Selective Recall and the Story-Statistics Gap in Stock Market Misreaction

Weiting Hong*
Yale School of Management

This draft: June 8, 2026
First draft: August 18, 2025

Abstract

Using 1.3 million corporate news announcements between 2003 and 2022 from the Capital IQ Key Developments dataset, I document a strong story-statistics gap in stock market misreaction to corporate news: whereas qualitative announcements exhibit mild overreaction, quantitative announcements generate pronounced underreaction. The story-statistics gap manifests both across and within news categories. This pattern cannot be fully explained by signal-strength uncertainty, diagnostic expectation, news sentiment, or investor attention. It is robust to the exclusion of known anomalies, such as post-earnings announcement drift and post-merger reversal. A characteristic-managed portfolio based on the story-statistics gap delivers an annualized Sharpe ratio of 1.31 and earns a statistically significant alpha of approximately 110 basis points per month against a variety of factor models. Motivated by experimental evidence, I apply selective recall to the stock market setting and propose distinct similarity structures for stories and statistics. I hypothesize that narrative content characterizes story-like news, which leads to intra-category interference; numerical precision defines statistical news, which leads to inter-category interference. I construct embedding-based similarity measures and show that theory-predicted misreactions strongly align with empirically observed misreactions across news categories. I additionally propose a measurement error-free approach to testing the model and find strong support for selective recall in announcement-level data. In line with the hypothesized similarity structure, I show that news announcements that are more similar in quantitativeness generate more comparable stock price impacts.

*I am very grateful to Nicholas Barberis for his guidance, teaching, and support in advising this paper. I thank Xugan Chen, James Choi, Ben Gardner, Pranav Garg, Stefano Giglio, Paul Goldsmith-Pinkham, William Goetzmann, Oliver Hellum, Dong Huang, Tom Huang, Bryan Kelly, Spencer Kwon, Lauren Lai, Tianshu Lyu, Song Ma, Cameron Peng, Mo Pourmohammadi, Jamil Rahman, Sasha Rothstein, Suproteem Sarkar, Tudor Schlanger, Andrei Shleifer, Kelly Shue, Alp Simsek, Hongyu Wu, Nicholas Wuthenow, and Dolly Yu for helpful comments.

1 Introduction

Past experiences help people evaluate future decisions. However, in making decisions, people may not consider all memories with equal probability. Recent experimental evidence from Graeber, Roth, and Zimmermann (2024) shows that memory interacts with qualitative and quantitative information in different ways, such that belief impact from stories is longer-lasting than that from statistics. They identify selective recall as a key driver of this story-statistics gap in their laboratory setting. In the stock market, corporate news announcements are a prominent source of investor information that exhibit substantial variation in quantitativeness. While executive change and strategic alliance news often uses story-like descriptions, earnings announcements and impairment notices typically feature statistical updates. Does the quantitativeness of a corporate news announcement affect the stock market response to it? And if the information type of news (qualitative vs. quantitative) influences investor behavior, does selective recall explain this connection?

I study 1.3 million corporate news announcements between 2003 and 2022 from the Capital IQ Key Developments dataset and provide affirmative evidence for both questions. In the first part of this paper, I document a strong story-statistics gap in stock market misreaction to firm-specific news: while individual stocks mildly overreact to qualitative news, they strongly underreact to quantitative news. I measure the quantitativeness of a news announcement with its *statistics ratio*, which divides the number of statistical tokens in its news text by its total word count. For each announcement, I define *subsequent return* as the cumulative abnormal return on the announcing firm from two to 90 days after the announcement date and *immediate return* as the cumulative abnormal return from one day before to one day after. In the spirit of Coibion and Gorodnichenko (2015), I measure misreaction in a sample of news announcements by regressing subsequent returns on immediate returns with announcement year, firm, and news category fixed effects. A positive regression coefficient indicates post-announcement return continuation and underreaction, while a negative regression coefficient indicates post-announcement return reversal and overreaction.

Labeling news with a statistics ratio of zero as stories and the remainder as statistics, I find that the misreaction regression coefficient is mildly negative for stories and significantly positive for statistics. I additionally sort statistical news into four equal-sized bins based on the statistics ratio, such that the fourth bin contains the most quantitative news. I find that the estimated coefficient monotonically increases with the statistical intensity of the bins, which suggests more underreaction to more quantitative news. To better account for cross-category variation in stock market misreaction as documented by Kwon and Tang (2026), I regress subsequent returns on immediate returns interacted with the statistics ratio at the news announcement level with the announcement year, firm, and news category fixed effects. The regression coefficient on the interaction term is significantly positive, which is consistent with the monotonic pattern across bins.

These systematic differences in post-announcement return behavior suggest a story-statistics gap in stock market misreaction. I show that the story-statistics gap manifests both across and within news categories. Signal strength uncertainty, diagnostic expectations, news sentiment, and investor attention are not able to account for the story-statistics gap. The empirical pattern is robust to the exclusion of known anomalies, such as post-earnings announcement drift and post-merger reversal. The key results do not hinge on the use of news summaries from the Key Developments dataset: repeating the same analytical procedure on full-length news articles from the Wall Street Journal replicates the same story-statistics gap.

To assess the economic significance of the story-statistics gap, I construct a characteristic-managed portfolio that uses the statistics ratio as a proxy for stock market misreaction. The story-statistics gap predicts that stock market underreaction increases in the quantitativeness of a news announcement. For each firm-month, I calculate the average statistics ratio and the average immediate return across the firm's news announcements in the previous month.

I standardize the two characteristics across firms within the previous month and take their firm-level products as the portfolio weights for the current month. The long leg contains the most underreacted (statistical) firms with the highest immediate returns (positive news), as well as the most overreacted (story-like) firms with the lowest immediate returns (negative news). The short leg contains the most underreacted (statistical) firms with the lowest immediate returns (negative news), as well as the most overreacted (story-like) firms with the highest immediate returns (positive news). Over the sample period, the misreaction-managed portfolio delivers an annualized Sharpe ratio of 1.31. It earns a statistically significant alpha of approximately 110 basis points per month against a variety of factor models. These results corroborate the economic significance of the story-statistics gap as a meaningful driver of stock market misreaction.

In the second part of this paper, I find that selective recall helps explain the story-statistics gap in stock market misreaction. I adapt the model of selective recall from Bordalo et al. (2023) to the stock market setting. My model formalizes the idea that, when investors call upon past memories to estimate the stock price impact of a corporate news announcement, they are more likely to recall experiences that are more similar to the current situation. For example, to assess the likelihood that an earnings beat would result in a large stock price uptick for the announcing firm, investors would like to review the distribution of valuation impact across past earnings beats. However, in addition to past earnings beats, they may accidentally recall similar yet irrelevant experiences that are uninformative about the valuation impact of earnings beats, such as past instances of successful product launches. These irrelevant recalls (i.e., successful past product launches) occupy limited cognitive resources, which interferes with the recall of relevant experiences (i.e., past earnings beats). Investors misestimate the magnitude of news when the similarity-based recall probabilities for relevant experiences at different valuation impact levels deviate from their objective frequencies. Overreaction arises when investors overestimate news magnitude, and underreaction ensues when investors underestimate news magnitude.

Building on the experimental evidence from Graeber, Roth, and Zimmermann (2024), I hypothesize that memory evaluates similarity based on different characteristics for stories and for statistics. I propose that two stories are alike because they share similar narrative content, while two statistics are alike because they convey similar valuation impact. Because stories lack numerical precision while statistics lack semantic meaning, I assume that stories are dissimilar to statistics. Moreover, given the salience of news direction (positive vs. negative) for both stories and statistics, I assume that only news in the same direction can be similar.

Consequently, only stories in the same direction and news category are similar, even if they carry different valuation impacts. For example, merger announcements are generally story-like. Both high- and low-impact merger announcements emphasize synergy-driven investment theses that focus on value creation and strategic fit. These common narrative features distinguish merger announcements from other story-like categories, such as lawsuits and client announcements. As a result, stories primarily face *intra*-category interference.

At the same time, only statistics with the same direction and valuation impact are similar, even if they belong to different news categories. For example, earnings announcements are typically statistical. Statistics are precise, so high-impact updates (e.g., consensus-defying top-line growth) greatly contrast with low-impact updates (e.g., consensus-following top-line growth), even when they belong to the same news category. Nevertheless, a positive statistical announcement from another news category, such as a successful product launch or business expansion, may appear similar to a positive earnings announcement if they deliver comparable valuation impacts. Therefore, statistics primarily face *inter*-category interference.

Ultimately, only past news in the same news category as the current news announcement is directly relevant to assessing the valuation impact of the present situation. The mismatch between the determinants of similarity and the standard of relevance leads to different investor behaviors for stories and for statistics. While the model does not predict a story-statistics gap ex ante, it offers concrete predictions for stock market misreaction as functions of the similarity relationships and relative frequencies among corporate news announcements.

To test whether the model generates the story-statistics gap in stock market misreaction, I empirically measure the similarity structure of corporate news announcements. I use a sentence transformer to extract embedding vectors from the news summaries, which are numerical representations of textual data that capture information content and semantic meaning. I estimate the average pairwise similarity between two groups of news announcements as the inner product between their mean normalized embedding vectors. I obtain the similarity measure by flooring the inner product at zero. The resulting similarity measure ranges from zero to one, where zero represents complete dissimilarity and one indicates full equivalence.

Following my hypothesized similarity structure, I restrict the similarity between groups of news announcements that are dissimilar in my theory to zero. For example, merger stories would have a similarity of zero with merger statistics in the restricted model because stories are dissimilar to statistics. Merger stories would also have a similarity of zero with lawsuit stories in the restricted model because stories are dissimilar across news categories. Positive merger stories would also have a similarity of zero with negative merger stories because positive news is dissimilar to negative news. However, positive, high-impact merger stories could be similar to positive, low-impact merger stories in the restricted model. In such cases where the similarity relationships are not restricted to zero, I use the empirically estimated similarity relationships between these groups of news announcements.

Based on the estimated similarity relationships and the theoretical restrictions, the restricted model predicts a significant story-statistics gap across news categories, such that more statistical news categories exhibit more underreaction. This prediction aligns well with the empirically observed story-statistics gap. At the news category level, the model-predicted, theoretical misreactions are significantly positively associated with the empirically estimated misreaction regression coefficients. Moreover, regressing empirical on theoretical misreactions yields an R^2 of 34%, which suggests that selective recall is an important driver of cross-category variations in stock market misreaction.

My proposed restrictions on the similarity structure for stories and statistics are critical for the model's explanatory power. Using the same similarity estimates, I calculate the theoretical misreactions under the unrestricted model, which does not include the hypothesized differences in interference patterns between stories and statistics. Theoretical misreactions from the unrestricted model do not generate a story-statistics gap, nor do they align with the empirical misreactions.

To address potential measurement error from measuring similarity, I additionally derive six testable implications from the model that relate variation in predicted misreaction solely to variation in the relative frequencies among past news announcements, which are precisely observed. They rely on two assumptions: the similarity function is non-negative, and a group of experiences is more similar to itself than it is to a disjoint group of experiences. These model implications do not depend on the similarity estimates and offer a measurement error-free alternative to testing the model. I empirically verify all six model implications and demonstrate their robustness under a variety of parameter choices.

Lastly, my model of selective recall generates a testable implication for pairwise differences in subjective valuation impact across news announcements. In my model, investors' subjective valuation impact depends in part on the similarity between the current news announcement and past experiences. The hypothesized story-statistics division in similarity structure implies that announcements with more similar quantitativeness should elicit more similar subjective valuation impacts. Because the main sample contains more than one million news announcements, estimating regressions on the full set of more than one trillion announcement pairs is computationally infeasible. I therefore estimate pairwise regressions on randomly sampled announcement pairs and use a cluster bootstrap for inference. Empirically, I find that pairs of news announcements with larger differences in statistics ratios also exhibit larger differences in the magnitudes of immediate stock return responses. This finding provides further support for the hypothesized similarity structure and the model of selective recall.

This paper makes four key contributions. First, I document a novel story-statistics gap in stock market misreaction, which confirms the experimental finding that people tend to underweight quantitative information. Second, I construct a novel predictor of stock returns based on the story-statistics gap. Third, I adapt the model of selective recall from Bordalo et al. (2023) to the stock market setting and derive its implications for stock market misreaction. Fourth, I validate these predictions and demonstrate that selective recall plays a material role in explaining the story-statistics gap, as well as variation in stock market misreaction at large. To my knowledge, this is the first paper to document a story-statistics gap in stock market misreaction and to explain this empirical pattern with selective recall.

Related Literature

My work primarily relates to four strands of literature.

First, this paper contributes to the growing evidence that people process qualitative and quantitative information differently. Graeber, Roth, and Zimmermann (2024) pioneer the study of how information type shapes belief dynamics. They experimentally document a story-statistics gap in belief impact, showing that stories exert more persistent influence on beliefs than statistics. Ke (2025) extends this pattern to analyst settings and finds that equity analysts overreact more to qualitative, story-like narratives than to quantitative data. I build on these studies and provide stock market evidence of the same phenomenon. I show that stock returns exhibit stronger underreaction to more quantitative corporate news announcements. By linking these behavioral differences to realized price dynamics, my work documents the asset-pricing implications of the story-statistics gap and highlights its economic significance.

Second, I provide field evidence on the economic relevance of selective recall outside of laboratory environments. A growing body of experimental work shows that selective recall materially shapes belief formation. For example, Graeber, Roth, and Zimmermann (2024) attribute the story-statistics gap in belief impact to selective recall. They argue that stories are more similar to memory prompts because both are conveyed through words that carry semantic associations. This semantic overlap makes past stories easier to retrieve at decision points than past statistics. In a series of betting-market experiments, Enke, Schwerter, and Zimmermann (2024) show that associative recall leads investors to overreact to financial news. They demonstrate that, when information is presented with memorable contexts, such as recurring narratives or images, participants disproportionately recall similar past news. This leads to amplified belief updating and more extreme market prices. My work extends these insights to the stock market. I show that selective recall can systematically affect investor behavior and generate predictable misreaction patterns in asset prices.

A recent field study by Jiang et al. (2025) provides complementary evidence on selective recall using survey responses and administrative trading data from retail investors in China. They document that, when the stock market rises on the survey day, investors are more likely to recall past bullish episodes and view their own past performance more positively. They demonstrate the link from memory to belief formation by finding a strong positive relationship between return expectations and retrieved memories. Finally, they show that more optimistic investors increase their equity holdings after the survey, completing the chain from memory to belief and action. Their work suggests that similarity-based recall shapes belief formation and individual trading behavior in the field. Their work also highlights return as an important determinant of similarity among stock market episodes. This finding supports my model assumption that positive-return and negative-return news announcements are dissimilar. Different from but complementary to their work, I study cross-sectional return patterns instead of individual portfolio decisions. Moreover, I emphasize the role and implication of interference as a core mechanism of selective recall.

Third, this paper introduces a new cognitive channel for variation in stock market misreaction. Existing behavioral finance research identifies several explanations for overreaction and underreaction. Augenblick, Lazarus, and Thaler (2025) argue that signal-strength uncertainty causes overreaction to weak signals and underreaction to strong ones. Kwon and Tang (2026) propose that diagnostic expectations lead investors to overreact more to news categories with more extreme historical outcomes. Tetlock, Saar-Tsechansky, and Macskassy (2008) show that negative-sentiment words in firm-specific news announcements predict stock market underreaction. Hou et al. (2025) emphasize investor attention. They find that while low attention induces underreaction, high attention exacerbates extrapolative behavior and overreaction. While these mechanisms capture important aspects of misreaction, they do not consider how memory shapes information processing. By introducing selective recall as a cognitive foundation for variation in investor responses, my work complements these frameworks and identifies a distinct, memory-based source of misreaction in the stock market.

Fourth, I identify empirical characteristics of corporate news announcements that meaningfully interact with selective recall. Theoretical models such as Bordalo et al. (2023) and Wachter and Kahana (2024) highlight that recall-based decision-making depends on similarity and relevance. However, these studies abstract from specifying what constitutes similarity and relevance in applied settings. I bridge this gap by linking these psychological constructs to observable features of corporate news announcements. I argue that stories are similar within each news category due to shared narrative structures, whereas statistics are similar if they convey the same valuation impact due to their numerical precision. This distinction clarifies which aspects of information activate selective recall and why investors respond differently to stories and to statistics.

The rest of the paper proceeds as follows. Section 2 describes the data and variable construction. Section 3 documents the story-statistics gap in stock market misreaction. Section 4 adapts the model of selective recall from Bordalo et al. (2023) to the stock market setting and tests its empirical implications under the hypothesized similarity structure for stories and statistics. Section 5 concludes.

2 Data and Methodology

In this section, I describe the data and construction of key variables for documenting the story-statistics gap in stock market misreaction. Section 2.1 describes the main data sources. Section 2.2 explains how I measure the quantitativeness of each corporate news announcement. Section 2.3 reviews how I measure stock market misreaction. Section 2.4 introduces controls for alternative sources of stock market misreaction.

2.1 Data sources

I retrieve 1.3 million textual summaries of corporate news announcements between 2003 and 2022 from the Capital IQ Key Developments dataset. This dataset aggregates announcements from more than 20,000 news sources, including press releases, news articles, regulatory filings, stock exchanges, company websites, call transcripts, and web-mined content. For tractability, I retain only the first 1,500 characters of each summary for analysis. This truncation affects roughly 10% of the summaries. The dataset provides the firm identifier, announcement date, news category, and a structured textual summary generated by Capital IQ for each announcement. I restrict the sample to CRSP firms with common U.S.

equity traded on NYSE, NASDAQ, or AMEX. Daily stock return data come from CRSP. All of the above data are accessed via WRDS. To compute abnormal returns, I obtain the 13 thematic factors of Jensen, Kelly, and Pedersen (2023), which summarize 153 known return anomalies, from jkpfactors.com.

To assess whether the empirical results rely on the use of news summaries from the Key Developments dataset, I additionally obtain 834,935 full-length news articles published by the Wall Street Journal between 1996 and 2024 via ProQuest TDM Studio. One challenge in using this data is that ProQuest does not provide a direct mapping from the articles to the CRSP firms. However, ProQuest does provide a list of mentioned companies in the metadata of each news article. 347,985 news articles mention at least one company in their metadata. I match the provided company names to a list of parent and subsidiary company names for CRSP firms obtained from WRDS. This procedure assigns 168,686 news articles (48% of firm-specific news) to at least one CRSP firm, and 5,391 CRSP firms are matched to at least one article. Another challenge is that, while my baseline analysis involves news category fixed effects, ProQuest does not provide a similar classification scheme for the news articles. To replicate my analysis as closely as possible, I employ transfer learning by fine-tuning FinBERT on the news summary and news category data from the Key Developments dataset. FinBERT is a pre-trained, BERT-based language model that specializes in capturing the semantic meaning of financial text (Araci, 2019). The fine-tuning process allows the specialized model to further learn the mapping from financial text to the set of category labels used by the Key Developments dataset. For validation, I evaluate the classifier's performance over a stratified random sample of 500 observations per news category from the Key Developments dataset. The FinBERT-based classifier achieves an overall precision of 0.90, a recall of 0.89, and an F1-score of 0.89 in the validation sample. I deploy the classifier on the full-length Wall Street Journal articles to obtain their news categories.

2.2 Measuring quantitativeness

I measure the amount of statistical content in each announcement using a "statistics ratio" based on the frequency of numerical tokens. First, I clean each textual summary by removing non-UTF-8 characters and English stop words. Second, I replace dates, addresses, ordinal numbers, and other common non-statistical tokens containing numeric symbols with placeholders. I detect and replace these tokens using regular expressions. For instance, "Jan 3, 2024" becomes "[DATE]" and "165 Whitney Ave, New Haven, CT 06511" becomes "[ADDR]". Third, I count the remaining numeric tokens, such as "5%" or "$1,000,000", and divide that count by the total word count. This ratio serves as a continuous measure of quantitativeness for each news announcement. For subsample analysis in later sections, I classify an announcement as a statistic if it contains at least one statistical token and as a story otherwise.

The quantitativeness of news announcements varies widely across news categories. Figure A.1 plots the 5th, 50th, and 95th percentiles of statistics ratios by category to illustrate these distributional differences. The categories are ordered from highest to lowest by their median statistics ratio. An intuitive examination hints at a potential story-statistics gap in stock market misreaction. Earnings announcements contain many statistical updates and are known to elicit underreaction (Ball and Brown, 1968; Bernard and Thomas, 1989; DellaVigna and Pollet, 2009). Mergers are generally more narrative-driven and are known to invoke overreaction (Jensen and Ruback, 1983; Agrawal, Jaffe, and Mandelker, 1992; Agrawal and Jaffe, 2000; Shleifer and Vishny, 2003). This coincidence suggests a negative relationship between statistical content and stock market misreaction, consistent with findings from Graeber, Roth, and Zimmermann (2024) and Ke (2025) that people tend to place less weight on statistics than on stories. I provide graphical and regression evidence on this negative relationship in Section 3.[1]

[1] For a complete list of news categories and their median statistics ratios, see Figure A.1 in the Appendix.

2.3 Measuring misreaction

I measure stock market misreaction toward a group of news announcements as the average post-announcement drift or reversal in the abnormal return of the announcing firm. For each news announcement, I define the immediate return as the buy-and-hold abnormal return of the announcing firm over the three-day event window centered on the announcement date. I include the returns from one day before and one day after the announcement date to account for pre-announcement leakage and after-hours releases (Kishore et al., 2008; Ho, Hong, and Zhang, 2025). I define the subsequent return as the buy-and-hold abnormal return of the announcing firm from the end of the event window to 90 days post-announcement, consistent with the baseline horizon in Kwon and Tang (2026). To compute abnormal returns, I estimate the factor loadings of the announcing firm for each news announcement by regressing its raw excess returns from 90 to two days before the announcement on the returns of the 13 thematic factors from Jensen, Kelly, and Pedersen (2023). I then generate predicted returns for the post-estimation period using these factor loadings. For both the event window and the post-announcement window, I calculate buy-and-hold abnormal returns as the difference between the raw and predicted buy-and-hold returns.

I regress subsequent return on immediate return while controlling for announcement year, firm, and news category fixed effects:

$$\textit{Subsequent return}_e = \textit{Fixed effects}_e + \beta \cdot \textit{Immediate return}_e + \varepsilon_e. \tag{1}$$

I include these fixed effects to reduce concerns about omitted variable bias. I cluster standard errors on announcement year, firm, and news category to allow for residual dependence along these dimensions. The regression coefficient β captures the average post-announcement drift

Table 1 shows the summary statistics for the key variables. Table A.3 in the Appendix shows the summary statistics for the key variables by information type. For each news category, Table A.4 in the Appendix shows its frequency in the sample, its share of all stories and statistics, and its own composition of information type.

or reversal in the abnormal returns of the announcing firm. I interpret this coefficient as the average extent of stock market misreaction to the announcements in the sample: a positive β indicates drift and underreaction, whereas a negative β indicates reversal and overreaction.

This interpretation of drift and reversal in abnormal returns as evidence of stock market misreaction has two important caveats. First, my approach does not strictly adhere to how Coibion and Gorodnichenko (2015) measure deviations from the full-information rational expectations (FIRE) benchmark. I use abnormal stock returns rather than forecast errors and forecast revisions. Nevertheless, Equation (1) closely parallels the regression of forecast error on forecast revision in their framework. I assume an environment in which the stock price can deviate from the FIRE benchmark in the short run but will eventually converge toward fundamentals over longer horizons (e.g., Shleifer and Vishny, 1997; Daniel, Hirshleifer, and Subrahmanyam, 1998). In this setting, the immediate return of a news announcement reflects the revision in investors' (potentially biased) best estimate of the firm's value. The subsequent return reflects the forecast error beyond the initial revision, because the stock price gradually converges to the rational benchmark in expectation. Second, I assume that the risk profile of the announcing firm is approximately constant around the news announcement, so that the estimated factor loadings do not meaningfully differ between the pre-announcement period, the event window, and the post-announcement period. Under this assumption, I account for risk-based explanations for return fluctuations by adjusting stock returns using a comprehensive set of risk factors.

To estimate how a variable affects stock market misreaction, I introduce an interaction term between the immediate return and the variable of interest. For example, to examine how the quantitativeness of a news announcement affects stock market misreaction, I standardize the statistics ratio and estimate the following regression:

$$
\begin{aligned}
\textit{Subsequent return}_e = \textit{Fixed effects}_e &+ \beta \cdot \textit{Immediate return}_e \\
&+ \gamma_{Stat} \cdot (\textit{Immediate return}_e \times \textit{Statistics ratio}_e) \\
&+ \textit{Additional controls}_e + \varepsilon_e.
\end{aligned} \tag{2}
$$

The regression coefficient on the interaction term γ_{Stat} measures the average change in the misreaction coefficient β following a one-standard-deviation increase in the statistics ratio.

2.4 Controls for alternative explanations

To assess whether the story-statistics gap adds to our understanding of stock market misreaction, I construct several variables to control for some alternative hypotheses of over- and underreaction.

Augenblick, Lazarus, and Thaler (2025) propose that people tend to overinfer from weak signals and underinfer from strong signals. They hypothesize that, while decision makers can accurately perceive the direction of signals, they cannot precisely assess their magnitude. This compresses perceived signal strength toward an intermediate level. Because the objective valuation impact of each news announcement is unobservable, it is difficult to measure its true signal strength. However, as long as investors form a noisy estimate of signal strength, the magnitude of the immediate return should reflect their potentially attenuated assessment. Consequently, to control for signal strength uncertainty, I use the absolute value of the immediate return as a proxy for the unobserved signal strength of the news announcement.

Kwon and Tang (2026) propose that diagnostic expectations are an important driver of both over- and underreaction. They find that a news announcement elicits stronger overreaction when prior news in the same category has been more extreme. Their interpretation is that categories with more extreme historical outcomes become more diagnostic of extreme events. This elevates investors' perceived likelihood of extreme outcomes in those categories, generating overreaction. Conversely, categories with milder histories appear more representative of non-tail outcomes, which leads to underreaction. To control for diagnostic expectations, I construct the extremeness of prior news ("prior extremeness") for each announcement. Following Kwon and Tang (2026), I fit a power-law distribution to the top decile of the absolute immediate returns of prior news in the same category over the preceding five years.

Tetlock, Saar-Tsechansky, and Macskassy (2008) demonstrate that news sentiment predicts stock market misreaction. They find that stock prices underreact to negative-sentiment words in firm-specific news announcements. Other studies also document meaningful influence of sentiment over investor belief formation (e.g., Hirshleifer and Shumway, 2003; Kamstra, Kramer, and Levi, 2003; Edmans, Garcia, and Norli, 2007). As summarized by Barberis (2018), these studies find that "an exogenous stimulus that improves (worsens) mood leads to more positive (negative) beliefs about [stock market performance]." To control for news sentiment, following the recent advent of large language models, I use FinBERT to generate sentiment labels (i.e., positive, negative, neutral) for each news summary.

Hou et al. (2025) propose that investor attention helps explain the heterogeneity in stock market misreaction. They show that post-earnings announcement drift arises under low investor attention due to insufficient reaction, whereas high attention amplifies price momentum by exacerbating behavioral biases such as overconfidence and extrapolative expectations. They also provide a comprehensive collection of attention measures. From this set, I control for monthly average daily share turnover and the average monthly number of RavenPack news articles over the prior 12 months. These proxies capture active trading and news supply, which are two key sources of investor attention. Because many attention proxies strongly correlate with firm size, I further control for the market capitalization of

Variable	Count	Mean	Std	p10	p25	p50	p75	p90
Subsequent return	1,325,878	0.033	0.277	-0.278	-0.109	0.028	0.164	0.334
Immediate return	1,325,878	0.003	0.064	-0.065	-0.024	0.002	0.030	0.073
Statistics ratio	1,325,878	0.066	0.069	0.000	0.000	0.038	0.130	0.171
Abs. immediate return	1,325,878	0.044	0.047	0.004	0.012	0.027	0.058	0.109
Prior extremeness	982,294	0.276	0.086	0.131	0.205	0.321	0.343	0.357
Positive sentiment	1,325,878	0.224	0.417	0.000	0.000	0.000	0.000	1.000
Neutral sentiment	1,325,878	0.540	0.498	0.000	0.000	1.000	1.000	1.000
Negative sentiment	1,325,878	0.236	0.425	0.000	0.000	0.000	0.000	1.000
Turnover	1,325,878	0.010	0.009	0.002	0.004	0.007	0.012	0.019
News articles	1,325,878	146	549	0	0	20	70	248
Market value (USD millions)	1,325,878	16,309	46,508	86	308	1,446	7,359	36,956

Table 1: Summary statistics

Note. Means and standard deviations are computed after winsorizing each variable at the 0.5th and 99.5th percentiles. Subsequent return is the buy-and-hold abnormal return for the announcing firm, relative to the 13 thematic factors from Jensen, Kelly, and Pedersen (2023), from two to 90 days post-announcement. Immediate return is the buy-and-hold abnormal return over the three-day window centered on the announcement date. Statistics ratio equals the number of statistical tokens in the news summary divided by its total word count. Prior extremeness corresponds to $\hat{\zeta}^{-1}$ in Kwon and Tang (2026) and is estimated over a rolling five-year window within each news category. Positive, neutral, and negative sentiment labels are generated from applying FinBERT on news summaries. Following Hou et al. (2025), turnover and news articles are the monthly average daily share turnover and the average monthly number of RavenPack news articles over the prior 12 months, respectively. Market value is the announcing firm's market capitalization on the announcement date.

the announcing firm on the announcement date.

After merging, cleaning, and constructing all variables, the final sample consists of 1,325,878 corporate news announcements spanning 38 news categories and 8,862 firms between 2003 and 2022. Table 1 reports summary statistics for the main variables used in the analysis.

3 The Story-Statistics Gap

In this section, I document a pronounced story-statistics gap in stock market misreaction. Section 3.1 groups corporate news announcements by their statistics ratio and shows that the tendency toward stock market underreaction rises with the quantitativeness of the announcement. Section 3.2 presents complementary regression evidence and demonstrates that the pattern is robust to controlling for alternative explanations for over- and underreaction. Because selective recall plays a central role in generating the story-statistics gap in laboratory settings, these empirical findings strongly motivate examining whether the same mechanism helps explain stock market misreaction. Section 3.3 evaluates the economic significance of the story-statistics gap by constructing a characteristic-managed portfolio. Finally, Section 3.4 connects these results to previously documented instances of the story-statistics gap across different settings.

3.1 Graphical evidence

To investigate whether the quantitativeness of a corporate news announcement affects its stock market response, I divide the sample into five bins based on statistics ratio and estimate the extent of misreaction within each bin using Equation (1). I assign all announcements that contain no statistical tokens to the "All Stories" bin, which accounts for roughly 30% of the sample. I then evenly divide the remaining announcements into quartiles by their statistics ratios, such that the "Stat-Q4" and "Stat-Q1" bins contain, respectively, the most and least quantitative announcements. For comparison with the "All Stories" bin, I also include an "All Statistics" bin that pools all announcements containing at least one statistical token. In other words, the "All Statistics" bin is the union of "Stat-Q1" through "Stat-Q4." I first study how within-category misreaction varies across different levels of statistical intensity. Within each bin, I estimate the misreaction regression with announcement-year, firm, and news-category fixed effects to control for unobserved heterogeneity along these dimensions. Standard errors are clustered at the announcement-year, firm, and news-category levels to allow for residual dependence along these dimensions.

Figure 1 presents the estimated stock market misreaction and its 95% confidence interval for each bin. These results illustrate a clear story-statistics gap in stock market misreaction: the estimated misreaction coefficient increases monotonically in the quantitativeness of the

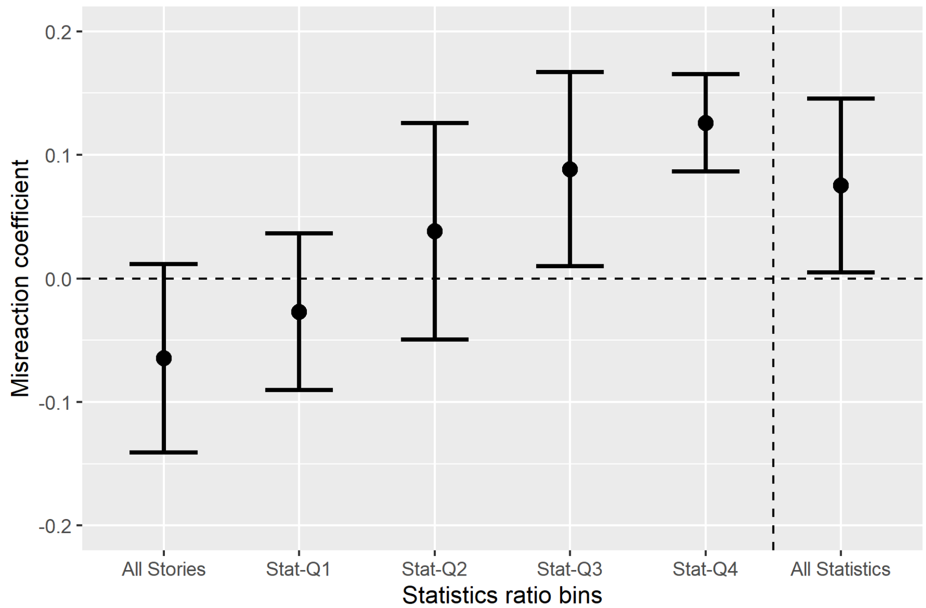


Figure 1: Within-category misreaction coefficients by statistics-ratio bins

Note. I divide the sample of corporate news announcements into bins based on their statistics ratio. The leftmost bin ("All Stories") contains any announcement that does not include a statistical token. The rightmost bin ("All Statistics") contains any announcement that includes at least one statistical token. Among announcements with at least one statistical token, I further sort the sample into quartiles by their statistics ratio. The four middle bins correspond to these quartiles in increasing order from left to right, such that "Stat-Q4" contains the most quantitative announcements. The misreaction coefficients are obtained by estimating Equation (1) separately within each bin. Positive (negative) coefficients indicate underreaction (overreaction). Error bars denote 95% confidence intervals. Standard errors are clustered at the announcement-year, firm, and news-category levels.

binned announcements. Whereas the "All Stories" bin exhibits borderline overreaction, the "All Statistics" bin shows significant underreaction. Moreover, while misreaction in the "Stat-Q1" and "Stat-Q2" bins is statistically insignificant, the "Stat-Q3" and "Stat-Q4" bins display significant and progressively stronger underreaction. The monotonicity in misreaction across both the extensive and intensive margins of quantitativeness suggests that information type shapes investor behavior, with investors tending to underreact more to statistics than to stories.

While Figure 1 is consistent with a story-statistics gap in stock market misreaction, news categories are not uniformly distributed across the statistics-ratio bins. As first documented by Kwon and Tang (2026), there is significant cross-category variation in stock market misreaction. Even with news-category fixed effects in the bin-specific regressions, Figure 1 could be driven by a few category-based misreaction patterns that cluster at the extremes of the quantitativeness spectrum. For example, earnings announcements, which are known to induce subsequent drift, are highly quantitative. At the same time, merger announcements, which are known to precede subsequent reversal, are highly qualitative.

To demonstrate that the story-statistics gap holds across news categories and contributes to the cross-category variation in stock market misreaction, I examine how misreaction relates to quantitativeness at the news category level. Figure 2 plots each news category by its misreaction coefficient and its statistics share. I estimate the misreaction coefficient of a news category by fitting Equation (1) on all the news announcements within that news category. I include only announcement-year and firm fixed effects because I estimate the misreaction coefficient separately for each news category. The statistics share of a category is the fraction of its announcements that contain at least one statistical token. To limit the influence of outliers, I exclude news categories with fewer than 100 observations per year-in-sample. Eight out of the 38 news categories are omitted as a result. Fitting an OLS regression across the news categories yields a significantly positive slope. This pattern indicates that more quantitative news categories tend to exhibit stronger stock market underreaction, consistent with the presence of a story-statistics gap.

3.2 Regression evidence

Having established the graphical intuition for the story-statistics gap in stock market misreaction, I examine its statistical significance using the regression framework in Equation (2). Specifically, I regress subsequent return on immediate return, the statistics ratio, and their interaction term using the full sample of corporate news announcements. As before, I include announcement year, firm, and news category fixed effects to mitigate omitted-variable concerns along these dimensions and cluster standard errors on the same variables to al-

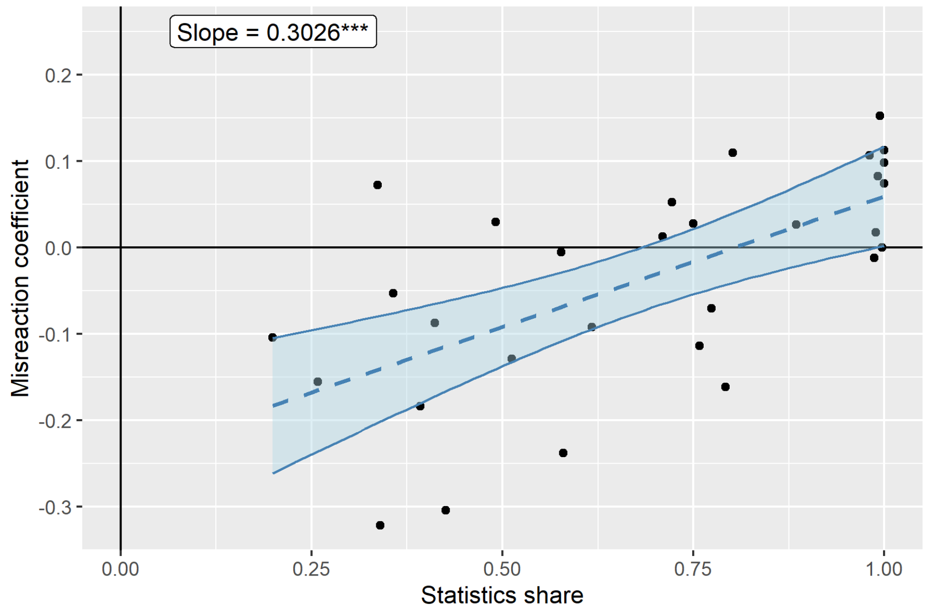


Figure 2: Misreaction coefficients vs. statistics shares across news categories

Note. Each point represents the misreaction coefficient and the statistics share of a news category. The misreaction coefficient of a news category is computed by estimating Equation (1) with announcement-year and firm fixed effects within the news category. The statistics share of a news category is the percentage of news announcements containing any statistics within that news category. News categories with fewer than 100 observations per year-in-sample are excluded to limit the impact of outliers. Eight out of the 38 news categories are omitted as a result. The dashed line is fit with OLS. The shaded region indicates the 95% confidence interval. Significance code: 0.1 * 0.05 ** 0.01 ***

low for residual dependence. To assess whether information type meaningfully affects stock market response, I test whether the coefficient on the interaction term differs from zero. A significantly positive (negative) coefficient indicates that the misreaction coefficient increases (decreases) with the statistics ratio, corresponding to stronger underreaction (overreaction). The graphical evidence in Section 3.1 suggests that the tendency toward underreaction rises with the quantitativeness of a news announcement. Accordingly, I expect a significantly positive coefficient on the interaction between immediate return and the statistics ratio.

Column (1) of Table 2 reports the baseline regression results. Although they are included in the regression, I do not report the main effects of the explanatory variables because only the interaction terms are relevant for understanding the source of variation in stock market misreaction. As expected, the estimated coefficient on the interaction between the immedi-

Controls	None (1)	Uncertainty (2)	Diagnostic (3)	Sentiment (4)	Attention (5)	All (6)
	Subsequent return					
Immediate return	0.009	-0.110***	-0.030	-0.002	-0.016	-0.146***
	(0.038)	(0.033)	(0.041)	(0.033)	(0.030)	(0.025)
Immediate return × Statistics ratio	0.062***	0.061***	0.045***	0.063***	0.055***	0.039**
	(0.003)	(0.007)	(0.015)	(0.011)	(0.005)	(0.015)
Immediate return × Abs. immediate ret.		0.047**				0.041**
		(0.020)				(0.019)
Immediate return × Prior extremeness			-0.027			-0.024
			(0.016)			(0.018)
Immediate return × Neg. sentiment				-0.006		-0.004
				(0.028)		(0.028)
Immediate return × Pos. sentiment				0.063***		0.045
				(0.019)		(0.028)
Immediate return × Turnover					-0.019**	-0.020**
					(0.008)	(0.008)
Immediate return × News articles					-0.003	0.000
					(0.010)	(0.008)
Immediate return × Log market value					-0.051**	-0.019
					(0.021)	(0.019)
Observations	1,325,878	1,325,878	982,294	1,325,878	1,325,878	982,294
R^2	0.084	0.086	0.082	0.084	0.114	0.117
Main effects of controls	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Firm fixed effects	✓	✓	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓	✓	✓
Clustered standard errors	✓	✓	✓	✓	✓	✓

Table 2: Misreaction regression with interacted statistics ratio

Note. Each column reports the regression coefficients from Equation (2) estimated on the full sample of corporate news announcements. Variables other than subsequent and immediate returns are standardized for interpretability. The main effects of the control variables are included in the regression but omitted from display. Standard errors, shown in parentheses, are clustered at the announcement year, firm, and news category levels. Significance code: 0.1 * 0.05 ** 0.01 ***.

ate return and the statistics ratio is positive and highly statistically significant. This result indicates that more quantitative news announcements tend to generate stronger underreaction, confirming the graphical patterns in Figures 1 and 2. For interpretability, I standardize the statistics ratio as well as any additional explanatory variables, except for immediate and subsequent returns. The significant point estimate of 0.06 on the interaction coefficient implies that, on average, the misreaction coefficient increases by six basis points for a one-standard-deviation increase in the statistics ratio.

Next, I investigate whether signal strength uncertainty, diagnostic expectations, news sentiment, and investor attention can explain the story-statistics gap in stock market misreaction. In theory, each of these mechanisms could generate a story-statistics gap. Table A.3 in the Appendix reports the summary statistics by information type. As a proxy for unobserved signal strength, the absolute value of the immediate return is larger for statistics, which could generate stronger underreaction to statistics in a setting with signal strength uncertainty. Past news within the same category tends to be more extreme for stories, which could induce greater overreaction to stories under diagnostic expectations. Moreover, news sentiment tends to be more negative for statistics, which could translate into more underreaction to statistics according to Tetlock, Saar-Tsechansky, and Macskassy (2008). At the same time, all three attention proxies—share turnover, news coverage, and firm size—are higher for stories, which could rationalize the story-statistics gap through the amplification of extrapolative biases under high investor attention.

To demonstrate that these control variables function as expected, Table A.6 in the Appendix reports the regression results under the framework of Equation (2) for each control variable, taken separately. The coefficients on the interacted control variables behave as expected. Augenblick, Lazarus, and Thaler (2025) find that signal strength uncertainty can lead to overinference from weak signals and underinference from strong signals. Accordingly, the interaction term for the absolute value of immediate return has a significantly positive coefficient, which suggests that stronger signal strength tends to increase stock market underreaction. Kwon and Tang (2026) document that overreaction is more severe following higher extremeness of prior news and attribute this result to diagnostic expectations. The significantly negative coefficient on the interaction term for prior extremeness confirms their finding in my sample. Tetlock, Saar-Tsechansky, and Macskassy (2008) demonstrate that underreaction to firm-specific news is more pronounced following news announcements with negative-sentiment words. The significantly positive coefficient on the negative-sentiment dummy aligns with their conclusion. I also find underreaction to positive-sentiment news announcements in my sample, which echoes their observation of some delayed reaction to positive news. Hou et al. (2025) report more underreaction during episodes of lower investor attention and vice versa. In line with their results, the interaction coefficients for turnover, number of RavenPack news articles, and firm size are all significantly negative. Notably, the interaction coefficient for statistics ratio in column (1) of Table 2 has the same order of magnitude as those for the control variables in Table A.6 in the Appendix. This suggests that the story-statistics gap and the three controlled hypotheses share similar economic significance in explaining the heterogeneity in stock market misreaction.

To determine whether these mechanisms of stock market misreaction explain away the story-statistics gap, I gradually introduce the control variables and their interactions with the immediate return to the regression in column (1). I test whether the regression coefficient on the interaction between the immediate return and the statistics ratio becomes statistically indistinguishable from zero after controlling for these known drivers of over- and underreaction. Columns (2) through (5) of Table 2 separately control for signal strength uncertainty, diagnostic expectations, news sentiment, and investor attention. Column (6) simultaneously controls for all four alternative hypotheses. In each specification, the interaction coefficient for statistics ratio remains significantly positive. This robustness suggests that these mechanisms are not the primary drivers of the story-statistics gap in stock market misreaction.

In addition to the robustness of the story-statistics gap, Table 2 shows that the interacted coefficients on the absolute value of immediate return and turnover are significant predictors of stock market misreaction across various regression settings. Together with the regression results from Table A.6 in the Appendix, these findings provide empirical support for the signal strength uncertainty and attention mechanisms proposed by Augenblick, Lazarus, and Thaler (2025) and Hou et al. (2025), respectively.

Furthermore, while the extremeness of prior news in the same category does not fully explain the story-statistics gap, the magnitude of the interaction coefficient for statistics ratio moderately declines after controlling for it. This finding suggests potential commonalities between diagnostic expectations and the main driver of the story-statistics gap. The experimental evidence from Graeber, Roth, and Zimmermann (2024) strongly motivates investigating selective recall when people react differently to stories and statistics. Indeed, Kwon and Tang (2026) note that similarity-based recall is a shared fundamental premise for both selective recall and diagnostic expectations. Bordalo et al. (2023) additionally argue that selective recall provides a microfoundation for diagnostic expectations. In light of the empirical results, these theoretical links further motivate examining stock market misreaction through the lens of selective recall.

To demonstrate that the story-statistics gap does not rely on the use of news summaries from the Key Developments database, I replicate Table 2 with full-length Wall Street Journal articles. Table A.7 in the Appendix presents the regression results. As in Table 2, the interacted coefficient on the statistics ratio is significantly positive across all regression specifications.

In Table A.8 in the Appendix, I additionally demonstrate that the story-statistics gap in the regression setting does not rely on well-documented anomalies, such as post-earnings announcement drift or post-merger reversal. I re-run the regression specification from column (1) of Table 2 over the sample of corporate news announcements outside of these news categories. The interacted coefficient on the statistics ratio remains significantly positive across all subsamples. These results echo the cross-category evidence from Figure 2: the story-statistics gap in stock market misreaction is a general phenomenon that goes beyond well-documented anomalies.

3.3 Economic significance

To assess the economic significance of the story-statistics gap, I construct a characteristic-managed portfolio that uses the statistics ratio as a proxy for stock market misreaction. The story-statistics gap predicts that stock market underreaction increases in the quantitativeness of a news announcement. A firm will have high returns this month if it has many positive statistics, negative stories, or both in the previous month. By the same logic, a firm will have low returns this month if it has many negative statistics, positive stories, or both in the previous month. Therefore, for each firm-month, I calculate the average statistics ratio and the average immediate return across the firm's news announcements in the previous month. The average statistics ratio helps predict the direction and extent of stock market misreaction, and the average immediate return helps determine the overall direction of news. I standardize the two characteristics across firms within the previous month and take their firm-level products as the portfolio weights for the current month. These portfolio weights define the misreaction-managed portfolio. The long leg contains the most underreacted (statistical) firms with the highest immediate returns (positive news), as well as the most overreacted (story-like) firms with the lowest immediate returns (negative news). The short leg contains the most underreacted (statistical) firms with the lowest immediate returns (negative news), as well as the most overreacted (story-like) firms with the highest immediate returns (positive news). The total return of the misreaction-managed portfolio is the buy-and-hold return on the monthly difference between the long and short legs.

Figure 3 plots the cumulative excess returns of the misreaction-managed portfolio as well as those for the long and short legs. All portfolios are scaled to 10% annual volatility to facilitate interpretability. Over the sample period, the portfolio delivers an annualized Sharpe ratio of 1.31. As a reference point, the annualized Sharpe ratio on the market factor is 0.65 over the same sample period. I test for abnormal returns by regressing the monthly excess returns of the portfolio against a variety of factor models over the same period. Table 3 presents the parameter estimates. The misreaction-managed portfolio earns approximately 110 basis points per month in abnormal returns against each of the five tested factor models. The estimated alphas are also highly statistically significant. These results corroborate the economic significance of the story-statistics gap as an important driver of stock market

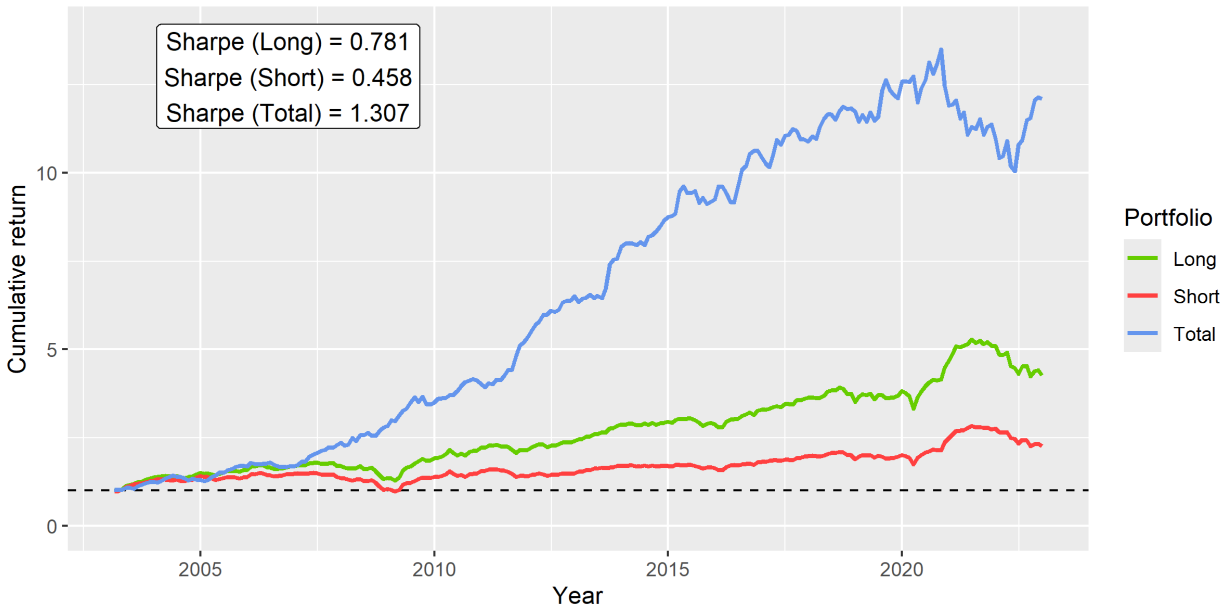


Figure 3: Cumulative returns from the misreaction-managed portfolio on statistics ratio

Note. For each firm-month, I calculate the average statistics ratio and the average immediate return across the firm's news announcements in the previous month. I standardize the two characteristics across firms within each month and take their firm-level products as the portfolio weights for the current month. The long leg contains the most underreacted (statistical) firms with the highest immediate returns, as well as the least underreacted (story-like) firms with the lowest immediate returns. The short leg contains the most underreacted (statistical) firms with the lowest immediate returns, as well as the least underreacted (story-like) firms with the highest immediate returns. The total return of the misreaction-managed portfolio is the buy-and-hold return on the monthly difference between the long and short legs. All Sharpe ratios are annualized. All portfolios are scaled to 10% annual volatility to facilitate interpretability.

misreaction.

To ensure the robustness of these results to the exclusion of small firms, I construct the misreaction-managed portfolio with only large firms. I focus on firms whose average market capitalization across their news announcements in the previous month is above the 70th percentile across all CRSP firms in the same month. Figure A.9 and Table A.10 in the Appendix document an annualized Sharpe ratio of 0.85 and a highly significant monthly abnormal return of 70 basis points. The significantly positive alphas across both settings further underscore the economic significance of the story-statistics gap in stock market misreaction.

Furthermore, I note that the misreaction-managed portfolio resembles a news momentum

Factor model	Portfolio excess return FF3 (1)	FF3+Mom. (2)	FF5 (3)	FF5+Mom. (4)	JKP13 (5)
Constant	0.011***	0.011***	0.012***	0.012***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Market excess return	-0.042	-0.035	-0.056	-0.047	
	(0.045)	(0.049)	(0.046)	(0.049)	
Small minus big	0.149*	0.150*	0.112	0.112	
	(0.082)	(0.082)	(0.088)	(0.088)	
High minus low	-0.242***	-0.235***	-0.179**	-0.166**	
	(0.061)	(0.064)	(0.077)	(0.081)	
Up minus down		0.017		0.025	
		(0.049)		(0.049)	
Conservative minus aggressive			-0.151	-0.159	
			(0.125)	(0.126)	
Robust minus weak			-0.105	-0.107	
			(0.103)	(0.103)	
Observations	239	239	239	239	239
R^2	0.068	0.069	0.077	0.078	0.093

Table 3: Regression of misreaction-managed portfolio excess returns on factor models

Note. Misreaction-managed portfolio excess returns and factor returns are measured at the monthly frequency. Coefficient estimates for the 13 thematic factors from Jensen, Kelly, and Pedersen (2023) are omitted for brevity. Significance code: 0.1 * 0.05 ** 0.01 ***.

strategy (Jiang, Li, and Wang, 2021). Whereas a news momentum strategy would long the firms with the most positive news announcements and short those with the most negative ones, the misreaction-managed portfolio additionally redirects trades based on the quantitativeness of the news announcements. If the story-statistics gap is economically significant, then the misreaction-managed portfolio should outperform a comparable news momentum strategy by reallocating trades more efficiently. Therefore, comparing the returns between the misreaction-managed portfolio and a comparable news momentum strategy further informs the economic importance of the story-statistics gap in stock market misreaction. To construct the comparable news momentum strategy, I change the portfolio weights from the product between the two standardized firm-level characteristics to only the standardized average immediate returns across news announcements in the previous month. The two strategies are otherwise identical. Figure A.11 in the Appendix plots the cumulative excess returns for both the misreaction-managed portfolio and the comparable news momentum strategy. Their annualized Sharpe ratios are 1.31 and 0.85, respectively. Table A.12 in the Appendix reports the factor model regressions for the comparable news momentum strategy. Whereas the misreaction-managed portfolio generates 110 basis points of abnormal returns per month, the comparable news momentum strategy earns approximately 70 basis points of monthly alpha. The superior performance of the misreaction-managed portfolio supports the economic significance of the story-statistics gap in stock market misreaction.

3.4 Connection to existing literature

The story-statistics gap in stock market misreaction aligns with the experimental evidence from Graeber, Roth, and Zimmermann (2024). Using a series of controlled experiments in a product review setting, they find that the effect of stories on beliefs is longer-lasting than that of statistics. My results complement their findings by providing field evidence on how people interact differently with qualitative and quantitative information. While they do not directly examine misreaction patterns across information types, both sets of results suggest that people tend to place less decision weight on statistics than on stories. In terms of mechanism, they experimentally demonstrate that selective recall can explain the story-statistics gap in belief impact through the semantic association between memory prompts and stories. Their results strongly motivate examining the story-statistics gap in stock market misreaction through the lens of selective recall and semantic association. Building on their insights, I apply the model of selective recall from Bordalo et al. (2023) to the stock market setting. Based on the importance of semantic association for recalling stories in their laboratory setting, I hypothesize distinct similarity structures for qualitative and quantitative corporate news announcements.

Moreover, the story-statistics gap in stock market misreaction strongly echoes the qualitative-quantitative gap in analyst misreaction documented by Ke (2025). Using Coibion and Gorodnichenko (2015) regressions instrumented with LLM embeddings from sell-side reports, Ke (2025) finds that equity analysts tend to underreact more to statistics than to stories. My results on the story-statistics gap in stock market misreaction complement his discovery by documenting the same empirical pattern in the stock market setting. The recurrence of this pattern across both individual and aggregate settings speaks to its economic significance. It also highlights the tight connection between stock market dynamics and analyst forecasts. Despite these similarities in the empirical patterns, Ke (2025) and I propose different mechanisms. Ke (2025) explains the story-statistics gap in analyst misreaction with overconfidence for stories and herding for statistics. I propose that, in theory, selective recall alone can explain the gap. Having derived the predictions of the model in Section 4.4, I now proceed to test the empirical implications of selective recall in the stock market setting.

4 The Model of Selective Recall

In this section, I incorporate the experimental findings from Graeber, Roth, and Zimmermann (2024) into a model of selective recall and derive its implications for stock market misreaction. In Section 4.1, I apply the model of selective recall from Bordalo et al. (2023) to the stock market setting. In Section 4.2, I propose distinct criteria of similarity for stories and for statistics based on the experimental findings from Graeber, Roth, and Zimmermann (2024). I hypothesize that stories within the same news category are similar because they share common semantic and narrative features, whereas statistics with the same level of valuation impact are similar because they are defined by numerical precision. I formalize these assumptions and derive the theoretical misreactions for stories and for statistics as functions of similarity relationships and relative frequencies among past news announcements. In Section 4.3, I test whether the model predicts the story-statistics gap in stock market misreaction based on the measured similarity relationships and observed relative frequencies among corporate news announcements. In Section 4.4, to address potential errors in measuring similarity, I derive several testable, measurement error-free empirical implications from the model that relate variation in predicted misreaction under the model solely to variation in the relative frequencies among past news announcements, which are precisely observed. In Section 4.5, I derive and test an additional model implication for pairwise differences in subjective valuation impact across news announcements.

4.1 Model setup

Consider an investor who encounters a corporate news announcement e that belongs to news category $c(e) \in \mathcal{C}$ and information type $t(e) \in \{Story, Statistic\}$. The news carries a valuation implication $\pi(e) \in \mathbb{R}$ for the stock price of the announcing firm $P_{i(e),t}$ such that $P_{i(e),t+1}/P_{i(e),t} = 1 + \pi(e)$. The direction of the news is either positive $sign(\pi(e)) = +1$ or negative $sign(\pi(e)) = -1$. The *valuation impact* of the news is the magnitude of its valuation implication $|\pi(e)|$. The news is either *high-impact* ($|\pi(e)| = \pi_1$ for some $\pi_1 > 0$) or *low-impact* ($|\pi(e)| = \pi_2$ for some $\pi_2 \in [0, \pi_1)$). Each news category has a fixed, unobserved probability of producing high-impact events. The investor observes the news category $c(e)$, information type $t(e)$, and direction $sign(\pi(e))$ of the current news announcement, but he does not observe its valuation impact $|\pi(e)|$. He therefore seeks to infer the unobserved magnitude to inform his trading decision.[2]

The investor relies on memory to form an assessment. Let E denote the experience database that contains all past news announcements available to the investor's memory. For any past news in the experience database $e' \in E$, the investor observes the news category $c(e')$, information type $t(e')$, direction $sign(\pi(e'))$, and valuation impact $|\pi(e')|$. A

[2] One microfoundation for investors' inability to observe stories' valuation impact is that qualitative information tends to be imprecise. One microfoundation for investors' inability to observe statistics' valuation impact is that, even though investors observe the same numerical update, they may adopt different valuation models that assign different levels of valuation impact to the same information input. If investors do not perfectly observe the prevalence of different valuation models in the economy, then they cannot derive the valuation impact of the statistical news with certainty, even when the information is precise.

past news announcement is *relevant* to assessing the valuation impact of the current news announcement if it belongs to the same news category $c(e') = c(e)$. Let $D \subseteq E$ denote the set of relevant past news announcements and $D' = E \backslash D$ the set of irrelevant ones. Additionally, a past news announcement is *unrelated* to the current news announcement if they differ in information type $t(e') \neq t(e)$ or news direction $sign(\pi(e')) \neq sign(\pi(e))$. Let $G \subseteq E$ denote the set of related past news announcements and $G' = E \backslash G$ the set of unrelated ones. Moreover, a past news announcement is high-impact if $\pi(e') = \pi_1$. Let $H_1 \subseteq E$ denote the set of high-impact past announcements and $H_2 = E \backslash H_1$ the set of low-impact ones. Accordingly, $H_i \cap D \cap G$ and $H_i \cap D' \cap G$ are the subsets of relevant and irrelevant news announcements within the valuation impact level H_i that are related to the current news announcement, respectively. For simplicity, I write $H_i^D = H_i \cap D \cap G$ and $H_i^{D'} = H_i \cap D' \cap G$ for $i \in \{1, 2\}$. I denote the number of past announcements that belong to an experience set $A \subseteq E$ by its cardinality $|A|$. These experience subsets partition the experience database into mutually exclusive components, such that $E = H_1^D \cup H_1^{D'} \cup H_2^D \cup H_2^{D'} \cup G'$ and $|E| = |H_1^D| + |H_1^{D'}| + |H_2^D| + |H_2^{D'}| + |G'|$.

The investor attempts to separately recall high- and low-impact experiences within the relevant news category to infer the outcome distribution of the current news announcement. He focuses on recalling high-impact, relevant experiences H_1^D in one attempt and on recalling low-impact, relevant experiences H_2^D in another. Let the *target subset* be the set of experiences the investor intends to retrieve, which is either H_1^D or H_2^D. In each of the two attempts, he retrieves a limited number $M \in \mathbb{N}$ of past news announcements from the memory database E with replacement. Owing to the stochastic nature of memory, the investor cannot control which specific experiences come to mind. All memories in the experience database simultaneously compete for retrieval. According to similarity-based recall, past experiences that are more *similar* to the target subset are more likely to be retrieved. Let $S(\cdot, \cdot)$ be a non-negative and symmetric similarity function. $S(e', e'')$ denotes the conceptual similarity between past experiences $e', e'' \in E$, and $S(e', A)$ denotes the average similarity between e' and each experience in a subset of experiences $A \subseteq E$. When the investor focuses on recalling experiences from a particular target subset H_i^D for $i \in \{1, 2\}$, the probability of recalling some past experience $e' \in E$ in a single draw is

$$r(e', H_i^D) = \frac{S(e', H_i^D)}{\sum_{e'' \in E} S(e'', H_i^D)}. \tag{3}$$

The investor can end up recalling any experience $e' \in E$, provided that it is not entirely dissimilar to the target subset $S(e', H_i^D) > 0$. This means he can accidentally retrieve past news announcements outside of the target subset, such as those in different news categories or impact levels.

The probability that a single recall belongs to the target subset H_i^D is the sum of recall probability for all past experiences in the target subset:

$$\begin{gathered} r(H_i^D) = \sum_{e' \in H_i^D} r(e', H_i^D) = \frac{\sum_{e' \in H_i^D} S(e', H_i^D)}{\sum_{e'' \in E} S(e'', H_i^D)} = \\ \frac{\sum_{e' \in H_i^D} S(e', H_i^D)}{\left(\begin{gathered} \sum_{e'' \in H_i^D} S(e'', H_i^D) + \sum_{e'' \in H_{-i}^D} S(e'', H_i^D) + \sum_{e'' \in H_i^{D'}} S(e'', H_i^D) \\ + \sum_{e'' \in H_{-i}^{D'}} S(e'', H_i^D) + \sum_{e'' \in G'} S(e'', H_i^D) \end{gathered} \right)}. \end{gathered} \tag{4}$$

To simplify the notation, for experience sets $A, B \subseteq E$, let $S_{AB} = S(A, B)$ denote the cross-similarity between A and B, which represents the average pairwise similarity across A and B. Additionally, let $S_A = S(A, A)$ denote the self-similarity of A, defined as the average pairwise similarity within A. Using this notation, I rewrite the probability that a single recall belongs to the target subset H_i^D as

$$r(H_i^D) = \frac{S_{H_i^D} |H_i^D|}{S_{H_i^D} |H_i^D| + S_{H_{-i}^D H_i^D} |H_{-i}^D| + S_{H_i^{D'} H_i^D} |H_i^{D'}| + S_{H_{-i}^{D'} H_i^D} |H_{-i}^{D'}| + S_{G' H_i^D} |G'|}. \tag{5}$$

While the investor does not control the retrieval process, he identifies which recalled experiences belong to the target subset. He discards any accidental retrievals from outside of the target subset. Importantly, even though the non-target experiences do not themselves enter the decision-making process, they still affect investor decisions by occupying limited cognitive resources. If K out of the M recalled experiences do not belong to the target subset, the investor does not redraw another K experiences from the experience database. He simply admits the $M-K$ successful recalls from the target subset into the decision-making process. Through the investor's cognitive limit, the accidentally retrieved, non-target experiences *interfere* with the recall of the target subset and affect decision outcome.

After the retrieval process, the investor compares the number of relevant recalls at each impact level to estimate the odds that the current situation is high-impact, denoted by p. Because experiences are drawn with replacement, the expected number of recalls from H_i^D is $n(H_i^D) = r(H_i^D)M$. He calculates

$$p = \frac{n(H_1^D)}{n(H_2^D)} = \frac{r(H_1^D)}{r(H_2^D)}. \tag{6}$$

The investor bases his trading decision on these subjective odds. Ideally, the investor would like to retrieve experiences across impact levels in proportion to their true frequencies within the relevant news category. Conditional on his experience set E, the best estimate for the odds that the current situation is high-impact equals the objective relative frequency of high- versus low-impact experiences among the relevant and related past news announcements:

$$p^* = \frac{\Pr(|\pi(e)| = \pi_1 \mid e \in D \cap G)}{\Pr(|\pi(e)| = \pi_2 \mid e \in D \cap G)} = \frac{|H_1^D|}{|H_2^D|}. \tag{7}$$

These objective odds simply reflect how often high-impact events have occurred relative to low-impact events among past announcements in the relevant news category. Therefore, misreaction occurs when the subjective odds deviate from the conditional best estimate:

$$\%\Delta p = \frac{p - p^*}{p^*} = \frac{p}{p^*} - 1 = \frac{r(H_1^D)}{r(H_2^D)} \frac{|H_2^D|}{|H_1^D|} - 1 \neq 0. \tag{8}$$

The model provides a continuous prediction for misreaction. Specifically, overreaction occurs when the investor overestimates the odds that the news announcement is high-impact $\%\Delta p > 0$, and underreaction occurs when the investor underestimates the odds that the news announcement is high-impact $\%\Delta p < 0$.

Similarity-based recall is the key mechanism that allows misreactions to occur. In the model, misreaction arises because recall probabilities diverge from the true frequencies of experiences, which can happen when similarity varies across different pairs of experiences. To see this, consider the benchmark case in which every pair of experiences in E is equally similar, such that $S(e', e'') = a$ for all $e', e'' \in E$ for some $a > 0$. In this setting, the average pairwise similarity between any two experience sets also equals a. Because all past experiences are equally likely to be retrieved, the similarity mechanism is effectively turned off. The recall probability for H_i^D then becomes

$$r(H_i^D) = \frac{a|H_i^D|}{a|H_i^D| + a|H_{-i}^D| + a|H_i^{D'}| + a|H_{-i}^{D'}| + a|G'|} = \frac{|H_i^D|}{|E|}, \tag{9}$$

and the subjective odds that the current situation is high-impact are

$$p = \frac{r(H_1^D)}{r(H_2^D)} = \frac{|H_1^D|/|E|}{|H_2^D|/|E|} = \frac{|H_1^D|}{|H_2^D|} = p^*. \tag{10}$$

In this benchmark, the investor correctly infers the magnitude of the news because the subjective odds equal the conditional best estimate, so $\%\Delta p = 0$.

This equivalence does not necessarily hold when similarity differs across experience pairs. For example, suppose experiences in H_i^D are completely distinct from all others, such that

$S(e', H_i^D) = 0$ for all $e' \in E \backslash H_i^D$. Let $S_{H_i^D} = a$ for some $a > 0$. The recall probability for H_i^D becomes

$$r(H_i^D) = \frac{a|H_i^D|}{a|H_i^D| + 0|H_{-i}^D| + 0|H_i^{D'}| + 0|H_{-i}^{D'}| + 0|G'|} = 1, \tag{11}$$

and the subjective odds that the current situation is high-impact are

$$p = \frac{r(H_1^D)}{r(H_2^D)} = 1. \tag{12}$$

In this case, the investor recalls $n(H_1^D) = r(H_1^D)M = M$ instances of high-impact news announcements and $n(H_2^D) = r(H_2^D)M = M$ instances of low-impact news announcements in the relevant news category. Even in a world where the low-impact news announcements are far more numerous than the high-impact news announcements in the relevant news category, the investor still ends up recalling M instances from each impact level due to his cognitive limit. As a result, he perceives high- and low-impact events as equally likely, regardless of their true frequencies across the entire experience database. Consequently, he overreacts to the current news announcement when low-impact news are more frequent $p^* < 1$ and underreacts when the reverse is true $p^* > 1$.

These examples illustrate the key idea that interference does not always induce the subjective estimate to deviate from the objective benchmark. When our limited cognition fails to retrieve the entire experience database at the point of decision-making, we require interference from non-target experiences to form an approximate sense of the overall distribution. It is similarity-based recall that leads to the misreactions. When similarity-based recall is turned off, as in the benchmark case, interference from non-target experiences helps us arrive at the correct odds estimate despite our limited cognitive capacity. When interference from non-target experiences is turned off, as in the second example, we lose sight of the target subset's true frequency relative to the entire experience database. In this extreme case, our subjective estimate becomes invariant to objective frequencies, which can lead to large, persistent misreaction.

4.2 Similarity patterns

Graeber, Roth, and Zimmermann (2024) highlight that memory interacts with stories and statistics in distinct ways. Through controlled experiments, they document a pronounced difference in belief impact between the two, showing that the influence of statistics fades much more quickly than that of stories. Consistent with the model of selective recall in Bordalo et al. (2023), they find that the persistence of belief impact from a story increases in its *semantic* similarity with the memory prompt. They also show that the persistence of belief impact from a story decreases in the semantic similarity between non-target stories and the memory prompt, which lends support to the interference mechanism. These findings underscore the central role of semantic associations in how memory evaluates similarity for stories. Graeber, Roth, and Zimmermann (2024) suggest that the story-statistics gap arises in their experimental setting because both memory prompts and stories are communicated through words, which carry semantic meaning. In contrast, the absence of semantic meaning for statistics removes an important dimension along which similarity can be assessed. As a result, past stories are easier to retrieve at future decision points than past statistics, leading to more persistent belief impact from qualitative information.

Building on the insight that the presence of semantic meaning drives the distinction between stories and statistics, I hypothesize that memory applies distinct criteria of similarity to these two types of information. I propose that stories are similar only within each news category, whereas statistics are similar only within each valuation-impact level. Figure 4 provides the graphical intuition for the resulting interference patterns. For stories, semantic features such as narrative structure and diction vary substantially across news categories. For example, press releases announcing CEO changes emphasize credentials and track records, whereas merger announcements focus on synergy and strategic vision. At the same time, regardless of their valuation impact, announcements within a given news category share similar vocabulary, tone, and communication goals. In light of the graphical framework, this implies that the recall of past stories is primarily subject to *vertical interference*, in which the investor accidentally retrieves non-target stories within the same news category that belong to the opposite impact level. For statistics, numerical precision is the defining feature. In the absence of semantic meaning, statistics across news categories are similar so long as they imply the same valuation impact. For example, a one-time, upfront loss of $50 million from a legal defeat resembles a one-time, upfront loss of $50 million from a factory fire, even though each contrasts sharply with a one-time, upfront loss of $1 million within its own news category. Accordingly, I argue that the recall of past statistics is primarily subject to *horizontal interference*, where the investor inadvertently retrieves non-target statistics from irrelevant news categories that nonetheless belong to the correct impact level.

Under the hypothesized similarity criteria, stories and statistics are dissimilar to each other. Stories associate with other experiences through semantic similarity, but statistics are comprised of numbers that possess little semantic meaning. Statistics associate with other experiences that carry the same level of valuation impact, but stories lack the numerical precision to convey a clear valuation impact. Consequently, stories and statistics only possess features that appeal to their respective similarity criteria, while lacking the attributes to satisfy those for the other. This makes past experiences of different information types dissimilar in memory. Because similarity drives recall probability, investors are unlikely to retrieve past experiences that do not share the same information type as the current news announcement.[3]

I further hypothesize that positive and negative news announcements are dissimilar for both stories and statistics. For stories, semantic features such as tone and sentiment create a clear distinction between positive and negative news. For statistics, this distinction is

[3]Admittedly, empirically, quantitative updates often simultaneously contain qualitative descriptions. In my model, I assume that the numerical figures are key objects in such releases that dominate investors' memory of the information release. Investors' ultimate goal is to assess the unknown valuation impact of future news announcements. Since both valuation impact and statistical figures are numerical objects, I assume that memory prioritizes the encoding of statistics over that of stories when both are present within the same experience.

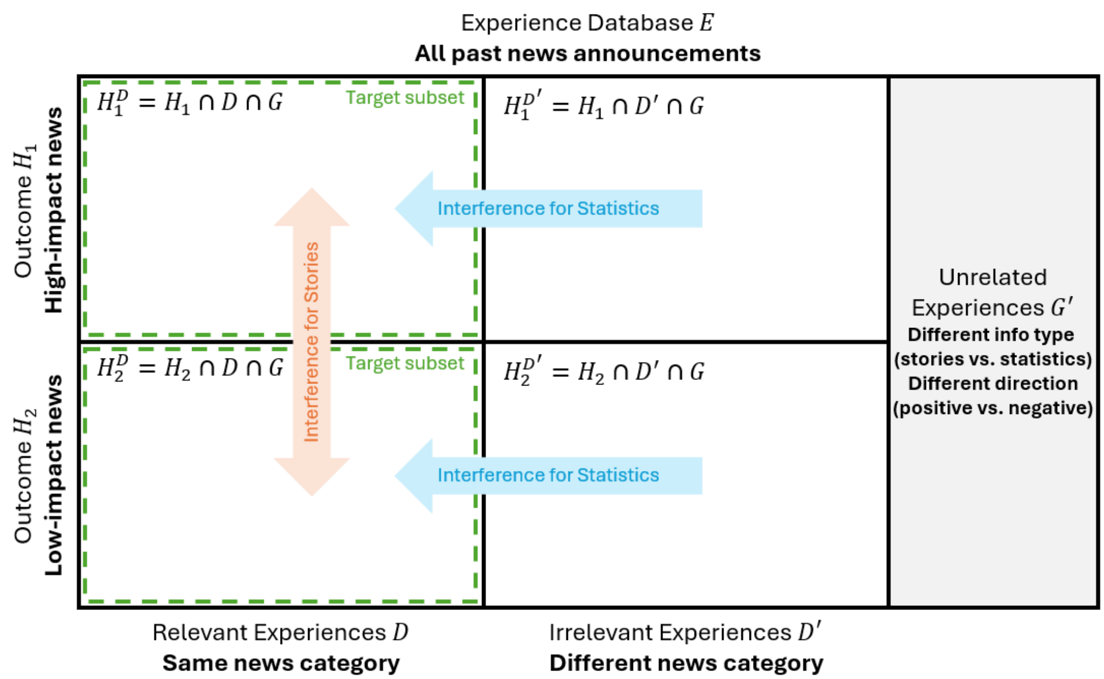


Figure 4: Hypothesized model of selective recall in the stock market context

Note. This diagram illustrates the hypothesized behavior of investors' memory when they estimate the valuation impact of a corporate news announcement. Investors draw from an experience database of past news announcements (E). They recall a limited number of previous news announcements that share the same information type (stories vs. statistics) and direction (positive vs. negative) as the current news announcement ($E \backslash G'$). They consider two possible valuation impact outcomes: the news announcement is either high-impact (H_1) or low-impact (H_2). Previous news announcements within the same news category as the current news announcement is relevant (D), while those outside of that news category are irrelevant (D'). Investors aim to recall previous news announcements under each target subset, which represents a possible valuation impact outcome within the relevant news category. Because similarity drives recall, non-target but similar experiences can interfere with the recall of targeted experiences by occupying limited cognitive resources. Distortions in recall probability due to variations in similarity relationships can lead to misreaction. I hypothesize that stories are not subject to interference across news categories due to semantic distinctions, whereas statistics are not subject to interference across valuation impact levels due to numerical precision. The arrows illustrate the hypothesized interference patterns.

even sharper because numeric data unambiguously include the sign of the valuation impact. Therefore, the investor should experience little interference from past news announcements of the opposite sign.

I incorporate the hypothesized interference patterns into the model. For stories, large semantic differences make stories from distinct news categories dissimilar. Under the similarity-driven interference mechanism, irrelevant stories do not interfere with the recall of the target subset. For statistics, conspicuous magnitude differences make statistics from distinct impact levels dissimilar. This implies that statistics with different valuation impact do not interfere with the recall of the target subset. In the model, these interference patterns translate into $S_{H_i^D H_i^{D'}} = 0$ when $t(e) = \mathit{Story}$ and $S_{H_i^D H_{-i}^D} = 0$ when $t(e) = \mathit{Statistic}$ for $i \in \{1, 2\}$. To reflect the dissimilarity across information types, I set the similarity between any story and any statistic to zero, so that $S(e', e'') = 0$ if $t(e') \neq t(e'')$. Similarly, to reflect the dissimilarity between positive and negative news, I set the similarity between any two announcements with opposite valuation impact to zero, such that $S(e', e'') = 0$ if $\mathit{sign}(\pi(e')) \neq \mathit{sign}(\pi(e''))$. In the model, because the unrelated experience set G' contains all past news announcements that have a different information type or news direction than the current situation, these restrictions imply $S_{G' H_i^D} = 0$ for $i \in \{1, 2\}$.

Given the division in similarity criteria by information type, I separately consider recall probabilities and misreactions for stories and for statistics. For clarity, let $S_{AB}^{t(e)}$ be the average pairwise similarity between $A, B \subseteq E$ when the information type of the current news announcement is $t(e)$. Substituting the above similarity restrictions into Equation (5) yields the recall probabilities for stories and statistics:

$$r_{H_i^D}^{\mathit{Stories}} = \frac{S_{H_i^D}^{\mathit{Stories}} |H_i^D|}{S_{H_i^D}^{\mathit{Stories}} |H_i^D| + S_{H_i^D H_{-i}^D}^{\mathit{Stories}} |H_{-i}^D|}, \tag{13}$$

$$r_{H_i^D}^{\mathit{Statistics}} = \frac{S_{H_i^D}^{\mathit{Statistics}} |H_i^D|}{S_{H_i^D}^{\mathit{Statistics}} |H_i^D| + S_{H_i^D H_i^{D'}}^{\mathit{Statistics}} |H_i^{D'}|}. \tag{14}$$

Misreactions for stories and statistics are then given by Equations (15) and (16), respectively:

$$\%\Delta p^{\mathit{Stories}} = \frac{r_{H_1^D}^{\mathit{Stories}}}{r_{H_2^D}^{\mathit{Stories}}} \frac{|H_2^D|}{|H_1^D|} - 1 = \frac{S_{H_1^D}^{\mathit{Stories}}}{S_{H_2^D}^{\mathit{Stories}}} \cdot \left(\frac{S_{H_2^D}^{\mathit{Stories}} + S_{H_2^D H_1^D}^{\mathit{Stories}} \frac{|H_1^D|}{|H_2^D|}}{S_{H_1^D}^{\mathit{Stories}} \frac{|H_1^D|}{|H_2^D|} + S_{H_1^D H_2^D}^{\mathit{Stories}}} \right) - 1, \tag{15}$$

$$\%\Delta p^{\mathit{Statistics}} = \frac{r_{H_1^D}^{\mathit{Statistics}}}{r_{H_2^D}^{\mathit{Statistics}}} \frac{|H_2^D|}{|H_1^D|} - 1 = \frac{S_{H_1^D}^{\mathit{Statistics}}}{S_{H_2^D}^{\mathit{Statistics}}} \cdot \frac{|H_2^D|}{|H_1^D|} \cdot \left(\frac{S_{H_2^D}^{\mathit{Statistics}} + S_{H_2^D H_2^{D'}}^{\mathit{Statistics}} \frac{|H_2^{D'}|}{|H_2^D|}}{S_{H_1^D}^{\mathit{Statistics}} + S_{H_1^D H_1^{D'}}^{\mathit{Statistics}} \frac{|H_1^{D'}|}{|H_1^D|}} \right) - 1. \tag{16}$$

4.3 Model-implied misreaction patterns

Depending on the similarity structure of corporate news announcements, the distinct similarity criteria can generate different misreaction patterns for stories and for statistics. My empirical findings on the story-statistics gap in stock market misreaction in Section 3, along with existing evidence from Graeber, Roth, and Zimmermann (2024) and Ke (2025), suggest that people tend to underreact more to statistics than they do to stories. I do not assume these misreaction patterns in my model, nor does it produce these misreaction patterns ex-ante. To test whether my model offers a compelling explanation for the story-statistics gap, I directly measure the similarity structure of corporate news announcements and check whether the model-implied misreaction patterns align with the empirical results.

To measure the similarity relationships among corporate news announcements, I download the `all-MiniLM-L6-v2` sentence transformer from Hugging Face to extract embedding vectors from the Capital IQ Key Development news summaries. `all-MiniLM-L6-v2` is an efficient, lightweight sentence transformer that offers good embedding quality with only 22.7 million parameters, making it a suitable option for analyzing 1.3 million news summaries. The output embedding vectors from this transformer are 384-dimensional representations of textual data that capture both information content and semantic meaning of the news summaries. I normalize the embedding vectors to unit length for each news summary. I estimate the average pairwise similarity between two groups of news announcements as the inner product between their mean normalized embedding vectors. I obtain the similarity measure by flooring the inner product at zero. The resulting similarity measure ranges from zero to one, where zero represents complete dissimilarity and one indicates full equivalence.[4,5]

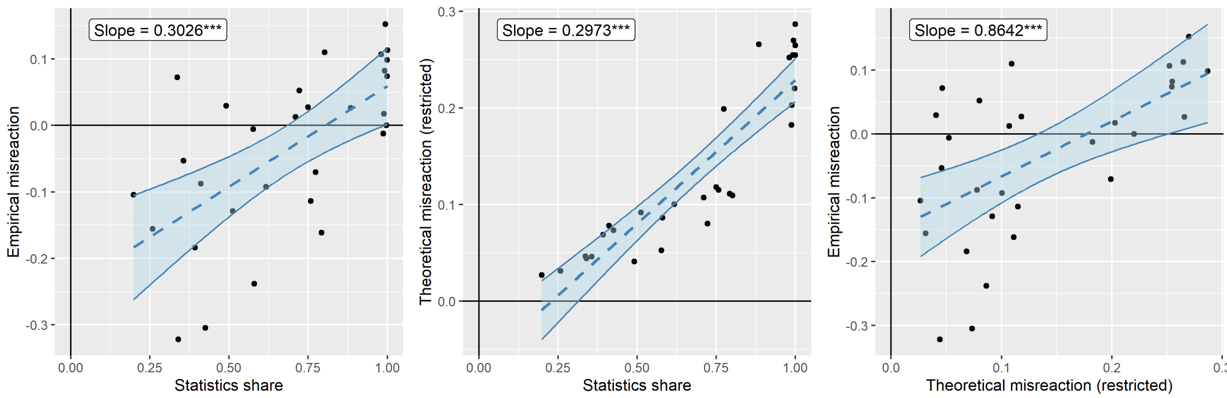
[4]For evaluation results of the `all-MiniLM-L6-v2` sentence transformer, see sbert.net, supermemory.ai, and huggingface.co.

[5]In practice, all inner products are already strictly positive before the flooring step. In other words, flooring the inner product does not make a difference beyond theoretical consistency.

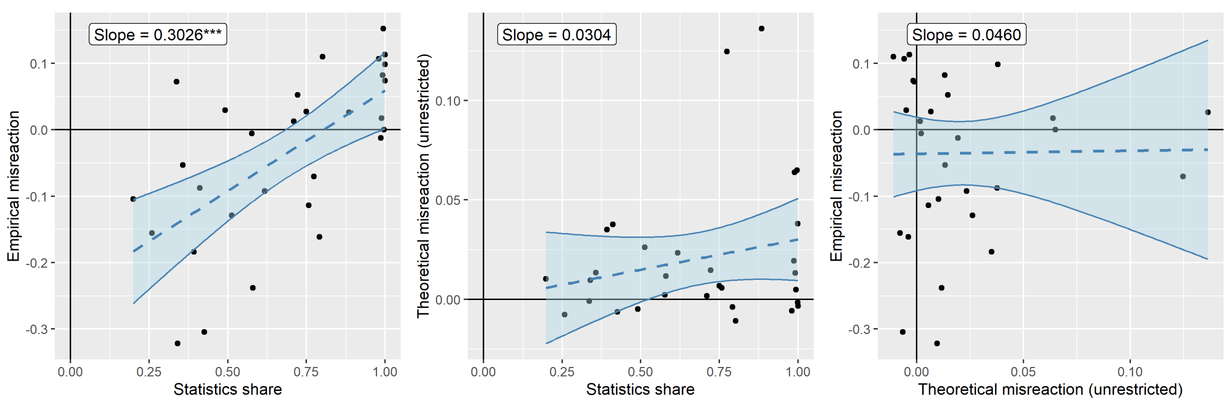
I empirically measure the similarity relationships in Equations (15) and (16) for each information type-news direction-news category triple. First, I identify all news announcements in my sample that are related to, or sharing the same information type and news direction with, the current triple. Second, I classify related news announcements within the same news category as the current triple as relevant news. Third, I classify related news announcements with an absolute value of immediate return above the sample median, or 2.7%, as high-impact news. These steps partition the related news announcements into the four key theoretical objects H_1^D, H_2^D, $H_1^{D'}$, and $H_2^{D'}$. Fourth, using the mean normalized embedding vectors from the key experience subsets, I calculate the similarity relationships in Equations (15) and (16). For simplicity, I assume that the similarity relationships do not vary over time and estimate them over the entire sample. Fifth, I plug the similarity measures and the cardinality of the key experience subsets into Equations (15) or (16), depending on the information type of the triple. This produces a model-predicted, theoretical misreaction for each information type-news direction-news category combination. Lastly, I explore how well the theoretical misreactions explain the cross-category variation in empirical misreaction documented by Figure 2 in Section 3.1. I take an average of the theoretical misreactions within each news category weighted by the frequency of news announcements under each aggregated triple. As a result, I obtain a model-predicted, theoretical misreaction for each news category.

Panel (a) of Figure 5 presents the prediction results from the *restricted* model, which includes the information type-specific interference patterns from Section 4.2 that culminate in Equations (15) and (16). The dashed lines are fit with OLS. The shaded regions indicate 95% confidence intervals. The left graph plots the news categories by their empirical misreaction and statistics share. This replicates Figure 2 under an additional requirement that the news category has at least one relevant, low-impact event to avoid division by zero. This filter does not remove any news categories in practice, so the left panel exactly replicates the cross-category story-statistics gap from Figure 2. The middle graph plots the news categories by their model-predicted, theoretical misreaction and statistics share. The theoretical misreactions also produce a significant story-statistics gap, where more statistical news categories see stronger underreaction. The empirical and theoretical story-statistics gaps are closely aligned in magnitude, which is reflected by the estimated slopes of the fitted lines. The right graph plots the news categories by their empirical and theoretical misreaction. The significantly positive relationship between empirical and theoretical misreactions provides compelling evidence that selective recall is a key driver of cross-category variations in stock misreaction. Furthermore, regressing empirical on theoretical misreactions yields an R^2 of 34%. This suggests that selective recall explains a significant share of the cross-category variation in stock market misreaction. The story-statistics gap in theoretical misreaction, the strong, positive relationship between the empirical estimates and the theoretical predictions, and the large R^2 suggest that selective recall materially contributes to explaining the story-statistics gap across news categories.

The intuition behind these results lies in the mismatch between the interference pattern for statistics and the relative frequency for the objective odds. The objective odds correspond to the *intra-category* distribution of valuation impact. But statistics only face *inter-category* interference. Because interference helps us form an approximate sense of the overall distribution, this mismatch desensitizes investors to the *intra-category* distribution of valuation impact for statistics. Consequently, investors attenuate their subjective odds for statistics toward an intermediate level. However, statistical news is more likely to deliver a high valuation impact than not in the news data. This makes investors underestimate the valuation impact of statistical news, which leads to systematic underreaction. At the same time, stories face *intra-category* interference. The alignment between the interference pattern and the objective odds for stories make investors sensitive to the objective odds, leading to weaker misreaction. Taken together, these patterns predict stronger underreaction to more quantitative news, which is in line with my empirical findings on the story-statistics gap in stock market misreaction.



(a) Restricted model



(b) Unrestricted model

Figure 5: Cross-category comparison of theoretical and empirical misreaction

Note. Each point represents a news category. The empirical misreaction, or misreaction coefficient, of a news category is computed by estimating Equation (1) with announcement-year and firm fixed effects within the news category. Under the restricted model, the theoretical misreaction of a news category is calculated by plugging the estimated similarity relationships into the model developed in Section 4.2. Under the unrestricted model, the theoretical misreaction of a news category is calculated by plugging the estimated similarity relationships into the model developed in Section 4.1, without the hypothesized interference patterns discussed in Section 4.2. The statistics share of a news category is the percentage of news announcements containing any statistics within that news category. News categories with fewer than 100 observations per year-in-sample are excluded to limit the impact of outliers. Eight out of the 38 news categories are omitted as a result. While news categories with no relevant, low-impact event would be excluded to avoid division by zero, no remaining news categories are excluded in practice. The dashed lines are fit with OLS. The shaded regions indicate 95% confidence intervals. Significance code: 0.1 * 0.05 ** 0.01 ***

To demonstrate the importance of the hypothesized interference patterns in Section 4.2, I additionally present the prediction results from the *unrestricted* model, which calculates misreaction according to Equation (8) from Section 4.1. The unrestricted model does not distinguish between stories and statistics and does not incorporate the interference patterns described in Section 4.2. In other words, it does not require $S_{H_i^D H_i^{D'}} = 0$ when $t(e) = \textit{Story}$ or $S_{H_i^D H_{-i}^D} = 0$ when $t(e) = \textit{Statistic}$ for $i \in \{1, 2\}$. Panel (b) of Figure 5 presents the prediction results from the unrestricted model. As in Panel (a), the left graph replicates Figure 2. The middle graph, which plots news categories by their theoretical misreaction and statistics share, shows no clear relationship between the plotted variables. This means that the unrestricted model does not predict a story-statistics gap. The right graph, which plots news categories by their empirical and theoretical misreaction, again shows no clear relationship between the plotted variables. This means that the theoretical misreaction from the unrestricted model does not align well with their empirical counterparts. In light of the significant results from the restricted model, the non-results from the unrestricted model reflect the critical role of the hypothesized interference patterns to explaining stock market misreaction with selective recall.

4.4 Measurement error-free implications for misreaction

The derivations from Section 4.2 highlight two complementary approaches to testing the model's empirical implications for stock market misreaction. Equations (15) and (16) show that misreaction varies with both the similarity structure and the relative frequencies of past experiences. On the one hand, directly testing how similarity relationships affect misreaction, as I have done in Section 4.3, speaks most clearly to the core mechanism of similarity-based recall. However, there is no consensus on how to convincingly measure conceptual similarity between abstract, high-dimensional objects such as corporate news announcements. As a result, evaluating the model by directly estimating similarity inevitably amounts to a joint test of both the model and the validity of the proposed similarity measure.

On the other hand, I can readily and precisely observe the relative frequencies of past news announcements. While similarity plays a more salient role than event frequency in the theory of selective recall, Equations (5), (13), and (14) show that the two enter symmetrically into the sampling weights of the various experience groups. These sampling weights are the key objects that determine the extent of distortion induced by similarity-based recall via the interference mechanism. Moreover, the hypothesized interference patterns restrict which experience sets can interfere with one another in ways that are distinctive to the model. As a result, testing the frequency-based implications for stock market misreaction remains informative about the model's economic relevance because these implications hinge directly on the similarity assumptions. Therefore, in addition to evaluating the model based on measured similarity in Section 4.3, I test whether variation in relative frequencies affects misreaction in the manner implied by the model.

Equations (15) and (16) reveal that misreactions to stories and statistics depend on three key relative frequencies under the restricted model. These three relative frequencies are the objective odds of high impact given relevance, $f_1 = \frac{|H_1^D|}{|H_2^D|}$; the objective odds of relevance given high impact, $f_2 = \frac{|H_1^D|}{|H_1^{D'}|}$; and the objective odds of relevance given low impact, $f_3 = \frac{|H_2^D|}{|H_2^{D'}|}$. These relative frequencies capture, respectively, the proportion of relevant experiences that are high-impact, the proportion of high-impact experiences within the relevant category, and the proportion of low-impact experiences within the relevant category. They map directly into the three potential channels of interference in the model, which correspond to the three arrows in Figure 4. Whereas the objective odds of high impact given relevance f_1 corresponds to vertical interference, the other two relative frequencies f_2 and f_3 correspond to horizontal interference for high- and low-impact events, respectively. The objective odds of high impact given relevance f_1 also coincide with the investor's conditional best estimate of the probability that the current situation is high-impact, p^*, as defined in Equation (7).

To obtain sign predictions for how these relative frequencies should affect stock market misreaction, I take the partial derivatives of Equations (15) and (16) with respect to each of the three relative frequencies. In signing the partial derivatives, I follow Bordalo et al. (2023) by assuming that similarity is non-negative and that cross-similarity between $A, B \subseteq E$ is less than self-similarity for either A or B.

For statistics, the partial derivative of misreaction with respect to the objective odds of high impact given relevance is negative:

Prediction 1. *For statistics, stock market underreaction increases in the objective odds of high impact given relevance, such that*

$$\frac{\partial}{\partial f_1} \% \Delta p^{Statistics} = - \frac{S_{H_1^D}^{Statistics}}{S_{H_2^D}^{Statistics}} \cdot \left(\frac{S_{H_2^D}^{Statistics} + S_{H_2^D H_2^{D'}}^{Statistics} f_3^{-1}}{S_{H_1^D}^{Statistics} + S_{H_1^D H_1^{D'}}^{Statistics} f_2^{-1}} \right) f_1^{-2} < 0. \tag{17}$$

To illustrate the intuition behind this prediction, I use earnings as an example of statistical news. The model predicts that, holding the other relative frequencies constant, investors would underreact more to an upcoming earnings announcement if high-impact news is more prevalent among historical earnings announcements. Because statistics are not subject to vertical interference within the same news category, the recall probabilities for high- and low-impact earnings announcements only depend on the other two relative frequencies, which are held constant. Investors take the ratio of the recall probabilities as their estimated prevalence of high-impact earnings announcements, which also remains unchanged (p stays the same). Because investors' estimated, subjective prevalence does not increase in the historical, objective prevalence of high-impact earnings, their underestimation for the probability that the upcoming earnings announcement is high-impact becomes stronger when high-impact earnings are historically more prevalent (p^* increases). Underestimating the probability that the upcoming earnings announcement is high-impact leads to underreaction in trading decisions, which translates to subsequent drift in the stock price of the announcing firm.

The derivatives of misreaction to statistics with respect to the objective odds of relevance given high and low impact are positive and negative, respectively:

Prediction 2. *For statistics, stock market underreaction decreases in the objective odds of relevance given high impact, such that*

$$\frac{\partial}{\partial f_2}\%\Delta p^{Statistics} = \frac{\left(S^{Statistics}_{H_1^D} + \frac{S^{Statistics}_{H_1^D}}{S^{Statistics}_{H_2^D}} S^{Statistics}_{H_2^D H_2^{D'}} f_3^{-1}\right)\left(S^{Statistics}_{H_1^D H_1^{D'}} f_2^{-2} f_1^{-1}\right)}{\left(S^{Statistics}_{H_1^D} + S^{Statistics}_{H_1^D H_1^{D'}} f_2^{-1}\right)^2} > 0. \tag{18}$$

Prediction 3. *For statistics, stock market underreaction increases in the objective odds of relevance given low impact, such that*

$$\frac{\partial}{\partial f_3}\%\Delta p^{Statistics} = -\left(\frac{\frac{S^{Statistics}_{H_1^D}}{S^{Statistics}_{H_2^D}} S^{Statistics}_{H_2^D H_2^{D'}} f_3^{-2} f_1^{-1}}{S^{Statistics}_{H_1^D} + S^{Statistics}_{H_1^D H_1^{D'}} f_2^{-1}}\right) < 0. \tag{19}$$

Both results trace back to the horizontal interference mechanism. In the earnings example, as the objective odds of relevance given high impact f_2 increase, statistics from non-earnings news categories are historically less prevalent among high-impact statistics. Thus, there is less interference when recalling high-impact earnings announcements, so the recall probability for high-impact earnings announcements increases. Holding the prevalence of non-earnings news categories among low-impact statistics (f_3) constant, the recall probability for low-impact earnings announcements remains unchanged. Investors take the ratio of the recall probabilities as their estimated prevalence of high-impact earnings announcements, which increases. Consequently, investors assign higher probability to the event that the upcoming earnings announcement is high-impact (p increases), even though the objective prevalence of high-impact earnings is held constant (p^* stays the same). This tilts investors toward overreaction. As the objective odds of relevance given low impact f_3 increase, the opposite occurs, which tilts investors toward underreaction.

For stories, the derivative of misreaction with respect to the objective odds of high impact given relevance is negative:

Prediction 4. *For stories, stock market underreaction increases in the objective odds of high impact given relevance, such that*

$$\frac{\partial}{\partial f_1}\%\Delta p^{Stories} = \frac{S^{Stories}_{H^D_1}}{S^{Stories}_{H^D_2}} \frac{\left(S^{Stories}_{H^D_1 H^D_2}{}^2 - S^{Stories}_{H^D_1} S^{Stories}_{H^D_2}\right)}{\left(S^{Stories}_{H^D_1} f_1 + S^{Stories}_{H^D_1 H^D_2}\right)^2} < 0. \tag{20}$$

To illustrate the intuition behind this prediction, I use mergers as an example of story-like news. The model predicts that, holding the other relative frequencies constant, investors will underreact more to an upcoming merger announcement if high-impact news is more prevalent among historical merger announcements. Because stories are subject to vertical interference within the same news category, the recall probability for high-impact mergers increases in the historical prevalence of high-impact mergers. By the same mechanism, the recall probability for low-impact mergers decreases in the historical prevalence of high-impact mergers. Investors take the ratio of the recall probabilities as their estimated prevalence of high-impact mergers, which increases. While the historical, objective prevalence of high-impact mergers also increases (p^* increases), vertical interference dampens the increase in investors' estimated, subjective prevalence through the assumption that cross-similarity is lower than self-similarity (p mildly increases). Because the increase in the subjective odds is less than the increase in the objective odds, the investor tilts toward underreaction.

The derivatives of misreaction to stories with respect to the objective odds of relevance given high and low impact are both zero:

Prediction 5. *For stories, the objective odds of relevance given high impact have no effect* *on stock market misreaction, such that*

$$\frac{\partial}{\partial f_2} \% \Delta p^{Stories} = 0. \tag{21}$$

Prediction 6. *For stories, the objective odds of relevance given low impact have no effect on stock market misreaction, such that*

$$\frac{\partial}{\partial f_3} \% \Delta p^{Stories} = 0. \tag{22}$$

Both results come from stories' immunity to horizontal interference. In the merger example, as the objective odds of relevance given high impact f_2 increase, stories from non-merger news categories are historically less prevalent among high-impact statistics. However, because stories are not subject to horizontal interference, this change does not alter the recall probability for either high- or low-impact mergers. Investors take the ratio of the recall probabilities as their estimated prevalence of high-impact mergers, which also remains unchanged. Consequently, investors do not change their estimated probability that the upcoming merger announcement is high-impact (p stays the same) when the objective prevalence of high-impact mergers is also held constant (p^* stays the same). As a result, investors' misreaction remains unchanged. The same occurs as the objective odds of relevance given low impact f_3 increase.

Together, these six predictions summarize the model's empirical implications. The numerical precision of statistics prevents vertical interference across valuation impact levels, but the absence of vertical interference desensitizes investors' subjective odds to the objective prevalence of high-impact news within the relevant news category. At the same time, variation in horizontal interference for statistics changes the recall probability for each valuation impact level. This affects investors' subjective odds when the objective prevalence of high-impact news remains unchanged, which can lead to more misreaction in either direction. In contrast, semantic links shield stories from horizontal interference. As a result, experiences from irrelevant news categories have no effect on misreaction to stories. However, under minimal assumptions on similarity structure, vertical interference dampens the increase in investors' subjective odds in response to an increase in the historical prevalence of high-impact news. These distinct predictions for stories and statistics provide testable implications for evaluating the model's economic relevance.

To test these empirical predictions, I construct the three relative frequencies for each news announcement in the Key Development database. For each news announcement e, I take a lookback window of 10 years in the baseline specification. First, I classify all news announcements within the lookback window that share the same information type and news direction with the current news announcement as related news. Next, as in Section 4.3, I classify related news announcements within the same news category as the current news announcement as relevant news. I also classify related news announcements with an absolute value of immediate return above the sample median, or 2.7%, as high-impact news. For each news announcement, these steps partition the related news announcements within its lookback window into the four experience sets: $H^D_{1,e}$, $H^D_{2,e}$, $H^{D'}_{1,e}$, and $H^{D'}_{2,e}$. From the size of these experience sets, I directly compute $f_{1,e} = |H^D_{1,e}|/|H^D_{2,e}|$, $f_{2,e} = |H^D_{1,e}|/|H^{D'}_{1,e}|$, and $f_{3,e} = |H^D_{2,e}|/|H^{D'}_{2,e}|$ for each news announcement.

To measure whether stock market misreaction varies as predicted in the three relative frequencies, I expand the regression framework of Equation (2):

$$\begin{aligned} \textit{Subsequent return}_e = &\ \textit{Fixed effects}_e + \beta \cdot \textit{Immediate return}_e \\ &+ \gamma_{f_1} \cdot (\textit{Immediate return}_e \times f_{1,e}) \\ &+ \gamma_{f_2} \cdot (\textit{Immediate return}_e \times f_{2,e}) \\ &+ \gamma_{f_3} \cdot (\textit{Immediate return}_e \times f_{3,e}) \\ &+ \textit{Additional Controls}_e + \varepsilon_e. \end{aligned} \tag{23}$$

While positive and negative misreaction in the model translates to stock market overreaction and underreaction, respectively, signs are flipped in the regression setting. In the model, greater underreaction corresponds to a more *negative* value of misreaction as defined by Equations (15) and (16). In the regression setting, greater underreaction implies a stronger positive drift in the subsequent return, which appears as a more *positive* coefficient on the immediate return. Consequently, Predictions 1 through 6 directly translate to the following sign restrictions for the regression coefficients within the stories and statistics subsamples:

$$\begin{gathered}\gamma_{f_1}^{Statistics} > 0, \quad \gamma_{f_2}^{Statistics} < 0, \quad \gamma_{f_3}^{Statistics} > 0, \\ \gamma_{f_1}^{Stories} > 0, \quad \gamma_{f_2}^{Stories} = 0, \quad \gamma_{f_3}^{Stories} = 0.\end{gathered} \tag{24}$$

Table 4 presents the regression results. Column (1) reports the baseline regression results for the statistics subsample. All three interaction coefficients on the objective frequencies carry the correct signs and are statistically significantly different from zero. Column (2) reports the results for the stories subsample. As predicted, $\hat{\gamma}_{f_1}^{Stories}$ is significantly positive, albeit at the 10% level. Moreover, neither $\hat{\gamma}_{f_2}^{Stories}$ nor $\hat{\gamma}_{f_2}^{Stories}$ are significantly different from zero. These findings fully align with the sign predictions for the stories subsample. Taken together, the regression results confirm all six measurement error-free implications of the model, which strongly support the relevance of selective recall for understanding stock market misreaction.

Columns (3) and (4) of Table 4 investigate whether selective recall explains the story-statistics gap in stock market misreaction. Because Table 4 focuses on the second half of the sample due to the 10-year lookback window, column (3) first confirms the presence of the story-statistics gap in this sample period by replicating column (1) from Table 2. The significantly positive interacted coefficient on the statistics ratio demonstrates that the story-statistics gap in stock market misreaction remains robust in this sample period. Following the same regression specification for Columns (2) through (5) in Table 2, column (4) regresses

Subsample	Statistics	Stories	All	
	Subsequent return			
	(1)	(2)	(3)	(4)
Immediate return	-0.009	-0.040	0.005	-0.019
	(0.037)	(0.058)	(0.035)	(0.036)
Immediate return × f_1	0.161***	0.267*		0.145***
	(0.020)	(0.119)		(0.038)
Immediate return × f_2	-0.057***	-0.307		-0.052***
	(0.005)	(0.195)		(0.004)
Immediate return × f_3	0.080***	0.265		0.060***
	(0.001)	(0.157)		(0.010)
Immediate return × Statistics ratio			0.055***	0.030
			(0.002)	(0.023)
Observations	457,916	184,145	642,061	642,061
R^2	0.085	0.114	0.083	0.083
Main effects of controls	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Firm fixed effects	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓
Clustered standard errors	✓	✓	✓	✓

Table 4: Misreaction regression with interacted relative frequencies

Note. For each news announcement, I classify all news announcements within a lookback window of 10 years that share the same information type and news direction as related news. Related news announcements within the same news category as the current news announcement are relevant news. Related news announcements with an absolute value of immediate return above the sample median, or 2.7%, are high-impact news. These steps partition the related news announcements within the lookback window into four experience sets: $H^D_{1,e}$, $H^D_{2,e}$, $H^{D'}_{1,e}$, and $H^{D'}_{2,e}$. From the cardinality of these experience sets, I directly compute $f_{1,e} = |H^D_{1,e}|/|H^D_{2,e}|$, $f_{2,e} = |H^D_{1,e}|/|H^{D'}_{1,e}|$, and $f_{3,e} = |H^D_{2,e}|/|H^{D'}_{2,e}|$ for each news announcement. Variables other than subsequent and immediate returns are standardized for interpretability. The main effects of the control variables are included in the regression but omitted from display. Standard errors, shown in parentheses, are clustered at the announcement year, firm, and news category levels. Significance code: 0.1 * 0.05 ** 0.01 ***.

subsequent return on immediate return interacted with the statistics ratio and the three objective frequencies from the model of selective recall. The three objective frequencies fully absorb the predictive power of the statistics ratio for stock market misreaction, such that the interaction coefficient on the statistics ratio becomes indistinguishable from zero. This finding suggests that selective recall offers a competitive explanation for the story-statistics gap in stock market misreaction. In light of the results in Table 2, where the alternative hypotheses of misreaction do not explain away the story-statistics gap, this result further suggests that selective recall plays a material role in driving the different patterns in stock

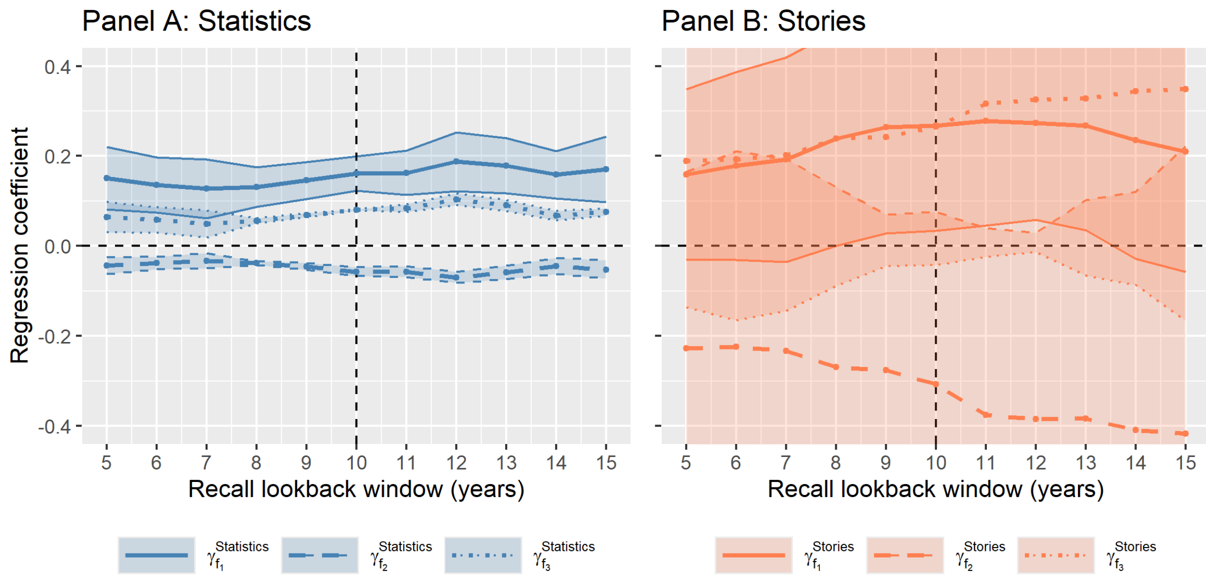


Figure 6: Key regression coefficients from Equation (23) over various lookback windows

Note. Positive (negative) interaction coefficients indicate that the stock market tends to underreact more (less) to a corporate news announcement as the variable increases. Panels A and B report the regression results for the statistics and stories subsamples, respectively, estimated over different lookback periods. The vertical dashed line marks the baseline specification, which uses a 10-year lookback window. Thick lines show the point estimates of the interaction coefficients across lookback periods. Thin lines and shaded regions depict the corresponding 95% confidence intervals. Solid, dashed, and dotted lines correspond to γ_{f_1}, γ_{f_2}, and γ_{f_3}, respectively.

market misreaction.

To assess the sensitivity of my empirical results to different parameter choices, I re-estimate Columns (1) and (2) of Table 4 over lookback windows ranging from five to 15 years. I stop at 15 years because each additional year of lookback removes one year of observations from the regression sample, and the overall sample spans only 20 years from 2003 to 2022. Figure 6 plots the key interaction coefficients and their 95% confidence intervals for both subsamples. Panel A shows that the estimated coefficients in the statistics subsample maintain the predicted signs and remain highly significant across all lookback windows. Panel B largely confirms the model's empirical implications for the stories subsample. $\gamma_{f_1}^{Stories}$ remains significantly positive over lookback windows ranging from eight to 13 years. Neither $\gamma_{f_2}^{Stories}$ nor $\gamma_{f_3}^{Stories}$ is statistically distinguishable from zero across all lookback windows.

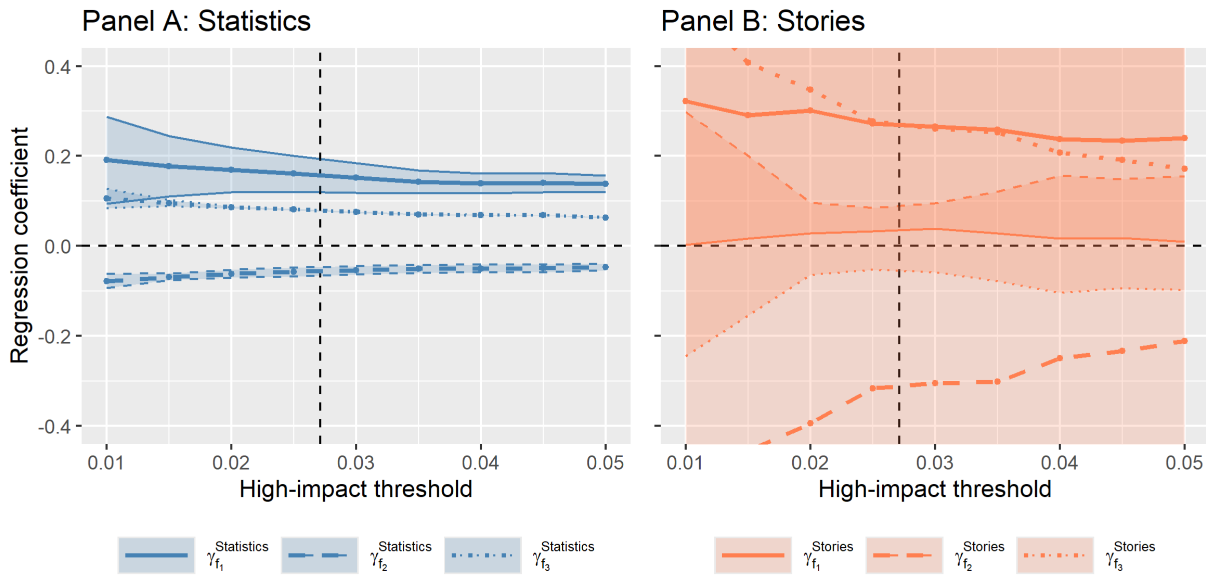


Figure 7: Key regression coefficients from Equation (23) over various high-impact thresholds

Note. Positive (negative) interaction coefficients indicate that the stock market tends to underreact more (less) to a corporate news announcement as the variable increases. Panels A and B report the regression results for the statistics and stories subsamples, respectively, estimated over different high-impact thresholds. The vertical dashed line marks the baseline specification, which uses the median of the absolute value of immediate return, or 2.7%. Thick lines show the point estimates of the interaction coefficients across high-impact thresholds. Thin lines and shaded regions depict the corresponding 95% confidence intervals. Solid, dashed, and dotted lines correspond to γ_{f_1}, γ_{f_2}, and γ_{f_3}, respectively.

I also evaluate the sensitivity of the coefficient estimates to different return thresholds for high-impact news announcements. I re-estimate Columns (1) and (2) of Table 4 using high-impact thresholds ranging from 1% to 5% in the absolute value of the immediate return. Figure 7 plots the key coefficients with their corresponding 95% confidence intervals for both subsamples. Panel A shows that the estimated coefficients for the statistics subsample maintain the predicted signs and remain highly significant across all high-impact thresholds. Panel B similarly confirms the empirical implications of the model for the stories subsample. $\gamma_{f_1}^{Stories}$ remains significantly positive over all high-impact thresholds. Neither $\gamma_{f_2}^{Stories}$ nor $\gamma_{f_3}^{Stories}$ is statistically distinguishable from zero across all high-impact thresholds.

These results suggest that the coefficient estimates from Table 4, along with the conclusions drawn from them, are highly robust to alternative parameter choices. This robustness further speaks to the material role of selective recall in driving stock market misreaction.

4.5 Model implication for pairwise differences

Beyond the misreaction patterns, my model of selective recall has an additional implication for pairwise differences in subjective valuation impact across news announcements. In my model, the investor estimates the valuation impact of a news announcement as an average of the valuation impact of relevant recalled experiences. His estimated valuation impact $\tilde{\pi}$ is a recall-weighted average of the two possible impact levels:

$$\begin{aligned}\tilde{\pi} &= \left(\frac{n(H_1^D)}{n(H_1^D)+n(H_2^D)}\right)\pi_1 + \left(\frac{n(H_2^D)}{n(H_1^D)+n(H_2^D)}\right)\pi_2 \\ &= \left(\frac{1}{1+p^{-1}}\right)\pi_1 + \left(\frac{p^{-1}}{1+p^{-1}}\right)\pi_2.\end{aligned} \tag{25}$$

As highlighted in Section 4.4, the investor's subjective odds of high impact p depends on (1) the similarity relationship between the current news announcement and the various partitions of the experience database as well as (2) the relative frequency between those partitions. Consequently, the investor should estimate closer subjective valuation impacts for news announcements that either (1) exhibit closer similarity relationships with the experience database or (2) have experience databases with closer relative frequencies.

Under my central hypothesis that stories and statistics share different similarity structures, my model implies that investors should estimate closer subjective valuation impacts for news announcements with more comparable quantitativeness. Formally, let $e, e' \in E$ be two news announcements. My model implies that the difference between the subjective valuation impacts of the news pair $\Delta\tilde{\pi}_{e,e'} = |\tilde{\pi}_e - \tilde{\pi}_{e'}|$ increases in the difference between the statistics ratios of the news pair $\Delta \mathit{Statistics\ ratio}_{e,e'} = |\mathit{Statistics\ ratio}_e - \mathit{Statistics\ ratio}_{e'}|$. Because investors' subjective valuation impact drives the magnitude of stock price response, I proxy the unobserved subjective valuation impact $\tilde{\pi}_e$ with the absolute value of immediate return $|Immediate\ return_e|$. My model implies $\zeta_{Stat} > 0$ in the following pair regression specification:

$$\begin{aligned}\Delta|Immediate\ return|_{e,e'} &= Fixed\ effects_e + Fixed\ effects_{e'} \\ &+ \zeta_{Stat} \cdot \Delta Statistics\ ratio_{e,e'} \\ &+ Additional\ pair\ controls_{e,e'} + \varepsilon_{e,e'}.\end{aligned} \tag{26}$$

With 1.3 million news announcements in my main sample, constructing all $\binom{1.3\times10^6}{2} \approx 0.8 \times 10^{12}$ news pairs in my sample and running regressions on them is not computationally feasible. I therefore estimate the pairwise regressions on randomly sampled news pairs and use a cluster bootstrap for inference. Because news announcements are likely correlated within announcement year and firm, randomly sampling news pairs as if they are independent could understate sampling uncertainty. I implement a two-way cluster bootstrap at the announcement year and firm levels. For each bootstrap sample, I first resample unique announcement years and firms independently with replacement. Each news announcement receives a bootstrap multiplicity equal to the product of its announcement year multiplicity and firm multiplicity. I then draw 100,000 random news pairs from the multiplicity-weighted sample of news announcements, compute pairwise differences, and estimate the regression equation.

To alleviate omitted-variable concerns, I include announcement year and firm fixed effects from both sides of the news pair. I do not include news-category fixed effects because the pairwise test is designed to capture both within- and cross-category variation in valuation-impact differences, and cross-category variation is part of the hypothesized story-statistics similarity structure. Instead, in specifications that include other news characteristics, I control for whether the two announcements belong to different news categories. I repeat this process 1,000 times. I report the mean of the bootstrapped distribution of each regression coefficient as the coefficient estimate. I use the 2.5th and 97.5th percentiles of the bootstrapped distribution of each regression coefficient as its 95% confidence interval.

Table 5 presents the results of this exercise. Column (1) reports the baseline results. Columns (2), (3), and (4) introduce differences in relative frequencies, other behavioral proxies, and other news characteristics, respectively, as additional pair controls. Column (5) includes all pair controls. ζ_{Stat} is significantly positive across all regression specifications. In other words, pairs of news announcements with larger differences in statistics ratios also exhibit larger differences in the magnitudes of stock return responses. This effect is not fully explained by pairwise differences in relative frequencies, other behavioral proxies, or other news characteristics. These empirical findings support the hypothesized division in similarity structure between stories and statistics. As the difference in quantitativeness between news announcements grows, they share increasingly different similarity relationships with the experience database. While more story-like news announcements receive more intra-category interference, more statistical news announcements face more inter-category interference. The difference in similarity relationships translates to different recall probabilities and subjective odds of high impact, which contribute to the difference in the magnitude of stock return response to each news announcement.

In addition to supporting the pairwise valuation-impact implication of my hypothesis and model, Column (5) of Table 5 provides new insights into other possible drivers of cognitive similarity among news announcements. Because the regression specification of Column (5) controls for pairwise differences in the relative frequencies, my model attributes the remaining variation in the difference in subjective valuation impact—as proxied by the magnitude of immediate return—to the differences in the news pair's similarity relationships with the experience database. This means that, beyond statistics ratio, stronger alignment in prior extremeness, turnover, and market capitalization makes two news announcements more aligned in their similarity with past experiences. If cognitive similarity among news announcements

	Δ |Immediate return|				
Controls	None (1)	Relative Freq. (2)	Behavioral (3)	Other (4)	All (5)
Δ Statistics ratio	0.075*** [0.053, 0.096]	0.022*** [0.008, 0.038]	0.022*** [0.006, 0.036]	0.079*** [0.056, 0.100]	0.015** [0.002, 0.029]
Δ f_1		0.066*** [0.052, 0.079]			0.038*** [0.009, 0.072]
Δ f_2		0.050*** [0.037, 0.063]			0.040*** [0.027, 0.053]
Δ f_3		-0.045*** [-0.056, -0.033]			-0.030*** [-0.041, -0.019]
Δ Prior extremeness			0.081*** [0.065, 0.101]		0.049*** [0.012, 0.085]
Different news sentiment			0.021** [0.003, 0.040]		0.017** [0.000, 0.034]
Δ Turnover			0.025** [0.005, 0.047]		0.025** [0.005, 0.047]
Δ News articles			-0.007 [-0.022, 0.008]		-0.007 [-0.023, 0.008]
Δ Log market value			0.051*** [0.040, 0.063]		0.051*** [0.040, 0.063]
Δ Word count				0.002 [-0.006, 0.010]	0.004 [-0.004, 0.011]
Different industry				0.016 [-0.005, 0.036]	0.015 [-0.006, 0.036]
Δ Announcement date				0.006 [-0.004, 0.016]	0.006 [-0.005, 0.016]
Different news direction				-0.004 [-0.016, 0.008]	-0.006 [-0.018, 0.006]
Different news category				-0.064*** [-0.091, -0.038]	-0.127*** [-0.158, -0.095]
R^2	0.182 [0.168, 0.201]	0.187 [0.172, 0.206]	0.188 [0.174, 0.207]	0.183 [0.168, 0.201]	0.190 [0.176, 0.210]
Year fixed effects	✓	✓	✓	✓	✓
Firm fixed effects	✓	✓	✓	✓	✓
Two-way cluster bootstrap	✓	✓	✓	✓	✓

Table 5: Bootstrapped regression coefficients of the pairwise difference in the absolute value of immediate return on the pairwise difference in statistics ratio and various controls

Note. I draw 1,000 bootstrap samples of 100,000 pairs of news announcements with two-way clustering at the announcement year and firm levels. For each sample, I regress the pairwise difference in the absolute value of immediate return on the pairwise difference in statistics ratio and various controls. I include announcement year and firm fixed effects from both sides of the news pairs. Pairwise differences between continuous variables are indicated by the Δ prefix. All Δ variables are standardized for interpretability. Pairwise differences between categorical variables are dummy variables that indicate one if the news pair differs on that dimension. The coefficient estimates are the means of the bootstrap distributions of the regression coefficients. 95% confidence intervals from the bootstrap distributions are indicated in brackets below each coefficient estimate. Significance code: 0.1 * 0.05 ** 0.01 ***.

is weakly transitive in the sense that

$$\frac{\partial}{\partial y}\mathbb{E}\left[S(e,e') \mid S(e,e'')=x, S(e',e'')=y\right] > 0, \tag{27}$$

then stronger alignment in these characteristics makes two news announcements more similar to each other. A candidate functional form for similarity that exhibits this property is $S(e, e') = \delta^{\sum_i |f_i(e) - f_i(e')|}$ from Bordalo et al. (2023), where $f_i(e)$ represents feature i of news announcement e. Given this functional form, these results suggest that statistics ratio, prior extremeness, turnover, and market capitalization enter the set of similarity-driving features $\{f_i\}$ for corporate news announcements.

Interestingly, the coefficient estimate on the dummy variable for different news category is significantly negative. This means that, on average, a pair of news announcements in different news categories has closer stock price impact than a pair in the same news category. To understand this seemingly counter-intuitive result through the lens of my model, I examine how this effect varies by the information types of the news pair in Appendix Table A.13. I find that the negative coefficient does not reflect a uniform relationship between difference in news category and difference in valuation impact. Once I separately control for whether the pair consists of two stories or two statistics and interact these indicators with the different-category dummy, the standalone coefficient on the different-category dummy becomes statistically insignificant. Instead, the negative effect is concentrated among pairs of statistical news announcements. Pairs of statistical news announcements within the same news category exhibit larger differences in the magnitudes of immediate returns, but this difference is substantially attenuated when the two statistics belong to different news categories. This pattern is consistent with my hypothesis that statistics can be similar across news categories when they convey similar valuation impact.

5 Conclusion

In this paper, I document a pronounced story-statistics gap in stock market misreaction that is both statistically and economically significant. This pattern manifests both across and within news categories. Neither signal strength uncertainty, diagnostic expectations, news sentiment, nor investor attention accounts for this pattern. A characteristic-managed portfolio based on the story-statistics gap earns an annualized Sharpe ratio of 1.31 and a statistically significant alpha of approximately 110 basis points per month against a variety of factor models.

Building on the experimental evidence from Graeber, Roth, and Zimmermann (2024), I extend the theoretical framework from Bordalo et al. (2023) and hypothesize that memory applies distinct similarity criteria for stories and for statistics. Stories are similar within each news category due to semantic associations, whereas statistics are similar within the same valuation impact level due to their numerical precision. Conditional on the similarity structure of corporate news announcements, my model offers concrete predictions for stock market misreaction at the news category level. I use embedding vectors from news summaries to quantify the similarity structure of corporate news announcements. Under the hypothesized similarity criteria, model-predicted misreactions at the news category level exhibit both a story-statistics gap and a significantly positive relationship with empirical misreactions. In contrast, without the hypothesized similarity patterns, my model no longer explains the cross-category variation in stock market misreaction.

I additionally derive six measurement error-free implications of my model for misreaction that do not rely on similarity measurements. I find support for all six implications across a wide range of parameter choices. Furthermore, in line with the hypothesized similarity structure, I show that news announcements that are closer in quantitativeness generate more comparable stock price impacts. Taken together, these empirical results suggest that my model of selective recall offers a useful theoretical framework for understanding stock market misreaction to corporate news announcements. To my knowledge, this is the first paper to document a story-statistics gap in stock market misreaction and to explain this empirical pattern through selective recall.

References

Agrawal, Anup and Jeffrey F Jaffe (2000), "The post-merger performance puzzle," *Advances in Mergers and Acquisitions*, pp. 7–41.

Agrawal, Anup, Jeffrey F Jaffe, and Gershon N Mandelker (1992), "The post-merger performance of acquiring firms: a re-examination of an anomaly," *The Journal of Finance* 47(4), pp. 1605–1621.

Araci, Dogu (2019), "Finbert: Financial sentiment analysis with pre-trained language models," *ArXiv preprint 1908.10063*.

Augenblick, Ned, Eben Lazarus, and Michael Thaler (2025), "Overinference from weak signals and underinference from strong signals," *The Quarterly Journal of Economics* 140(1), pp. 335–401.

Ball, Ray and Philip Brown (1968), "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research* 6(2), pp. 159–178.

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

Bernard, Victor L and Jacob K Thomas (1989), "Post-earnings-announcement drift: delayed price response or risk premium?," *Journal of Accounting Research* 27, pp. 1–36.

Bordalo, Pedro et al. (2023), "Memory and probability," *The Quarterly Journal of Economics* 138(1), pp. 265–311.

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

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998), "Investor psychology and security market under-and overreactions," *the Journal of Finance* 53(6), pp. 1839–1885.

DellaVigna, Stefano and Joshua M Pollet (2009), "Investor inattention and Friday earnings announcements," *The Journal of Finance* 64(2), pp. 709–749.

Edmans, Alex, Diego Garcia, and Øyvind Norli (2007), "Sports sentiment and stock returns," *The Journal of Finance* 62(4), pp. 1967–1998.

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

Graeber, Thomas, Christopher Roth, and Florian Zimmermann (2024), "Stories, statistics, and memory," *The Quarterly Journal of Economics* 139(4), pp. 2181–2225.

Hirshleifer, David and Tyler Shumway (2003), "Good day sunshine: Stock returns and the weather," *The Journal of Finance* 58(3), pp. 1009–1032.

Ho, Steven Wei, Weiting Hong, and Mingrui Ray Zhang (2025), "Information Leakage Prior to SEC Form Filings–Evidence from TAQ Millisecond Data," *Working Paper*.

Hou, Kewei et al. (2025), "A tale of two anomalies: The implications of investor attention for price and earnings momentum," *Working Paper*.

Jensen, Michael C and Richard S Ruback (1983), "The market for corporate control: The scientific evidence," *Journal of Financial Economics* 11(1-4), pp. 5–50.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen (2023), "Is there a replication crisis in finance?," *The Journal of Finance* 78(5), pp. 2465–2518.

Jiang, Hao, Sophia Zhengzi Li, and Hao Wang (2021), "Pervasive underreaction: Evidence from high-frequency data," *Journal of Financial Economics* 141(2), pp. 573–599.

Jiang, Zhengyang et al. (2025), "Investor memory and biased beliefs: Evidence from the field," *The Quarterly Journal of Economics* 140(4), pp. 2749–2804.

Kamstra, Mark J, Lisa A Kramer, and Maurice D Levi (2003), "Winter blues: A SAD stock market cycle," *American Economic Review* 93(1), pp. 324–343.

Ke, Shikun (2025), "Analysts' Belief Formation in Their Own Words," *Available at SSRN 5025830*.

Kishore, Runeet et al. (2008), "Earnings announcements are full of surprises," *Available at SSRN 909563*.

Kwon, Spencer Y and Johnny Tang (2026), "Extreme categories and overreaction to news," *Review of Economic Studies* 93(2), pp. 1137–1166.

Shleifer, Andrei and Robert W Vishny (2003), "Stock market driven acquisitions," *Journal of Financial Economics* 70(3), pp. 295–311.

Shleifer, Andrei and Robert W Vishny (1997), "The limits of arbitrage," *The Journal of Finance* 52(1), pp. 35–55.

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy (2008), "More than words: Quantifying language to measure firms' fundamentals," *The Journal of Finance* 63(3), pp. 1437–1467.

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

A.1 Statistics ratio by news category

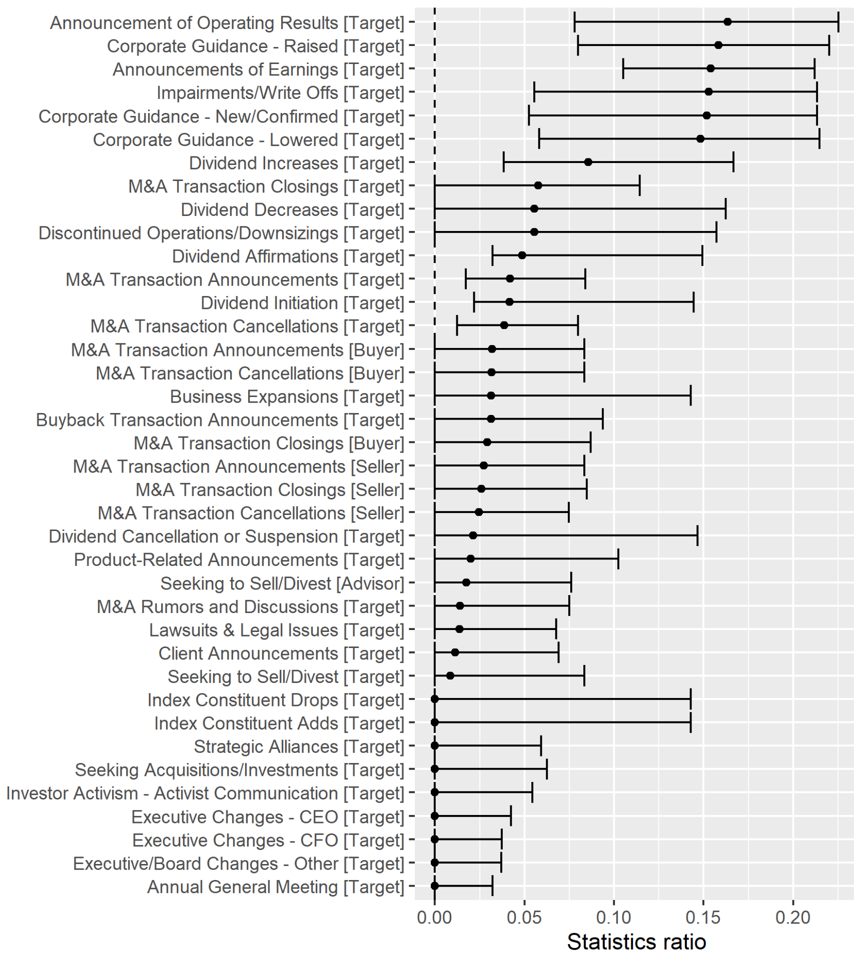


Figure A.1: Statistics ratio by news category

Note. Dots represent the median statistics ratio across all news items in the indicated news category, while the error bars represent the 5th and 95th percentiles of the corresponding distributions. Categories are arranged in descending order by their median statistics ratios. Ties are broken using the mean.

A.2 Statistics ratio by news category over time

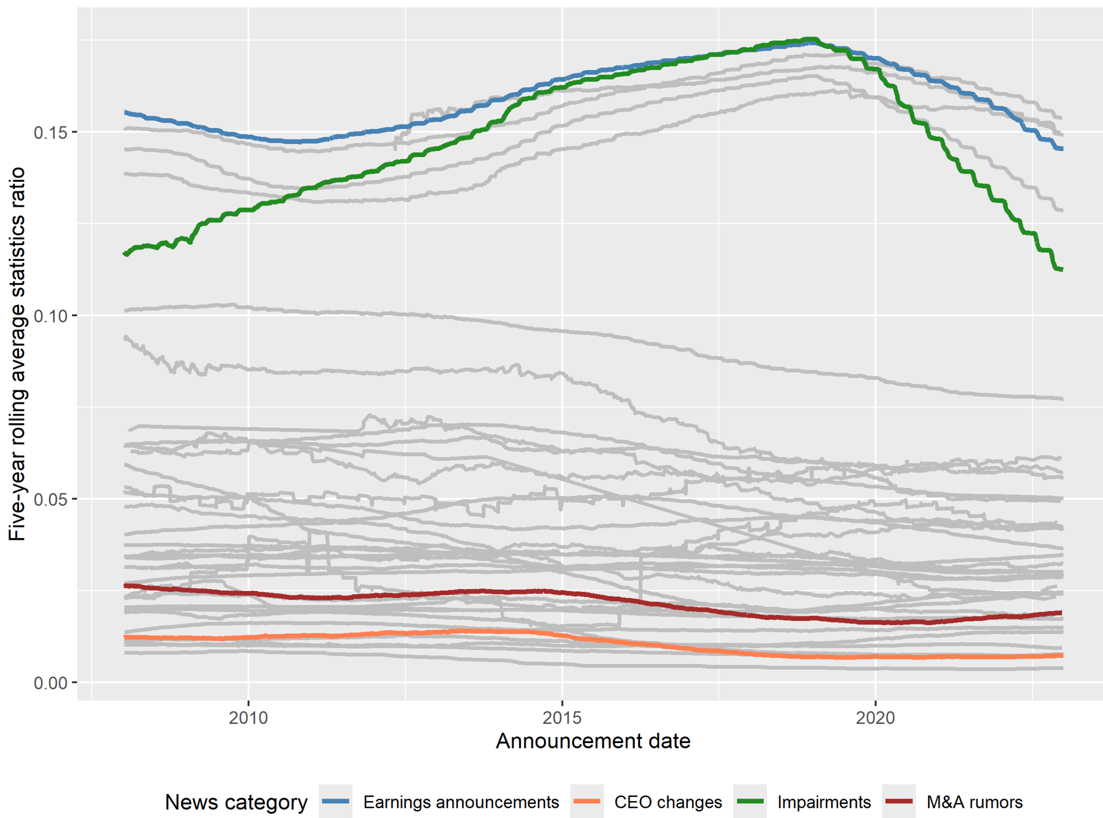


Figure A.2: Statistics ratio by news category over time

Note. Each line plots the average statistics ratio across all corporate news announcements within a news category, computed using five-year rolling windows. Earnings announcements, CEO changes, impairments, and M&A rumors are highlighted for illustration, while the remaining news categories are shown in grey.

A.3 Summary statistics by information type

Variable	Count	Mean	Std	p10	p25	p50	p75	p90
Subsequent return	380,194	0.031	0.285	-0.291	-0.116	0.027	0.166	0.341
Immediate return	380,194	0.003	0.054	-0.053	-0.021	0.001	0.025	0.058
Statistics ratio	380,194	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Abs. immediate return	380,194	0.037	0.041	0.004	0.010	0.023	0.047	0.087
Prior extremeness	277,827	0.322	0.042	0.255	0.315	0.332	0.347	0.360
Positive sentiment	380,194	0.189	0.392	0.000	0.000	0.000	0.000	1.000
Neutral sentiment	380,194	0.736	0.441	0.000	0.000	1.000	1.000	1.000
Negative sentiment	380,194	0.075	0.263	0.000	0.000	0.000	0.000	0.000
Turnover	380,194	0.010	0.010	0.002	0.004	0.007	0.012	0.019
News articles	380,194	209	974	0	0	20	76	292
Market value (USD millions)	380,194	19,150	53,271	85	306	1,511	8,240	45,666

(a) Summary statistics within the story subsample

Variable	Count	Mean	Std	p10	p25	p50	p75	p90
Subsequent return	945,684	0.034	0.273	-0.273	-0.107	0.029	0.164	0.331
Immediate return	945,684	0.004	0.067	-0.070	-0.026	0.002	0.032	0.079
Statistics ratio	945,684	0.092	0.064	0.016	0.032	0.078	0.149	0.182
Abs. immediate return	945,684	0.047	0.049	0.005	0.012	0.029	0.062	0.118
Prior extremeness	704,467	0.258	0.092	0.122	0.168	0.303	0.339	0.356
Positive sentiment	945,684	0.238	0.426	0.000	0.000	0.000	0.000	1.000
Neutral sentiment	945,684	0.461	0.498	0.000	0.000	0.000	1.000	1.000
Negative sentiment	945,684	0.301	0.459	0.000	0.000	0.000	1.000	1.000
Turnover	945,684	0.009	0.009	0.002	0.004	0.007	0.012	0.019
News articles	945,684	128	455	0	0	20	68	233
Market value (USD millions)	945,684	15,178	43,586	86	309	1,422	7,022	33,985

(b) Summary statistics within the statistics subsample

Table A.3: Summary statistics by information type

Note. Means and standard deviations are computed after winsorizing each variable at the 0.5th and 99.5th percentiles. Subsequent return is the buy-and-hold abnormal return for the announcing firm, relative to the 13 thematic factors from Jensen, Kelly, and Pedersen (2023), from two to 90 days post-announcement. Immediate return is the buy-and-hold abnormal return over the three-day window centered on the announcement date. Statistics ratio equals the number of statistical tokens in the news summary divided by its total word count. Prior extremeness corresponds to $\hat{\zeta}^{-1}$ in Kwon and Tang (2026) and is estimated over a rolling five-year window within each news category. Positive, neutral, and negative sentiment labels are generated from applying FinBERT on news summaries. Following Hou et al. (2025), turnover and news articles are the monthly average daily share turnover and the average monthly number of RavenPack news articles over the prior 12 months, respectively. Market value is the announcing firm's market capitalization on the announcement date.

A.4 News category frequency by information type

News Category	# Stat	# Story	# Obs	% of Stat	% of Story	% of All
Announcement of Operating Results [Target]	5,013	2	5,015	0.53	0.00	0.38
Corporate Guidance - Raised [Target]	15,602	92	15,694	1.65	0.02	1.18
Announcements of Earnings [Target]	244,597	4	244,601	25.86	0.00	18.45
Impairments/Write Offs [Target]	19,003	7	19,010	2.01	0.00	1.43
Corporate Guidance - New/Confirmed [Target]	112,359	2,244	114,603	11.88	0.59	8.64
Corporate Guidance - Lowered [Target]	7,856	64	7,920	0.83	0.02	0.60
Dividend Increases [Target]	21,345	63	21,408	2.26	0.02	1.61
M&A Transaction Closings [Target]	399	84	483	0.04	0.02	0.04
Dividend Decreases [Target]	1,268	77	1,345	0.13	0.02	0.10
Discontinued Operations/Downsizings [Target]	10,496	1,367	11,863	1.11	0.36	0.89
Dividend Affirmations [Target]	77,618	872	78,490	8.21	0.23	5.92
M&A Transaction Announcements [Target]	1,992	26	2,018	0.21	0.01	0.15
Dividend Initiation [Target]	407	3	410	0.04	0.00	0.03
M&A Transaction Cancellations [Target]	457	13	470	0.05	0.00	0.04
M&A Transaction Announcements [Buyer]	14,756	3,650	18,406	1.56	0.96	1.39
M&A Transaction Cancellations [Buyer]	660	151	811	0.07	0.04	0.06
Business Expansions [Target]	34,438	10,065	44,503	3.64	2.65	3.36
Buyback Transaction Announcements [Target]	11,285	4,613	15,898	1.19	1.21	1.20
M&A Transaction Closings [Buyer]	24,242	8,092	32,334	2.56	2.13	2.44
M&A Transaction Announcements [Seller]	6,605	1,734	8,339	0.70	0.46	0.63
M&A Transaction Closings [Seller]	9,570	3,058	12,628	1.01	0.80	0.95
M&A Transaction Cancellations [Seller]	295	95	390	0.03	0.02	0.03
Dividend Cancellation or Suspension [Target]	76	69	145	0.01	0.02	0.01
Product-Related Announcements [Target]	100,658	38,755	139,413	10.64	10.19	10.51
Seeking to Sell/Divest [Advisor]	145	78	223	0.02	0.02	0.02
M&A Rumors and Discussions [Target]	8,734	6,339	15,073	0.92	1.67	1.14
Lawsuits & Legal Issues [Target]	16,854	10,442	27,296	1.78	2.75	2.06
Client Announcements [Target]	83,711	61,400	145,111	8.85	16.15	10.94
Seeking to Sell/Divest [Target]	1,744	1,660	3,404	0.18	0.44	0.26
Index Constituent Drops [Target]	6,967	9,405	16,372	0.74	2.47	1.23
Index Constituent Adds [Target]	9,773	28,059	37,832	1.03	7.38	2.85
Strategic Alliances [Target]	10,053	10,418	20,471	1.06	2.74	1.54
Seeking Acquisitions/Investments [Target]	9,692	19,092	28,784	1.02	5.02	2.17
Investor Activism - Activist Communication [Target]	2,000	3,884	5,884	0.21	1.02	0.44
Executive Changes - CEO [Target]	4,154	6,432	10,586	0.44	1.69	0.80
Executive Changes - CFO [Target]	5,919	8,476	14,395	0.63	2.23	1.09
Executive/Board Changes - Other [Target]	54,968	99,118	154,086	5.81	26.07	11.62
Annual General Meeting [Target]	9,973	40,191	50,164	1.05	10.57	3.78
Total	945,684	380,194	1,325,878	100.00	100.00	100.00

Table A.4: Relative frequency of news categories within each information type

Note. This table reports the relative frequency of news categories within each information type, as defined in Section 2.2. Each column corresponds to a statistics-ratio bin in the order presented in Figure 1. All values are percentages.

A.5 News category frequency by statistics-ratio bins

News Category	% of Story	% of Stat Q1	Q2	Q3	Q4
Announcement of Operating Results [Target]	0.00	0.01	0.09	0.66	1.35
Corporate Guidance - Raised [Target]	0.02	0.02	0.24	2.44	3.89
Announcements of Earnings [Target]	0.00	0.02	1.07	44.06	58.25
Impairments/Write Offs [Target]	0.00	0.05	0.81	2.93	4.24
Corporate Guidance - New/Confirmed [Target]	0.59	0.41	3.21	18.60	25.26
Corporate Guidance - Lowered [Target]	0.02	0.03	0.27	1.39	1.63
Dividend Increases [Target]	0.02	0.15	3.59	4.38	0.88
M&A Transaction Closings [Target]	0.02	0.02	0.08	0.06	0.00
Dividend Decreases [Target]	0.02	0.03	0.32	0.14	0.05
Discontinued Operations/Downsizings [Target]	0.36	0.78	2.05	1.28	0.31
Dividend Affirmations [Target]	0.23	1.28	26.02	3.57	1.66
M&A Transaction Announcements [Target]	0.01	0.24	0.54	0.06	0.00
Dividend Initiation [Target]	0.00	0.04	0.11	0.02	0.01
M&A Transaction Cancellations [Target]	0.00	0.07	0.11	0.01	0.00
M&A Transaction Announcements [Buyer]	0.96	2.41	3.31	0.50	0.01
M&A Transaction Cancellations [Buyer]	0.04	0.11	0.15	0.02	0.00
Business Expansions [Target]	2.65	5.35	5.95	2.41	0.84
Buyback Transaction Announcements [Target]	1.21	1.50	2.67	0.55	0.04
M&A Transaction Closings [Buyer]	2.13	3.98	5.23	1.01	0.02
M&A Transaction Announcements [Seller]	0.46	1.27	1.32	0.20	0.01
M&A Transaction Closings [Seller]	0.80	1.87	1.82	0.35	0.01
M&A Transaction Cancellations [Seller]	0.02	0.06	0.06	0.01	0.00
Dividend Cancellation or Suspension [Target]	0.02	0.00	0.01	0.01	0.00
Product-Related Announcements [Target]	10.19	21.12	15.56	5.37	0.60
Seeking to Sell/Divest [Advisor]	0.02	0.04	0.02	0.00	0.00
M&A Rumors and Discussions [Target]	1.67	2.06	1.37	0.26	0.01
Lawsuits & Legal Issues [Target]	2.75	4.28	2.53	0.33	0.01
Client Announcements [Target]	16.15	21.46	11.93	2.01	0.12
Seeking to Sell/Divest [Target]	0.44	0.33	0.32	0.08	0.00
Index Constituent Drops [Target]	2.47	0.00	0.07	2.70	0.19
Index Constituent Adds [Target]	7.38	0.00	0.33	3.39	0.42
Strategic Alliances [Target]	2.74	2.85	1.24	0.17	0.00
Seeking Acquisitions/Investments [Target]	5.02	2.09	1.70	0.30	0.02
Investor Activism - Activist Communication [Target]	1.02	0.51	0.30	0.04	0.00
Executive Changes - CEO [Target]	1.69	1.39	0.30	0.06	0.02
Executive Changes - CFO [Target]	2.23	2.10	0.34	0.07	0.02
Executive/Board Changes - Other [Target]	26.07	18.89	3.98	0.45	0.12
Annual General Meeting [Target]	10.57	3.16	1.01	0.07	0.00

Table A.5: Relative frequency of news categories within each statistics-ratio bin

Note. This table reports the relative frequency of news categories within each statistics-ratio bin constructed in Section 3.1. Each column corresponds to a statistics-ratio bin, ordered as in Figure 1. Values are percentages.

A.6 Misreaction results for alternative hypotheses

Controls	Uncertainty (1)	Diagnostic (2)	Subsequent return Sentiment (3)	(4)	Attention (5)	(6)
Immediate return	-0.082**	-0.024	-0.009	0.049	0.042	0.012
	(0.035)	(0.041)	(0.039)	(0.041)	(0.041)	(0.029)
Immediate return × Abs. immediate ret.	0.050**					
	(0.020)					
Immediate return × Prior extremeness		-0.058***				
		(0.006)				
Immediate return × Neg. sentiment			0.081***			
			(0.011)			
Immediate return × Pos. sentiment			0.111***			
			(0.020)			
Immediate return × Turnover				-0.025***		
				(0.008)		
Immediate return × News articles					-0.031***	
					(0.007)	
Immediate return × Log market value						-0.052**
						(0.022)
Observations	1,325,878	982,294	1,325,878	1,325,878	1,325,878	1,325,878
R^2	0.085	0.082	0.084	0.084	0.084	0.114
Main effects of controls	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Firm fixed effects	✓	✓	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓	✓	✓
Clustered standard errors	✓	✓	✓	✓	✓	✓

Table A.6: Misreaction results for alternative hypotheses

Note. Each column reports the regression results for a given control variable using the framework of Equation (2) in isolation. Using the absolute value of the immediate return as a proxy for unobserved signal strength, column (1) reports the regression results for signal strength uncertainty (Augenblick, Lazarus, and Thaler, 2025). Column (2) reports the regression results for diagnostic expectations (Kwon and Tang, 2026). Column (3) reports the regression results for news sentiment using FinBERT (Tetlock, Saar-Tsechansky, and Macskassy, 2008). Columns (4) through (6) report the regression results for the three investor attention variables from Hou et al. (2025). All control variables are standardized. Significance code: 0.1 * 0.05 ** 0.01 ***.

A.7 Misreaction results with WSJ articles

Controls	None	Uncertainty	Diagnostic	Sentiment	Attention	All
			Subsequent return			
	(1)	(2)	(3)	(4)	(5)	(6)
Immediate return	-0.122***	-0.163***	-0.137***	0.003	-0.090**	-0.153*
	(0.032)	(0.055)	(0.019)	(0.053)	(0.044)	(0.077)
Immediate return × Statistics ratio	0.078***	0.094***	0.118**	0.098***	0.095**	0.112**
	(0.027)	(0.030)	(0.048)	(0.026)	(0.037)	(0.046)
Immediate return × Abs. immediate ret.		-0.002				0.007
		(0.002)				(0.006)
Immediate return × Prior extremeness			0.052			0.030
			(0.045)			(0.044)
Immediate return × Negative sentiment				-0.221**		-0.026
				(0.085)		(0.083)
Immediate return × Positive sentiment				-0.035		0.130
				(0.043)		(0.089)
Immediate return × Turnover					-0.025**	-0.051
					(0.010)	(0.034)
Immediate return × News articles					0.017	0.045
					(0.032)	(0.040)
Immediate return × Log market value					0.004	0.012
					(0.057)	(0.051)
Observations	121,667	121,667	85,260	121,667	86,971	84,363
R^2	0.249	0.251	0.138	0.250	0.156	0.156
Main effects of controls	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Firm fixed effects	✓	✓	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓	✓	✓
Clustered standard errors	✓	✓	✓	✓	✓	✓

Table A.7: Misreaction regression with interacted statistics ratio, WSJ articles

Note. Each column reports the regression coefficients from Equation (2) estimated on the sample of full-length Wall Street Journal articles as described in Section 2.1. Variables other than subsequent and immediate returns are standardized for interpretability. The main effects of the control variables are included in the regression but omitted from display. Standard errors, shown in parentheses, are clustered at the announcement year, firm, and news category levels. Significance code: 0.1 * 0.05 ** 0.01 ***.

A.8 Misreaction results without known anomalies

Specification	Subsequent return Baseline (1)	 No earnings (2)	 No M&A (3)	 Neither (4)
Immediate return	-0.146***	-0.117***	-0.148***	-0.120***
	(0.025)	(0.030)	(0.025)	(0.030)
Immediate return × Statistics ratio	0.039**	0.057***	0.042**	0.059***
	(0.015)	(0.017)	(0.016)	(0.017)
Immediate return × Abs. immediate ret.	0.041**	0.038*	0.041**	0.039*
	(0.019)	(0.021)	(0.019)	(0.021)
Immediate return × Prior extremeness	-0.024	-0.063***	-0.019	-0.054**
	(0.018)	(0.014)	(0.019)	(0.018)
Immediate return × Neg. sentiment	-0.004	-0.023	0.001	-0.016
	(0.028)	(0.072)	(0.025)	(0.070)
Immediate return × Pos. sentiment	0.045	0.085***	0.046*	0.083***
	(0.028)	(0.014)	(0.026)	(0.012)
Immediate return × Turnover	-0.020**	-0.016*	-0.018**	-0.013
	(0.008)	(0.009)	(0.008)	(0.009)
Immediate return × News articles	0.000	-0.012	0.001	-0.014
	(0.008)	(0.008)	(0.010)	(0.011)
Immediate return × Log market value	-0.019	0.011	-0.017	0.018
	(0.019)	(0.013)	(0.020)	(0.012)
Observations	982,294	693,140	915,749	626,595
R^2	0.117	0.127	0.118	0.129
Main effects of controls	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Firm fixed effects	✓	✓	✓	✓
Category fixed effects	✓	✓	✓	✓
Clustered standard errors	✓	✓	✓	✓

Table A.8: Misreaction regression with interacted statistics ratio, excluding known anomalies

Note. Each column reports the regression coefficients from Equation (2) estimated on the sample of corporate news announcements outside of the indicated news categories. "No earnings" column excludes earnings, operating results, and corporate guidances. "No M&A" column excludes M&A rumors, announcements, closings, and cancellations. "Neither" column excludes both sets of news categories. Variables other than subsequent and immediate returns are standardized for interpretability. The main effects of the control variables are included in the regression but omitted from display. Standard errors, shown in parentheses, are clustered at the announcement year, firm, and news category levels. Significance code: 0.1 * 0.05 ** 0.01 ***.

A.9 Misreaction-managed portfolio, large firms

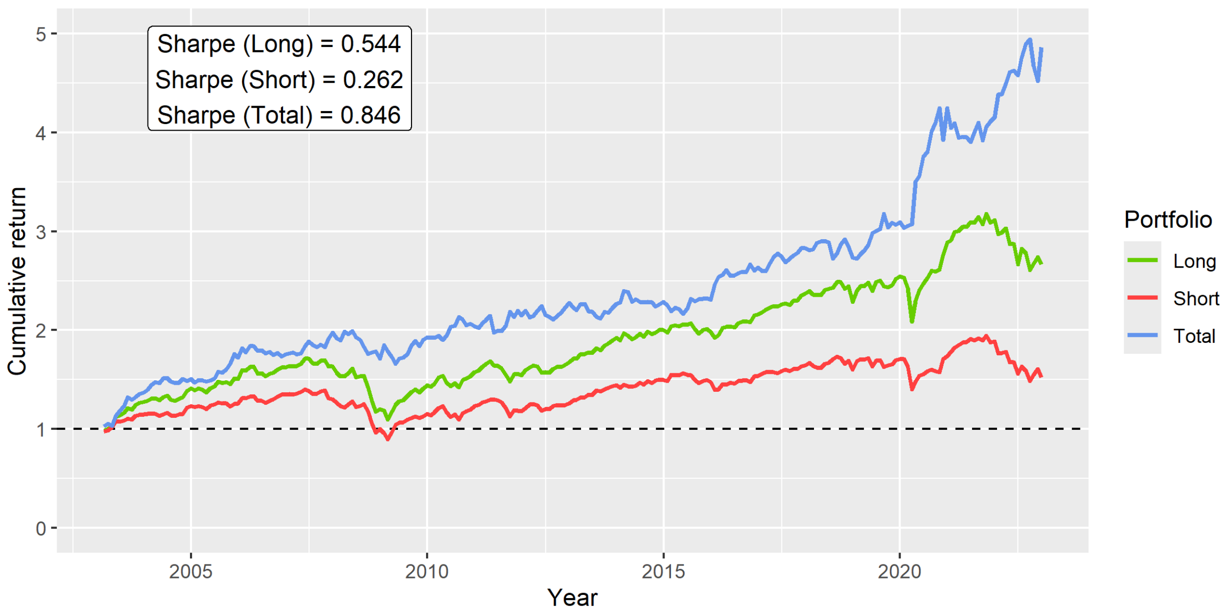


Figure A.9: Cumulative returns from the misreaction-managed portfolio on statistics ratio, large firms

Note. I focus on firms whose average market capitalization, computed over their news announcements in the previous month, exceeds the 70th percentile of all CRSP firms in the same month. For each firm-month, I calculate the average statistics ratio and the average immediate return across the firm's news announcements in the previous month. I standardize the two characteristics across firms within each month and take their firm-level products as the portfolio weights for the current month. The long leg contains the most underreacted (statistical) firms with the highest immediate returns, as well as the least underreacted (story-like) firms with the lowest immediate returns. The short leg contains the most underreacted (statistical) firms with the lowest immediate returns, as well as the least underreacted (story-like) firms with the highest immediate returns. The total return of the misreaction-managed portfolio is the buy-and-hold return on the monthly difference between the long and short legs. All Sharpe ratios are annualized. All portfolios are scaled to 10% annual volatility to facilitate interpretability.

A.10 Misreaction-managed portfolio, large firms

	Portfolio excess return				
Factor model	FF3	FF3+Mom.	FF5	FF5+Mom.	JKP13
	(1)	(2)	(3)	(4)	(5)
Constant	0.007***	0.007***	0.007***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Market excess return	-0.007	0.004	0.002	0.010	
	(0.046)	(0.050)	(0.047)	(0.051)	
Small minus big	0.043	0.043	0.025	0.025	
	(0.084)	(0.084)	(0.090)	(0.090)	
High minus low	-0.144**	-0.132**	-0.196**	-0.184**	
	(0.063)	(0.066)	(0.079)	(0.083)	
Up minus down		0.030		0.024	
		(0.050)		(0.050)	
Conservative minus aggressive			0.164	0.156	
			(0.128)	(0.129)	
Robust minus weak			-0.075	-0.077	
			(0.106)	(0.106)	
Observations	239	239	239	239	239
R^2	0.022	0.024	0.032	0.033	0.085

Table A.10: Regression of misreaction-managed portfolio excess returns on factor models, large firms

Note. I focus on firms whose average market capitalization, computed over their news announcements in the previous month, exceeds the 70th percentile of all CRSP firms in the same month. Misreaction-managed portfolio excess returns and factor returns are measured at the monthly frequency. Coefficient estimates for the 13 thematic factors from Jensen, Kelly, and Pedersen (2023) are omitted for brevity. Significance code: 0.1 * 0.05 ** 0.01 ***.

A.11 News momentum strategy

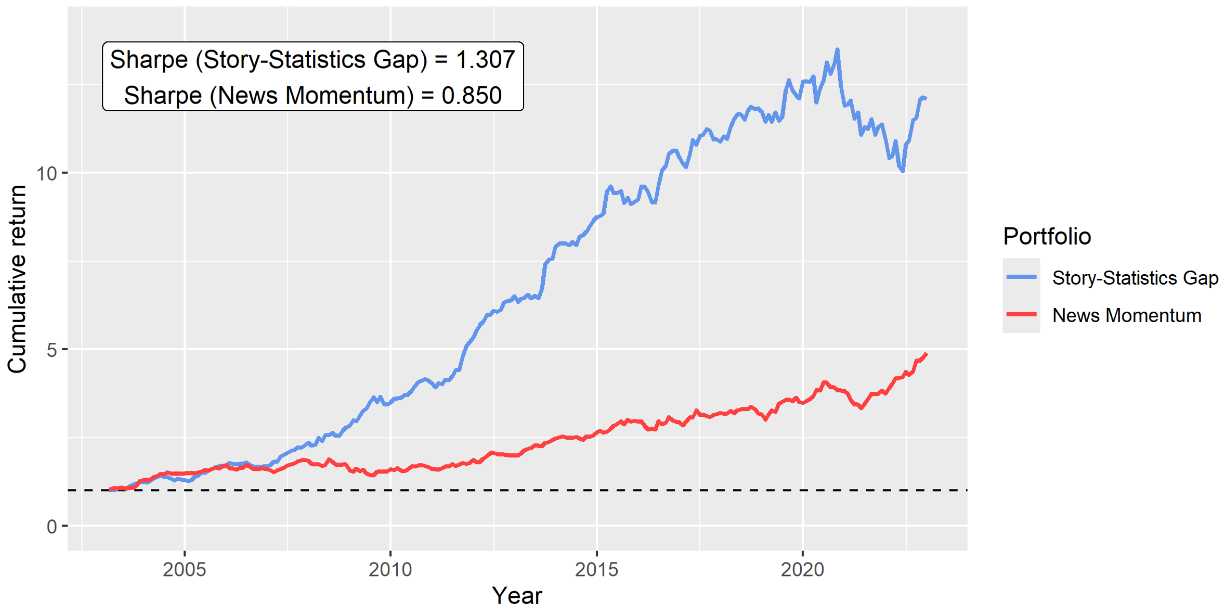


Figure A.11: Cumulative excess returns from the misreaction-managed portfolio and the news momentum strategy

Note. The misreaction-managed portfolio is constructed following the procedure described in the notes accompanying Figure 3. To construct the comparable news momentum strategy, I change the portfolio weights from the product between the two standardized firm-level characteristics with only the standardized average immediate returns across news announcements in the previous month. All Sharpe ratios are annualized. All portfolios are scaled to 10% annual volatility to facilitate interpretability.

A.12 News momentum strategy

Factor model	FF3 (1)	FF3+Mom. (2)	FF5 (3)	FF5+Mom. (4)	JKP13 (5)
	News momentum excess return				
Constant	0.008***	0.007***	0.007***	0.007***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Market excess return	-0.103**	-0.036	-0.087*	-0.026	
	(0.045)	(0.047)	(0.046)	(0.048)	
Small minus big	-0.033	-0.028	0.000	-0.001	
	(0.082)	(0.079)	(0.088)	(0.085)	
High minus low	-0.152**	-0.079	-0.225***	-0.135*	
	(0.061)	(0.063)	(0.077)	(0.079)	
Up minus down		0.182***		0.176***	
		(0.048)		(0.048)	
Conservative minus aggressive			0.185	0.132	
			(0.125)	(0.123)	
Robust minus weak			0.087	0.077	
			(0.103)	(0.100)	
Observations	239	239	239	239	239
R^2	0.068	0.123	0.078	0.129	0.250

Table A.12: Regression of news momentum returns on factor models

Note. News momentum strategy excess returns and factor returns are measured at the monthly frequency. Coefficient estimates for the 13 thematic factors from Jensen, Kelly, and Pedersen (2023) are omitted for brevity. Significance code: 0.1 * 0.05 ** 0.01 ***.

A.13 Pair regression diff. category effect heterogeneity

	Δ |Immediate return|		
Controls	None	Information type	With interaction
	(1)	(2)	(3)
Δ Statistics ratio	0.015**	0.018**	0.016**
	[0.002, 0.029]	[0.002, 0.034]	[0.001, 0.032]
...	...	...	...
Different news category	-0.127***	-0.115***	0.032
	[-0.158, -0.095]	[-0.145, -0.082]	[-0.013, 0.076]
Both story		-0.059***	0.003
		[-0.084, -0.035]	[-0.067, 0.069]
Both statistic		0.096***	0.298***
		[0.076, 0.117]	[0.224, 0.377]
Diff. category × Both story			-0.058
			[-0.124, 0.017]
Diff. category × Both statistic			-0.216***
			[-0.283, -0.147]
R^2	0.190	0.193	0.193
	[0.176, 0.210]	[0.178, 0.212]	[0.179, 0.213]
All pair controls	✓	✓	✓
Year fixed effects	✓	✓	✓
Firm fixed effects	✓	✓	✓
Two-way cluster bootstrap	✓	✓	✓

Table A.13: Bootstrapped regression coefficients of the pairwise difference in the absolute value of immediate return on the pairwise difference in statistics ratio and various controls

Note. I draw 1,000 bootstrap samples of 100,000 pairs of news announcements with two-way clustering at the announcement year and firm levels. For each sample, I regress the pairwise difference in the absolute value of immediate return on the pairwise difference in statistics ratio and various controls. I include announcement year and firm fixed effects from both sides of the news pairs. Pairwise differences between continuous variables are indicated by the Δ prefix. All Δ variables are standardized for interpretability. Pairwise differences between categorical variables are dummy variables that indicate one if the news pair differs on that dimension. The coefficient estimates are the means of the bootstrap distributions of the regression coefficients. 95% confidence intervals from the bootstrap distributions are indicated in brackets below each coefficient estimate. Significance code: 0.1 * 0.05 ** 0.01 ***.

A.14 Example summary for each news category

News Category	Example Summary
Announcement of Operating Results [Target]	AEP Industries Inc. reported unaudited consolidated earnings results for the second quarter and six months ended April 30, 2015. For the quarter, the company reported net sales of $285,722,000 against $294,942,000 a year ago. Operating income was $20,843,000 against $1,927,000 a year ago. Income before benefit for income taxes was $16,181,000 against loss before benefit for income taxes of $3,022,000 a year ago. Net income was $11,128,000 against net loss of $2,724,000 a year ago. Diluted earnings per share were $2.18 against diluted loss per share of $0.49 per share a year ago. Adjusted EBITDA was $23,051,000 against $14,825,000 a year ago. The decrease in net sales of a 2% decrease in average selling prices primarily due to the pass through of lower resin costs compared to the same period in the prior fiscal year.For the six months, the company reported net sales of $561,161,000 against $567,459,000 a year ago. Operating income was $26,416,000 against $259,000 a year ago. Income before benefit for income taxes was $16,838,000 against loss before benefit for income taxes of $9,386,000 a year ago. Net income was $11,604,000 against net loss of $6,451,000 a year ago. Diluted earnings per share were $2.27 against diluted loss per share of $1.16 per share a year ago. Adjusted EBITDA was $29,602,000 against $22,357,000 a year ago. The decrease in net sales were the result of a 1% decrease in sales volume partially offset by a 0.8% increase in average selling prices compared to th
Corporate Guidance - Raised [Target]	AAC Holdings, Inc. provided revised earnings guidance for the full year 2018. The company is revising its guidance for the full year 2018 to include AdCare from the March 1, 2018 acquisition date. Results for 2018 full year guidance have been provided for the 10 months inclusive of ASC 606, and for comparability only, 2018 full year guidance has also been provided excluding the impact of ASC 606. The company currently expects total revenues to be $360 million to $375 million. Adjusted EBITDA are expected to be $68 million to $72 million. Adjusted Earnings per Diluted Common Share are expected to be $0.75 to $0.80. Adjusted for Adoption of ASC 606 on January 1, 2018, the company expects total revenues to be $325 million to $340 million. Adjusted EBITDA are expected to be $68 million to $72 million. Adjusted Earnings per Diluted Common Share are expected to be $0.75 to $0.80. AAC expects an annual effective tax rate of 24% to 26% and diluted weighted-average common shares outstanding of approximately 24.5 million for the year.
Announcements of Earnings [Target]	AAC Holdings, Inc. announced unaudited consolidated earnings results for the fourth quarter and full year ended December 31, 2014. For the quarter, the company announced revenue of $37,166,000 compared to $28,060,000 for the same period a year ago. Income from operations was $4,452,000 compared to loss from operations of $2,829,000 for the same period a year ago. Income before income tax was $3,458,000 compared to loss before income tax of $3,001,000 for the same period a year ago. Net income was $3,273,000 compared to net loss of $1,968,000 for the same period a year ago. Net income attributable to the company stockholders was $3,354,000 compared to net loss attributable to the company stockholders of $1,945,000 for the same period a year ago. Net income attributable to the company common stockholders was $3,109,000 or 0.15 per basic and diluted share compared to net loss attributable to the company common stockholders of $1,945,000 or 0.14 per basic and diluted share for the same period a year ago. Adjusted EBITDA was $6,559,000 compared to adjusted LBITDA of $1,081,000 for the same period a year ago. Capital expenditures in the fourth quarter of 2014 totaled $3.6 million. Cash flows provided by operations totaled $2.2 million for the fourth quarter of 2014 compared with $2.1 million in the prior-year period and $5.8 million in the third quarter of 2014.For the full year, the company announced revenue of $132,968,000 compared to $115,741,000 for the same period a year ago.
Impairments/Write Offs [Target]	Aadi Bioscience, Inc. reported impairment for the third quarter ended September 30, 2021. For the quarter, the company reported impairment of intangible asset of $74,156,000.

Corporate Guidance - New/Confirmed [Target]	AAC Holdings, Inc. announced unaudited consolidated earnings results for the fourth quarter and full year ended December 31, 2014. For the quarter, the company announced revenue of $37,166,000 compared to $28,060,000 for the same period a year ago. Income from operations was $4,452,000 compared to loss from operations of $2,829,000 for the same period a year ago. Income before income tax was $3,458,000 compared to loss before income tax of $3,001,000 for the same period a year ago. Net income was $3,273,000 compared to net loss of $1,968,000 for the same period a year ago. Net income attributable to the company stockholders was $3,354,000 compared to net loss attributable to the company stockholders of $1,945,000 for the same period a year ago. Net income attributable to the company common stockholders was $3,109,000 or 0.15 per basic and diluted share compared to net loss attributable to the company common stockholders of $1,945,000 or 0.14 per basic and diluted share for the same period a year ago. Adjusted EBITDA was $6,559,000 compared to adjusted LBITDA of $1,081,000 for the same period a year ago. Capital expenditures in the fourth quarter of 2014 totaled $3.6 million. Cash flows provided by operations totaled $2.2 million for the fourth quarter of 2014 compared with $2.1 million in the prior-year period and $5.8 million in the third quarter of 2014.For the full year, the company announced revenue of $132,968,000 compared to $115,741,000 for the same period a year ago.
Corporate Guidance - Lowered [Target]	AAR Corp. reported unaudited consolidated earnings results for the second quarter and six months ended November 30, 2013. For the quarter, the company reported sales of $540.7 million against $512.8 million last year. Operating income was $40.7 million against $37.8 million last year. Income before income tax expense was $30.5 million against $27.2 million last year. Net income attributable to AAR was $20.0 million or $0.50 per diluted share against $17.8 million or $0.44 per diluted share last year. Net cash provided from operating activities was $38.8 million. Investment on property, plant and equipment expenditures was $7.1 million. Aviation Services segment performed well in the quarter, with strong revenue growth, while Technology Products declined slightly in comparison to prior year. During the second quarter, the company generated $38.8 million of cash from operations. The company had $31.7 million of free cash flow. For the six months, the company reported sales of $1,055.2 million against $1,063.3 million last year. Operating income was $78.9 million against $76.2 million last year. Income before income tax expense was $58.0 million against $55.2 million last year. Net income attributable to AAR was $37.9 million or $0.95 per diluted share against $36.0 million or $0.89 per diluted share last year. Book value per share was $24.42 against $23.34 at May 31, 2013. Net debt was $582.3 million against $709.7 million at November 30, 2012.The company is adjusting its full
Dividend Increases [Target]	AAON Inc. announced that the amount of the Company's semi-annual cash dividend has been increased to $0.16 per share or $0.32 annually.
M&A Transaction Closings [Target]	Michael Cartwright and one of the Directors bought up a large chunk of shares in AAC Holdings, Inc. (NYSE:AAC) on August 4, 2015. Michael Cartwright and one of the Directors acquired a large chunk of shares in AAC Holdings, Inc. (NYSE:AAC) on August 4, 2015.
Dividend Decreases [Target]	The board of directors of AAON Inc. has declared a semi-annual cash dividend of $0.12 per share to the holders of the outstanding common stock of the company as of the close of business on December 1, 2011, the record date, and payable on December 22, 2011.
Discontinued Operations/Downsizings [Target]	AAC Holdings, Inc. announced a cost reduction initiative to improve operating performance across the organization. AAC's implemented cost reduction program is currently expected to reduce the Company's total expenses on an annualized basis by approximately $15 million and includes an aggregate reduction of approximately 200 positions across the Company. The actions taken include: The consolidation of the San Diego outpatient and sober-living facilities into the Laguna operations in California to provide more efficient operations and stronger financial performance. The Company expects this consolidation to be fully executed prior to December 31, 2018; A strategic alternative for Townsend operations in Louisiana; and Effective November 30, 2018, a reduction of approximately 100 positions, including corporate functions, which is intended to align staffing levels with current occupancy.
Dividend Affirmations [Target]	The Board of Directors of AAON Inc. has declared a semi-annual cash dividend of $0.20 per share to the holders of the outstanding Common Stock of the Company as of the close of business on December 13, 2006, the record date, payable on January 3, 2007.

M&A Transaction Announcements [Target]	Tyco Electronics, Ltd. (NYSE: TEL) entered into a definitive agreement to acquire ADC Telecommunications Inc. (NasdaqGS: ADCT) for $1.2 billion in cash on July 12, 2010. As reported under the terms of the transaction, Tyco Electronics will launch a tender offer to acquire all the issued and outstanding common stock of ADC for $12.75 per share in cash. Tyco Electronics, Ltd. has sufficient cash, available lines of credit or other sources of immediately available funds to enable it to consummate the offer. The agreement also provides a break up fee of $38 million to be paid by ADC in the event of termination of the transaction. ADC Telecommunications granted Tyco Electronics with top up option to acquire the minimum number of new shares which when added to the tendered shares constitutes 90% plus one share of the total issued and outstanding common stock of ADC. The tender offer commenced on July 26, 2010 and will expire on August 23, 2010. The transaction is subject to customary closing conditions, including the tender of a majority of ADC shares, approval by the shareholders of ADC Telecommunications, regulatory approvals and applicable anti-trust reviews. The transaction is not subject to any financing condition. The Board formed a special committee which included all directors other than Switz. The Special Committee unanimously approved the transaction. The transaction was unanimously approved by the Board of Directors of ADC and Tyco Electronics. The deal is expected to
Dividend Initiation [Target]	AAON Inc. announced that, the Board of Directors has voted to initiate a semi-annual cash dividend of $0.20 per share to the holders of the outstanding Common Stock of the Company as of the close of business on June 12, 2006, the record date, payable on July 3, 2006.
M&A Transaction Cancellations [Target]	Argonne Capital Group, LLC entered into a definitive merger agreement to acquire ALCO Stores, Inc. (NasdaqGM:ALCS) from MFP Investors LLC, Dimensional Fund Advisors LP, Heartland Advisors, Inc., Aegis Financial Corporation, Everbright Development Overseas Ltd. and others for $45.9 million in cash on July 25, 2013. Under the terms of the transaction each common stock holder and restricted stock holder will receive $14 in cash and each option will be paid an amount equal to the excess, if any, of $14 over the exercise price for each option held. Wells Fargo, N.A. and GB Credit Partners, Inc. will provide debt financing on the terms and subject to the conditions set forth in their respective debt commitment letters. Mallard Investment Partners, LLC will provide equity financing for the transaction. Upon termination of the agreement by Alco Stores, it has to pay a termination fee of $2.25 million which amount is reduced to $1.75 million under specified conditions, or reimburse buyer and its affiliates for transaction expenses up to a maximum of $1 million, subject to specified exceptions. Upon termination of the transaction, buyer will be required to pay Alco a termination fee of $3.5 million. Argonne Capital Group will make a deposit of $3.5 million within two business day of signing the agreement.In addition, subject to specified exceptions and limitations, either party may terminate the transaction if the merger is not consummated by December 22, 2013. Alco Stores can exercise
M&A Transaction Announcements [Buyer]	AAON, Inc. announced its acquisition of Air Wise Inc.
M&A Transaction Cancellations [Buyer]	ADC Telecommunications Inc. (Nasdaq: ADCT) signed a definitive agreement to acquire Andrew Corporation (Nasdaq: ANDW) for a reported value of approximately $2.1 billion in stock on May 30, 2006. As per the terms of the agreement, Andrew shareholders will receive 0.57 of an ADC common share for each common share of Andrew they hold and each outstanding option and warrant will be converted into ADC options and warrants respectively. Also, each share of Andrew's restricted stock will be converted into ADC shares in the same ratio as the common stock exchange ratio. Upon completion of the transaction, ADC shareholders will own approximately 56 percent of the combined company and Andrew shareholders will own approximately 44 percent of the combined company. ADC will assume all debt of Andrew and Andrew's convertible notes will become convertible into ADC shares. Pequot Ventures also sold its stake in the transaction. Post closing, Andrew will become a wholly owned subsidiary of ADC. In the event of termination, either of, ADC or Andrew will be obligated to pay a termination fee of approximately $75 million. The combined company will be based at ADC's world headquarters in Minnesota with ADC's John A. Blanchard continuing as non-executive chairman, and ADC's Robert E. Switz continuing as its president and CEO. Ralph E. Faison, president and CEO of Andrew, will serve as a consultant to the combined company to facilitate an efficient transition. The board of directors of the co
Business Expansions [Target]	AAC Holdings, Inc. announced that it has commenced construction on a new 11,000-square-foot in-network lab in Slidell that is expected to be completed in the third quarter of 2016 and replace the existing Townsend in-network lab.
Buyback Transaction Announcements [Target]	The Board of Directors of AAON Inc. (NasdaqNM: AAON) initiated a stock buyback program to repurchase up to 10% of its outstanding stock, representing 1,325,000 shares on October 17, 2002. The company will fund the stock purchases out of existing cash and future earnings.

M&A Transaction Closings [Buyer]	AAC Holdings, Inc. (NYSE:AAC) signed a definitive agreement to acquire Clinical Services of Rhode Island, Inc. from Reinhard Straub and Peter Letendre for $1.2 million in cash and stock on January 23, 2015. Under the terms of the agreement, AAC Holdings, Inc. will pay $0.665 million in cash and will issue 0.04 million restricted shares of $1.2 million worth of restricted shares of AAC Holdings' common stock. Following the acquisition, Reinhard Straub will continue as Clinical and Business Development Liaison for American Addiction Centers, Inc., and Peter Letendre will serve as Chief Executive Officer and Clinical Director Rhode Island Operations. Clinical Services of Rhode Island, Inc. generated revenues of approximately $1.2 million and adjusted EBITDA of approximately $0.2 million for the year ended December 31, 2014. The acquisition, which is subject to certain closing conditions such as the assignment of certain contracts and the receipt of certain licenses necessary to operate the business, is expected to close during the second quarter of 2015. AAC Holdings, Inc. (NYSE:AAC) completed the acquisition of Clinical Services of Rhode Island, Inc. from Reinhard Straub and Peter Letendre on April 17, 2015.
M&A Transaction Announcements [Seller]	Chromalloy Gas Turbine Corporation acquired Engine Component Repair business unit of AAR Corp. (AIR) for $7.7 million on February 17, 2005. AAR will record an after-tax charge of approximately $2.1 million associated with the sale in its fiscal third quarter ending February 28, 2005. As part of the transaction, AAR will enter into a multi-year parts supply agreement with Chromalloy. It will also acquire certain assets from Chromalloy's inventory and receive cash consideration of $7.7 million. AAR expects the transaction to positively impact earnings beginning in its fiscal 2005 fourth quarter. Jefferies & Company, Inc. advised AAR Corp. on the deal.
M&A Transaction Closings [Seller]	Dixie Aerospace, Inc. agreed to acquire parts manufacturer approval line of business (PMA) from AAR Corp. (NYSE: AIR) on May 26, 2011. The purchase includes all PMA inventory and associated intellectual property. Dixie Aerospace also entered into a long-term sales, distribution and new parts development agreement whereby Dixie Aerospace will provide PMA parts on a preferred supplier basis. Dixie Aerospace will work with AAR's customers on new PMA development opportunities. Dixie Aerospace, Inc. completed the acquisition of parts manufacturer approval line of business (PMA) from AAR Corp. (NYSE: AIR) on May 31, 2011.
M&A Transaction Cancellations [Seller]	An unkown buyer signed an agreement to acquire Certain Contracts and Assets of Contractor-Operated Business of AAR Corp. (NYSE:AIR) Subsequent to the end of third quarter of AAR. The transaction is subject to certain regulatory approvals and is expected to close before the end of calendar 2019.An unkown buyer cancelled the acquisition of Certain Contracts and Assets of Contractor-Operated Business of AAR Corp. (NYSE:AIR) on February 28, 2020.
Dividend Cancellation or Suspension [Target]	ARC Document Solutions, Inc. announced that dividend payments have been suspended for the balance of the year 2020 due to the uncertainty surrounding the COVID-19 pandemic.
Product-Related Announcements [Target]	Addiction Labs of America, the laboratory division of American Addiction Centers is in the early stages of developing a pharmacogenetics test for cannabis and other cannabis products, such as cannabidiol (CBD). It's estimated a quarter of U.S. adults are taking some form of CBD, and a new study in the American Journal of Psychiatry is now touting its potential benefits for opioid use disorder. Addiction Labs specializes in pharmacogenetics and already offers testing on more than 150 prescribed medications, mostly in the mental health arena. This information helps doctors prescribe personalized medicine for those with substance use disorder and co-occurring mental health conditions, ultimately improving patient outcomes. The cannabis testing would be the first time the lab has investigated the pharmacogenetics of a formerly illicit drug. As the cannabis market and products like CBD continue to grow in popularity, Dr. Calarco also cautions that it's unlikely CBD will be the panacea for opioid use disorder that many are hoping.
Seeking to Sell/Divest [Advisor]	Marylebone Warwick Balfour Group Plc (MWB) on July 2 announced that it had appointed Bank of America (BoA) to secure a buyer for the Malmaison and Hotel du Vin property assets following the failure of Vector Hospitality Plc to float and complete its proposed acquisition of those assets. The appointment of BoA is in line with MWB's stated strategy of returning cash to shareholders through its Cash Distribution Program. The Board has appointed BoA after reviewing proposals from a number of investment banks. When a sale of the Property Portfolio has been agreed, MWB's Board will seek shareholders' approval for the disposal, in exactly the same manner as it did on 4 May 2007 when seeking approval for the proposed sale to Vector. At the Company's EGM, held on 22 May 2007, over 90% of those voting in person or by proxy were in favor of the sale to Vector for £495.1 million, representing a £109 million surplus over the independent valuation of the Property Portfolio at 31st December 2006, undertaken for MWB's year end accounts.

M&A Rumors and Discussions [Target]	AAR Corp. (NYSE:AIR) has begun the sale of Telair International GmbH in a deal that could value the German business at up to €800 million, reported Reuters citing three sources familiar with the matter. The sources said that Cinven Limited and EQT Partners AB are among those through to the second round of the process being run by Citi. One of the sources, declining to be identified because the matter is private, said that second-round bids are due early 2015. A source familiar with the process said that Telair could also appeal to rivals such as TransDigm Group Incorporated (NYSE:TDG) while it remains unclear which strategic players will submit final bids. The sources said that annual core earnings (EBITDA) at Telair are about €80 million and the company could fetch up to ten times that. The sources also said that no deal was certain and the process could fall through. AAR, Citi and Cinven declined to comment to Reuters. Telair, TransDigm and EQT were not immediately available for comment.
Lawsuits & Legal Issues [Target]	Block & Leviton LLP announced that it has filed a class action lawsuit alleging violations of the federal securities laws against AAC Holdings, Inc. and certain of its current and former officers and directors. The case is pending in the United District Court for the Middle District of Tennessee, No. 3:15-cv-00923. The complaint was brought on behalf of all investors who purchased or otherwise acquired the company's securities between October 2, 2014 and August 3, 2015 (class period). The complaint alleges that, throughout the class period, defendants made false and misleading statements and failed to disclose material information, including with respect to legal proceedings brought against subsidiaries of the company and several former and one current employees, including its President at the time, Jerrod N. Menz. On August 3, 2015, the company revealed that second degree murder and dependent adult abuse indictments had been brought in California. This news caused the company's share price to drop substantially, harming investors. The drop in the days following the disclosure of the indictments equaled approximately $153 million in market capitalization losses, or 39%, of the company's value. The lawsuit seeks to recover damages on behalf of all class members.
Client Announcements [Target]	On January 13, 2022, Aadi Bioscience, Inc. (the "Company") entered into that certain Negotiated Purchase Order Terms and Conditions for Clinical and Commercial Product (the "Agreement") with Fresenius Kabi, LLC pursuant to which Fresenius Kabi will manufacture Fyarro (sirolimus protein-bound particles for injectable suspension) (albumin-bound), for intravenous use ("Fyarro") for the Company, and the Company will purchase Fyarro as a finished drug product from Fresenius Kabi, on a purchase-order basis. In addition to manufacturing, Fresenius Kabi will perform specified labeling, packaging and serialization activities in respect of Fyarro for the benefit of the Company.
Seeking to Sell/Divest [Target]	AAON Inc. plans to close its Burlington facility in Ontario in July 2009. Norm Asbjornson, AAON's President and CEO said, "The manufacture of Canada's custom equipment will be relocated to the company's Tulsa, Okla., and Longview, Texas plants. We've been thinking whether we should keep it open and it's looking more dismal every day." AAON plans to put the Burlington, Ontario, facility up for sale.
Index Constituent Drops [Target]	AAC Holdings, Inc.(NYSE:AAC) dropped from S&P Global BMI Index
Index Constituent Adds [Target]	AAC Holdings, Inc.(NYSE:AAC) added to MSCI Acwi + Frontier Markets(acwi Fm) All Cap Index
Strategic Alliances [Target]	Mirati Therapeutics, Inc. and Aadi Bioscience, Inc. announced a clinical collaboration to evaluate the combination of adagrasib, a KRASG12C selective inhibitor, and nab-sirolimus, a small molecule mTOR inhibitor complexed with human albumin in KRAS G12C mutant non-small cell lung cancer (NSCLC) and other solid tumors. The primary objective of this multi-center, single-arm, open-label Phase 1/2 trial is to determine the optimal dose and recommended Phase 2 dose for the combination of adagrasib and nab-sirolimus in patients with KRASG12C - mutant solid tumors. In addition, the study will investigate the safety, tolerability and efficacy of adagrasib and nab-sirolimus in combination in patients both with and without prior exposure to a KRASG12C inhibitor. The trial will build on preclinical data showing enhanced anti-tumor efficacy with the combination of adagrasib and nab-sirolimus relative to either agent alone. Under the terms of the collaboration agreement, Mirati will be responsible for sponsoring and operating the Phase 1/2 study, and jointly with Aadi, will oversee and share the cost of the study.
Seeking Acquisitions/Investments [Target]	AAC Holdings, Inc. (NYSE:AAC) is looking for acquisition opportunities. AAC completed the closing of two financing facilities with Deerfield Management Company, L.P. ("Deerfield"), AAC's largest institutional shareholder of record and a leading healthcare specific investor with over $6 billion under management. The capital commitment consists of $25 million of subordinated convertible debt and up to $25 million of subordinated debt, together with an incremental facility of up to an additional $50 million of subordinated convertible debt (subject to certain conditions). AAC drew down $25 million of convertible debt at closing and will use the proceeds to fund its active acquisition strategy, its de novo pipeline and for other corporate purposes.

Investor Activism - Activist Communication [Target]	On February 4, 2022, Caligan Partners has bought a stake in ADMA Biologics Inc and may urge the biotech Company to consider a potential strategic review or explore a sale.
Executive Changes - CEO [Target]	Aadi Bioscience, Inc. announced the appointment of Neil Desai, Founder, President and CEO to Executive Chairman and Brendan Delaney, Chief Operating Officer, to CEO, effective as of January 1, 2023. Dr. Desai is the founder of Aadi Bioscience and has served as its President and CEO since 2011. He is the inventor of nab technology and the mTOR inhibitor nab-sirolimus (FYARRO) and helped guide the organization through the drug development process leading to approval of FYARRO, the first FDA approved therapy for advanced malignant perivascular epithelioid cell tumor (PEComa). The Company raised $155 million concurrent with a public listing in August 2021 and since that time, has grown to more than 80 employees including the recruitment of a top-notch leadership team. He has helped to guide the organization through the inaugural commercialization of FYARRO for advanced malignant PEComa, and the launch of the PRECISION 1 tumor-agnostic trial targeting TSC1 and TSC2 inactivating alterations. Most recently, following completion of a $72.5 million financing in September 2022, the Company launched a new clinical collaboration exploring the potential of the combination of nab-sirolimus with a KRAS inhibitor in NSCLC and other cancers. Prior to joining Aadi in September 2021, Mr. Delaney served as the Chief Commercial Officer (CCO) of Constellation Pharma which was acquired by MorphoSys for $1.7 billion. Prior to joining Constellation, Mr. Delaney was the CCO at Immunomedics, where he l
Executive Changes - CFO [Target]	AAC Holdings, Inc. announced that Kirk R. Manz, will resign as Chief Financial Officer effective December 31, 2017. Andrew McWilliams, who currently serves as the company's Chief Accounting Officer, will be appointed as Chief Financial Officer immediately following Manz's departure. After over a 15-year career with Ernst & Young LLP, Mr. McWilliams joined AAC in August 2014 to assist with the company's initial public offering and to build a public company accounting and finance department. During his tenure with Ernst & Young, Mr. McWilliams served multiple healthcare clients and also gained experience across a variety of corporate transactions, including debt financings and public equity offerings, as well as and mergers and acquisitions.
Executive/Board Changes - Other [Target]	On July 27, 2015, Jerrod N. Menz, the former President of AAC Holdings, Inc. and a former member of the company's board of directors, presented to the board a letter of resignation whereby he resigned from his position of President and as a member of the Board. Mr. Menz's resignation was conditioned upon, and effective as of, the release of an indictment by a grand jury in connection with an investigation by the California Department of Justice (the Indictment). The Indictment was unsealed on July 29, 2015, at which time Mr. Menz's resignation became effective. Mr. Menz remains an employee of the company.
Annual General Meeting [Target]	AAC Holdings, Inc., Annual General Meeting, May 19, 2015., at 14:00 Central Standard Time. Location: 150 Third Avenue South, Suite 2800. Agenda: To elect Jerry D. Bostelman, Lucius E. Burch, III, Michael T. Cartwright, Darrell S. Freeman, Sr., David C. Kloeppel, Jerrod N. Menz and Richard E. Ragsdale as directors to hold office until the 2016 annual meeting of stockholders and until their respective successors are elected and qualified; to approve the AAC Holdings, Inc. Employee Stock Purchase Plan; to ratify the appointment of BDO USA, LLP as independent registered public accounting firm for the fiscal year ending December 31, 2015; and to transact any other business that properly comes before the Annual Meeting or any adjournments or postponements thereof.

Table A.14: Example summaries for each news category