades are described. Overall elaboration differs little between purchases and sales. Instead, the distinction lies in narrative structure. Sale narratives are more specific (−0.10 for purchases vs. +0.14 for sales, $\Delta = -0.24$, $p = .004$): they contain more transaction mechanics, quantitative anchors, and contemporaneous mental states, whereas purchase narratives contain more contextual framing. The largest difference is Episodicity (−0.20 for purchases vs. +0.28 for sales, $\Delta = -0.48$, $p < .001$). Although both conditions describe a specific recalled trade, sale narratives are told as completed episodes, while purchase narratives resemble broader investment theses about the stock.

Purchase narratives are also significantly more forward-looking and less backward-looking than sale narratives. Investors in the *PURCHASE* condition use future-oriented language in 37% of narratives, compared with 22% in the *SALE* condition ($\Delta = +15$ pp, $p < .001$). The pattern reverses for past-oriented language: 16% of purchase narratives reference the past versus 25% of sale narratives ($\Delta = -9$ pp, $p = .002$). The forward-looking tilt is not a matter of grammar alone: purchase narratives are both more likely to be written in the future tense (12% vs. 6%, $p = .002$) and more likely to mention future prices, developments, or expectations (24% vs. 14%, $p < .001$).

4.3 Within-Person Comparison of Purchase and Sale Narratives

We next compare each respondent's narrative of why they generally buy stocks with their narrative of why they generally sell them. Because the comparison is within person, it holds constant individual differences in writing style, such as vocabulary and fluency. Every respondent answers two open-text questions: "What usually prompts you to buy a particular stock?" and "What usually prompts you to sell a particular stock?"

The content asymmetries closely parallel those in the randomized comparison (Table 8). Purchase narratives are more likely to emphasize *Fundamentals* (45% versus 22%) and *Recommendation* (17% versus 4%), whereas sale narratives instead emphasize *Price/Rule* (57% versus 33%) and *Circumstance* (28% versus 20%; all $p < .001$). The paper's central contrasts are robust to a Bonferroni correction for the eighteen feature tests reported in Tables 7 and 8.[23]

The within-person comparison reinforces the differences in narrative form. The same investor describes selling more specifically than buying (Specificity $\Delta = -0.22$, $p < .001$) and adopts a more episodic style when selling (Episodicity $\Delta = -0.18$, $p = .002$). Descriptions of selling center on concrete transactions, prices, and implementation, whereas descriptions of buying are framed more broadly around context and general investment

[23]The similarity in magnitudes across the two designs suggests that the differences are not driven by the particular trades investors happened to recall. Nor are they explained by narrative length: purchase and sale narratives differ by only about one word on average (17.9 versus 16.9 words), and the standardized Elaboration index does not differ significantly ($p = .105$), yet the content differences are virtually identical across the two designs. Under the correction, all four content contrasts remain significant in both designs, as do the episodicity contrasts and the between-person temporal-orientation contrasts.

Table 8: Features of General Purchase vs. Sale Narratives

Feature	Gen. Buy (n = 1,185)	Gen. Sale (n = 1,185)	Δ	p
Panel A: Content features				
Fundamentals	45%	22%	+23	< .001
Price/Rule	33%	57%	−23	< .001
Recommendation	17%	4%	+12	< .001
Circumstance	20%	28%	−8	< .001
Panel B: Narrative form				
Elaboration	+0.02	−0.02	+0.04	.105
word count	17.9	16.9	+1.0	.015
Specificity	−0.11	+0.11	−0.22	< .001
event-specific detail	0.28	0.44	−0.17	< .001
incidental detail	0.39	0.36	+0.02	.300
Episodicity	−0.09	+0.09	−0.18	.002
episodic markers (per 100w)	0.45	1.06	−0.61	< .001
habitual markers (per 100w)	0.35	0.45	−0.10	.187
Panel C: Temporal orientation				
Future orientation	36%	32%	+4	.037
Past orientation	5%	4%	+1	.159

Notes: Content and temporal features are binary indicators; values show the share of respondents whose narrative contains at least one word-list entry for the feature. Δ is the percentage-point difference (Purchase narrative minus Sale narrative), computed from unrounded values; McNemar's test compares discordant pairs within the same respondent. In Panel B, Elaboration, Specificity, and Episodicity are the pre-specified narrative-form indices of Table 7, standardized within the pooled general-purchase and general-sale sample; the indented rows report each index's components in raw units and decompose the indices and are not counted as separate tests; for the indices, Δ is the raw difference in index scores and p is from a within-person paired t-test. n = 1,185 respondents.

reasoning.

Temporal orientation differs less strongly than in the randomized comparison. General purchase narratives remain more likely to use future-oriented language than general sale narratives (36% versus 32%, Δ = +4 pp, p = .037), whereas the use of past-oriented language is statistically indistinguishable (5% versus 4%, p = .159). This weaker contrast is consistent with respondents describing their general investment approach in largely timeless terms, while still framing buying somewhat more in terms of future prospects.

5 Reconstruction of Trading Narratives

Sections 3 and 4 showed which trading experiences investors retrieve from memory and how they represent them. We now ask whether these representations remain stable

over time or are reconstructed as investors revisit the same trade. Because the follow-up survey re-elicits a narrative about the same recalled trade approximately ten months later, we observe how each investor's account evolves. We document two main findings. First, trading narratives persist over time but become substantially shorter and less detailed. Second, they are selectively reconstructed rather than merely forgotten: as subsequent outcomes unfold, investors retain different elements of the original narrative depending on whether those outcomes confirm or disconfirm the initial decision.

5.1 Decay and Persistence

Narratives become much shorter over time, but they do not disappear. Ten months after the initial survey, investors describe the same trade using fewer words, fewer sentences, and fewer coded features (Panel A of Table 9). Narrative length falls by 56%, the number of sentences by 31%, and the number of coded features by 30%. These patterns are consistent with ordinary forgetting over a ten-month interval. Some of the shortening may also reflect lower response effort in the shorter follow-up survey. Our analysis therefore focuses not on how much investors write, but on which elements of the original narrative survive and how they are represented.

Despite this compression, trading narratives remain recognizably related. Narrative length is moderately persistent within investors ($\rho = 0.40$), and five of the six word-list features exhibit significant within-person persistence. Overall, 53% of narratives that mention at least one coded feature at t_0 still contain at least one of the same features ten months later. Thus, although investors remember fewer details, they often retain the central elements of the original narrative.

The shortening of narratives reflects a loss of detail rather than a change in their overall structure. Elaboration falls mechanically with length (-0.79 for purchases, -0.55 for sales, both $p < .001$). By contrast, Specificity is essentially unchanged: event-specific and incidental details decline together, so the re-told narratives are no more or less context-heavy than the originals.

Table 9: Decay and Persistence of Trading Narratives

Panel A: Narrative volume and content (n = 478)

	t_0	t_1	Δ	Within-person ρ
Words	17.3	7.6	-9.7^{***}	0.40^{***}
Sentences	1.65	1.14	-0.50^{***}	0.13^{***}
Word-list features (of 6)	1.61	1.13	-0.49^{***}	0.14^{***}

Panel B: Narrative form (n = 477)

	Elaboration			Specificity			Episodicity		
	t_0	t_1	Δ	t_0	t_1	Δ	t_0	t_1	Δ
Purchase (n = 331)	+0.04	−0.75	-0.79^{***}	−0.05	−0.14	−0.09	−0.18	−0.31	−0.13
Sell (n = 146)	−0.08	−0.63	-0.55^{***}	+0.11	+0.05	−0.06	+0.42	+0.00	-0.42^{**}

Notes: Panel A reports means of the paired narratives at t_0 and t_1, the within-person change (Δ), and the within-person Spearman correlation between t_0 and t_1 values. In Panel B, Elaboration, Specificity, and Episodicity are the indices of Section 4.1, standardized on the original (t_0) narratives, with the follow-up narratives scored on the same scale. t_0 is the original narrative; t_1 is the narrative at the follow-up survey; one paired respondent lacks form scores (n = 477). Δ is the mean within-person change from t_0 to t_1; p-values are from paired t-tests. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

The substantive change concerns Episodicity. Sale narratives lose much of their episodic form over time, whereas purchase narratives change little. As a result, the difference in the Episodicity index between purchases and sales is roughly halved over the ten-month interval (from 0.60 to 0.31). This pattern suggests that the event-like quality of sale narratives fades more rapidly than the thesis-like structure of purchase narratives. Even as narratives become shorter, purchases continue to be represented primarily as investment theses rather than completed episodes.

5.2 Motivated Narrative Reconstruction

We examine whether investors selectively retain narrative elements depending on their investment outcome. We measure *retention* as whether a content feature mentioned in the original narrative at t_0 is mentioned again at t_1.[24] We measure the outcome of the recalled trade using the return on the stock between the respondent's main and follow-up survey dates—the window over which new information arrives between the two tellings.[25]

[24]The *acquisition* of content absent from the original narrative is examined in Appendix Table AIII.

[25]The realized return of the recalled stock is computed using daily stock-price data, which extend beyond the January 2024 end of the transaction records. A self-reported performance measure elicited in the follow-

Categorizing the trades based on whether that return is positive (Up) or negative (Down) and separating them by the type of the recalled trade yields four cells: Buy-Up, Buy-Down, Sell-Up, and Sell-Down. Appreciation confirms a purchase but disconfirms a sale; depreciation does the reverse.[26]

Table 10 shows that retention varies systematically with whether the subsequent outcome confirms the original decision. Among purchases followed by a price increase, *Fundamentals* exhibit the highest retention rate: 61% of respondents who mentioned fundamentals at t_0 mention them again at t_1. When a purchase is followed by a price decline, *Fundamentals* are retained less often (41%), while *Recommendation* has the highest retention rate (50%, based on twelve initial mentions). Thus, when a purchase performs well, investors disproportionately retain the fundamental thesis supporting the investment; when it performs poorly, external attribution (to an analyst tip or a friend's advice) becomes relatively more persistent.

The corresponding sale patterns point in the same direction. When a sale is followed by a price increase—an outcome that disconfirms the decision to sell—investors retain *Price/Rule* content at a rate of 61% and *Circumstance* content at 40%. These features provide reasons for the sale that remain defensible even though the stock subsequently appreciated. When the stock instead declines after the sale, thereby confirming the decision, *Price/Rule* again dominates retention, but at a lower level (41%), which is in line with ordinary decay of a price-anchored narrative.

Because individual cell comparisons are based on relatively small samples, we use a joint test to assess whether retention reorganizes with outcome confirmation. The test pools information across features while controlling for trade direction and narrative length and clustering standard errors by respondent. The feature-by-confirmation interaction is

up survey yields similar results.

[26] Our analysis uses the sign of the realized return rather than its magnitude. Theoretically, the hypothesis concerns whether subsequent performance confirms or disconfirms the original decision, consistent with evidence that many investor behaviors are organized around the gain–loss boundary (Shefrin and Statman, 1985; Odean, 1998; Barberis and Xiong, 2012). Empirically, a sign-based classification provides substantially greater statistical power than modeling continuous return interactions given the binary retention outcomes and sample size. Of the 478 paired respondents, 46 whose recalled stock cannot be matched to price data and 4 whose return over the window is exactly zero are excluded, leaving 428 respondents. Results are similar when returns within ±5% or ±10% of zero are excluded; Appendix C reports these and additional robustness checks.

Table 10: Retention of Content Features, by Outcome Confirmation

	Confirmed		Disconfirmed		
	Buy-Up	Sell-Down	Buy-Down	Sell-Up	Δ Conf. − Disc.
Panel A: Objective outcome (return on the recalled stock)					
Fundamentals	61% (90)	18% (11)	41% (46)	12% (17)	+23***
Price/Rule	28% (67)	41% (37)	32% (34)	61% (44)	−16**
Recommendation	19% (21)	0% (2)	50% (12)	0% (1)	−29
Circumstance	19% (27)	25% (12)	24% (17)	40% (25)	−13
Panel B: Recalled outcome (self-reported performance since the trade)					
Fundamentals	62% (82)	18% (11)	37% (49)	12% (16)	+26***
Price/Rule	38% (58)	41% (29)	29% (35)	52% (44)	−3
Recommendation	33% (21)	0% (3)	33% (15)	– (0)	−4
Circumstance	14% (29)	17% (6)	28% (18)	38% (24)	−19*

Notes: Entries are retention rates—the share of respondents who mentioned the feature at t_0 and continue to mention it at t_1—with the t_0 base in parentheses. Columns are grouped by whether the realized outcome confirmed the original decision: appreciation confirms a purchase (Buy-Up) and disconfirms a sale (Sell-Up); depreciation the reverse. The Δ column reports the difference between the pooled confirmed and pooled disconfirmed cells; significance is from Fisher's exact test. Acquisition rates are in Appendix Table AIII. In Panel A, the outcome is the sign of the stock's return between the respondent's two survey dates (cell sizes: Buy-Up 197, Buy-Down 96, Sell-Up 76, Sell-Down 59); in Panel B, it is the sign of the respondent's own report of how the stock performed since the trade (cell sizes: 176, 105, 73, and 44; respondents reporting an unchanged price, $n = 80$, are excluded from Panel B). Cells with fewer than five initial mentions (the two recommendation sale cells) are shown for completeness and not interpreted. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

strongly significant ($F = 6.70$, $p < .001$). Thus, the composition of what survives is systematically related to whether subsequent performance confirms or disconfirms the original trading decision.

Several findings suggest that this confirmation pattern is not an artifact of the survey design (Appendix C). Within trade direction, respondents whose stocks subsequently appreciated and those whose stocks depreciated are balanced on observable t_0 characteristics. The pattern also appears separately in both randomized follow-up elicitation orders, including the narrative-first arm in which respondents re-tell the narrative before the survey mentions the stock's subsequent performance. Finally, acquisition of new narrative content shows no comparable outcome contingency (Appendix Table AIII), suggesting that re-telling operates primarily through selective retention rather than the addition of new justifications.

The investor's perceived outcome provides a complementary benchmark. Panel B of Table 10 reconstructs the four cells using respondents' own reports of how the stock per-

formed. The same basic pattern emerges. *Fundamentals* are retained most strongly when a purchase is believed to have performed well (62%), whereas *Price/Rule* and *Circumstance* content is particularly persistent when the investor believes the stock rose after the sale (52% and 38%). The joint confirmation contrast is again significant ($p = .002$). The pattern is strongest when perceived performance is elicited before the narrative is re-elicited, consistent with narrative reconstruction responding to the outcome as the investor perceives it.

Two examples illustrate the pattern. A Buy-Down respondent whose holding declined by approximately 14% moved from a multi-sentence investment thesis at t_0—centered on a product launch, friends' enthusiasm, and the success of an earlier installment—to the phrase "Der Hype um das Videospiel [...]" ("The hype around the video game [...]") at t_1. The detailed firm-level thesis disappears, while external attribution remains. A Sell-Up respondent initially explained the sale by a prior price increase and a desire to reduce risk and take profits. Ten months later, after the stock had appreciated, the respondent acknowledged that holding would have been preferable but retained "Gewinnmitnahme" ("profit-taking") as the explanation for the original decision. Appendix D reports examples from all four outcome cells.

6 From Trading Narratives to Investor Beliefs and Behavior

The preceding sections documented which trades investors retrieve from memory, how they mentally represent those trades through narratives, and how those narratives are selectively reconstructed over time. A natural question is whether these cognitive representations are linked to investor beliefs and behavior. We show that investors' assessments of their own investment ability line up with remembered trading experiences—specifically, with which trades come to mind. Further, trading narratives capture economically meaningful differences in investment behavior. Investors who describe their decisions in terms of company fundamentals trade differently from investors who emphasize prices or trading rules, or who articulate no reason at all. These differences are reflected not only in trading styles, such as extrapolative buying and distinct approaches to selling, but also in

the performance of trading decisions.

6.1 Trading Narratives and Investor Confidence

Are cognitive representations linked to how investors form or update beliefs about their own investment ability? We compare investors' actual investment performance to their elicited self-assessed performance rank. Investors have little knowledge of their relative performance: mean self-assessed ranks vary narrowly between 47 and 52 across quintiles of actual performance, while actual ranks span 11 to 89, yielding a correlation of only 0.08 (Appendix L). On average, investors are well calibrated (mean overconfidence = −0.8 percentile points), but calibration is poor at the individual level (Appendix L). Investors in the bottom performance quintile overestimate their rank by 37 percentile points, whereas those in the top quintile underestimate it by 38. The standard deviation of overconfidence is more than 35 percentile points (Appendix Table AXIII), indicating substantial heterogeneity in self-assessments.[27]

If investors do not infer their relative ability from realized performance, what information do they use instead? Our evidence points to remembered trading experiences. Table 11 revisits the *FREE* condition, where respondents recalled whichever trade came to mind. Investors who spontaneously recall a purchase assess their own performance about 9 percentile points higher than investors who recall a sale, controlling for their actual share of purchases and trading activity. At the same time, buy-recallers have somewhat lower realized performance. As a result, they are approximately 18 percentile points more overconfident. The type of trade that memory retrieves therefore predicts investors' beliefs about their own investment ability. The association is not an artifact of elicitation: randomly assigning respondents to recall a purchase rather than a sale leaves confidence unchanged (−1.9 points, $p = .19$), so it reflects which trade memory volunteers, not the act of recalling it.[28] One interpretation is that self-assessments sample from what memory

[27] Overconfidence increases with trading activity, directionally consistent with overconfidence-based accounts of excessive trading (Barber and Odean, 2000, 2001), though the association is marginal under the daily measure (Spearman $\rho = 0.06$, $p = .10$; $\rho = 0.08$, $p = .02$ under the static variant).

[28] Among respondents for whom this is their main brokerage account, where the actual rank is best measured, the estimate is similar (+16.2, s.e. 7.3, $n = 184$). Appendix M asks whether the association varies with the narrative told about the recalled trade; suggestively, it is concentrated among respondents whose

makes available: a still-held purchase is an open, forward-looking position whose thesis has not yet been falsified, whereas a recalled sale carries a realized verdict. An investor whose memory serves up open theses rather than closed outcomes may thus maintain a more favorable self-image than the record supports.

The form of the narratives—though not their content—also tracks confidence. Investors who provide longer trading narratives assess their own investment ability more pessimistically (−3.6 percentile points per standard deviation of log narrative length, $p < .001$), whereas none of the four narrative-content features predicts self-assessment (joint test $p = .89$). Longer narratives are also associated with slightly lower actual ranks (−2.2 percentile points, $p = .06$): because length lowers self-assessed and actual ranks together, it predicts confidence rather than miscalibration. Consistent with this interpretation, investors who retain more trade-specific detail between the two survey waves subsequently revise their confidence upward (+1.90 percentile points per standard deviation of retained specificity, $p = .003$).

6.2 Trading Narratives and Investment Decisions

Do the narrative types documented in Section 4 predict how investors actually trade? Respondents' general purchase and sale narratives provide concise descriptions of how they think about trading, while the administrative records contain their complete trading histories, comprising approximately 104,000 stock trades (about 88,000 after aggregating same-day trades in the same stock). We examine whether these narratives are associated with differences in trading style and trading performance. The analysis is cross-sectional and based on a single market period.[29]

Table 12 relates trading narratives to trading styles on both sides of the trade. Panel A examines extrapolative buying (Barberis, Greenwood, Jin, and Shleifer, 2018; Liao, Peng, and Zhu, 2022), a well-documented pattern in retail buying behavior. Investors whose purchase narratives emphasize company fundamentals are the most contrarian: they are

narratives are short and unelaborated.

[29]Post-trade performance is market-adjusted (the traded stock's total return minus the same-window return of the DAX performance index; alternative benchmarks in Appendix N). Investor-level measures require at least five priced purchases (or sales); construction details are in Appendix N.

Table 11: Recall of a Stock Purchase, Confidence, and Overconfidence

	(1) Confidence	(2) Self-assessed percentile	(3) Actual percentile	(4) Overconfidence
Recall purchase (=1)	7.60**	8.70**	−9.37	18.07***
	(3.38)	(3.94)	(5.86)	(6.89)
Share of purchases	−16.38***	−7.40	−6.56	−0.84
	(5.49)	(7.05)	(10.03)	(11.59)
Log trade count	1.17	1.05	−2.13	3.17*
	(0.79)	(1.09)	(1.29)	(1.65)
Demographic FEs	Yes	Yes	Yes	Yes
N	343	254	254	254
R^2	0.208	0.202	0.129	0.161

Notes: OLS with HC1 standard errors (in parentheses). The sample is restricted to respondents in the *FREE* condition; columns (2)–(4) further require a computable actual percentile rank. Dependent variables: columns (1) and (2), the respondent's self-assessed percentile rank (= 100− *Fraction better clients*, 0–100), with column (1) estimated on the full *FREE*-condition sample and column (2) on the subsample with a computable actual rank; column (3), the actual percentile rank of the respondent's twelve-month portfolio return among the broker's clients (Figure 4); and column (4), overconfidence, their difference. *Recall purchase* equals one if the respondent recalled a stock purchase; zero if a sale. *Share of purchases* is the historical fraction of stock trades that were purchases. Demographic FEs absorb gender, age, education, risk tolerance, finance knowledge, and marital status. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

significantly less likely than other investors to buy stocks after large recent price increases (a price increase of 10% or more), and their purchases follow the largest prior price declines (a median pre-purchase return of −5.8%). By contrast, investors whose narratives contain none of the four content features are the least contrarian and exhibit the strongest tendency to chase recent winners. Price- and rule-based narratives, despite their emphasis on prices, are not associated with more extrapolative buying.

Sale narratives likewise distinguish selling styles (Panel B). Investors whose sale narrative uses profit-taking vocabulary realize gains at nearly double the rate of other investors, and investors with price/rule narratives disproportionately sell after price run-ups. Investors whose narratives are more episodic hold positions roughly twice as long before selling—winners and losers alike. These narrative differences predict selling styles, but not selling skill: once returns are market-adjusted, no narrative feature is robustly associated with the subsequent performance of sold stocks.

Beyond trading style, narratives are associated with the quality of investors' trades.

Table 12: Trading Narratives and Trading Styles

Panel A: Extrapolative buying (by purchase narrative; investor level)

		Share of buys after +10%		Pre-purchase return (3m, %)	
Narrative mentions. . .	*N*	share	vs. rest	median	vs. rest
Fundamentals	318	0.20	−0.04***	−5.8	−2.2**
Price/Rule	246	0.20	−0.02**	−5.5	−1.4
Recommendation	117	0.21	−0.01	−4.4	+0.2
Circumstance	140	0.20	−0.01	−4.3	+0.3
None of the four	152	0.29	+0.09***	−2.2	+3.2**
All	714	0.21		−4.5	

Panel B: Selling styles (by sale narrative; investor level)

	N		Group median		
Marker: measure	with	without	with	without	Diff.
Profit-taking: prop. of gains realized	30	331	0.21	0.12	+0.09**
Price/Rule: sales after +10% run-up (share)	251	197	0.29	0.24	+0.05**
Episodic: holding days, sold winners	33	300	88	43	+45**
Episodic: holding days, sold losers	33	303	94	44.5	+49.5***

Notes: Panel A: investor-level measures of extrapolative buying—the share of the investor's purchases following a +10% return over the preceding three months, and the median return of the purchased stock over the three months before the purchase (winsorized at the 1st/99th percentiles)—reported as group medians by whether the general purchase narrative mentions the feature; investors with at least five priced purchases. Panel B: investor-level selling-style measures by whether the general sale narrative carries the marker: the proportion of gains realized (average-cost basis, following Odean 1998; investors with at least five gain and five loss opportunities), the share of sales following a three-month return above +10% (at least five priced sales), and the median holding period, in days, of sold winners and losers (at least five classified sales). *Profit-taking vocabulary* and *price/rule* are word-list flags on the general sale narrative; the *episodic register* flag equals one if the Episodicity index (Section 4.1) on the same narrative is strictly above the sample median; the median is zero, so the flag selects the most episodic investors. Because the records begin in October 2021, the level of the proportion of gains realized is not comparable to complete-history estimates. All entries are group medians; significance from two-sided Mann–Whitney tests of feature-present against feature-absent investors. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Following Odean (1999), we compare the subsequent performance of purchased and sold stocks over the three months following each trade; the classic finding for U.S. retail investors is that this differential is negative—the stocks they buy subsequently underperform the stocks they sell. In our sample, purchases and sales perform similarly, implying little information in the average trading decision. However, substantial heterogeneity emerges across narrative types (Table 13). Investors whose narratives emphasize company fundamentals are the only group whose trades predict future returns: the stocks they buy outperform those they sell by 1.9 percentage points over the subsequent quarter ($p = .001$). By contrast, investors whose narratives contain none of the four content features exhibit

Table 13: Trading Narratives and the Informativeness of Trades

	Post-trade 3m return, DAX-adj. (%)					
Narrative mentions...	Purchases	Sales	Purchases−Sales	p	N purchases	N sales
Fundamentals	−2.4	−4.3	+1.9***	.001	23,460	10,671
Price/Rule	−3.1	−3.6	+0.5	.342	19,134	21,970
Recommendation	−2.2	−1.7	−0.5	.694	5,669	1,015
Circumstance	−1.5	−2.5	+1.1	.227	11,030	11,645
None of the four	−4.3	−2.8	−1.5	.171	10,209	9,043
All trades	−3.2	−3.5	+0.3	.124	49,122	36,931

Notes: Trade-level mean returns of the traded stock over the three months following the trade, in excess of the DAX performance-index return over the same window, winsorized at the 1st/99th percentiles; pre-survey trades only. Rows classify trades by whether the respondent's general purchase (for purchases) or sale (for sales) narrative mentions the feature; rows overlap. *Purchases−Sales* is the return differential between purchased and sold stocks from a trade-level regression with standard errors clustered by investor; zero for uninformed trading. The fundamentals differential exceeds that of the group mentioning none of the four features by 3.4 points ($p = .007$, interaction test) and survives two-way clustering by investor and calendar month ($p < .001$). For sales, a *lower* entry means the sold stock subsequently underperformed, i.e., selling was well timed. All groups trail the DAX on both legs, consistent with Barber and Odean (2000). Benchmark and horizon robustness in Table AXV; construction details in Appendix N. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

the opposite pattern, with purchases tending to underperform sales.

The difference in performance is economically meaningful: it appears after one month, increases over longer horizons, remains robust across alternative benchmarks, and survives two-way clustering by investor and calendar month (Appendix N).[30] Because fundamental reasoning is disproportionately applied to buying rather than selling (45% versus 22% of general narratives; Section 4), the representational asymmetry documented in this paper may help explain the buy–sell performance asymmetries observed in market data (Odean, 1998; Akepanidtaworn et al., 2023). Two findings argue against interpreting the association as reflecting only reverse causality from realized performance to self-description. First, measured at the trade level, the gap in extrapolative buying between the fundamentals and none-of-the-four groups is identical for trades executed before and after the survey (2.9 percentage points in both periods). Second, during the post-survey market upswing, investors whose narratives emphasize fundamentals continue to buy stocks that had underperformed over the previous twelve months, whereas investors

[30] Two qualifications remain: trades are weighted per decision rather than per euro (euro-weighting leaves the point estimate unchanged but widens its confidence band), and the comparison rests on a single market period.

whose narratives mention none of the four content features shift toward buying prior winners ($p < .01$).[31]

7 Conclusion

Investment experience has traditionally been measured through realized or recalled outcomes. This paper argues that experience also includes how investors remember and interpret the trading decisions that generated those outcomes. We study these remembered experiences directly through *trading narratives*: investors' own accounts of why they bought or sold a stock. Combining randomized recall prompts, within-person comparisons, repeated narrative elicitation, and administrative transaction records, we document how investors retrieve, represent, and later reconstruct their past trades.

First, retrieval is selective rather than exhaustive: purchases are substantially more likely than sales to be recalled, a pattern organized above all by current ownership, whereas sale recall depends almost entirely on the size and recency of the trade. Second, retrieved purchases and sales are represented through systematically different narratives: purchase narratives are forward-looking and centered on company fundamentals, whereas sale narratives are backward-looking and centered on prices, trading rules, and the details of the transaction—suggesting that purchases and sales occupy distinct mental representations rather than opposite sides of the same portfolio decision. Third, these narratives are not static: as the consequences of their trades unfold, they are selectively reconstructed in ways that justify the original decision. Finally, trading narratives are linked to future beliefs and behavior: investors who predominantly retrieve purchases rather than sales are more confident in their investment ability without performing better ex post, and the way investors describe their trades predicts how they buy, how they sell, and how informative their trades are—economically meaningful differences in how investors trade going forward.

Methodologically, our results illustrate how open-text narratives can be used to study

[31]The narratives are not purely stable traits either: mentions of fundamentals increase with realized past performance (39% in the bottom performance tercile versus 50% in the top), consistent with some self-serving adjustment in how investors describe their own trading style.

investor cognition directly rather than infer it solely from realized choices. Combining narrative elicitation with administrative data opens a new empirical approach to measuring the cognitive representations through which experiences shape beliefs and decisions. More generally, this approach may prove useful for studying memory, belief formation, and decision-making in many economic settings in which experiences are remembered, interpreted, and reconstructed over time.

Investment experience, then, is best understood as a cognitive representation rather than as a simple record of realized outcomes. The experiences that influence future beliefs and behavior are not identical to the experiences investors actually lived through: they are first filtered through selective retrieval, then organized into narratives, and later revised in light of subsequent outcomes. Experience effects in financial markets therefore depend not only on what happened, but also on what investors continue to remember and how they reconstruct it.

References

Akepanidtaworn, Klakow, Rick Di Mascio, Alex Imas, and Lawrence D. W. Schmidt, 2023, Selling Fast and Buying Slow: Heuristics and Trading Performance of Institutional Investors, *Journal of Finance* 78, 3055–3098.

Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai, 2021, Learning from Noise: Evidence from India's IPO lotteries, *Journal of Financial Economics* 140, 965–986.

Andersen, Steffen, Tobin Hanspal, and Kasper Meisner Nielsen, 2019, Once Bitten, Twice Shy: The Power of Personal Experiences in Risk Taking, *Journal of Financial Economics* 132, 97–117.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart, 2024a, Narratives About the Macroeconomy, Working Paper.

Andre, Peter, Philipp Schirmer, and Johannes Wohlfart, 2024b, Mental Models of the Stock Market, Working Paper.

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

Barber, Brad M., and Terrance Odean, 2001, Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment, *The Quarterly Journal of Economics* 116, 261–292.

Barber, Brad M., and Terrance Odean, 2008, All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *Review of Financial Studies* 21, 785–818.

Barber, Brad M., and Terrance Odean, 2013, The Behavior of Individual Investors, in *Handbook of the Economics of Finance*, volume 2, 1533–1570 (Elsevier).

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2018, Extrapolation and bubbles, *Journal of Financial Economics* 129, 203–227.

Barberis, Nicholas, and Wei Xiong, 2012, Realization utility, *Journal of Financial Economics* 104, 251–271.

Bénabou, Roland, and Jean Tirole, 2002, Self-Confidence and Personal Motivation, *Quarterly Journal of Economics* 117, 871–915.

Bénabou, Roland, and Jean Tirole, 2016, Mindful Economics: The Production, Consumption, and Value of Beliefs, *Journal of Economic Perspectives* 30, 141–64.

Bordalo, Pedro, John J. Conlon, Nicola Gennaioli, Spencer Y. Kwon, and Andrei Shleifer, 2023, Memory and Probability, *Quarterly Journal of Economics* 138, 265–311.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2020, Memory, Attention, and Choice, *Quarterly Journal of Economics* 135, 1399–1442.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick, 2009, Reinforcement Learning and Savings Behavior, *Journal of Finance* 64, 2515–2534.

Enke, Benjamin, Frederik Schwerter, and Florian Zimmermann, 2024, Associative Memory, Beliefs and Market Interactions, in *Journal of Financial Economics*.

Ferrario, Beatrice, and Stefanie Stantcheva, 2022, Eliciting People's First-Order Concerns: Text Analysis of Open-Ended Survey Questions, *AEA Papers and Proceedings* 112, 163–169.

Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack, 2024, Selective-memory equilibrium, *Journal of Political Economy* 132, 3978–4020.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as Data, *Journal of Economic Literature* 57, 535–574.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2021, Five Facts About Beliefs and Portfolios, *American Economic Review* 111, 1481–1522.

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli, 2023, ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks, *Proceedings of the National Academy of Sciences* 120, e2305016120.

Graeber, Thomas, Christopher Roth, and Florian Zimmermann, forthcoming, Stories, Statistics, and Memory, in *Quarterly Journal of Economics*.

Grimmer, Justin, and Brandon M. Stewart, 2013, Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, *Political Analysis* 21, 267–297.

Gödker, Katrin, Peiran Jiao, and Paul Smeets, 2025, Investor Memory, in *Review of Financial Studies*.

Haaland, Ingar, Christopher Roth, Stefanie Stantcheva, and Johannes Wohlfart, 2025, Understanding Economic Behavior Using Open-ended Survey Data, *Journal of Economic Literature* 63, 1244–1280.

Haaland, Ingar K, Christopher Roth, Stefanie Stantcheva, and Johannes Wohlfart, 2024, Measuring What Is Top of Mind, Working paper, National Bureau of Economic Research.

Jiang, Zhengyang, Hongqi Liu, Cameron Peng, and Hongjun Yan, 2025, Investor Memory and Biased Beliefs: Evidence From the Field, *Quarterly Journal of Economics* 140, 2749–2804.

Kaustia, Markku, and Samuli Knüpfer, 2008, Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions, *Journal of Finance* 63, 2679–2702.

Knüpfer, Samuli, Elias Rantapuska, and Matti Sarvimäki, 2017, Formative Experiences and Portfolio Choice: Evidence from the Finnish Great Depression, *Journal of Finance* 72, 133–166.

Korinek, Anton, 2023, Generative AI for Economic Research: Use Cases and Implications for Economists, *Journal of Economic Literature* 61, 1281–1317.

Krippendorff, Klaus, 2019, *Content Analysis: An Introduction to Its Methodology*, fourth edition (Sage Publications).

Lee, David S., 2009, Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects, *Review of Economic Studies* 76, 1071–1102.

Levine, Brian, Eva Svoboda, Janine F. Hay, Gordon Winocur, and Morris Moscovitch, 2002, Aging and Autobiographical Memory: Dissociating Episodic from Semantic Retrieval, *Psychology and Aging* 17, 677–689.

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

Loughran, Tim, and Bill McDonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

Malmendier, Ulrike, 2021, Experience effects in finance: Foundations, applications, and future directions, *Review of Finance* 25, 1339–1363.

Malmendier, Ulrike, and Stefan Nagel, 2011, Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?, *Quarterly Journal of Economics* 126, 373–416.

Malmendier, Ulrike, and Stefan Nagel, 2016, Learning from Inflation Experiences, *The Quarterly Journal of Economics* 131, 53–87.

Malmendier, Ulrike, and Jessica A Wachter, 2024, Memory of past experiences and economic decisions, in *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*, 2228–2266 (Oxford University Press).

Manski, Charles F, 2004, Measuring expectations, *Econometrica* 72, 1329–1376.

Odean, Terrance, 1998, Are Investors Reluctant to Realize Their Losses?, *Journal of Finance* 53, 1775–1798.

Odean, Terrance, 1999, Do investors trade too much?, *American Economic Review* 89, 1279–1298.

Shefrin, Hersh, and Meir Statman, 1985, The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance* 40, 777–790.

Shiller, Robert J., 2017, Narrative Economics, *American Economic Review* 107, 967–1004.

Stantcheva, Stefanie, 2021, Understanding Tax Policy: How Do People Reason?, *The Quarterly Journal of Economics* 136, 2309–2369.

Stantcheva, Stefanie, 2023, How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible, *Annual Review of Economics* 15, 205–234.

Strahilevitz, Michal A., Terrance Odean, and Brad M. Barber, 2011, Once Burned, Twice Shy: How Naive Learning, Counterfactuals, and Regret Affect the Repurchase of Stocks Previously Sold, *Journal of Marketing Research* 48, 102–120.

Tetlock, Paul C., 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance* 62, 1139–1168.

Wachter, Jessica A., and Michael Jacob Kahana, 2024, A Retrieved-context Theory of Financial Decisions, in *Quarterly Journal of Economics*.

Zimmermann, Florian, 2020, The Dynamics of Motivated Beliefs, *American Economic Review* 110, 337–361.

Appendices

A Trading Activity in the Sample

Table AI describes the amount of stock trading activity in the sample, reported as the cross-investor distribution (one value per investor, so a few highly active accounts do not dominate). Panel A covers stock trading over the observation window; Panel B reports the share of each investor's trading and portfolio that is in stocks. Activity is right-skewed—the median investor makes six stock trades and holds four distinct stocks over the window—and the stock share of the portfolio is dispersed, with mass at both extremes, though stocks are the median investor's largest single component (76% of portfolio value).

Table AI: Stock trading activity in the sample

	Mean	SD	P10	P25	Median	P75	P90
Panel A. Stock trading activity, Oct 2021–Jan 2024							
Stock trades	57	219	0	1	6	22	124
Purchases	33	115	0	1	4	14	74
Sales	25	105	0	0	1	7	52
Distinct stocks traded	13	31	0	1	4	11	33
Median trade size (EUR)	3,910	12,775	107	356	1,136	3,134	7,952
Total traded (EUR)	789,692	6,063,624	0	169	5,645	44,541	450,143
Panel B. Stock share of the investor's activity and portfolio							
Share of trades in stocks	0.54	0.42	0.00	0.06	0.58	1.00	1.00
Stock share of portfolio value	0.58	0.43	0.00	0.01	0.76	1.00	1.00
Stock share of positions	0.61	0.40	0.00	0.15	0.75	1.00	1.00

Notes: Cross-investor distribution (one observation per survey investor) of stock trading activity and portfolio composition; the mean, standard deviation, and 10th–90th percentiles are shown. Trades are aggregated to the same-day, same-stock, same-direction level, as in Section 2. Panel A spans all 1,524 survey investors, with a non-trader counted as zero, so the trade counts match the full-sample mean and median reported alongside Table 1; "Median trade size" is defined only for the 1,203 investors with at least one stock trade. "Stock trades", "Purchases", and "Sales" count transactions in the broker records over October 2021–January 2024, "Distinct stocks traded" counts ISINs, and "Total traded" sums stock-trade value. Panel B covers the 1,420 investors present in the January 2024 holdings snapshot: the share of the investor's trades that are in stocks, and the stock share of portfolio value (weighting positions by euro value) and of positions (weighting them equally).

B Summary Statistics of Elicited Narratives

Table AII reports summary statistics for the narratives respondents provide in the surveys. The average general narrative consists of 16–17 words, the average trade-specific narrative of 14 words. Response to the open-text questions is universal across treatments. The re-elicited narratives are significantly shorter, with 8 words on average. The length of the narratives motivates binary content measures and explicit richness measures rather than frequency-based statistics.

Table AII: Summary Statistics: Narratives

	N	Non-empty	Mean	SD	P10	P25	Median	P75	P90
General purchase narrative	1,524	1,520	17.2	18.0	3	6	12	22	37
General sale narrative	1,524	1,520	16.3	16.7	3	6	12	21	35
Trade-specific narrative (t_0)	1,524	1,509	14.4	18.4	2	4	8	18	32
PURCHASE condition	567	562	15.6	20.5	2	4	9	20	37
SALE condition	472	467	11.6	14.3	2	4	7	14	27
FREE condition	485	480	15.6	19.2	2	5	10	20	35
Re-elicited narrative (t_1)	478	478	7.6	8.5	1	3	5	9	18

Notes: Word counts of the non-empty narratives (a blank or "." response counts as empty). The distribution columns report the mean, standard deviation, and 10th–90th percentiles of words per narrative. Mean sentences per narrative: 1.7 (general buy), 1.6 (general sale), 1.5 (trade-specific), 1.1 (re-elicited). The general and trade-specific narratives cover all 1,524 main-survey respondents; the re-elicited narrative covers the 478 follow-up returners, who wrote longer t_0 narratives than non-returners (Appendix E).

C Robustness of the Outcome-Contingent Retention

This appendix probes the Section 5 finding that which narrative content is retained depends on the interaction of trade type (purchase or sale) and realized outcome. One concern is that our finding reflects features of the survey design rather than narrative reconstruction itself. Three possibilities are particularly relevant. First, realized outcomes are not randomly assigned, so respondents whose recalled stocks appreciated may differ systematically from those whose recalled stocks depreciated. Second, because the follow-up survey elicited the stock's subsequent performance and the re-told narrative in sequence, the elicitation order might mechanically influence which narrative elements are restated. Third, respondents might infer that the researchers are interested in the subsequent per-

formance of the stock and tailor their re-telling accordingly—an experimenter-demand effect.

The first concern does not explain the results. Within trade direction, respondents whose recalled stocks subsequently appreciated and those whose stocks depreciated are balanced on the observable characteristics of the original (t_0) narrative and the underlying trade. The confirmation contrast is also robust to permutation inference and to alternative definitions of outcome based on excluding returns close to zero (Appendix C).

The randomized elicitation order addresses the second and third concerns. The confirmation pattern appears separately in both elicitation-order treatments ($p = .040$ and .026). Importantly, it also appears in the narrative-first arm, in which respondents re-tell the narrative before the survey asks about the stock's subsequent performance. Thus, the pattern does not require that the realized outcome be made salient immediately before the narrative is produced.

The acquisition analysis provides a further check against experimenter demand. If respondents merely tailored their narratives to fit the observed outcome, they would primarily do so by adding outcome-consistent justifications. Instead, acquisition of new content shows no significant outcome contingency.[32] The evidence therefore suggests that narrative reconstruction operates primarily through selective retention of existing content rather than the addition of new outcome-consistent explanations.

Balance. The outcome split is as good as random conditional on the recalled trade: comparing respondents whose recalled stock subsequently appreciated with those whose stock declined, within direction, none of the nine t_0 characteristics differs at the 5% level in either trade direction (narrative length, the four content features, stock-return expectations, confidence, age, gender; 18 tests in total).

Follow-up block order. The follow-up survey randomly assigned respondents to one of two elicitation orders: one group reported how the stock had performed since the trade

[32] Acquisition, the mirror image of retention, exhibits the same direction-level asymmetries ($p < .001$) but no significant confirmation effect ($p = .47$). The only pooled contrast reaching conventional significance (circumstance, −6 percentage points, $p = .046$) is directionally consistent with the motivated account but should be interpreted cautiously given the insignificant joint test. Table AIII reports the full results.

and then re-told the narrative for the recalled trade (performance-first); the other re-told the narrative first and reported the performance afterwards (narrative-first). Block order matters for levels, and in a direction opposite to naive priming: performance-first respondents mention *Price/Rule* at t_1 substantially *less* often (25% vs. 34%, $p = .027$; retention 29% vs. 51%, $p = .002$). Having just reported the stock's performance, respondents appear not to restate price content in the narrative—a repetition-avoidance effect, and a caution for anyone using the t_1 narratives in isolation. It does not, however, affect the cell contrasts: the order arm is balanced across the four outcome cells (χ^2 $p = .247$), and the feature $\times$ direction $\times$ outcome interaction survives controlling for feature $\times$ order-arm effects ($F = 3.85$, $p = .010$). Estimated within each arm separately, the objective-outcome pattern replicates on both halves at half the power (confirmation contrast $p = .040$ and $p = .026$). The recalled-outcome pattern of Panel B of Table 10 is concentrated in the performance-first arm ($p = .004$ vs. $p = .27$ in the narrative-first arm)—the arm in which the believed outcome was elicited before the re-telling and therefore could not have been shaped by it, which is the cleaner sequence for measuring belief-contingent reconstruction.

Specification. The primary test in the text is the confirmation contrast, which replaces the four cells with a single confirmed-versus-disconfirmed indicator (Buy-Up and Sell-Down against the rest); it is the prediction the motivated-reconstruction account makes directly, and it is robust throughout ($F = 6.70$, $p = .0002$; within each order arm, $p = .040$ and $p = .026$; under the self-reported outcome, $p = .002$). The finer feature $\times$ direction $\times$ outcome parameterization ($p = .003$) is more demanding of the data and more fragile: it is estimated with controls for t_0 and t_1 narrative length, and because t_1 length is itself an outcome of the re-telling, we report the sensitivity—without the t_1 control it weakens ($p = .069$ with the t_0 control only, $p = .057$ with none). Permutation inference—reshuffling realized outcomes across respondents within direction 500 times—yields $p = .042$. Dropping any single cell leaves it significant (p from .012 to $< .001$). The acquisition-side null is unchanged by adding a control for the number of t_0 features ($p = .478$).

Acquisition. Table AIII reports acquisition rates in the layout of Table 10. As the joint tests indicate (acquisition does not vary significantly by realized outcome, $p = .47$), the pooled differences are small; the only one to reach significance is circumstance, acquired more often after disconfirming outcomes (−6 pp, $p = .046$), directionally in line with the motivated account but read as suggestive given the joint null. The direction-level pattern—price/rule and circumstance acquired on the sell side, fundamentals on the buy side—mirrors the retention results.

Table AIII: Acquisition of Content Features, by Outcome Confirmation

	Confirmed		Disconfirmed		
	Buy-Up	Sell-Down	Buy-Down	Sell-Up	Δ Conf. − Disc.
Panel A: Objective outcome					
Fundamentals	21% (107)	6% (48)	20% (50)	14% (59)	+0
Price/Rule	15% (130)	36% (22)	16% (62)	47% (32)	−8
Recommendation	5% (176)	2% (57)	4% (84)	3% (75)	+1
Circumstance	5% (170)	9% (47)	10% (79)	16% (51)	−6**
Panel B: Recalled outcome					
Fundamentals	20% (94)	9% (33)	23% (56)	16% (57)	−2
Price/Rule	14% (118)	40% (15)	17% (70)	38% (29)	−6
Recommendation	5% (155)	0% (41)	1% (90)	4% (73)	+1
Circumstance	7% (147)	11% (38)	7% (87)	16% (49)	−2

Notes: Entries are acquisition rates—the share of respondents who did not mention the feature at t_0 but did so at t_1—with the t_0 non-mentioner base in parentheses; layout and conventions as in Table 10. The Δ column reports the pooled confirmed-minus-disconfirmed difference (Fisher's exact test). Joint tests: acquisition differs by feature and direction ($p < .001$) but not by realized outcome ($p = .47$). $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Tone. Re-tellings also shift in tone. Table AIV codes each narrative for hedging words ("vielleicht," "eventuell," "unsicher") and assertive words ("sicher," "klar," "definitiv") and tracks, by cell, the change in each and the hedge *share*—hedges over hedges plus assertions, which is invariant to the overall shrinkage of the narratives. As narratives shrink, mentions of both kinds fall, but assertions fall faster, so the hedge share rises in every cell (by 3 to 10 percentage points). Sell-Up is the only cell in which hedging rises absolutely (+2.6 pp) while assertions barely move: disconfirmed sales are re-told in a distinctly more tentative register, echoing the content result. These shifts are descriptive; we report no tests on them.

Table AIV: Tone change by outcome cell

Cell	n	Δ Hedge	Δ Assert	Hedge share $t_0 \to t_1$	Δ share
Buy-Up	197	−11.7	−18.3	40% → 43%	+3
Buy-Down	96	−14.6	−16.7	53% → 62%	+9
Sell-Up	76	+2.6	−1.3	40% → 50%	+10
Sell-Down	59	−13.6	−15.3	50% → 60%	+10

Notes: Hedge and assertion levels are the percentage of narratives containing at least one hedging or assertive word from fixed German word lists. Hedge share equals hedges/(hedges+assertions); Δ values are in percentage points. Sample and cells as in Table 10.

Outcome classification. The sign-based classification is not sensitive to the cutoff. Excluding returns within ±5% or ±10% of zero (dropping 47 and 93 respondents) leaves the headline rates essentially unchanged: Buy-Up fundamentals retention is 61%, 62%, and 62% under the three definitions; Sell-Up price/rule retention is 61%, 60%, and 60%, and its acquisition 47%, 47%, and 50%; Sell-Up circumstance retention is 40%, 40%, and 38%; Sell-Down price/rule retention is 41%, 34%, and 41%. The only unstable figure is the Buy-Down recommendation rate, whose small t_0 base falls from twelve mentions to eight. Magnitude points the same way as sign: above versus below the median gain (+31%), fundamentals retention among buys is 67% versus 55%, and Sell-Up price/rule acquisition is 55% versus 33%—directionally consistent with outcome-driven reconstruction strengthening with the dose, but underpowered at these cell sizes (Fisher $p = .23$–$.29$), and we treat the gradients as suggestive.

D Example Narrative Pairs

This appendix reproduces four original–re-told narrative pairs, one from each outcome cell of Section 5. Selection is mechanical, to guard against curation: within each objective-outcome cell, one pair was drawn at random from pairs with at least ten words at t_0 and at least three words at t_1, excluding the examples quoted in the text. The German originals are verbatim, including typographical errors; stock and issuer names are replaced by bracketed placeholders; translations are ours.

Buy-Up (return +52%). t_0: "Warren Buffet [sic] hat ebenfalls investiert, ich habe vorher zudem über keine Position in Lateinamerika / Südamerika verfügt. Neben den Wachtsumszahlen [sic] haben mich auch die aktuellen Zahlen darin bestätigt einen Kauf durchzuführen" ("Warren Buffett has also invested; moreover, I previously held no position in Latin America / South America. Besides the growth figures, the current numbers also confirmed my decision to buy."). t_1: "Günstige Bewertung, Kauf von Berkshire, Interesse am Bankenmarkt, in Südamerika war ich bisher nicht investiert" ("Cheap valuation, Berkshire's purchase, interest in the banking market, not previously invested in South America."). A confirmed buy thesis survives nearly intact: authority, diversification, and numbers at t_0; the same pillars, compressed, ten months later.

Buy-Down (return −29%). t_0: "aussichtsreiche Aktie die meiner Meinung nach zu Unrecht abgeschossen wurde, ich habe hier Bestandspositionen verbilligt." ("A promising stock that in my view was unjustly shot down; I averaged down my existing positions here."). t_1: "Ein gutes Unternehmen zu einem günstigen Kurs zu kaufen." ("Buying a good company at a cheap price."). A specific, first-person episode becomes a general policy stated in the habitual register, with the intervening loss absent from the re-telling.

Sell-Up (return +2%). t_0: "Der Aktienkurs war an einem Allzeithoch und es gab Meldungen zu Gewinnwarnungen." ("The share price was at an all-time high and there were reports of profit warnings."). t_1: "Der Kursanstieg ubd [sic] eher schlechtere Aussichten." ("The price rise and rather worse prospects."). The price/rule justification of the sale survives re-telling almost verbatim.

Sell-Down (return −7%). t_0: "Wurden mir ursprünglich aus Aktienanleihen angedient und hatten seitdem eine sehr gute Performance. Ich habe den Gewinn realisiert und wieder in Aktienanleihen auf [Emittent] mit entsprechenden Puffer investiert." ("These were originally allotted to me from reverse convertibles and had since performed very well. I realized the gain and reinvested in reverse convertibles on [issuer] with an appropriate buffer."). t_1: "Verkauf und gleichzeitiger Kauf einer Aktienanleihe auf den

gleichen Emittenten" ("Sale and simultaneous purchase of a reverse convertible on the same issuer."). The confirmed sale is re-told as a pure transaction record—mechanics and nothing else.

E Selection into the Follow-up Survey

Participation in the follow-up survey is voluntary: of the 1,524 main-survey respondents, 914 agreed to be re-contacted and 478 completed the follow-up. Table AV compares the 478 returners with the remaining 1,046 respondents on demographics, survey engagement, elicited beliefs, and actual trading activity.

Selection is present but modest. Returners are on average about three years older and seven percentage points more likely to be male, and they are more engaged with the survey: in the main survey they wrote open-text narratives that were roughly a quarter longer (about twelve more words in total). They also report somewhat more optimistic return expectations and somewhat higher confidence. Reassuringly, however, returners and non-returners do not differ in their actual trading activity over the sample period, whether measured by the number of purchases, sales, or total trades. A joint logit of follow-up participation on all observables is statistically significant (pseudo-$R^2 = 0.10$), driven mainly by age, market-return expectations, and confidence, but it explains only a small share of the variation in who returns.

Table AV: Selection into the Follow-up Survey

	Non-returners (n = 1,046)	Returners (n = 478)	Difference	S.E.
Panel A: Demographics				
Male (= 1)	0.81	0.88	+0.07***	(0.02)
Age (years)	43.9	46.8	+2.9***	(0.8)
Education (1–9)	6.18	6.32	+0.13	(0.11)
Main account at broker (= 1)	0.66	0.62	−0.05*	(0.03)
Panel B: Survey engagement (main-survey narrative length, words)				
General purchase narrative	15.8	20.1	+4.3***	(1.1)
General sale narrative	15.4	18.2	+2.8***	(0.9)
Trade-specific narrative	13.0	17.3	+4.3***	(1.1)
All narratives	44.0	55.6	+11.7***	(2.5)
Panel C: Beliefs (expectations 1–10; confidence 1–7)				
Exp. stock return (1y)	6.36	6.80	+0.44***	(0.14)
Exp. portfolio return (1y)	6.60	6.92	+0.32***	(0.11)
Exp. market return (1y)	6.04	6.35	+0.31***	(0.09)
Confidence, stock return	3.79	4.03	+0.24***	(0.09)
Confidence, portfolio return	3.77	3.98	+0.21**	(0.08)
Confidence, market return	3.38	3.56	+0.18**	(0.08)
Panel D: Actual trading activity (2023, broker records)				
Number of purchases	31.4	34.8	+3.4	(13.4)
Number of sales	26.1	29.6	+3.5	(14.9)
Number of trades	57.5	64.4	+6.9	(28.2)

Notes: The sample is the 1,524 main-survey respondents; "Returners" completed the follow-up survey (n = 478) and "Non-returners" did not (n = 1,046). "Difference" is returners minus non-returners and "S.E." its standard error, from Welch's t-tests (continuous variables) or two-proportion z-tests (binary variables). Expectations are on a 1–10 categorical scale and confidence on a 1–7 scale. A joint logit of follow-up participation on all listed variables yields pseudo-R^2 = 0.10 (likelihood-ratio $p < .001$); the significant predictors are age, market-return expectations, and portfolio-return confidence. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

F Recall Asymmetry: Robustness Across Subsamples

The purchase tilt in *FREE*-condition recall (Section 3) is stable across subsamples. Table AVI reports the share of purchases among recalled trades and the investor-weighted average of respondents' own purchase share, split by whether the respondent's main brokerage account is at this broker and by whether they report trading in the four weeks before the survey.

Table AVI: Recall asymmetry across subsamples (*FREE* condition)

Subsample	*N*	Recalled purchases	Actual purchases	Difference (pp)
All respondents	379	80.2%	70.4%	+9.8
Main account at this brokerage	242	78.5%	70.8%	+7.7
Main account elsewhere	137	83.2%	69.6%	+13.6
Traded in the last four weeks	207	81.6%	67.7%	+13.9
Did not trade in the last four weeks	172	78.5%	74.0%	+4.5

Notes: The sample is the 379 *FREE*-condition respondents who recalled a stock purchase or sale. "Recalled purchases" is the share of respondents whose recalled trade was a purchase; "Actual purchases" is the investor-weighted average of each respondent's own purchase share in the transaction records (one value per investor, matching the recalled share); "Difference" is the former minus the latter, in percentage points. This is the same per-investor benchmark used in the main text (Section 3).

A recalled trade that we cannot find in the records need not be a false memory: some respondents hold accounts at other brokers, so an unmatched recall could be a real trade we do not observe. Table AVII repeats the accuracy table of Section 3.2 on the subsample of respondents who report that this broker is their main account, for whom unobserved trading is far less likely. Match rates rise modestly, consistent with a small role for outside accounts, but the central pattern is undisturbed: purchase illusions still far exceed sale illusions (40% versus 17%), and most purchase illusions still match no recorded trade of that direction on any date. The recall asymmetry is therefore not an artifact of trades placed elsewhere.

Table AVII: Accuracy of recall in guided recall conditions (main-account holders)

	PURCHASE cond.	*SALE* cond.
Answered survey	352	315
Panel A. Is the recalled trade real?		
Recalled a stock trade	289 (82.1%)	231 (73.3%)
Matched, fuzzy	201 (57.1%)	124 (39.4%)
Matched, strict	110 (31.2%)	68 (21.6%)
Panel B. Right sense of timing? (traded in last four weeks)		
Traded, reported	87 (95.6%)	68 (91.9%)
Traded, not reported (forgetfulness)	4 (4.4%)	6 (8.1%)
Not traded, reported (illusion)	105 (40.2%)	42 (17.4%)
Not traded, not reported	156 (59.8%)	199 (82.6%)
Panel C. What are the illusions?		
Named a stock	70 (66.7%)	36 (85.7%)
Matched, fuzzy	43 (41.0%)	17 (40.5%)
Matched, strict	18 (17.1%)	6 (14.3%)

Notes: Table 4 recomputed on respondents who report this broker as their main account (Demographics: main account = 1; $N = 352$ *PURCHASE*, 315 *SALE*). Panel and match definitions are identical to Table 4. Percentages in Panel A are of the column total; percentages in Panel B are of the relevant traded / not-traded subtotal; percentages in Panel C are of the illusion count.

Section 3.1 rules out two mechanical explanations for the purchase tilt in *FREE*-condition recall; Table AVIII reports the supporting benchmarks. Panel A compares the observed purchase share among recalled trades to two random-recall nulls drawn from each investor's own trades: a uniform null, and a recency-weighted null in which each trade is drawn in proportion to its empirical recall propensity (matched recalled trades over available trades within its age bin). The recency weighting is steep, but even the recency-weighted null reaches only 71.0%, far short of the observed 81.0% ($p < .001$). The reason is in Panel B: the purchase share of investors' available trades barely moves across time since trade, so tilting recall toward recent trades does little to raise the benchmark. Automated saving is ruled out separately, as savings plans at this broker are confined to ETFs and mutual funds and the stock-trade record contains no savings-plan trades.

Table AVIII: Mechanical benchmarks for the purchase tilt in recall (*FREE* condition)

	Purchase share	
Panel A. Recalled vs. random-recall benchmarks		
Recalled trades (observed)		0.810
Random recall, uniform	0.725	[0.687, 0.763]
Random recall, recency-weighted	0.710	[0.673, 0.746]
Panel B. Purchase share of available trades, by time since trade (days)		
0–28		0.625
29–90		0.594
91–182		0.567
183–365		0.589
366–730		0.572
>730		0.570

Notes: The sample is the *FREE*-condition respondents linked to the stock-trade records. Panel A: "observed" is the purchase share among recalled trades; the two benchmarks are the mean purchase share under 10,000 simulated random-recall draws, one trade per investor from that investor's own available trades, under a uniform null and under a recency-weighted null in which each trade's draw probability is proportional to its empirical recall propensity (matched recalled trades divided by available trades, within its age bin). Brackets give the 2.5th and 97.5th percentiles of the simulated distribution; the p-value reported in the text tests the observed share against the recency-weighted null. Panel B reports the purchase share of investors' available trades within each trade-age bin (days before the survey), showing that the buy share is nearly flat across trade age. Savings-plan trades in the stock-trade record: none.

G Content Asymmetry Across Measurement Methodologies

Table AIX compares the purchase–sale differences of Section 4 across the three measurement methodologies, in both elicitation designs. Each cell reports the purchase-minus-sale difference under one methodology's own test; a dash marks a feature the methodology does not measure.

Table AIX: The content asymmetry across measurement methodologies

	Between-person (*PURCHASE* vs. *SALE*)			Within-person (general narratives)		
	Word list (Δ pp)	Human (Δ pp)	LLM (Cohen's d)	Word list (Δ pp)	Human (Δ pp)	LLM (Cohen's d)
Fundamentals	$+21^{***}$	$+20^{***}$	$+0.92^{***}$	$+23^{***}$	$+16^{***}$	$+0.64^{***}$
Price/Rule	-29^{***}	-14^{***}	—	-23^{***}	-12^{***}	—
Recommendation	$+8^{***}$	$+7^{***}$	$+0.21^{***}$	$+12^{***}$	$+13^{***}$	$+0.27^{***}$
Circumstance	-9^{***}	—	—	-8^{***}	—	—
Future orientation	$+15^{***}$	$+12^{***}$	$+0.88^{***}$	$+4^{**}$	-1	$+0.19^{***}$
Past orientation	-9^{***}	-2	—	$+1$	$+5^{***}$	—

Notes: Purchase-minus-sale differences by measurement methodology and design. *Word list*: the deterministic coder of Section 4.1 (percentage points; Welch's t between persons, McNemar within persons). *Human*: two trained coders applying the underlying coding manual, strict aggregation (a feature counts only if both coders agree); the mapped categories are corporate news (Fundamentals), stock price/return mentions (Price/Rule), recommendation followed (Recommendation), and the future- and past-reference categories (time orientation); the past category has low inter-coder agreement. *LLM*: hypothesis-blind continuous dimensions scored without observing the purchase/sale labels (Cohen's d; Welch's t / paired t); the mapped dimensions are fundamental-vs-technical (Fundamentals; a bipolar axis whose negative pole is price/technical reasoning), external influence (Recommendation), and forward-looking orientation (Future). A dash marks a feature the methodology does not measure separately: circumstance is measured by the word-list coder alone, the LLM pipeline scores no past dimension, and price/rule has no dedicated LLM dimension because the bipolar fundamental-vs-technical axis reports it on the Fundamentals row. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

H Bounding the Content Asymmetry Under Selective Recall

The between-person comparison of Section 4.2 conditions on recalling a trade, and respondents in the two conditions recall trades at different rates: 83.2% of *PURCHASE*-condition respondents report a trade against 70.8% in the *SALE* condition. Table AX bounds the four content gaps under worst-case selection, following the trimming logic of Lee (2009). Under the assumption that anyone who would recall a trade under the *SALE* prompt would also recall one under the *PURCHASE* prompt, the excess *PURCHASE* recallers amount to 15.0% of the *PURCHASE*-condition sample; reassigning their narratives entirely to one extreme of a binary feature gives sharp bounds on the *PURCHASE* mean among always-recallers, and hence on the gap. The fundamentals, price/rule, and circumstance gaps keep their sign under either extreme. The recommendation gap, the smallest of the four, is not sign-identified under worst-case selection, though the worst case requires

every excess recaller's narrative to omit (or contain) the feature simultaneously.

Table AX: Worst-case bounds on the content gaps under differential recall

Feature	Observed gap	Lower bound	Upper bound
Fundamentals	+21.3	+10.4	+28.0
Price/Rule	−29.1	−41.9	−24.2
Recommendation	+7.7	−8.1	+9.5
Circumstance	−9.4	−24.4	−6.8

Notes: Purchase-minus-sale gaps in percentage points. Bounds trim the *PURCHASE*-condition sample by the excess response share $p = (0.832 - 0.708)/0.832 = 0.150$, assigning the trimmed mass entirely to one (upper bound for buy-tilted features) or zero (lower bound) of each binary feature, under the monotonicity assumption that recall under the *SALE* prompt implies recall under the *PURCHASE* prompt.

I Measurement Appendix

Narrative-form dimensions. All narrative-form dimensions are computed in Python with no model and no human judgment in the loop, from fixed German word lists and counts documented entry by entry in the replication package. The construction was specified and frozen before any purchase–sale contrast was computed; vocabulary coverage was completed by a corpus recall audit that inspects only the frequency of unmatched tokens, never outcomes. Five categories count event-specific detail, adapting the internal-detail categories of the Autobiographical Interview (Levine et al., 2002) to the trade episode: *transaction mechanics* (order and execution vocabulary, quantities), *episode anchoring* (calendar and market-event references, e.g. months, years, "Corona"), *market context* (market, index, and sector vocabulary), *quantitative anchors* (tokens containing digits, percentage and currency marks), and *contemporaneous mental state* (first-person past-tense mental verbs and phrases). *Incidental detail* counts sentences of at least three tokens containing no trade-related term (count capped at five). *Episodic narration* counts past-tense auxiliaries and singular-episode demonstratives per 100 words; *habitual phrasing* counts habitual adverbs and phrases per 100 words. The three indices are Elaboration $= z(\log \text{word count})$, Specificity $= z(\text{event-specific detail}) - z(\text{incidental detail})$, and Episodicity $= z(\text{episodic}) - z(\text{habitual})$, standardized within each design's estimation

sample. Two independent derivations of the word lists—one drafted from the construct definitions alone, one reusing previously built project lexicons—produce nearly identical purchase–sale contrasts on all three indices.

LLM validation of the narrative-form constructs. A pinned model (`claude-sonnet-4-6`, temperature 0) scored all 4,549 narratives (trade-specific, general purchase, and general sale), in the original German, on the same construct definitions; the prompt contains the definitions only—no trade-direction labels and no hypotheses—and the output is schema-enforced. Agreement with the deterministic coder on the trade-specific narratives is $r = 0.60$ for total event-specific detail (0.78 for the quantitative category), Spearman $\rho = 0.55$ for episodicity, and $\rho = 0.63$ for hedging-related language, and the scoring is stable across models ($r = 0.73$–0.91 across dimensions against an independent annotator on a 160-narrative subsample). The LLM instrument reproduces the direction and significance of the narrative-form contrasts in both designs: event-specific detail is higher in sale narratives (-0.31, $p = .028$ between persons; -0.25, $p < .001$ within), and episodicity, scored directly on a 1–5 scale, is likewise sale-tilted (-0.73, $p < .001$ between; -0.06, $p < .001$ within). One dimension does not converge: incidental detail correlates at only $r = 0.04$ between the deterministic and LLM instruments (and weakly between all instrument pairs), so we base no claims on incidental detail alone; under the LLM instrument, which measures incidental content by semantic judgment rather than keyword absence, the specificity contrast is carried entirely by its event-specific leg.

Temporal-orientation word lists. The two temporal features are coded from the following fixed German lists; entries ending in $*$ match by stem prefix, and multi-word entries match as phrases.

Future	wird, werden, zukunft∗, künftig∗, potenzial∗, potential∗, erwart∗, hoff∗, glaube, glauben, prognose∗, aussicht∗, perspektive∗, dürfte, soll, sollte, wachstum∗, wachsen, steigen, steigend∗, fallen, fallend∗, in zukunft, langfristig positiv, ausgehe, ausgegangen, ausgeh∗, rechnete, rechnen, denke, davon ausgeh∗
Past	war, waren, hatte, hatten, damals, früher, letzten, letzte, vergangen∗, historisch, schon lange, schon immer, seit jahren, seit langer zeit, bisher, in der vergangenheit, vor jahren, vor langer zeit

The content-feature word lists are longer and are reproduced in the replication package together with the coding manual.

Splitting future orientation: tense versus content. The headline future list above mixes three kinds of entries: grammatical markers (the auxiliaries and modals *wird, werden, dürfte, soll, sollte*), content terms (expectation, forecast, and future stems), and growth and price-movement vocabulary (*wachstum∗, steigen, fallen, aussicht∗*). The refined split reported in Section 4.2 separates the first two. *Uses future tense* matches the future-tense forms of *werden* (werde, wirst, wird, werden). *Mentions future prices/expectations* matches content terms only, derived from the coding manual's future category: erwart∗, zukunft∗, zukünftig∗, künftig∗, potenzial∗, potential∗, prognose∗, gehe davon aus, rechne mit, glaube, denke, hoffe, langfristig∗. Growth and price-movement terms are assigned to neither component, so the two components refine the headline measure without partitioning it exactly; jointly they cover 78% of the trade-specific narratives the headline measure flags. In the between-person design the two components agree: purchase narratives are more likely both to use the future tense (12% vs. 6%, +6 pp, $p = .002$) and to mention future prices or expectations (24% vs. 14%, +10 pp, $p < .001$). In the general narratives of Section 4.3 the content component remains buy-tilted (25% vs. 19%, +6 pp, $p < .001$), while the tense component reverses (8% vs. 13%, −5 pp, $p < .001$)—not because general sale accounts are prospective, but because *wird* and *werden* also serve as the passive auxiliary in German, and general sale accounts are often phrased as standing rules of the form "wenn . . . , wird verkauft." In the general narratives the tense component therefore picks up rule-like

phrasing rather than temporal stance, and we rely on the content component there.

Human coding. Two trained coders independently applied the investor-focused coding manual (v6; reproduced in the replication package) to the general purchase and sale narratives and the trade-specific narrative of each of the 1,185 respondents in the within-person sample—3,555 texts in total—coding roughly 38 binary categories nested under five umbrellas (financial rationale, non-pecuniary rationale, variable mentioned, recommendation followed, time dimension). Coders read the original German narratives in a purpose-built coding interface; the manual and category labels are in English, with German-language glosses where the mapping matters (e.g., that "umschichten" is not per se diversification). Coders were blind to the study's hypotheses. The headline human-coding results use the strict aggregation, counting a category only when both coders agree. Table AXI reports coder prevalence and inter-coder agreement for the categories mapped to the paper's features. Agreement is substantial for the content categories; the past-reference category is an outlier ($\kappa = 0.20$), which is why the paper treats human-coded past orientation with caution. The manual spreads circumstance-type motives—liquidity needs, taxes, rebalancing, identity, familiarity—across its rationale umbrellas rather than collecting them in a single category, so the word-list circumstance feature, which consolidates these motives, has no one-to-one human-coded analog in Table AIX.

Table AXI: Inter-coder agreement on the mapped human-coded categories

Human-coded category	Coder A (%)	Coder B (%)	Cohen's κ
Corporate news (fundamentals)	22.7	21.5	0.74
Stock price/return mentions (price/rule)	56.5	40.7	0.64
Recommendation followed	8.5	8.8	0.85
Future reference (temporal)	26.7	18.9	0.62
Past reference (temporal)	27.9	5.6	0.20

Notes: Share of the 3,555 narratives flagged by each coder and Cohen's κ, for the human-coded categories mapped to the paper's features in Table AIX.

The hypothesis-blind LLM pipeline. The LLM dimensions were constructed in two stages, designed so that neither stage could see the purchase–sale comparison. In the induction stage, candidate dimensions were discovered from pooled narrative samples

across twelve independent runs (three random samples × two models × two prompt variants); dimensions that recurred across runs were promoted to a frozen rubric of six: fundamental vs. technical, external influence, forward-looking orientation, time horizon, emotional vs. rational, and quantitative specificity. A tautological trade-direction dimension that also recurred was explicitly excluded, and prompts were audited against a list of forbidden direction-revealing tokens. In the scoring stage, a pinned model (`claude-sonnet-4-6`) scored each of the 3,555 narratives, in the original German, on the frozen rubric, returning one continuous value in $[0, 1]$ per dimension; the prompt contains dimension definitions only—no trade-direction labels and no hypotheses. Table AXII reports validation: agreement with an independent model (`claude-opus-4-7`), test–retest stability, and correlation with the strict human coding. The three dimensions used in Table AIX (fundamental vs. technical, external influence, forward-looking orientation) are the ones with substantial human-coding correlations (ρ = 0.45–0.51); the remaining dimensions are reported for completeness.

Table AXII: Validation of the LLM dimensions

LLM dimension	Cross-model ρ (n = 200)	Test–retest ρ (n = 50)	Human coding ρ (n = 3,555)
Fundamental vs. technical	0.87	1.00	0.48
External influence	0.72	0.95	0.45
Forward-looking orientation	0.83	1.00	0.51
Time horizon	0.92	0.99	0.28
Emotional vs. rational	0.59	0.91	0.04
Quantitative specificity	0.55	0.98	0.00

Notes: Spearman correlations. *Cross-model*: primary scores (`claude-sonnet-4-6`) against an independent model (`claude-opus-4-7`) on 200 random narratives. *Test–retest*: two independent scoring passes of the primary model on 50 narratives. *Human coding*: correlation with the strict human-coded category mapped to each dimension, across all 3,555 narratives.

J Robustness of Narrative Measures

The content asymmetry is not an artifact of our measurement strategy. Appendix Table AIX repeats the purchase–sale comparisons under the three methodologies introduced in Section 4.1—the deterministic word-list coder, independent human coding by two trained coders, and the hypothesis-blind LLM scoring pipeline—in both designs. For the content core the picture is uniform: fundamentals and recommendation tilt toward buys, and price/rule toward sales, under

every methodology that measures them and in both designs; magnitudes differ across methods, but direction and significance do not. The forward-looking asymmetry is method-robust in the between-person design, where all three methodologies find purchase narratives more future-oriented. The past-orientation contrast is not: between persons it survives the multiple-testing correction but is significant only under the word-list coder, and in the general narratives the human coding, if anything, finds *more* past reference in purchase narratives (+5 pp, though inter-coder agreement on this category is low). We therefore read the temporal asymmetry as established between persons for forward-looking language, and as suggestive elsewhere. The narrative-form indices have a cross-method check of their own: a hypothesis-blind LLM instrument scoring the same constructs on the original German text confirms the direction of the specificity and episodicity contrasts in both designs (Appendix I).

Circumstance deserves its own scrutiny. It has no analog outside the word-list coder and rests on that measure alone—and its asymmetry is largely definitional. Some motives structurally exist only on the sell side (liquidity needs, profit-taking, stop-losses, taxes): dropping narratives that invoke direction-specific motives eliminates the circumstance gap while leaving the fundamentals and recommendation asymmetries essentially unchanged and most of the price/rule asymmetry intact.[33]

A final check dispenses with predefined features altogether. We train an L_2-regularized logistic regression on TF-IDF representations of the raw text to distinguish the same respondent's general buy paragraph from their general sale paragraph, removing all explicit buy- and sell-direction words; some direction-adjacent vocabulary (*Gewinnmitnahme*) necessarily remains, so the accuracy is an upper bound on purely thematic separation. It separates the two paragraphs with a five-fold cross-validated accuracy of 77% (with folds grouped by respondent, so a respondent's two paragraphs never straddle the training–test split), and the words that drive the separation recover the coded features without being told about them: among the most buy-distinctive tokens are terms of company quality, valuation, and advice (*Unternehmen*, *günstig*, *unterbewertet*, *Dividende*, *Empfehlungen*); among the most sale-distinctive, terms for realized outcomes and targets (*Gewinn*, *Verlust*, *erreicht*, *Gewinnmitnahme*, *gestiegen*).

K Construction of the Overconfidence Measure

The self-assessed percentile rank is elicited in the main survey: respondents estimate the percentage of the broker's clients who achieved a higher portfolio return over the twelve months preceding the survey, and the self-assessed rank is 100 minus this figure (Section 2). The actual percentile rank is constructed to mirror this question as closely as the data allow.

For every client of the broker, we reconstruct the portfolio's composition on each trading day by rolling the January 2024 holdings snapshot back through the complete trade record. The daily portfolio return is the value-weighted total return of the positions held that day (split- and dividend-adjusted total returns; weights are the previous day's position values), and the twelve-month return is compounded from these daily returns over the 365 days preceding the Monday of the respondent's survey week—a time-weighted return of the kind a broker would report, which credits each investor only with returns earned while actually holding a position. The

[33] Estimated on word rates per 100 words (Cohen's d, purchase minus sale, original $\rightarrow$ restricted): fundamentals +0.37 $\rightarrow$ +0.34 between persons and +0.32 $\rightarrow$ +0.29 within; recommendation +0.14 $\rightarrow$ +0.13 and +0.21 $\rightarrow$ +0.20; price/rule $-0.50 \rightarrow -0.45$ and $-0.28 \rightarrow -0.24$; circumstance $-0.26 \rightarrow -0.07$ (n.s.) and $-0.16 \rightarrow -0.04$ (n.s.). The restriction drops 42 of 806 between-person and 171 of 1,185 within-person narrative pairs.

Table AXIII: Descriptive Statistics of the Section 6 Outcome Measures

	Mean	Median	St. Dev.	*N*
Investor level				
Actual percentile rank (0–100)	49.67	48.4	28.44	843
Overconfidence (self-assessed − actual)	−0.80	−0.3	35.13	843
Chasing: share of purchases after +10%	0.25	0.21	0.20	714
Return before purchase (3m, %)	−3.94	−4.5	15.05	714
Trade level (pre-survey trades)				
Post-purchase 3m return, DAX-adj. (%)	−3.18	−2.7	21.75	49,122
Post-sale 3m return, DAX-adj. (%)	−3.45	−2.9	21.51	36,931

Notes: The actual percentile rank and overconfidence are constructed as described in this appendix; the sample comprises respondents with a computable rank. Chasing and the pre-purchase return are investor-level measures (the investor's share of purchases following a +10% three-month return, and the median three-month return of purchased stocks before the purchase, winsorized 1/99) over investors with at least five priced purchases. Post-trade returns are trade-level returns of the traded stock over the three months after the trade, in excess of the DAX performance index, winsorized 1/99 (Appendix N).

respondent's actual percentile rank is their rank within their own survey week's distribution of client returns (≈ 5,400 clients per week). We require that at least half of a portfolio's positions be priced at the evaluation date and that the portfolio have at least 60 valid return days in the window, for respondents and reference clients alike; positions without total-return data (chiefly smaller ETFs and certificates) are excluded from the return. The measure is computable for 843 of 1,524 respondents; respondents with a computable measure are somewhat more active traders than the remainder.

Three properties deserve note. First, a simpler static variant—the survey-date portfolio held fixed, each position's twelve-month total return value-weighted at survey-date values—correlates at 0.66 with the daily measure and yields the same conclusions throughout (flat calibration, quintile-end miscalibration of ±38 points, near-zero mean overconfidence); the daily measure is preferred because the static version's end-of-window weights mechanically overweight winners and it credits positions bought during the year with pre-ownership returns. Second, the actual rank is measured with noise, which exaggerates apparent miscalibration at the extremes of Figure 4; the near-zero correlation between self-assessed and actual ranks is the robust statement. Third, for respondents whose main account is elsewhere, the actual rank reflects only the portfolio at this broker; among main-account respondents, where the rank is best measured, the overconfidence decomposition of Table 11 is unchanged (+16.2, s.e. 7.3).

Table AXIII reports descriptive statistics for the constructed outcome measures of Section 6—the survey-elicited outcomes (return expectations, the self-assessed rank) are summarized in Table 1.

A boundary observation—beliefs about assets, unlike beliefs about the self, appear to bypass the narrative layer—rests on two facts. The one-year return expectation for the recalled stock is strongly predicted by the investor's experienced return on that stock (+0.023 scale points per percentage point, $p < .001$), with the stock's calendar returns over one to twelve months jointly irrelevant once the experienced return is controlled ($p = .52$); and adding the narrative measures to a regression of stock return expectations on direction and demographics raises R^2 by only 0.002.

L Overconfidence in the Survey Sample

Figure 4 establishes a simple fact: investors have little knowledge of their relative performance. Mean self-assessed ranks vary only between 47 and 52 across quintiles of actual performance, while actual ranks span 11 to 89, yielding a correlation of only 0.08. On average, investors are well calibrated (mean overconfidence = −0.8 percentile points), but calibration is poor at the individual level. Investors in the bottom performance quintile overestimate their rank by 37 percentile points, whereas those in the top quintile underestimate it by 38. The standard deviation of overconfidence is more than 35 percentile points (Appendix Table AXIII), indicating substantial heterogeneity in self-assessments.

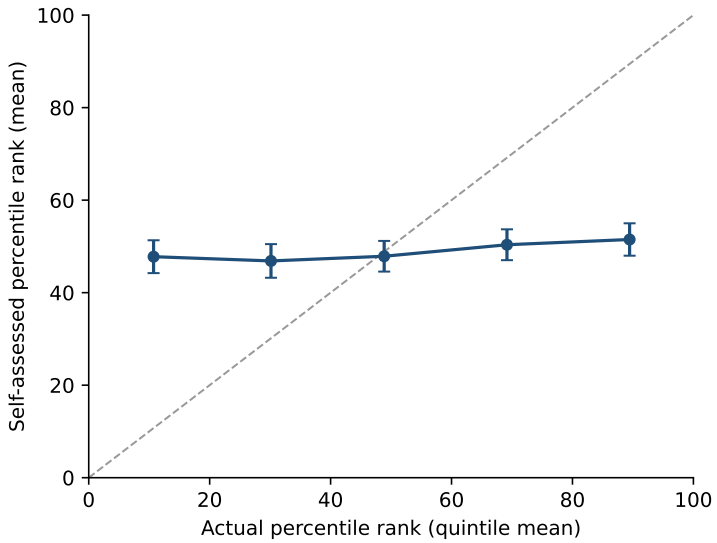


Figure 4: Self-assessed versus actual performance rank

Notes: Mean self-assessed percentile rank by quintile of the actual percentile rank, with 95% confidence intervals; the dashed line indicates perfect calibration. $N = 843$ respondents with a computable actual rank. Because the actual rank is measured with noise, the quintile means overstate miscalibration at the extremes; the near-zero correlation (0.08) is the robust statement.

M Narrative Features and Overconfidence

Section 6.1 shows that which trade memory volunteers marks overconfidence. This appendix asks whether the narrative told about that trade carries additional information. In levels, it does not: regressing overconfidence on the four content features and the three narrative-form indices of Section 4.1 (with the controls of Table 11 and treatment-condition fixed effects; $n = 742$) yields joint tests of $p = .45$ for content and $p = .79$ for form. As discussed in the main text, elaboration lowers the self-assessed rank and, under the daily performance measure, is also associated with a somewhat lower actual rank—modesty, not miscalibration.

Table AXIV interacts each narrative feature with the recall-direction indicator in the *FREE* condition. Two patterns stand out. The buy-recall association with overconfidence is large among respondents whose narrative contains none of the measured features (+32.5, s.e. 13.0), and it declines with elaboration (−12.9 per standard deviation, $p = .049$); the seven interactions are jointly significant ($p = .020$). Read literally, the miscalibration that recalled purchases mark is concentrated among investors who cannot say much about the trade they recall—consistent with the reading that self-assessment leans on internal evidence precisely where that evidence is thinnest.

We present this as a suggestive exhibit only. The sample contains 254 respondents, of whom only 43 recalled a sale; the recommendation interaction, though nominally the largest coefficient in the table, is identified by a single sale-recaller who mentioned a recommendation and should not be interpreted; and seven interactions were examined without pre-specification. A design with a larger free-recall arm would be needed to establish the pattern.

Table AXIV: Recall Direction, Narrative Features, and Overconfidence

	Main effect	× Recall purchase
Recall purchase (=1)	32.54** (13.02)	
Fundamentals	8.69 (18.51)	−12.47 (19.66)
Price/Rule	7.64 (12.73)	−13.65 (14.24)
Recommendation	55.17*** (21.14)	−64.62*** (24.51)
Circumstance	0.54 (14.57)	2.03 (16.59)
Elaboration	7.30 (5.98)	−12.92** (6.58)
Specificity	5.43 (4.99)	−4.12 (5.35)
Episodicity	5.22 (3.21)	−5.52 (3.48)
Share of purchases	−5.20 (11.43)	
Log trade count	3.94** (1.77)	
Joint test, all interactions	$F = 2.44$, $p = .020$	
N / R^2	254 / 0.220	

Notes: OLS with HC1 standard errors (in parentheses); the dependent variable is overconfidence (self-assessed minus actual percentile rank, daily measure). Sample: *FREE*-condition respondents with a computable actual rank. Content features are indicators from the recalled-trade narrative; form indices are standardized (Section 4.1). Demographic fixed effects as in Table 11. The recommendation row is identified by a single sale-recaller who mentioned a recommendation and should not be interpreted. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

N Construction of the Trading-Behavior Measures

Behavior is measured from respondents' complete stock-trade records (roughly 104,000 trades, or about 88,000 after same-day aggregation, by 1,203 respondents, October 2021–January 2024); 92% of these trades are matched to split- and dividend-adjusted total-return data. Trade-level returns are winsorized at the 1st and 99th percentiles, and investor-level measures are medians across the investor's trades; all group comparisons require at least five priced purchases (or, for selling measures, five classified sales).

Extrapolative buying is measured from the return of the purchased stock over the three months preceding the purchase: "chasing" is the share of an investor's purchases that follow a return above +10%, and "dip buying" the share following a return below −10%. Post-trade performance is the median market-adjusted return of the traded stock over the one to six months following the trade, where the market benchmark is a broad world-market index constructed from transaction-implied prices of the most heavily traded index fund in the records; subtracting instead the average same-direction trade return of all respondents in the same calendar month yields similar results.

The trade-informativeness exhibit (Table 13) is constructed at the trade level: each pre-survey trade's return over the three months following the trade date, in excess of the DAX performance-index return over the same window (daily index closes from public sources, validated against known values; dividends are included in both stock and index returns), winsorized at the 1st/99th percentiles within horizon. Differentials and the interaction test are OLS on a buy indicator with standard errors clustered by investor. The three-month display horizon matches the median life of a closed position (112 days from first purchase to sale, left-truncated by the October 2021 start of the records).

Table AXV probes the exhibit. Panel A varies the benchmark: raw returns with no adjustment, the median same-window return of the priced trading universe, a DAX benchmark implied by the transaction prices of the most-traded DAX index fund, and the official DAX. The fundamentals differential and its contrast with the no-reason group are essentially invariant—common market movements difference out of the purchase–sale comparison—while the small aggregate differential is benchmark-sensitive and loses significance under the DAX versions, which is why the text does not lean on it. Panel B varies the horizon on the identical trade sample: the differential is present at one month and grows monotonically through twelve, inconsistent with a transitory price-pressure account and consistent with the persistent decision-quality differences documented for institutional investors by Akepanidtaworn et al. (2023). Three further checks: weighting trades by euro amount leaves the fundamentals differential's point estimate unchanged (+2.1 under the median-stock benchmark) but roughly quadruples its standard error, as euro volume concentrates in fewer effective observations; restricting to respondents for whom this is their main brokerage account yields +1.7 ($p = .003$); and a within-investor version (median of each investor's own purchase–sale differential, at least five priced trades on each side) is signed the same way but less precise (+0.4 vs. −0.0 for the no-reason group, $p = .052$), so the exhibit's precision comes partly from between-investor variation.

Table AXVI repeats the extrapolation comparison of Table 12 with narratives classified into mutually exclusive types (exactly one theme mentioned, several, or none). Chasing depends only on whether *any* reason is articulated: every reason-giving group sits six to nine percentage points below the no-reason group, regardless of theme.

Selling measures follow Odean (1998): on each day an investor sells, every position held (with an average-cost basis reconstructed from the observed record) is classified as a realized or paper gain or loss, and the proportion of gains realized is computed per investor. Positions whose basis predates the records are excluded; a comparison with the holdings snapshot bounds residual

Table AXV: Robustness of the Trade-Informativeness Differential

Panel A: Benchmarks (3-month horizon)			
	Fundamentals	Fundamentals − None	All trades
No adjustment (raw)	+2.24*** (.001)	+2.77** (.037)	+0.40** (.046)
Median stock, same window	+2.14*** (.000)	+3.23** (.010)	+0.43** (.015)
DAX, transaction-implied	+1.97*** (.001)	+3.43*** (.006)	+0.31* (.078)
DAX, official index	+1.91*** (.001)	+3.39*** (.007)	+0.27 (.124)
Panel B: Horizons (official DAX, identical trade sample)			
	Fundamentals	Fundamentals − None	All trades
One month	+0.60** (.033)	+1.68** (.012)	−0.03 (.799)
Three months	+1.91*** (.001)	+3.39*** (.007)	+0.27 (.124)
Six months	+3.04*** (.002)	+3.59** (.037)	+0.44 (.116)
Twelve months	+4.78** (.029)	+7.61** (.019)	+0.56 (.294)

Notes: Purchase–sale return differentials as in Table 13 (p-values in parentheses, standard errors clustered by investor). *Fundamentals* is the differential among trades whose relevant narrative mentions fundamentals; *Fundamentals − None* is the interaction contrast against the no-reason group; *All trades* pools the sample. Panel B holds the trade sample fixed across horizons. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table AXVI: Extrapolative Buying: Mutually Exclusive Narrative Types

		Share of buys after +10%		Pre-purchase return (3m, %)	
Narrative mentions...	N	share	vs. none	median	vs. none
Only fundamentals	147	0.20	−0.09***	−5.8	−3.6*
Only price/rule	104	0.20	−0.09**	−4.4	−2.1
Only recommendation	66	0.21	−0.08*	−5.3	−3.1
Only circumstance	36	0.23	−0.06	−1.7	+0.5
Multiple themes	209	0.19	−0.09***	−5.5	−3.3**
None of the four	152	0.29		−2.2	

Notes: As in Panel A of Table 12, but with respondents classified by whether the relevant narrative mentions exactly one of the four content themes, several ("multiple themes"), or none. Comparisons are Mann–Whitney tests against the no-reason group. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

contamination at about 3% of classified sales. Because the records begin in October 2021, the level of the measure is not comparable to estimates from complete account histories, and we interpret only cross-investor differences. Profit-taking vocabulary is a word-list flag on the general sale narrative; the episodic register is an indicator for an Episodicity index (Section 4.1, computed on the general sale narrative) strictly above the sample median. The median narrative contains no episodic markers, so the indicator selects the most episodic investors.

O Survey Instructions and Items: Main Survey

[Translated from the German original.]

Welcome to the survey!

This survey is conducted by Goethe University Frankfurt and Bocconi University.

Thank you for taking the time to participate in our survey on the topic of financial decisions.

Completing the survey takes about 15 minutes. Your data will be stored and analysed exclusively in pseudonymised form. The results cannot be used to draw conclusions about you as a person. You can find further information in the data-protection notice.

As a thank-you, you will receive an Amazon voucher or a PayPal payout worth €5. To receive this, please provide an email address at the end of the survey. We will send you the payout/voucher code by email. The email address will be stored separately from your answers and used only for payment/documentation of the payment.

Do you have questions? Feel free to contact us at umfrage@safe-frankfurt.de.

–Next page–

Consent
The purpose of this page is to provide you with the information you need so that you can give your consent to participate in this research project.

Study description
The study examines financial decisions made by private households.

Procedure and duration
Completing the questionnaire takes about 15 minutes.

Benefits and foreseeable risks
This study is not associated with any known physical or emotional risk. The study serves scientific knowledge generation.

Voluntary participation
Your participation is voluntary and you may withdraw at any time without giving reasons. You do not have to answer any questions you do not wish to answer, and withdrawing is not associated with any penalty or loss of benefits.

Compensation
As a thank-you for your time and contribution, those who participate in the survey will receive an Amazon voucher or a PayPal payout worth €5 (optional). To receive this, please provide an email address at the end of the survey. We will send you the payout or voucher code by email. The email address will be stored separately from your answers and used only for payment / documentation of the payment.

Deception
This study does not involve any deception.

CONSENT DECLARATION
I confirm that I have received the above information and declare that I have read and understood its contents. I confirm that I am 18 years old or older and agree to participate in this study. I acknowledge that my data will be processed in full compliance with the law, and I voluntarily agree to the use of my data in the manner described and for the purposes described. I also declare that I have understood my rights and limitations as well as how to exercise them.
Options:

- I would like to participate
- I would not like to participate

–Next page–

In the last month (i.e., in the last four weeks), have you [*PURCHASE* condition: "bought"; *SALE* condition: "sold"; *FREE* condition: "traded"] stocks through your brokerage account at (*BROKER NAME*)?

- Yes
- No
- I don't know

–Next page–

[if respondents answered with "No" or "I don't know" to the previous question:]

Now please think of a stock trade that happened further in the past.

We would like to ask you some questions about a specific stock you have [*PURCHASE* condition: "bought"; *SALE* condition: "sold"; *FREE* condition: "traded"].

If you have made several [*PURCHASE* condition: "purchases"; *SALE* condition: "sales"; *FREE* condition: "trades"], please choose the stock [*PURCHASE* condition: "purchase"; *SALE* condition: "sale"; *FREE* condition: "trade"] that comes to mind first.

Which stock did you [*PURCHASE* condition: "buy"; *SALE* condition: "sell"; *FREE* condition: "trade"]?

Please enter the name of one stock.

–Next page–

[if respondents answered with "Yes" to the previous question:]

Now we would like to ask you some questions about a specific stock you have [*PURCHASE* condition: "bought"; *SALE* condition: "sold"; *FREE* condition: "traded"].

If you have made several [*PURCHASE* condition: "purchases"; *SALE* condition: "sales"; *FREE* condition: "trades"], please choose the stock [*PURCHASE* condition: "purchase"; *SALE* condition: "sale"; *FREE* condition: "trade"] that comes to mind first.

Which stock did you [*PURCHASE* condition: "buy"; *SALE* condition: "sell"; *FREE* condition: "trade"]?

Please enter the name of one stock.

–Next page–

[if respondents are in the *FREE* condition:]
Did you buy or sell the stock?

- Buy
- Sell

–Next page–

Now we ask you once again to describe your thoughts and views in your own words.

Your answers are very valuable for this research project. Please take the time and answer carefully. We are interested in what you think. Please therefore describe your views and opinions in detail.

Why did you [*PURCHASE* condition: "buy"; *SALE* condition: "sell"; *FREE* condition: either "buy" or "sell," depending on their answer to the previous question] this stock?

–Next page–

What was the trigger that prompted you to [*PURCHASE* condition: "buy"; *SALE* condition: "sell"; *FREE* condition: either "buy" or "sell," depending on their answer to the previous question] the stock at exactly this moment? Please answer as thoroughly as possible and in complete sentences.

How long did you think about the stock trade before you carried it out?

- Not at all
- A few minutes
- A few hours
- A few days
- A few weeks
- A few months
- Longer than a few months

How long do you currently plan to hold the shares?

- A few days
- A few weeks
- A few months
- A few years
- I haven't thought about that yet.

How confident are you in this assessment? Scale: not confident at all (1) — extremely confident (7)

When making the trade, did you choose between several stocks?

- Yes
- No

[if "Yes" to previous question:] Which other stocks did you also consider?

–Next page–

In your opinion, what return will this stock achieve over the next 12 months?

- less than -20%
- between -20% and -15%
- between -15% and -10%
- between -10% and -5%
- between -5% and 0%
- between 0% and +5%
- between +5% and +10%
- between +10% and +15%
- between +15% and +20%
- more than +20%

How confident are you in your estimate? Scale: not confident at all (1) — extremely confident (7)

–Next page–

On the next page you will find an open question in which we ask you to describe your thoughts and views in your own words.

Your answers are very valuable for this research project. Please take the time and answer carefully. We are interested in what you think. Please describe your views and opinions.

–Next page–

What usually prompts you to BUY a particular stock? Please answer as thoroughly as

possible and in complete sentences.

–Next page–

What usually prompts you to SELL a particular stock? Please answer as thoroughly as possible and in complete sentences.

–Next page–

In your opinion, how important is timing when trading stocks?

- Extremely important
- Very important
- Somewhat important
- Not very important
- Not at all important

–Next page–

Please give us your estimate: What proportion of retail investors at your broker (*BROKER NAME*) achieved a higher annual portfolio return than you in the last 12 months?

Please enter a percentage (between 0% and 100%).

–Next page–

Please give us your estimate: What return will your total portfolio at broker (*BROKER NAME*) achieve over the next 12 months?

- less than -20%
- between -20% and -15%

- between -15% and -10%
- between -10% and -5%
- between -5% and 0%
- between 0% and +5%
- between +5% and +10%
- between +10% and +15%
- between +15% and +20%
- more than +20%

How confident are you in your estimate? Scale: not confident at all (1) — extremely confident (7)

–Next page–

For our research project, a few more detailed pieces of information about the transaction would help. We would be very grateful if you could take the time to provide these. You stated that you traded the stock (*STOCK NAME*).

What is the ISIN or WKN of this security?

What total amount did you trade in this stock on that day (in euros)?

Do you still remember on which date you traded?

–Next page–

How often do you trade per year (across all your accounts, if you have more than one)?

- Daily
- Every week
- Every two weeks
- Every month
- Every 6 months
- Once a year
- Less than once a year

In your opinion, what return will the German stock index (DAX) achieve over the next 12 months?

- less than -20%
- between -20% and -15%
- between -15% and -10%
- between -10% and -5%
- between -5% and 0%
- between 0% and +5%

- between +5% and +10%
- between +10% and +15%
- between +15% and +20%
- more than +20%

How confident are you in your estimate? Scale: not confident at all (1) — extremely confident (7)

–Next page–

For how many years have you been trading securities?

- I do not trade securities
- less than 1 year
- 1–3 years
- 3–5 years
- 5–10 years
- 10–20 years
- more than 20 years

–Next page–

To what extent do you agree with the statement: "My personal financial knowledge is good." Scale: strongly disagree (1) — strongly agree (5)

–Next page–

How would you describe yourself personally: when it comes to investment decisions, are you a more risk-taking person, or do you try to avoid risks?

Please select a value on the scale, where 0 means "not at all willing to take risks" and 10 means "very willing to take risks". Scale: 0 (not at all risk-taking) — 10 (very risk-taking)

–Next page–

Is your brokerage account at (*BROKER NAME*) your main investment account?

- Yes
- No

–Next page–

Please select your year of birth (dropdown menu).

What gender are you?

- Male
- Female
- Non-binary
- Prefer not to say

What is your current marital status?

- Single
- Married
- Widowed

- Divorced
- Registered partnership
- Registered partner (in a registered partnership) deceased
- Registered partnership dissolved
- Prefer not to say

Do you have children?
- Yes
- No

What is your highest educational qualification?
- No school leaving qualification
- Lower secondary school leaving certificate (Grund-/Hauptschule)
- Realschule (intermediate school leaving certificate)
- Gymnasium (Abitur)
- Completed vocational training
- University of applied sciences degree
- University (Bachelor's)
- University (Master's)
- University (Doctorate/PhD)
- Prefer not to say

–Next page–

Thank you for your participation.

May we contact you again in the future to take part in a scientific survey?
- Yes
- No

If we may contact you again, please provide an email address at which we may contact you for a future survey. Email address for a future survey:

For answering our questions, you will receive an Amazon voucher or a PayPal payout (optional) in the amount of €5.

May we use your email address to send you an Amazon voucher or a PayPal payout?
- Yes
- No

If you would like to receive the payout/voucher, please provide an email address to which we may send the payout/voucher. Email address for the payout/voucher:

How would you like to receive your €5 for participating in the study?
- Gift voucher (Amazon)
- PayPal payout

Taxes and duties: the payment is “gross for net”.
Checkbox: I hereby confirm that I will pay any taxes and other duties that may be due myself.

You will receive the payout/voucher within the next 4 weeks. If you do not receive a payout/voucher within 4 weeks, please contact us at umfrage@safe-frankfurt.de.

–Next page–

End of the survey

Thank you for participating. You can now close the tab.

P Survey Instructions and Items: Follow-up Survey

[Translated from the German original.]

Welcome to the second part of our survey!

This survey is conducted by Goethe University Frankfurt and Bocconi University.

Thank you for taking the time to complete the second part of our survey on financial decision-making.

It only takes 5 minutes to complete.

Your data will be stored and analysed exclusively in pseudonymised form. The results do not allow any conclusions to be drawn about you as a person. You can find further information in the data protection notice.

As a thank-you, you will receive an Amazon voucher or a PayPal payout worth €5. To receive this, please enter an email address at the end of the survey. We will send you the payout/voucher code by email. Your email address will be stored separately from your answers and used only for payment and payment documentation.

–Next page–

In our previous survey (on *DATE*) we asked whether you have [*PURCHASE* condition: "bought"; *SALE* condition: "sold"; *FREE* condition: "traded"] stock shares in the past. Your answer was that you [*PURCHASE* condition: "bought"; *SALE* condition: "sold"; *FREE* condition: either "bought" or "sold," depending on their answer to the question whether they have bought or sold] the stock (*STOCK NAME*, ISIN: *XXX*).

Can you please confirm if that is correct?

- Yes
- No

–Next page–

How do you think the value of this stock developed over the last six months (07.07.2025 to 07.01.2026)?

- Increased
- Decreased
- Stayed the same

–Next page–

By how much has the value of the stock changed (in percent)?

–Next page–

Looking back at this transaction, what would you say motivated you at the time to [*PURCHASE* condition: "buy"; *SALE* condition: "sell"; *FREE* condition: either "buy" or "sell," depending on their answer to the question whether they have bought or sold] this stock?

–Next page–

[if respondents bought a stock:]

How long did you think about buying the stock before you made the purchase?

- Not at all
- A few minutes
- A few hours
- A few days
- A few weeks
- A few months
- Longer than a few months

How long did you plan to hold the stock?

- A few days
- A few weeks
- A few months
- A few years
- I had not thought about that yet.

–Next page–

[if respondents sold a stock:]

How long did you think about selling the stock before you made the sale?

- Not at all
- A few minutes
- A few hours
- A few days
- A few weeks
- A few months
- Longer than a few months

How long did you hold the shares?

- Longer than you had planned when buying
- About as long as you had planned when buying
- Shorter than you had planned when buying

- When buying, I did not have a specific idea of how long I would hold them

–Next page–

How do you think the value of this stock will develop over the next 12 months?

- less than -20%
- between -20% and -15%
- between -15% and -10%
- between -10% and -5%
- between -5% and 0%
- between 0% and +5%
- between +5% and +10%
- between +10% and +15%
- between +15% and +20%
- more than +20%

How confident are you in this estimate? Scale: 1 (not at all confident) — 7 (extremely confident)

–Next page–

Now we have a few general questions about your investments and the stock market.

What is your estimate: What percentage of retail investors at your broker (*BROKER NAME*) achieved a higher annual portfolio return than you in 2023?

Please enter a percentage (between 0% and 100%).

–Next page–

How do you think the value of YOUR PORTFOLIO with your broker (*BROKER NAME*) will develop over the next 12 months?

- less than -20%
- between -20% and -15%
- between -15% and -10%
- between -10% and -5%
- between -5% and 0%
- between 0% and +5%
- between +5% and +10%
- between +10% and +15%
- between +15% and +20%
- more than +20%

How confident are you in this estimate? Scale: 1 (not at all confident) — 7 (extremely confident)

–Next page–

How do you think the value of the German stock index (DAX) will develop over the next 12 months?

- less than -20%
- between -20% and -15%
- between -15% and -10%
- between -10% and -5%
- between -5% and 0%
- between 0% and +5%
- between +5% and +10%
- between +10% and +15%
- between +15% and +20%
- more than +20%

How confident are you in this estimate? Scale: 1 (not at all confident) — 7 (extremely confident)

–Next page–

Thank you for your participation.

For answering our questions, you will receive an Amazon voucher or a PayPal payout (optional) worth 5 euros.

How would you like to receive your 5 euros?
- Amazon voucher
- PayPal payout

If you would like to receive the payout/voucher, please provide an email address to which we may send the payout/voucher:

Taxes and duties:
[Checkbox:] I am aware that the amounts earned are paid out gross. I hereby confirm that I am not an entrepreneur within the meaning of §2(1) UStG and that I will pay any taxes and other charges that may be due myself.

You will receive the payout/voucher within the next 4 weeks. If you do not receive any payout/voucher within 4 weeks, please contact us at umfrage@safe-frankfurt.de.

–Next page–

End of the survey
Thank you for participating. You can close the tab now.