Journal of Financial Economics 149 (2023) 92–113

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

When can the market identify old news?

Anastassia Fedyk[a,*], James Hodson[b,c]

[a] University of California at Berkeley, Haas School of Business, 545 Student Services Building #1900, Berkeley, CA, 94720 USA
[b] Jozef Stefan International Postgraduate School, Artificial Intelligence Laboratory, Ljubljana, Slovenia
[c] AI for Good Foundation, Berkeley, CA, USA

ARTICLE INFO

Article history:
Received 3 December 2022
Revised 10 April 2023
Accepted 10 April 2023
Available online 6 May 2023

JEL classification:
G12
G14
G41
D83
C91

Keywords:
News
Recombination
Information processing
Correlation neglect
Inattention
Reversal

ABSTRACT

What drives the puzzle of market reactions to old news? Motivated by theories of correlation neglect, we conduct an experiment on finance professionals and show that even sophisticated investors have difficulty identifying old information that *recombines* content from multiple sources. We evaluate the market implications of this mechanism using a unique dataset of 17 million news articles from the Bloomberg terminal. Recombination of old information prompts larger price moves and subsequent reversals than direct reprints. This effect persists across news sentiment, ambiguity, and investor attention. Furthermore, while overall reactions to old information decline over time, differential reactions to recombinations increase.

1. Introduction

News, as the word implies, refers to information that is believed to be new. Yet not all financial "news" contains new information. The past two decades have seen a proliferation of information transmission mechanisms, increasing the volume of news and the variety of sources available to investors.[1] Given the voluminous news landscape, market participants' task of sifting through the news and identifying novel content is nontrivial and susceptible to limited attention. A long literature on limited attention shows that investors often *under*-react to initial information signals, leading to subsequent drift when they see the signals with a delay or receive information from other sources (Hirshleifer and Teoh (2003); Peng and Xiong (2006); Fedyk (2022)). However, there is also substantial empirical evidence of *over*-reaction in financial markets, especially in response to old news (Tetlock (2011); Gilbert et al. (2012)). We conjecture that this can also stem from (a specific type of) limited attention: "correlation neglect," whereby decision-makers fail to fully account for correlations across signals (DeMarzo et al. (2003); Ortoleva and Snowberg (2015)).[2] In this case, reactions to

* Corresponding author.
E-mail address: fedyk@berkeley.edu (A. Fedyk).

[1] For example, the Bloomberg terminal offers readers over a million news articles per day from thousands of different sources, RavenPack News Analytics draws from over 19,000 traditional and social media sources, the Factiva platform from Dow Jones includes information from over 30,000 sources, and EventRegistry.org, which aggregates and extracts events from online news, reflects over 150,000 distinct sources.

[2] Laboratory experiments confirm that individuals have difficulty accounting for correlations when those correlations are sufficiently complex (Eyster and Weizsacker (2010); Enke and Zimmermann (2019)).

https://doi.org/10.1016/j.jfineco.2023.04.008
0304-405X/

non-novel signals can overshoot, resulting in subsequent reversal.

We test theories of correlation neglect by showing that even sophisticated finance professionals have difficulty processing old news when it draws from multiple previous articles, and that this limitation has real asset pricing implications. First, we conduct an experiment directly on active finance professionals, eliciting their perceptions of news novelty. In this controlled setting, finance professionals perceive higher novelty in articles that recombine facts from several previous news, compared to direct reprints of single previous articles. Second, we test the asset pricing implications of our experimental results using a unique dataset of all financial news appearing on the Bloomberg terminal. The Bloomberg data are both (i) well-suited to measure repetition and recombination of information due to the inclusion of thousands of different sources; and (ii) ideal for testing the market relevance of the proposed cognitive limitation, given Bloomberg's user base of predominantly institutional investors. Consistent with investors' difficulty in processing recombination news, higher volumes of recombinations (as opposed to reprints) are associated with larger stock price moves that are more likely to reverse over the following week. These results help shed light on the empirical puzzle of market reactions to old information.[3]

Our experiment recruits 155 active finance professionals from a variety of relevant institutions, including broker dealers such as Goldman Sachs and Morgan Stanley, investment management firms such as State Street and PIMCO, and hedge funds such as Citadel and Point72. The participants are representative of the variety of roles and hierarchical positions in the financial services industry, ranging from analysts and traders to partners and managing directors. The experiment prompts participants to assess the novelty of a stream of headlines. The headlines are divided into two independent samples of twenty headlines each, covering two hypothetical firms, Argosy Logistics Inc. and Laker Pharmaceuticals LLC. Each participant sees one of these two sets of experimental headlines.

Both sets of experimental headlines are specifically designed to include ten novel headlines, five reprints, and five recombinations, with all three categories of news having the same length, and the two categories of old news (reprints and recombinations) having equal textual overlap with prior news. The only aspect on which recombinations differ from reprints is that recombinations draw information from multiple previously shown headlines, whereas reprints draw from a single previous source. As an illustration, participants in the Argosy arm of the experiment encounter the following four headlines at the start of the experiment:

1. "Argosy's misfit design business down, some tough questions to answer"
2. "Argosy Trucking Q3 results above expectations eps 1.2 vs 1.1"
3. "Argosy beats expectations: Q3 trucking results eps up 0.1 on 1.1"
4. "Argosy Q3 earnings beat expectations, but design business down"

Headline 3 is a direct reprint of headline 2, with 75% of its words already appearing previously. Headline 4 is a recombination: it also contains only old news, with 82.5% of its words appearing previously, but it combines equal parts from headlines 1 and 2. In addition to the headlines about Argosy, the participant also encounters twenty randomly interspersed filler headlines drawn from a sample of actual Reuters headlines. These serve to simulate the information overload that market participants face when reading real-world news.

The results of the experiment reveal that even experienced finance professionals are susceptible to recombination of old information. As a baseline, participants correctly perceive headlines containing new information to be more novel than headlines containing old news: for example, on a seven-point scale, novel headlines are marked with average novelty ratings of 4.52, compared to 2.82 for old headlines (both reprints and recombinations combined). However, within the set of old news, participants are significantly more likely to mistake recombinations for new news, compared to reprints: they rate recombinations with average novelty scores of 3.03, statistically significantly higher than 2.61 for reprints. These results are robust to varying the structural features of the experiment: the recombination effect is present on both the Argosy Logistics news sample and the Laker Pharmaceuticals news sample, and the results remain the same on the five-point novelty scale as on the seven-point scale. At the individual level, 68% of participants perceive recombinations to be, on average, more novel than their reprint counterparts. By contrast, only 19% of participants perceive reprints to be more novel than recombination news. When we repeat the experiment on 776 retail investors, we obtain qualitatively similar results, with one key difference: retail investors are more susceptible to all old news, including basic reprints.

We draw out and test the asset pricing implications of the documented susceptibility to information recombination. First, as a baseline, given that finance professionals generally perceive old news (especially reprints) as less novel than actual new news, we should observe smaller price reactions accompanying old news articles. Second, since recombinations are perceived as more novel than simple reprints, we expect to observe larger price moves associated with recombination news. Finally, to the extent that the larger price reactions to recombination news arise from a cognitive limitation on the investors' part (rather than inherent value in combining previously known facts), reactions to recombination news should be followed by larger subsequent reversals.

To test these predictions empirically, we exploit a novel dataset of 387 million news articles from the Bloomberg terminal, which is uniquely suited to study news novelty. The Bloomberg terminal aggregates information from a variety of news sources, encompassing a large volume of reprints and recombinations. Restricting attention to articles tagged with individual U.S. equity securities and

[3] See Huberman and Regev (2001), Tetlock (2011), Gilbert et al. (2012), and Drake et al. (2016).

marked as highly relevant by the editorial staff, our final sample covers over 17 million news articles published between January 2000 and December 2014.[4] Approximately 10% of the sample is comprised of news published directly by Bloomberg, with the bulk (roughly 60%) coming from key national and international news wires, and the remainder gathered from the web. This enables us to construct comprehensive measures of old news and trace out the etymology of any given piece of news.

We classify each news article in the Bloomberg news sample as either a novel piece of news, a reprint, or a recombination by assessing the percentage of unique words in that article that have already appeared in similar articles about the same firm during the preceding three days. In the main specification, we mark an article as novel if at least 40% of the unique words in its text have not appeared in the most textually similar five preceding articles about the same firm. All other news articles, whose content is at least 60% covered by preceding articles, are labeled as old news. Within old news, reprints are those articles whose text is spanned predominantly (at least 80% of the old text) by a *single* preceding article. Recombinations, on the other hand, contain content that is not novel, but also not spanned by the single closest neighbor. In untabulated analyses, we confirm that alternative methodologies such as the cosine similarity measure (Cohen et al. (2020); Fedyk (2022)) yield similar classification of news. For ease of interpretability, we use the more intuitive measure based on the percentage of spanned words throughout the paper.

We estimate the relationship between the daily firm-level returns and the proportion of the firm's news articles that are old and that are specifically recombinations. In the baseline, higher proportions of old news are associated with smaller absolute abnormal price changes: an additional 10% of the firm's news being old translates into an 11.5-basis-points-smaller absolute abnormal daily return on the same trading day. However, controlling for the overall level of old news, market reactions are larger when more of the old news is of the recombination type rather than reprints. Changing 10% of a firm's news articles from reprints to recombinations corresponds to a 17.6-basis-point-larger absolute abnormal return on the same trading day. Of course, outside of the experimental setting, recombination news is not randomly assigned, and it is possible that the choice to combine two or more facts into a single article can be triggered by especially important content. In this case, price responses to recombination news would constitute correct market reactions, rather than *over*reactions to old news. To differentiate the two stories, we examine the reversal of the initial reactions to recombination news.

Consistent with our conceptual predictions, the larger price responses following higher proportions of recombination news are more likely to reverse over the following week. Specifically, an additional 10% of the firm's news being old predicts an additional 13.9% of the *overall* daily return reversing over the following week; an additional 10% of old news being recombinations rather than reprints increases the reversal by another 16.9%. This suggests that, on average, the excess response to recombination news tends to fully reverse within a week.[5] In additional analyses, we confirm that our results are robust to alternative: (a) numbers of days from which we compare previous articles; (b) numbers of closest preceding articles included in the novelty comparisons; (c) thresholds for novelty, reprints, and recombinations; (d) measures of abnormal returns; and (d) continuous measures of old and recombination content instead of a discrete classification.

Our market evidence highlights the relevance of the cognitive limitations documented in our experiment, with difficulty in processing recombination news translating into sizable *over*reactions in returns. Interestingly, our results contrast with the more positive role that recombination of information plays in other domains. For example, Hirshleifer et al. (2018) find that patents that cite prior work from a wider range of technology classes predict persistent improvements in firm productivity and higher abnormal returns. Our results indicate that in the context of financial news, erroneous overreaction to recombination news due to correlation neglect outweighs potential positive effects such as recombination news providing novel insight by synthesizing individual pieces of news.

We extend our empirical analysis in several directions. First, to directly explore the role of limited attention, we slice our sample based on investor attention. We use the measure of investor attention on the Bloomberg terminal from Ben-Rephael et al. (2017), which is available at the security-date level. Consistent with limited attention, the market reacts less to old news when investor attention is high. But this is driven primarily by better screening out of simple reprints, and the recombination effect is present even during high investor attention. Second, inspired by the greater susceptibility of retail investors to stale news in the experiment, we examine institutional versus retail order imbalance in response to old and recombination news. We provide evidence of relatively greater retail trading (than institutional trading) in response to old news in general, but less so for recombination news. Third, we explore how news sentiment and ambiguity (hard quantitative versus soft subjective information) affect our results. We find that our empirical results are very similar for both positive and negative news and robust to controlling for the amount of hard quantitative information.

Finally, we exploit the time series of our large dataset to investigate how the documented effects vary over time.

[4] The sample ends in 2014 due to data availability from the data provider. Given this general restriction from the data provider, our sample of 387 million articles (of which 17 million are labelled as highly relevant to individual U.S. equity securities) between 2000 and 2014 is the most comprehensive dataset of news from the Bloomberg terminal that has been available for research to date.

[5] The average absolute daily return on securities in the sample over the sample period of 2000–2014 is 1.15%. Therefore, a 10% increase in recombination news translates into a predictable directional reversal of $16.9\% \times 1.15\% = 19$ basis points during the following week. Recalling that the initial impact of a 10% increase in recombination news on the daily abnormal return is 18 basis points, this means that the additional returns induced by recombination news tend to fully reverse within the following week.

We perform the tests separately for each full year in our sample, from 2001 until 2014. The time trends of the coefficients indicate that abnormal daily returns have become progressively smaller for firms with high levels of (overall) old news. However, the *differential* response to recombination news increased over time, with the coefficient on an additional 10% of the firm's news being recombinations (rather than reprints) increasing from a positive 5-basis-point effect in 2001 to a positive 24-basis-point effect in 2014. These results point to increased investor sophistication in screening out simple reprints but continuing susceptibility to articles that recombine previously available information.

This paper contributes to the literature on cognitive biases in financial markets by proposing a channel for the market failure to identify old news and providing evidence for this mechanism both experimentally and in empirical asset prices.[6] Our paper relates to the growing number of studies examining the effects of limited investor attention in processing financial information on asset prices.[7] In particular, motivated by the theoretical work on correlation neglect (DeMarzo et al. (2003); Ortoleva and Snowberg (2015)), we conjecture a specific type of limited attention, whereby market participants do not fully discount recombination of old information from multiple sources as old news. This builds on the evidence from experimental economics that human subjects tend to ignore correlations when analyzing a stream of reports representing different combinations of the same underlying signals (Enke and Zimmermann (2019)) and recent evidence of overweighting of repeated information in comparables pricing (Murfin and Pratt (2019)). Our study is unique in that we offer direct experimental evidence of correlation neglect in financial news and show that even experienced investors have difficulty processing the recombination of old news. In addition to documenting this cognitive limitation experimentally, we showcase its relevance by empirically confirming its implications for asset prices and demonstrating its persistent relevance over time.

The "recombination effect" that we document in this paper helps explain the puzzle of overreactions to old news, which has been highlighted by several previous studies. In early evidence by Huberman and Regev (2001), a front page article in the *New York Times* in May 1998, which largely repeated information from five months prior, prompted a 330% price increase for the covered firm. Tetlock (2011), in a systematic investigation of old news, observes that absolute abnormal returns are generally lower when more of the news about a firm is old but still finds evidence of overreaction to old news. Evidence from Gilbert et al. (2012) hints at the mechanism we develop in our paper: investors overreact to recombination of previously-released inputs into summary statistics in the form of the Leading Economic Indicator. Our paper sheds light on the type of inattention that can generate the observed overreactions to old news, namely the failure to identify recombination of previously-available information.

The remainder of the paper proceeds as follows. Section 2 presents the experimental investigation of finance professionals' perceptions of news novelty, showing that market participants are more susceptible to recombination of old information than to direct reprints. Section 3 describes the data and methodology for assessing the textual novelty on the large sample of Bloomberg news. Section 4 presents our main empirical results evaluating the effects of the proposed mechanism on asset price dynamics. Section 5 presents robustness and additional analyses, and Section 6 concludes.

2. Experimental evidence

In this section, we provide controlled experimental evidence directly from the setting of interest: finance professionals consuming financial news. We find that investors are less likely to identify old information in news articles that draw from multiple sources than in direct reprints of single previous articles. After documenting this limitation experimentally, we draw the empirical implications for trading volumes and asset price dynamics.

2.1. Design

We recruited 155 active finance professionals through the Harvard Business School alumni network to participate in an online experiment conducted in late 2018 and early 2019. The participants span the full landscape of the financial services industry (their firm affiliations are presented in Internet Appendix Table A.1). The vast majority of the participants come from large banks and broker dealers such as Goldman Sachs and Morgan Stanley, investment management firms such as Fidelity and State Street, private equity firms such as Bain Capital and Lindsay Goldberg, hedge funds such as Two Sigma and Point 72, and investment banks such as Barclays and Macquairie. The remaining 22% of the sample consists of individuals from financial news organizations (e.g., Financial Times), insurance companies (e.g., Liberty Mutual), government agencies (e.g., Federal Reserve Board), consulting firms (e.g., Deloitte), private investors, and finance-oriented employees of universities or tech companies such as Facebook. The sample includes key decision makers such as partners and managing directors, as well as active younger employees such as portfolio managers and traders.

Each participant in the experiment faced a stream of forty news headlines. This includes twenty headlines about one of two fictitious companies, Argosy Logistics Inc. and Laker Pharmaceuticals LLC.[8] We exploit two samples of

[6] For studies on the effects of financial news on security prices see, among others: Barber and Loeffler (1993), Busse and Green (2002), Tetlock (2007), Fang and Peress (2009), Engelberg and Parsons (2011), Carvalho et al. (2011), and Dougal et al. (2012).

[7] See, for example, Klibanoff et al. (1998), Engelberg (2008), DellaVigna and Pollet (2009), Hirshleifer et al. (2009), Da et al. (2011), Drake et al. (2016), Boulland et al. (2017), Ben-Rephael et al. (2017), Kaniel and Parham (2017), Hirshleifer and Sheng, and Fedyk (2022).

[8] The experimental design uses headlines rather than full news text, to keep the survey length manageable for the finance professionals. In untabulated analyses, we confirm that text-based measures of novelty are very similar based on the headlines and based on the full news text. To the extent that headlines are easier to process for human readers, our

experimental headlines about two different hypothetical firms in order to ensure that our findings are qualitatively similar across separate samples of news headlines and not driven by any one specific set of headlines. Each participant was independently randomly assigned to one of the two headline samples; out of the 155 participants, 76 were assigned to the Argosy news sample and 79 to the Laker news sample. The remaining twenty headlines presented to participants were filler headlines about other firms, aimed to simulate the type of information overload that market participants are likely to face while consuming real-world news. The filler headlines were drawn from a sample of thirty actual news headlines published by Reuters during August 2017. Each participant saw a subset of twenty randomly-selected filler headlines, randomly interspersed with the twenty experimental headlines about either Argosy or Laker. All headlines used in the experiment, including the two experimental sets and the filler headlines, are displayed in Table A.2 in the Internet Appendix.

The individual headlines about Argosy and Laker represent different informational content. Some are novel news, for which only about 20% of the words have appeared in previous headlines about the same firm. The rest are old news, for which approximately 80% of the words are covered by preceding headlines about the same firm. The old news headlines are further split into two categories: reprints, for which the bulk of the previously-seen content comes from a *single* preceding headline about the same firm, and recombinations, which draw from at least two prior headlines.

Importantly, the experimental setting allows us to test the proposed mechanism in a fully controlled environment, where we designed the wording and ordering of the experimental headlines in such a way as to ensure that the headlines do not differ along any dimensions other than their information structures. Both sets of experimental headlines (about Argosy and Laker) are specifically designed to be: (i) equal in average length across the three groups (novel news, reprints, and recombinations); (ii) equal in average order across reprints and recombination; and (iii) indistinguishable in average level of old content across reprints and recombinations. These balance checks are presented in Table 1. Overall, across the two samples (Argosy and Laker), both reprints and recombinations have exactly 81.1% of their words spanned by preceding headlines about the same firm.[9]

For concreteness, the following are examples of novel headlines about Laker, a fictitious pharmaceutical company:

1. “More scandal at Laker as CFO Russell fends off allegations of misconduct”
2. “Laker shares plummet, as CEO George ousts Russell on hardline”
3. “Laker’s AdventiMed releases major landmark in DP2 cures (PharmaToday)”

Table 1
Balance checks on the experimental headlines along three dimensions: **Panel 1** looks at the length (the average number of words), **Panel 2** looks at the order (i.e., the location of each headline in the sequence of 20 headlines), and **Panel 3** looks at the amount of old content (the percentage of words that have already appeared in sequentially-earlier headlines). These characteristics are computed separately for novel headlines, reprints, and recombinations across three samples: headlines about Argosy Logistics Inc., headlines about Laker Pharmaceuticals LLC, and the combined set of headlines.

Panel 1: Length (# words)			
	Novel	Reprint	Recombination
Argosy sample	7.8	7.8	7.8
Laker sample	7.6	7.4	7.4
Overall	7.7	7.6	7.6
Panel 2: Location (#1-#20)			
	Novel	Reprint	Recombination
Argosy sample	9.1	12.6	11.2
Laker sample	8.8	11.6	12.8
Overall	9.0	12.1	12.0
Panel 3: Old content (% words seen)			
	Novel	Reprint	Recombination
Argosy sample	22.2%	82.7%	81.1%
Laker sample	20.8%	79.4%	81.1%
Overall	21.5%	81.1%	81.1%

Below is an example of a piece of news that reprints the information revealed by headline 2 above:

- “Laker CEO George takes hardline to oust Russell, source says”

67% of the (stemmed) words in this headline already appear in preceding headlines about Laker, all of them from the novel headline 2.

By contrast, the following headline is a recombination:

- “Laker CFO leaves on scandal during DP2 release”

This headline mentions both the release of the new cure and the ousting of a chief executive. 62.5% of its words appear in preceding headlines, but the content is a *combination* of novel headlines 2 and 3.

The headlines were presented to participants sequentially, in the manner depicted in Fig. 1. Each headline was displayed in the middle of the screen, along with a tag indicating whether the headline pertained to a specific firm, such as Toyota, or topic, such as World News. In the instructions (provided verbatim in Internet Appendix B), the participants were told to evaluate each headline’s novelty against previous headlines that he or she has seen about the same firm during the experiment, ignoring any news he or she may have consumed in real life from other sources. The lack of contamination from any news consumed outside of the experimental setting was also ensured by the experimental design centering on news about fictitious firms (Argosy and Laker), which the participants could not have encountered outside of the experiment, by construction. Participants’ perceptions of novelty were

streamlined experimental setting may actually understate the full extent of cognitive limitations in processing recombination news.

[9] Prior to performing these comparisons, all words are stemmed so that, for example, “earned” and “earnings” are treated as the same term.

Fig. 1. Screenshot of the experimental survey, eliciting finance professionals' perception of the novelty of a given news headline.

elicited through the novelty scale below the headline. We used two setups for the scale: a seven-point scale, ranging from "Nothing New" to "Completely New" (as displayed in Fig. 1) and an analogous five-point scale.[10]

To simulate the type of cognitive overload that finance professionals are likely to face in real-world news environments, such as scrolling news on the Bloomberg terminal, we placed a time constraint on each question. Each participant was allotted exactly 10 seconds to mark the novelty of each headline. There was a timer above the headline, which showed the amount of time left. Once the 10 seconds elapsed, the experiment progressed to the next question. If the participant had not marked anything within that time, the participant received a pop-up notification that he or she missed that question; missing more than three questions resulted in being disqualified.

The participants were incentivized to detect novelty as well as they could: the five participants whose answers most correctly matched actual novelty of the articles received $50 bonuses. The entirety of the survey took seven minutes to complete, and all participants were offered a $10 gift card as a token of appreciation for their participation.

2.2. Results

The experiment achieved a relatively high completion rate. Out of the total of 155 finance professionals who participated, 18 individuals signed up but did not answer any questions before getting timed out of the survey. Another 25 were disqualified or volutantarily stopped the survey partway, leading to partial responses. The remaining 112 participants (72%) completed the entirety of the survey. Table 2 presents the results from the responses of all individuals who completed at least some questions (a total of 137 individuals). Table A.3 in the Internet Appendix shows that the results are very similar when excluding attrited participants and limiting to the 112 participants who completed the entire survey.

[10] The two scales were implemented sequentially. The 46 participants in the first run of the survey faced a seven-point scale; the remaining 109 participants faced a five-point scale.

Table 2
Experimental results. **Panel 1** presents the mean participant responses for novel news, reprints, and recombinations. **Panel 2** shows the differences in responses, with accompanying standard errors, across two comparisons: novel news versus old news (pooling both reprints and recombinations) and between the two sets of old news (recombinations vs. reprints). The results are computed across the two survey variants: using a seven-point scale and a five-point scale. The results include all participants who completed at least some of the survey questions, including attritors (a total of 137 participants).

Panel 1: Mean responses			
	Novel	Reprint	Recombination
7-point scale	4.52	2.61	3.03
5-point scale	3.84	2.40	2.63

Panel 2: Comparisons			
		Novel vs. Old	Recombination vs. Reprint
7-point scale	Diff	1.70***	0.42**
	(SE)	*(0.15)*	*(0.20)*
5-point scale	Diff	1.33***	0.23**
	(SE)	*(0.06)*	*(0.10)*

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The results in Table 2 reveal that, in general, participants correctly identify novel articles as containing more new information than either recombinations or reprints, but participants are more susceptible to recombination of information than to direct reprinting. As can be seen in Panel 1 of Table 2, on average, new articles receive novelty scores of 4.52 on the seven-point scale and 3.84 on the five-point scale. Both recombinations and reprints are identified as containing substantially less new information than their novel counterparts, but with recombinations rated consistently higher than reprints. On average, reprints receive novelty scores of 2.61 out of 7 (2.40 out of 5), while recombinations are rated as 3.03 out of 7 (2.63 out of 5).

The differences in novelty rankings across these three groups of news are significant and consistent across different survey designs. As displayed in Panel 2 of Table 2, novel articles are ranked, on average, 1.70 (1.33) points out of 7 (5) higher than their non-novel counterparts. This difference is highly statistically significant. More interestingly,

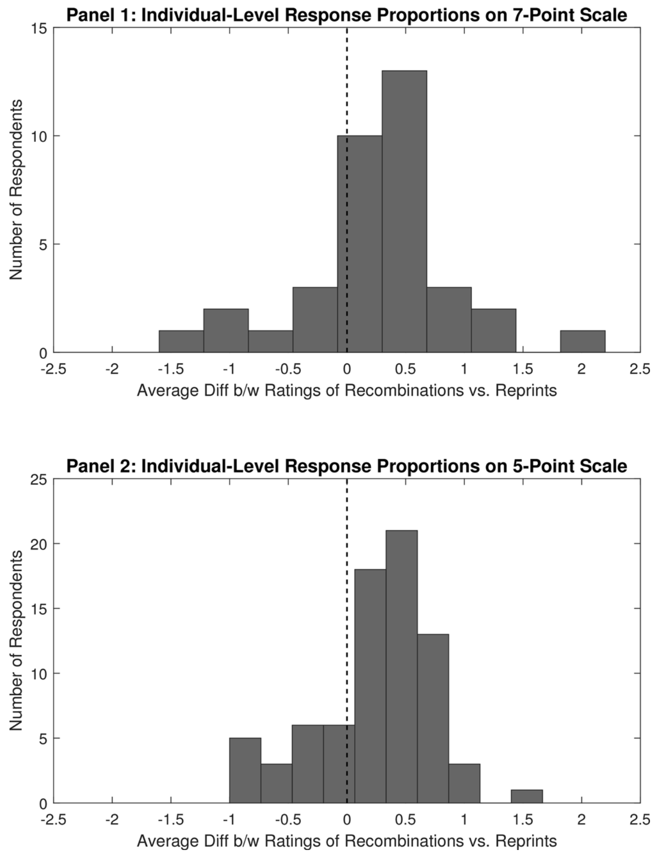

Fig. 2. Individual-level experimental responses: for each participant, we compute the difference in average novelty scores assigned to recombinations versus reprints. **Panel 1** considers the responses from the finance professionals asked to evaluate news novelty on a seven-point scale, while **Panel 2** tabulates the responses on a five-point scale.

the difference in average rankings between recombinations and reprints is economically meaningful at 0.42 (0.23) points out of the seven-point (five-point) novelty scale and statistically significant at the 5% level, with a t-statistic of 2.05 in the seven-point survey and a t-statistic of 2.42 in the five-point survey. This difference is robust to the exclusion of attritors and to the two subsamples of news headlines—about Argosy and about Laker. Table A.4 in the Internet Appendix shows that the effect is stronger for the Argosy headlines in the survey design with a seven-point scale and for the Laker headlines in the five-point survey, indicating no consistent difference in either direction.

The individual-level results, presented in Fig. 2, confirm the findings from the pooled analysis. We compute the difference in average responses regarding recombination versus reprints separately for each finance professional in the sample. Panel 1 of Fig. 2 displays the distribution of these individual-level differences for the survey design with a seven-point novelty scale, while Panel 2 presents the individual results from the five-point scale. The median difference is 0.40 (0.20) points out of the seven-point (five-point) scale. The difference is positive for 68% of the finance professionals who participated in the survey, meaning that 68% of participants rated recombinations as, on average, more novel than repritns. By contrast, only 19% of participants rated reprints higher, and 13% of participants perceived the two sets of news equally.

Finally, for additional context, in Internal Appendix Table A.5 we repeat the experiment with a five-point scale on a sample of 776 retail investors recruited through the online survey platform Cint; 527 respondents completed the entire survey. In general, retail investors are much more susceptible to stale news than institutional investors, which is consistent with their lower level of sophistication and with the evidence in Tetlock (2011). Retail investors mark novel news with average novelty scores of 3.51 out of 5, slightly lower than the 3.84 marked by institutional investors. However, they mistakenly mark both reprints and recombinations as quite novel, at 3.30 and 3.36 out of 5, respectively (compared to only 2.40 and 2.63, respectively, for institutional investors). These results highlight two patterns: (i) retail investors are more susceptible to stale news than institutional investors, mistaking even simple reprints as new, but (ii) the difference in retail investors' susceptibility to recombination versus reprint news is still statistically significant (with a t-statistic of 1.77 among all respondents and 1.90 among the 527 respondents who completed the entirety of the survey).

Overall, finance professionals are relatively better at screening out simple reprints than retail investors, but even sophisticated finance professionals are susceptible to recombination news. The experimental results show this to be the case even in a setting where these two sets of headlines are specifically designed to be equally "old news." This provides new evidence for theories of correlation neglect in an important context that includes sophisticated investors. Prior evidence on correlation neglect in settings such as comparables pricing (Murfin and Pratt (2019)) shows that individuals can overweight repeated signals even when the repetition is relatively straightforward. Our results highlight that cognitive limitations in identifying and discarding repeated information are much more persistent, impacting even sophisticated players, when the information structure is more complex.

2.3. Asset pricing implications

Conceptually, market participants' susceptibility to old news in the form of recombination of previously-available information should lead to larger market reactions to recombination news that subsequently reverse. We formally spell out the empirical predictions for asset prices and trading volumes, based on our experimental evidence.

First, as a baseline, the experimental results indicate that finance professionals do perceive new information to be, on average, more novel than old news. This means that old news articles should garner smaller market responses upon their publication.

Prediction 1. Compared to novel news, old news is associated with lower trading volumes and absolute price changes immediately following publication.

This prediction receives empirical support from existing studies including Tetlock (2011). We confirm that it holds likewise in our large dataset of Bloomberg news.

Second, central to the recombination mechanism, market participants perceive recombination headlines to be more novel than reprint headlines with the same length, relative location, and actual amount of old content. Outside of the lab, we expect finance professionals' susceptibility to this cognitive limitation to have the following market implications:

Prediction 2. Among old news, recombination articles are associated with larger trading volumes and absolute price changes immediately following publication than reprint articles.

Third, if market participants mistakenly overreact to old news, then we expect initial overreactions to old news to correct with time. In particular, the stronger initial reactions to recombination news, reflecting a cognitive limitation in the processing of more complex information structures, should be followed by larger subsequent price reversals than the milder responses to reprint news. We summarize this in the following empirical prediction:

Prediction 3. Initial reactions to old news are subject to subsequent reversal. In particular, during the days or weeks following news publication:

(a) The initial price moves after old news see more reversal than the initial price moves after novel news.
(b) The initial price moves after recombination articles see more reversal than the initial price moves after reprint articles.

We test these predictions in Section 4, using the large Bloomberg news dataset and textual analysis methodology summarized in Section 3 below.

3. Bloomberg news data and textual analysis

To test the asset pricing implications of finance professionals' limitations in processing complex information, we use news data from the Bloomberg terminal. Bloomberg boasts one of the most comprehensive news databases in the world and offers a view of the full landscape of financial media. Descriptive statistics of our sample and the various screens imposed on the data are tabulated in Table 3.

Most news databases, including the widely used Dow Jones Newswire, contain articles from a single publisher, which limits the potential to determine the extent of individual articles' novel content. Factiva, Dow Jones' database of news from more than 30,000 global sources, offers strong research and analysis tools, but is both smaller and less tailored to the real-time needs of financial professionals than news from the Bloomberg terminal. The Bloomberg news data, aggregated from a variety of sources, are reflected in the Bloomberg database almost instantaneously (typically within 100 milliseconds) upon original publication. The sources of news fall into three categories: news written and published by Bloomberg directly (roughly 10% of the sample); key national and international news wires from partner news organizations (60% of the sample); and content from web sources, including regional and local news, blogs, and social media (remaining 30%). The overall number of articles passing through the Bloomberg terminal reached 1 million articles per day in recent years, several times larger than other similar services. We impose several conditions on the news used in the analysis, in order to benefit from the breadth of coverage while limiting noise. First, financial news articles passing through the Bloomberg terminal are explicitly tagged with security codes, either manually or through a rules-based algorithm. We restrict our sample to news articles tagged with security codes corresponding to equities traded in the U.S. and exclude stocks with prices below $5 per share to minimize microstructure effects. This leaves approximately 29,500 news articles per day. Second, we limit our attention to articles whose tagged securities are especially relevant based on Bloomberg's relevance tags. The majority of articles are tagged with more than one security code, and prior work has used indirect proxies for relevance, such as limiting the sample to articles tagged with one or two securities. Bloomberg's explicit relevance markers offer a more direct way to screen relevant articles. For each article-security link, the Bloomberg database includes a relevance score (assigned either manually or through a rules-based algorithm). Articles with relevance scores around 90% are highly targeted to the tagged security, talking about that security's earnings, products, or strategy. Articles with relevance scores around 70% are less immediately tied to that security, but still relevant; for ex-

Table 3
Summary statistics of Bloomberg news data, consisting of the volume of articles and the density of relevant tags per article. Summary statistics for the full sample include the mean, median, and interquartile range of each variable; mean values are also provided separately for each year in the sample.

	Daily articles	Daily articles with security codes	Daily articles w/ sec. codes and relevance $\geq$ 70	Tags per article
Full Sample				
Mean	70,723	29,570	4,026	1.30
Median	60,783	25,212	2,675	1.26
25th Percentile	31,668	12,353	1,311	1.22
75th Percentile	94,010	41,785	6,032	1.37
By year (mean)				
2000	20,940	8,727	940	1.60
2001	23,844	9,692	1,062	1.40
2002	26,844	10,621	1,159	1.39
2003	30,190	12,458	1,309	1.32
2004	33,417	14,432	1,577	1.27
2005	38,251	16,372	1,582	1.26
2006	47,784	20,755	2,194	1.21
2007	60,024	24,775	2,596	1.22
2008	66,819	27,258	2,673	1.20
2009	75,303	28,976	2,777	1.18
2010	82,227	36,901	3,774	1.22
2011	93,058	42,208	5,966	1.25
2012	102,778	45,471	6,096	1.25
2013	170,409	69,775	12,582	1.35
2014	188,964	75,128	14,106	1.37

ample, they may talk about the firm's main competitor. Articles with relevance scores around 50% are only tangentially relevant for the security in question.[11]

Our analysis sample comprises all news articles that are assigned at least 70% relevance for at least one U.S.-traded equity security. For those articles deemed at least 70% relevant for more than one security, we include all security tags with relevance of 70% or above. This limits the sample to approximately 4000 news articles per day, with each article linked to an average of 1.3 securities.

3.1. Old news

We proxy for old information in news using the extent to which the textual content of each individual article is spanned by the text of preceding articles about the same security.

For each article s in the sample, we first extract the unique words (unigrams) in the text of the article. We exclude stop words (common words such as "a," "the," "in," "when," etc.) and stem all words into unique *terms* using the standard stemming algorithm by Porter (1980) (so that words such as "earned" and "earnings" are represented by the same term, "earn-"). We use the norm $||\cdot||$ to denote the number of unique terms in a set of articles. For example, $||s_1 \cap s_2||$ is the number of unique terms appearing in both s_1 and s_2.

We measure the old content of each news article by the extent to which it is spanned by previous articles. For each article s tagged with firm i, we look at all articles s' that are also tagged with i and published up to three days (72 hours) before the publication of s. We identify the five preceding articles $(s'_1, \ldots, s'_5)$ that individually span the largest fraction of terms in s and define the extent to which s contains old information as:

$$Old(s) = \frac{||s \cap (\cup_{i=1}^{5} s'_i(s))||}{||s||}, \tag{1}$$

Our measure of old content is similar to the methodology introduced by Tetlock (2011), with one key innovation. Instead of defining old content as the average intersection of an incoming article s with the closest previous articles about the same firm, we consider the overall percentage of s that is spanned by the previous articles. This is meant to differentiate between the following cases: (i) articles s, s', and s'' all cover some background information about the firm in their first paragraphs (taking up 50% of their text), but otherwise talk about completely different facts; (ii) articles s and s' cover exactly the same information (intersection of 100%), while s'' does not intersect with s at all. Using the average intersection metric, s would be considered equally old (50%) in these two cases, despite the fact that it contains new information in case (i) and nothing new whatsoever in case (ii). The metric based on the percentage of terms spanned *overall* circumvents this issue. In untabulated analyses, we also confirm that our empirical results are robust to alternative methods of capturing old news. For example, the measure is virtually unchanged when using bigrams (pairs of words) instead of unigrams and robust to an alternative construction based on the cosine similarity measure (Cohen et al. (2020); Fedyk (2022)), which compares the full vectors of words across articles.

The distribution of the old content measure is presented in Panel 1 of Fig. 3. A large percentage of articles, nearly 40%, are almost entirely (90% or more) spanned by

[11] For context, here are some sample headlines for AAPL, with their relevance scores: "Apple announces event on 3/17 to unveil new iPad" (95% relevance); "Android Grows U.S. Smartphone Market Shr to 50.1%" (70% relevance); "JCPenney lowered to BB at Standard & Poor's on new strategy" (50% relevance).

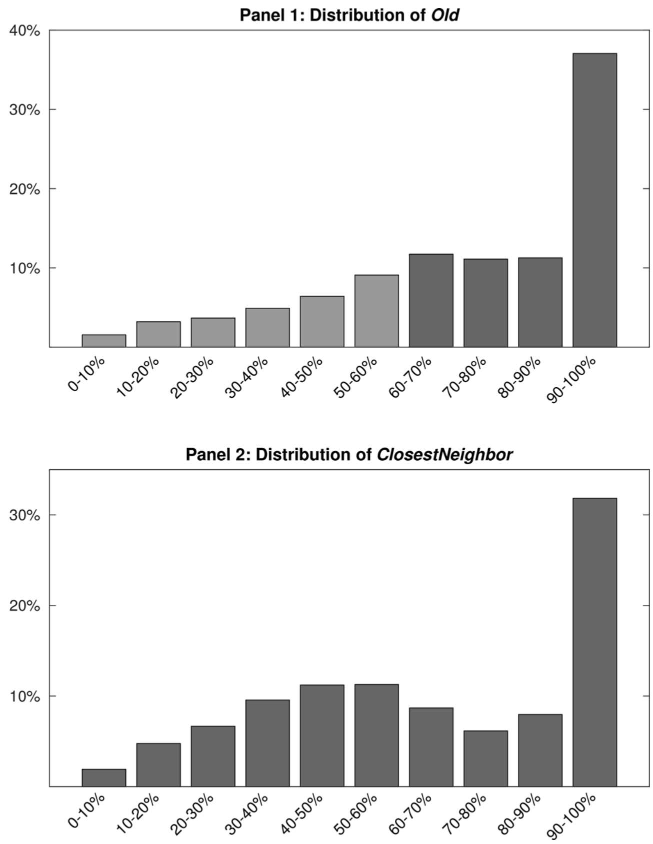

Fig. 3. Distributions of article-level textual novelty measures. **Panel 1** presents a histogram of the *Old* measure, while **Panel 2** displays the distribution of the *ClosestNeighbor* measure.

preceding news about the same firm. Another 30% of the sample is evenly distributed across 60–90% prior coverage, and the remainder has at least 40% novel content. Very few articles have less than 10% old content, as even unrelated news articles about the same security are likely to share common words.

We define old news (denoted by the binary indicator $OldNews(s)$) as having at least 60% of the content spanned by the closest previous articles about the same firm. Robustness checks in Section 5.5 consider alternative specifications: (i) looking at similar articles going back five days or ten days instead of three days; (ii) looking for the closest ten articles instead of five; and (iii) varying the threshold for old news. We also show that our results are robust to using the continuous measure of old content instead of the binary indicator variable.

3.2. Reprints and recombinations

We use textual analysis to distinguish two types of old news in the Bloomberg news data: reprints of single previous articles and recombinations that combine information from two or more previous sources. Fig. 4 illustrates our approach. Each panel of the figure displays a news article s (in dark grey), and the content of s that was already present in each of the five nearest articles $s'_1(s), \ldots, s'_5(s)$ (marked in solid or hatched gray). The bottom row displays $\cup_{i=1}^{5} s'_i(s)$ and likewise marks the content intersecting with s in hatched gray. This last row captures the measure of old information given by (1): the percentage of s's content that was already seen in at least one of the five most similar preceding articles about the same firm. In both panels, the measure $Old(s)$ is at 90%. However, the two cases are very different. The top panel of Fig. 4 illustrates a reprint: s is almost an exact copy of $s'_1(s)$, marked in solid shading for emphasis. By contrast, in the bottom panel, there is no single previous article that captures more than half of the content of s; instead, s is a recombination of $s'_1(s)$ and $s'_2(s)$ (whose intersections with s are also highlighted in solid shading).

We differentiate reprints from recombinations by looking at the extent to which the content of each article s is spanned by its *single* closest previous neighbor, $s'_1(s)$:

$$ClosestNeighbor(s) = \frac{\max_{s'} ||s \cap s'||}{||s||} = \frac{||s \cap s'_1(s)||}{||s||} \qquad (2)$$

The distribution of the *ClosestNeighbor* measure is displayed in Panel 2 of Fig. 3. There is a large number of articles almost all of whose words have appeared in the single closest neighbor (capturing exact reprints), and a large share of articles approximately half of whose words appeared in the single closest neighbor (potential recombinations).

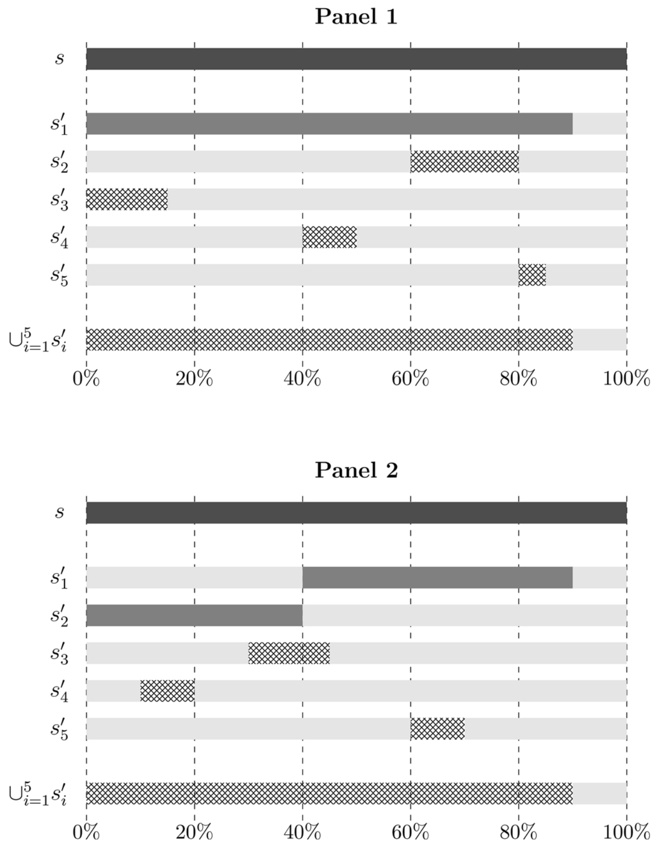

Fig. 4. Two news articles, both 90% old but displaying different types of old news. **Panel 1** features an example of a reprint: article s, 90% of whose content is spanned by the single closest neighbor, $s'_1(s)$. **Panel 2** provides an example of a recombination: article s that is also 90% old but that recombines information from two previous articles: $s'_1(s)$ and $s'_2(s)$.

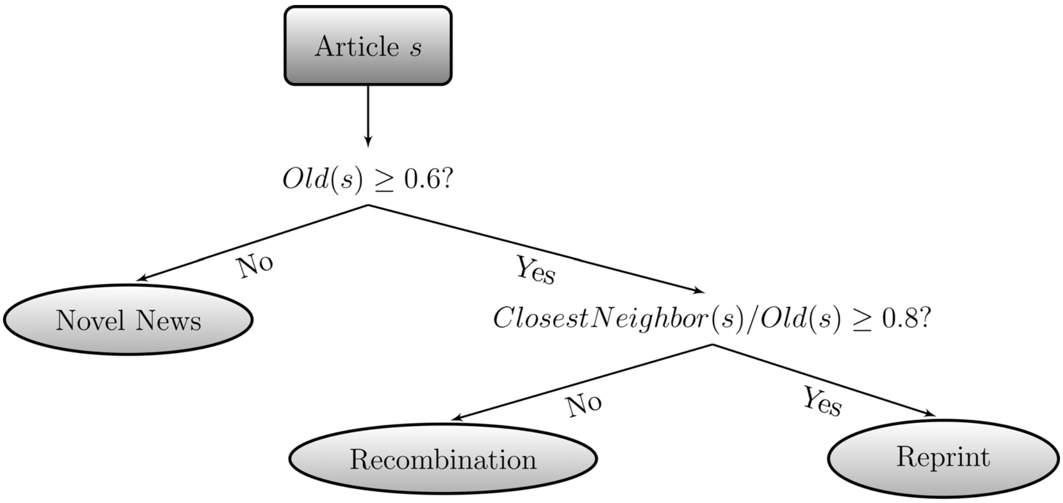

Fig. 5. Classification of articles as novel news, reprints, or recombinations.

We classify each article in two steps, first looking at its overall level of old information and then considering the extent to which this old information is spanned by the single closest neighbor. The classification process is illustrated in Fig. 5. First, an article is classified as old news if and only if at least 60% of its text has already been seen in the five closest previous articles about the same firm. Second, for each article in the old news category, we consider the share of that article's old content that is spanned by the single closest neighbor. If this share is higher than 80%, then we classify the article as a reprint (denoted by the indicator variable *Reprint*(s)); if this share is smaller than

Table 4
Summary statistics of extracted news text, including article-level and firm-level numbers of words per article and the computed measures of old and recombination content. Summary statistics for the full sample include the mean, median, and interquartile range of each variable; mean values are also provided separately for each year in the sample.

	Article-Level			Firm-Level		
	# Terms	Old	ClosestNeighbor	# Terms	% Old News	% Recombinations
Full Sample						
Mean	146	0.72	0.52	151	72%	21%
Median	138	0.77	0.56	146	69%	20%
25% Percentile	67	0.65	0.43	110	56%	14%
75% Percentile	201	0.94	0.92	173	82%	25%
By year (mean)						
2000	200	0.68	0.43	200	70%	23%
2001	200	0.68	0.43	199	71%	24%
2002	202	0.68	0.44	200	70%	21%
2003	194	0.68	0.43	188	69%	22%
2004	196	0.70	0.48	190	72%	23%
2005	202	0.69	0.46	196	72%	19%
2006	157	0.71	0.50	158	74%	20%
2007	161	0.72	0.52	160	75%	18%
2008	155	0.68	0.46	151	69%	18%
2009	133	0.67	0.47	129	68%	15%
2010	130	0.67	0.45	135	68%	16%
2011	120	0.72	0.54	130	71%	18%
2012	129	0.70	0.51	129	70%	17%
2013	138	0.79	0.54	144	75%	19%
2014	138	0.80	0.55	142	72%	20%

80%, then the article is considered to be a recombination (denoted by the indicator variable $Recombination(s)$).

Article-level summary statistics are presented in the first three columns of Table 4. We tabulate, for the full sample and each year in the sample, the average number of unique terms in each article, the percentage of articles that are classified as old news, and the percentage of articles that are classified as recombinations. Over the fourteen-year sample period, continuous article-level measures of both $Old(s)$ and $ClosestNeighbor(s)$ have steadily risen along with the overall increase in news volume, complicating the market's task of identifying novel content and increasing susceptibility to cognitive pitfalls explored in Section 2.

3.3. Firm-level measures

We aggregate individual article-level classifications to the firm-day level. For each firm i on date t, $PrcOld_{i,t}$ is the percentage of news articles tagged with i on date t that are old news, and $PrcRecombination_{i,t}$ is the percentage of articles that are recombinations.[12] The last two columns of Table 4 display the average values of these two firm-level measures over the full sample and separately for each year in the sample.

We screen out the effect of firm-specific news flow on firm-level measures of old information and recombinations. For each of the two measures ($PrcOld_{i,t}$ and $PrcRecombination_{i,t}$), we take the residuals from daily cross-sectional regressions of the measure on the log of the number of articles for firm i on date t, the log of the average number of unique terms per article, and the square of the log average number of unique terms per article. This results in $AbnPrcOld_{i,t}$ and $AbnPrcRecombination_{i,t}$, which capture abnormal proportions of old content and recombination content, respectively.

4. Empirical results

In this section, we test the asset pricing implications of the mechanism documented in Section 2. Consistent with our predictions, recombination news articles are associated with larger returns and trading volumes than reprint news. Importantly, these reactions to recombination of old information tend to reverse over the following week.

4.1. Market reactions to old news

We begin our empirical analysis by examining how old and recombination news content relates to measures of market activity—abnormal trading volumes and returns. In order to enhance comparability to the prior literature, we follow the empirical approach set out in Tetlock (2011) and estimate Fama and MacBeth (1973) regressions of daily abnormal returns and volumes against daily measures of old news:

$$|AbnRet|_{i,t} = a + b_1 AbnPrcOld_{i,t} + b_2 AbnPrcRecombination_{i,t} + gX_{i,t} + e_{i,t} \quad (3)$$

$$AbnVol_{i,t} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \quad (4)$$

where $|AbnRet|_{i,t}$ denotes the absolute value of the abnormal return for firm i on date t, calculated as the

[12] All empirical tests in the paper are restricted to trading days and performed close-to-close, so that any articles published on holidays are included in the measure for the next trading day.

Table 5
Results from Fama and MacBeth (1973) regressions of absolute abnormal returns and abnormal trading volumes on abnormal shares of old and recombination news:
$Column1: |AbnRet|_{i,t} = a + bAbnPrcOld_{i,t} + gX_{i,t} + e_{i,t}$
$Column2: |AbnRet|_{i,t} = a + b_1AbnPrcOld_{i,t} + b_2AbnPrcRecombination_{i,t} + gX_{i,t} + e_{i,t}$
$Column3: AbnVol_{i,t} = \alpha + \beta AbnPrcOld_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$
$Column4: AbnVol_{i,t} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$
Controls $X_{i,t}$ include news volume for firm i on date t ($Stories_{i,t}$), abnormal news volume over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average number of unique terms per article ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), and prior-week measures of abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal trading volume ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$), and illiquidity ($Illiq_{i,[t-5,t-1]}$). Abnormal variables are computed relative to a value-weighted index of all firms in the universe on the given date. Newey and West (1987) standard errors robust to heteroskedasticity and five autocorrelation lags are provided in parentheses. All coefficients are scaled to correspond to the effect sizes from a 10% increase in the explanatory variables. The coefficients are reported in percentage point units.

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		Dependent variable: $AbnVol_{i,t}$	
	(1) Old News Only	(2) Old News & Recombinations	(3) Old News Only	(4) Old News & Recombinations
$AbnPrcOld_{i,t}$	-0.098%***	-0.115%***	-0.062%***	-0.075%***
	(0.017%)	*(0.019%)*	*(0.006%)*	*(0.007%)*
$AbnPrcRecombination_{i,t}$		0.176%***		0.088%***
		(0.016%)		*(0.006%)*
Controls:				
$Stories_{i,t}$	X	X	X	X
$AbnStories_{i,[t-5,t-1]}$	X	X	X	X
$Terms_{i,t}$	X	X	X	X
$MCap_{i,t}$	X	X	X	X
$BM_{i,t}$	X	X	X	X
$AbnRet_{i,[t-5,t-1]}$	X	X	X	X
$AbnVol_{i,[t-5,t-1]}$	X	X	X	X
$AbnVolatility_{i,[t-5,t-1]}$	X	X	X	X
$Illiq_{i,[t-5,t-1]}$	X	X	X	X
R^2	0.243	0.252	0.185	0.197

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

difference between firm i's return on date t and the return on a value-weighted index of all firms in our universe on date t. $AbnVol_{i,t}$ is abnormal trading volume for firm i on date t, defined as the difference between the fraction of shares turned over for firm i on date t and the value-weighted average of the fraction of shares turned over for all firms in our sample. The controls $X_{i,t}$ include daily news volume for the firm ($Stories_{i,t}$), abnormal news volume during the preceding week relative to the preceding three months ($AbnStories_{i,[t-5,t-1]}$),[13] the average article length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), cumulative abnormal return during the preceding week ($AbnRet_{i,[t-5,t-1]}$), average abnormal trading volume during the preceding week ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility during the preceding week relative to value-weighted average volatility of all firms in the sample ($AbnVolatility_{i,[t-5,t-1]}$), and the log of the Amihud (2002) illiquidity measure during the preceding week ($Illiq_{i,[t-5,t-1]}$). We estimate the regressions over the full sample period; in untabulated analyses, we confirm that the results are robust to excluding earnings announcement dates, and in Section 5.4, we explore the dynamics of the effects over time.

Table 5 presents the results from specifications (3) and (4) in columns 2 and 4, respectively.[14] The coefficients reflect the *differential* market reactions to overall old news and specifically recombinations relative to novel news. Coefficient estimates are scaled to correspond to the effects of 10% increases in the *AbnPrcOld* and *AbnPrcRecombination* measures. First, we test the baseline intuition that, in general, the market reacts less to old information than to new information (Prediction 1). While this prediction is conceptually obvious in the clean experimental setting, in the empirical data, repetition of news is not randomly assigned. If more important news is repeated more frequently, repetition will be associated with stronger market responses. This does not appear to be the main driver of old news in the sample, as an additional 10% of the news about firm i being old corresponds to an 11-basis-points-*smaller* absolute abnormal return and 0.08% lower abnormal trading volume. Stated differently, going from a piece of news in the lowest quartile of old content to a piece of news in the top quartile corresponds to an approximately 30 basis points lower absolute abnormal return and 0.20% lower trading volume (since the interquartile range in the old news measure is 26%, per Table 4). This is an economically meaningful effect: given that the average absolute daily return for securities in our sample is 1.15%, the interquar-

[13] The construction of $AbnStories_{i,[t-5,t-1]}$ requires three months of preceding data. As a result, we begin computing all dependent variables, key measures, and control variables on April 1, 2000, leaving January 1—March 31, 2000, as a buffer for variable construction.

[14] Columns 1 and 3 reestimate equations (3) and (4) for *AbnPrcOld* alone, without including *AbnPrcRecombination* among the explanatory variables, akin to Tetlock (2011), and yield consistent results.

tile range of old news explains 26% of the average daily price move.

Second and most importantly, the results in Table 5 confirm the key empirical prediction from our experiment: that difficulty in processing more complex repackaging of old news implies larger price responses to recombinations than to reprints (Prediction 2). The positive coefficient on *AbnPrcRecombination* shows that, holding overall old news constant, the market reacts more strongly when the old news consists mostly of recombination articles rather than reprints. An additional 10% of recombination (rather than reprint) old news corresponds to an additional 18 basis points absolute abnormal return and an additional 0.09% abnormal trading volume. In other words, given that the interquartile range of the recombination measure is 11%, firms in the top quartile experience 19 basis points larger absolute abnormal returns and 0.10% larger trading volumes than firms in the bottom quartile. Combining the coefficients on *AbnPrcRecombination* and *AbnPrcOld*, we see that, all other things equal, if a firm has an additional 10% of recombination news on a given day, it experiences on average a 6-basis-points larger absolute abnormal return ($0.176\% - 0.115\% = 0.061\%$). This suggests that the market reacts quite strongly to recombination news, potentially even more so than to novel news. However, in the next subsection we see that these additional returns associated with recombination news tend to fully reverse over the following week.

Finally, it is worth noting that two factors may bias the analysis *against* our results. First, occasional news articles (especially breaking news) may receive updates in subsequent articles. Such updates would be classified as either novel news (if the additional information accounts for more than 40% of the content) or reprints of the original story (if the original story accounts for at least 60% of the content). This can push against our results by misclassifying new information as reprints. Second, recombination news tends to be published by sources with slightly lower quality than reprints and novel news, although the difference is very small. On Bloomberg's five-point scale (where 1 is best and 5 is worst), the average ranking of the sources publishing novel news, reprint news, and recombination news is 2.28, 2.26, and 2.35, respectively. These factors suggest that the estimated coefficient on *AbnPrcRecombination* in Table 5 may be a lower bound.

4.2. Return reversals

The preceding results confirm that the market reacts more strongly to recombination of information than to direct reprints. But do these stronger responses reflect overreactions?

The experimental results in Section 2 showcase imperfect processing of recombination news: since these news articles combine old information from several sources, they are more difficult for investors to identify as old news than simple reprints, even holding all else about the news headlines constant. However, in the non-experimental setting of the Bloomberg terminal, recombination news is not randomly assigned, and it is possible that the choice to combine two or more facts into a single article can be triggered by nuanced links and connections that are especially worthwhile to highlight. Such recombination news would then serve the valuable function of facilitating information processing through the juxtaposition of pre-existing but inadequately noted information. For example, this has been found to be the case in patent innovation (Hirshleifer et al. (2018)). Under this alternative, reactions to recombination news would constitute correct reactions to important insight, rather than merely overreactions to previously-seen old news.

Our aim is not to establish that all recombination news articles carry no value, but rather to quantify market *over*reactions to old content in recombinations. Financial professionals' susceptibility to actual old content in recombination news is established in the clean experimental setting in Section 2. Now, we show that this cognitive limitation has important implications for asset prices by demonstrating that a large part of initial market reactions to recombination news subsequently reverses, in line with Prediction 3. As a null hypothesis, if the reactions to recombination news are warranted, then we should observe no subsequent reversal. To this end, we estimate the following specification:

$$\begin{aligned} &AbnRet_{i,[t+t_1,t+t_2]} \\ &\quad = \alpha + \beta_1 AbnPrcOld_{i,t} \\ &\quad + \beta_2 AbnPrcOld_{i,t} \times AbnRet_{i,t} + \beta_3 AbnRet_{i,t} \\ &\quad + \delta_1 AbnPrcRecombination_{i,t} + \delta_2 AbnPrcRecombination_{i,t} \\ &\quad \times AbnRet_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{5}$$

where $AbnRet_{i,[t+t_1,t+t_2]}$ is the signed abnormal return for firm i over the dates $[t+t_1, t+t_2]$, and $AbnRet_{i,t}$ is the signed abnormal return for firm i on date t. Following Tetlock (2011), we focus on the delayed window of $[t+2, t+5]$, but we also consider $[t+1, t+5]$ and $[t+2, t+10]$. We include the same set of controls as in Section 4.1.

Specification (5) assesses (i) the extent to which market reactions to old content in general tend to reverse in subsequent trading and (ii) the extent to which these reversals are driven by recombination news. The coefficient on the first interaction term, β_2, captures the degree of differential reversal of abnormal returns after larger shares of (daily) old news relative to the reversal of returns after lower shares of old news (Prediction 3a). To the extent that there is any overreaction to simple reprints of old information, the coefficient β_2 should be negative. The coefficient on the second interaction term, δ_2, measures differential reversal following larger shares of recombination news (as opposed to straightforward reprints). Prediction 3b posits that this coefficient should be negative, indicating that the larger reactions to recombination news documented in the previous subsection reflect market overreactions that subsequently reverse.

The results in Table 6 show that the coefficients on both interaction terms, $AbnPrcOld_{i,t} \times AbnRet_{i,t}$ and $AbnPrcRecombination_{i,t} \times AbnRet_{i,t}$, are negative and statistically significant in all specifications. For example, a comparison of the coefficients on $AbnPrcOld_{i,t} \times AbnRet_{i,t}$ and $AbnRet_{i,t}$ in column 2 suggests that when an additional 10% of a firm's news on a given day is old, then that firm's next business day return is subject to a thrice larger reversal

Table 6
Results from Fama and MacBeth (1973) regressions of abnormal returns over $[t+t_1, t+t_2]$ on abnormal returns on t, interacted with abnormal shares of old and recombination news: $AbnRet_{i,[t+t_1,t+t_2]} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcOld_{i,t} \times AbnRet_{i,t} + \beta_3 AbnRet_{i,t} + \delta_1 AbnPrcRecombination_{i,t} + \delta_2 AbnPrcRecombination_{i,t} \times AbnRet_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$ The regressions are run over three horizons of delayed abnormal returns: $[t+2, t+5]$ in column 1, $[t+1, t+5]$ in column 2, and $[t+2, t+10]$ in column 3. Controls include daily news volume for the firm ($Stories_{i,t}$), abnormal news over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average number of unique terms per article ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$), and illiquidity ($Illiq_{i,[t-5,t-1]}$). Abnormal variables are computed relative to a value-weighted index of all firms in the universe on the given date. Newey and West (1987) standard errors robust to heteroskedasticity and five autocorrelation lags are reported in parentheses. All coefficients are scaled to correspond to the effect sizes from a 10% increase in the explanatory variables.

	(1) $AbnRet_{i,[t+2,t+5]}$	(2) $AbnRet_{i,[t+1,t+5]}$	(3) $AbnRet_{i,[t+2,t+10]}$
$AbnRet_{i,t}$	-0.044**	-0.051**	-0.039
	(0.020)	*(0.023)*	*(0.033)*
$AbnPrcOld_{i,t} * AbnRet_{i,t}$	-0.094***	-0.139***	-0.168*
	(0.025)	*(0.029)*	*(0.038)*
$AbnPrcRecombination_{i,t} * AbnRet_{i,t}$	-0.131***	-0.169***	-0.199**
	(0.039)	*(0.059)*	*(0.098)*
R^2	0.092	0.098	0.082
Controls:			
$Stories_{i,t}$	X	X	X
$AbnStories_{i,[t-5,t-1]}$	X	X	X
$Terms_{i,t}$	X	X	X
$MCap_{i,t}$	X	X	X
$BM_{i,t}$	X	X	X
$AbnRet_{i,[t-5,t-1]}$	X	X	X
$AbnVol_{i,[t-5,t-1]}$	X	X	X
$AbnVolatility_{i,[t-5,t-1]}$	X	X	X
$Illiq_{i,[t-5,t-1]}$	X	X	X

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

during $t+2$ to $t+5$. Relative to the baseline size of the daily price move in the sample, which is 1.15%, this means that an additional 10% of a firm's news being old systematically predicts an 11 basis point return in the opposite direction over $t+2$ to $t+5$.

Most importantly, the estimated coefficients on the incremental reversal following recombination news, $AbnPrcRecombination_{i,t} \times AbnRet_{i,t}$, are consistently negative, economically sizable, and statistically significant. In fact, the additional reversal in response to recombination news is even starker, both economically and statistically, than the baseline reversal following old news in general. An additional 10% of recombination news on a given day forecasts, on average, a systematic 16 (19) basis point return in the opposite direction during $t+2$ to $t+5$ ($t+1$ to $t+5$). Recalling that the initial (next day) effect of recombination news is 18 basis points, this means that the additional response to recombination news (relative to reprints) fully reverses within the following week. Combining these results with the findings in Section 4.1, we see that reactions to recombinations are not only stronger than those to reprints, but also subsequently reverse—consistent with the market overreacting to recombination of old news.

5. Additional analyses

We explore how important characteristics of the environment, including investor attention and news sentiment and ambiguity, modulate our results. We also examine potential differences in reactions by retail versus institutional investors and the evolution of the documented effects over time. At the end, we discuss the robustness of our main results to alternative specifications.

5.1. The role of investor attention

We examine whether the overreactions to stale and recombination news are stronger when investors are less attentive to news overall.

We use Bloomberg's measure of daily investor attention on the terminal, described in detail in Ben-Rephael et al. (2017). The measure (available from the Bloomberg terminal as "News Heat – Daily Max Readership") is at the security-day level and is based on the daily maximum of eight-hour news reads and searches for a given security. The measure takes values between 0 and 4. The majority of security-days have the baseline value of 0, indicating that no eight-hour period within that day ranks above the 80th percentile of attention for that security, compared to the preceding 30 days. Values of 1, 2, 3, and 4 are assigned to security-days whose maximal eight-hour counts rank above the 80th, 90th, 94th, and 96th percentiles, respectively. We slice the sample into a subsample of security-days with the value of 0 (no attention spikes, 61% of all security-days) and a subsample of security-days with values of 1 and above (at least some attention spikes, 39% of all security-days).

The results from estimating equations (3), (4), and (5) for the two subsamples of investor attention are reported in Table 7. Panel 1 considers contemporaneous market reactions (day t returns and trading volumes), while Panel 2 looks at return reversal over days $t+2$ through $t+5$. The results in Panel 1 reveal that the market reacts

Table 7
Differential effects based on investor attention. The sample is split into two subsamples: with attention spike measure at 0 (61% of the sample) and with attention spikes above 0 (39% of the sample). **Panel 1** estimates Fama and MacBeth (1973) regressions of the same-day absolute abnormal returns and abnormal trading volumes on *AbnExtentOld* and *AbnExtentRecombination*, separately for each subsample of investor attention. **Panel 2** estimates the reversal of the initial abnormal returns on day t during the subsequent period $[t+2, t+5]$, separately for each subsample of investor attention.
Controls include the daily news volume for the firm ($Stories_{i,t}$), abnormal news volume over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average article length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$), and illiquidity ($Illiq_{i,[t-5,t-1]}$). Abnormal variables are computed relative to a value-weighted index of all firms in the universe on the given date. Newey and West (1987) standard errors robust to heteroskedasticity and five autocorrelation lags are reported in parentheses. All coefficients are scaled to correspond to the effect sizes from a 10% increase in the explanatory variables.

Panel 1: Market Reaction

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		Dependent variable: $AbnVol_{i,t}$	
	(1) Low attention	(2) High attention	(3) Low attention	(4) High attention
$AbnPrcOld_{i,t}$	-0.086%***	-0.159%***	-0.004%	-0.123%***
	(0.020%)	*(0.027%)*	*(0.007%)*	*(0.008%)*
$AbnPrcRecombination_{i,t}$	0.164%***	0.180%***	0.028***	0.102%***
	(0.019%)	*(0.025%)*	*(0.007%)*	*(0.008%)*

Panel 2: Return Reversal

	Dependent variable: $AbnRet_{i,[t+2,t+5]}$	
	Low attention	High attention
$AbnExtentOld_{i,t} \times AbnRet_{i,t}$	-0.105***	-0.056
	(0.038)	*(0.049)*
$AbnExtentRecombination_{i,t} \times AbnRet_{i,t}$	-0.133**	-0.099*
	(0.052)	*(0.059)*

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

much less to stale news (compared to novel news) when investor attention is high. For example, a 10% increase in overall stale content corresponds to 16-basis-points-lower absolute abnormal returns when investor attention is high, and only 9-basis-points-lower abnormal returns when investor attention is low. Panel 2 shows that there is also less reversal of the initial return associated with stale news when investor attention is high. These results are consistent with increased investor attention helping to mitigate some of the overreactions to stale news.

Interestingly, there is no evidence of the recombination effect weakening with increased investor attention. In Panel 1, the *differential* returns and especially trading volumes associated with recombinations (vs. reprints) are higher in the sub-period with high investor attention. In Panel 2, the reversal of the returns associated with recombinations is lower in the high-attention subsample, but the difference is not significant. Overall, the combined set of results in Table 7 suggests that increased investor attention helps to mitigate overreaction to stale news, but this is driven primarily by better screening out of simple reprints. The recombination effect—the relative difference in market participants' susceptibility to recombinations versus reprints—is present in periods of both high and low investor attention.

5.2. Retail and institutional investors

We consider how our results vary across retail and institutional investors. In general, cognitive biases tend to be more pronounced in retail investors than institutional investors, including overreactions to stale news (Tetlock (2011)). Our experimental results also suggest that retail investors have more difficulty identifying novel news.

We examine this question in two ways. First, we split the sample based on retail versus institutional ownership. We collect data on firm-level institutional investor ownership from NASDAQ.[15] The median firm in the sample is 78% owned by institutions. Table 7 reports the results from estimating equations (3), (4), and (5) on two subsamples: below-median and above-median institutional ownership. Panel 1 shows that stale news is associated with larger reactions (absolute returns and especially trading volumes) in stocks with low institutional ownership. In fact, news staleness does not appear to reduce trading volumes at all in low-institutional-ownership stocks (a coefficient of -0.002% and statistically insignificant). By contrast, the relation between the extent of recombination news and market activity is similar in the two subsamples. In Panel 2, we see that there is slightly more correction of returns (reversal) associated with stale news in the low-institutional-ownership subsample, but slightly more correction of returns associated with recombination news in the high-institutional-ownership subsample. Overall, these results confirm that intitutional investors are better at screening out straightforward duplicates, but even they appear susceptible to more complex recombination of old news.

Second, Internet Appendix Table A.6 directly assesses daily institutional versus retail order imbalance (IROI). We identify institutional trades in the NYSE Trade and Quote

15 For tickers with no available data on the NASDAQ website, we supplement the dataset with information from Fidelity and MarketBeat.

Table 8
Differential effects based on institutional versus retail ownership. The sample is split into two subsamples: firms with above-median and below-median retail ownership. **Panel 1** estimates Fama and MacBeth (1973) regressions of the same-day absolute abnormal returns and abnormal trading volumes on *AbnExtentOld* and *AbnExtentRecombination*, separately for each subsample of institutional ownership. **Panel 2** estimates the reversal of the initial abnormal returns on day t during the subsequent period $[t+2, t+5]$, separately for each subsample of institutional ownership. Controls include daily news volume for the firm ($Stories_{i,t}$), abnormal news volume over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average article length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$), and illiquidity ($Illiq_{i,[t-5,t-1]}$). Abnormal variables are computed relative to a value-weighted index of all firms in the universe on the given date. Newey and West (1987) standard errors robust to heteroskedasticity and five autocorrelation lags are reported in parentheses. All coefficients are scaled to correspond to the effect sizes from a 10% increase in the explanatory variables.

Panel 1: Market Reaction

	Dependent variable: $\lvert AbnRet_{i,t}\rvert$		Dependent variable: $AbnVol_{i,t}$	
	(1) Low instit. ownership	(2) High instit. ownership	(3) Low instit. ownership	(4) High instit. ownership
$AbnPrcOld_{i,t}$	-0.076%***	-0.123%***	-0.002%	-0.077%***
	(0.022%)	(0.020%)	(0.005%)	(0.013%)
$AbnPrcRecombination_{i,t}$	0.168%***	0.175%***	0.079%***	0.055%***
	(0.024%)	(0.023%)	(0.008%)	(0.010%)

Panel 2: Return Reversal

	Dependent variable: $AbnRet_{i,[t+2,t+5]}$	
	Low institutional ownership	High institutional ownership
$AbnExtentOld_{i,t} \times AbnRet_{i,t}$	-0.104**	-0.085**
	(0.042)	(0.040)
$AbnExtentRecombination_{i,t} \times AbnRet_{i,t}$	-0.112**	-0.145***
	(0.049)	(0.051)

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

database (TAQ) following Bushee et al. (2020), as trades that are at least $50,000 in size. We identify retail trades following Boehmer et al. (2021): if the trade has a TAQ exchange code of "D" and a price of just above or below a round penny. We present the analysis as suggestive, due to the caveat that there is no perfect measure of retail and institutional trades; size-based approaches are limited by the extent to which institutions split their large trades, and sub-penny trades do not always identify retail trades (Barber et al. (2022)). Table A.6 Column 1 considers the aggregate measure of IROI: institutional shares traded minus retail shares traded, scaled by the sum of the two. Column 2 concentrates on buy-initiated trades, and column 3 considers sell-initiated trades.[16] Each measure is demeaned relative to the firm-level average, to account for the potential predisposition of retail investors to favor certain stocks. The results show that a 10% increase in overall stale news content on a given day corresponds to a 0.656 percentage point reduction in institutional-retail order imbalance (IROI). Recombinations are associated with slightly (statistically insignificantly) higher IROI than reprints. Overall, retail investors are more eager to trade on reprints, and the divergence appears to be smaller for recombinations.

5.3. News sentiment and ambiguity

We explore how our empirical results interplay with news content, along two key dimensions: sentiment (capturing positive versus negative news) and ambiguity (capturing the extent to which a given piece of news conveys clear factual information vs. more subjective information).

The methodology for measuring news sentiment and ambiguity follows Fedyk (2021). Human experts manually tag 10,000 news articles as positive, negative, or neutral, and as hard or soft information. The articles are then represented as vectors of features including story length, topics covered, indicators for specific unigrams, bigrams, and trigrams appearing in the text, syntactic complexity, and indicators for particular patterns of syntactic structure and semantic relationships. We use a machine learning method—support vector machine (Cortes and Vapnik, 1995)—to classify other news articles based on the attributes learned from the training data. Sentiment is classified into three classes: positive, negative, and neutral. Ambiguity is a continuous measure from 0 (most straightforward) to 1 (most ambiguous). The ambiguity score is an average of two components: (i) the confidence with which the method identifies the article's sentiment (i.e., how far the article is from the separating hyperplane for its sentiment class), and (ii) whether the article is classified as hard information, interacted with the confidence of this classification. For example, articles covering about earnings reports with hard numbers are classified as straightforward. By contrast, articles discussing employee satisfaction are likely to be classified as ambiguous.

News sentiment can interact in an important way with the recombination effect, if some recombination articles combine multiple previous articles with opposing sentiment, juxtaposing a positive and a negative piece of news about a given firm). Empirically, this is not the case. In 59.4% of cases, recombinations draw from two articles with the same sentiment (both positive, both negative, or

[16] The table focuses on the time period beginning in 2005, due to availability of the Boehmer et al. (2021) measure.

Table 9
Differential effects based on news sentiment. The sample is split into two subsamples: positive news and negative (and neutral) news. **Panel 1** estimates Fama and MacBeth (1973) regressions of the same-day absolute abnormal returns and abnormal trading volumes on *AbnExtentOld* and *AbnExtentRecombination*, separately for each subsample of news. **Panel 2** estimates the reversal of the initial abnormal returns on day t during the subsequent period $[t+2, t+5]$, separately for each subsample of news. Controls include daily news volume for the firm ($Stories_{i,t}$), abnormal news volume over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average article length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$), and illiquidity ($Illiq_{i,[t-5,t-1]}$). Abnormal variables are computed relative to a value-weighted index of all firms in the universe on the given date. Newey and West (1987) standard errors robust to heteroskedasticity and five autocorrelation lags are reported in parentheses. All coefficients are scaled to correspond to the effect sizes from a 10% increase in the explanatory variables.

Panel 1: Market Reaction

	Dependent var: $\lvert AbnRet_{i,t}\rvert$		Dependent var: $AbnVol_{i,t}$	
	(1) Positive news	(2) Negative news	(3) Positive news	(4) Negative news
$AbnPrcOld_{i,t}$	-0.123%***	-0.107%***	-0.086%***	-0.068%***
	(0.038%)	*(0.032%)*	*(0.013%)*	*(0.013%)*
$AbnPrcRecombination_{i,t}$	0.192%***	0.144%***	0.106***	0.089%***
	(0.039%)	*(0.038%)*	*(0.014%)*	*(0.010%)*

Panel 2: Return Reversal

	Dependent variable: $AbnRet_{i,[t+2,t+11]}$	
	Low attention	High attention
$AbnExtentOld_{i,t} \times AbnRet_{i,t}$	-0.096***	-0.087*
	(0.034)	*(0.048)*
$AbnExtentRecombination_{i,t} \times AbnRet_{i,t}$	-0.135**	-0.113**
	(0.054)	*(0.056)*

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

both neutral), and in 34.2% of cases recombinations mix a signed story (positive or negative) with a neutral one. Only 6.4% of cases reflect recombinations drawing from one positive and one negative previous article. Thus, mixed tone does not appear to be an issue for recombination news.

In Table 8, we split the sample based on news sentiment. For each ticker-date in the sample, we compute the average sentiment of the news published about that ticker on that date. We classify ticker-dates with *positive* average sentiment as "good news," and the remaining ticker-dates as "bad news." The results are very similar in both good news and bad news subsamples. Coefficient estimate are slightly larger in the good news subsample, but the differences are insignificant. Overall, news sentiment does not appear to be an important factor for our results.

We incorporate a control for news ambiguity in Table A.7 in the Internet Appendix. For each security on each day, ambiguity is measured as the average of the ambiguity scores of the articles tagged with that security on that day. The results are very similar to the baseline results reported in Tables 5 and 6, with slightly stronger coefficients for same-day returns and volumes and slightly lower coefficients for return reversal. Overall, our findings are consistent across good and bad news and robust to controlling for the amount of hard quantitative (unambiguous) information in the news about each security.

5.4. Time series of the effects

We investigate whether the presence of increasingly more sophisticated arbitrageurs has attenuated reactions to recombination news. Evidence in Section 2 pushes against this being the case: in a controlled experimental setting conducted during 2018–2019, active finance professionals remain susceptible to old news in recombination form. We now investigate whether this susceptibility is supported by persisting market effects.

Taking advantage of the long time period of our Bloomberg terminal data, we investigate the dynamics of the price responses to old and recombination news by considering the time series of the estimated coefficients. Specifically, we re-estimate regression specification (3) separately for each full year in our sample. The resulting coefficients are plotted in Fig. 6 with solid lines, with the 95% confidence intervals marked in dashed lines.

Market reactions to old news in general, compared to new news, declined from 2001 to 2014, as can be seen in Panel 1 of Fig. 6. For example, an additional 10% of a firm's news on a given day consisting of old content in 2001 corresponds to a statistically insignificant 4-basis-points-smaller absolute abnormal return, while an additional 10% of daily old news in 2014 translates into a precisely estimated 14-basis-points-smaller absolute abnormal return. This is consistent with advances in natural language processing enabling finance professionals to more readily screen out basic repetition such as direct reprints of old news.[17]

By contrast, the differential reaction to recombination of information (compared to reprints) has only increased over time. Panel 2 of Fig. 6 shows that the re-

[17] While the Bloomberg terminal did not offer functionality to screen out reprints in 2000–2014, there were efforts to employ natural language processing to group reprints under the original headline. It is likely that analogous methods were also applied by the more sophisticated institutional clients of Bloomberg.

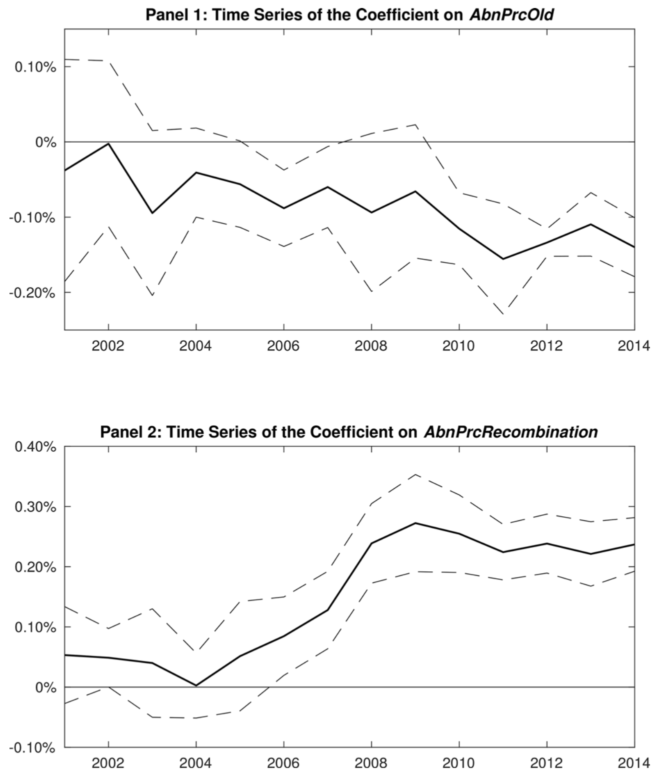

Fig. 6. Time series of the coefficients from regressions of absolute abnormal returns on old and recombination news. Daily Fama and MacBeth (1973) cross-sectional regressions are estimated for each year $T \in [2001, 2014]$: $|AbnRet|_{i,t} = \alpha_T + \beta_{1,T} AbnPrcOld_{i,t} + \beta_{2,T} AbnPrcRecombination_{i,t} + \gamma_T X_{i,t} + \epsilon_{i,t}$ The time series of the estimated coefficients $\{\hat{\beta}_{1,T}\}_{T\in[2001,2014]}$ is displayed in **Panel 1. Panel 2** presents the time series of $\{\hat{\beta}_{2,T}\}_{T\in[2001,2014]}$. Coefficient estimates are presented in solid lines, with 95% confidence bounds marked in dashed lines.

lationship between absolute abnormal returns and the *AbnPrcRecombination* measure, controlling for overall old news, is positive and significant at the 5% level in eleven out of the fourteen years in the sample period. The exception years (2001, 2003, 2004, and 2005) all fall at the beginning of the sample. Both the economic and the statistical significance of the larger absolute abnormal returns in response to recombination news are substantially greater in the second half of the sample. While the effect is statistically indistinguishable from 0 in 2001, by 2014 an additional 10% of a firm's news featuring recombination articles (as opposed to simple reprints) translates into 24-basis-points-larger absolute abnormal returns, statistically significant at the 1% level. The divergence in market reactions to recombinations versus reprints may have been amplified by algorithmic screening tools, which tend to consist of a locality-sensitive hashing applied to *pairwise* comparisons of articles (Petrović et al. (2010)). These tools aim to group reprints with the original news. However, they do not go through the multi-step necessary procedure for identifying recombinations, because they tend to prioritize speed over sophistication, for faster real-time processing.

These results suggest that the differential market reaction to recombination of old information is not a vestige of the earlier years in the sample. Our "recombination effect" mechanism is, if anything, even stronger in more recent years. While investors appear to be getting increasingly more sophisticated in disregarding old information contained in straightforward reprints, cognitive limitations in screening out recombination of old information continue to impact asset prices.

5.5. Robustness

We confirm that our results are robust to different constructions of both abnormal returns and old news. We also consider continuous measures of old and recombination content instead of a discrete classification of each article, and restrict our sample in order to screen out any differences in residual novel content between reprints and recombinations.

First, we repeat the baseline tests of same-day abnormal returns, same-day trading volumes, and subsequent return reversal around stale and recombination news. Instead of using simple abnormal returns, we compute characteristics-adjusted returns following the methodology in Daniel et al. (1997). Internet Appendix Table A.8 shows that our results are robust to this specification.

Second, our baseline measures of old news and recombinations consider, for any given article, the five most similar articles about the same firm over the preceding three business days. We test the robustness of our results to the discretionary choices in this specification by varying: (i) the look-back window (τ) to five and 10 previous business days; (ii) the number (n) of considered most similar articles about the same firm to 10; and (iii) imposing a minimum limit, whereby we only consider firm-dates that actually have at least n articles in the preceding τ days. The results of these different construction methods are presented in Internet Appendix Table A.9 (for same-day absolute abnormal returns and trading volumes) and Internet Appendix Table A.10 (for return reversal). The results are significant for every single construction method. The economic estimates are also robust to specification, with the baseline specification (highlighted) falling in the interior of the range.

Interestingly, the results do not weaken when we allow for similar news to occur further in the past ($t = 5$ and $t = 10$). This is consistent with experimental evidence on memory and associative recall. Enke et al. (2020) examine repeated signals (corresponding to direct reprints in our setting) and show overreaction to repetition that triggers associative recall of past signals through similar contextual cues. Their effect also does not decline over time and strengthens when the time lag between the initial signal and the repetition increases from 15 minute to three days.

Third, we confirm that our results are robust to an alternative approach based on each article's continuous measures of old and recombined content instead of a discrete classification. For each firm i on date t, we compute the firm-level measures, $ExtentOld_{i,t}$ and $ExtentRecombination_{i,t}$, by averaging the respective measures across the set of articles $S_{i,t}$ published on date t and tagged with firm i (with $|S_{i,t}|$ denoting the size of $S_{i,t}$). Specifically:

$$ExtentOld_{i,t} = \frac{1}{|S_{i,t}|} \sum_{s \in S_{i,t}} Old(s) \tag{6}$$

$$ExtentRecombination_{i,t} = \frac{1}{|S_{i,t}|} \sum_{s \in S_{i,t}} (Old(s) - ClosestNeighbor(s)), \tag{7}$$

As before, we take abnormal values of both measures, $AbnExtentOld_{i,t}$ and $AbnExtentRecombination_{i,t}$, by computing residuals from daily cross-sectional regressions of $ExtentOld_{i,t}$ and $ExtentRecombination_{i,t}$, respectively, against firm-level daily news volume, the log of the average news length, and the square of the log of the average news length. For ease of interpretability, we also normalize the continuous variables to have a mean zero and a standard deviation of one. Table A.11 in the Internet Appendix tabulates the results for abnormal returns and trading volumes (Panel 1) and return reversals (Panel 2), using $AbnExtentOld_{i,t}$ and $AbnExtentRecombination_{i,t}$ measures. The results in Panel 1 are consistent with those in Table 5, and the results in Panel 2 are comparable to those in Table 6.

Finally, we address the possibility that the differential market reactions documented in Section 4 reflect systematic differences in the quality of residual novel content in recombinations versus reprints. Both recombinations and reprints are subsets of old news, which we define as including any article at least 60% of whose textual content is spanned by the closest preceding articles about the same firm. Hence, these articles can contain up to 40% novel text. Given two articles with equal amounts of old content, one a recombination and the other a reprint, it could be the case that the recombination article is simply more likely to contain important information in the portion of its text that is novel.

We address this possibility in two ways. First, the return reversal results in Section 4.2 are inconsistent with the narrative of the reactions to recombinations arising from superior content. Second, as an additional check that the differences between recombinations and reprints are not due to differences in the articles' residual novel content, we repeat our main tests using only those old news articles that have practically no novel content. In particular, we reclassify articles into novel news, reprints, and recombinations using an old content threshold of 90%, rather than 60%. Table 9 displays the results for absolute abnormal returns and trading volumes in Panel 1, and for return reversal in Panel 2. The results in Panel 1 are similar in size and significance to their counterparts in Table 5 and remain statistically significant despite lower power from substantially restricting the set of old news. Likewise, return reversals documented in Panel 2 of Table 9 are comparable to those catalogued in Table 6. The robustness of the results to restricting the sample to news articles that have practically no novel content confirms that differences in residual novel content are not driving our empirical results.

Table 10
Replication of the key results restricting attention to news articles whose text contains at least 90% old content, with corresponding staleness and recombination measures $AbnPrcOld^{90\%}$ and $AbnPrcRecombination^{90\%}$. **Panel 1** estimates Fama and MacBeth (1973) regressions of the same-day absolute abnormal returns and abnormal trading volumes on $AbnPrcOld^{90\%}$ and $AbnPrcRecombination^{90\%}$. **Panel 2** estimates the reversal of the initial returns on day t during subsequently period $[t+2, t+5]$. Controls include daily news volume for the firm ($Stories_{i,t}$), abnormal news volume over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average article length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$), and illiquidity ($Illiq_{i,[t-5,t-1]}$). Abnormal variables are computed relative to a value-weighted index of all firms in the universe on the given date. Newey and West (1987) standard errors robust to heteroskedasticity and five autocorrelation lags are reported in parentheses.

Panel 1: Market Reaction

Explanatory variable	Dependent variable: $\lvert AbnRet \rvert_{i,t}$	Dependent variable: $AbnVol_{i,t}$
$AbnPrcOld^{90\%}_{i,t}$	-0.135%***	-0.080%***
	(0.034)	(0.008)
$AbnPrcRecombination^{90\%}_{i,t}$	0.150%***	0.087%***
	(0.038)	(0.008)

Panel 2: Return Reversal

Explanatory variable	Dependent variable: $AbnRet_{i,[t+2,t+5]}$
$AbnPrcOld^{90\%}_{i,t} \times AbnRet_{i,t}$	-0.061*
	(0.035)
$AbnPrcRecombination^{90\%}_{i,t} \times AbnRet_{i,t}$	-0.143**
	(0.060)

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

6. Conclusion

This paper shows that limited attention and neglect of correlations across media sources makes even sophisticated investors susceptible to repetition in financial news. While news articles that directly reprint information from individual previous articles are easily identifiable as old news, articles that combine information from several places are more difficult to distinguish from novel information. We document this in a randomized control trial conducted directly on finance professionals employed at key financial institutions including large banks such as Goldman Sachs, investment managers such as PIMCO, and hedge funds such as Two Sigma.

Investors' cognitive difficulty in dealing with complex information structures has direct implications for asset prices. Using a uniquely comprehensive database of news passing through the Bloomberg terminal, we confirm that market reactions are significantly larger in response to recombination articles than in response to simple reprints, and these extra reactions reverse during the following week. The recombination effect is very robust. The results are consistent across a variety of empirical specifications and prevalent across different news sentiment and ambiguity. Furthermore, unlike simple reprints, susceptibility to recombinations is visible even among institutional investors and during times of high investor attention. Finally, the time series of the estimated coefficients indicates that the recombination effect has strengthened over time, suggesting that investors are becoming increasingly more sophisticated in identifying reprints but remain susceptible to recombination of old information. Our findings shed light on the types of cognitive limitations that drive anomalies such as price responses to old news. We believe that further understanding cognitive biases and limitations and how they contribute to market stability and efficiency constitutes a productive direction for future work.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
James Hodson reports financial support was provided by the European Commission.

Data availability

Experimental data are available at doi:10.17632/dvp8y4tdt2.1. Bloomberg news data are unfortunately proprietary, and the authors do not have permission to share them.

Acknowledgements

David Hirshleifer and Nikolai Roussanov were the editors for this article. The authors are very grateful for their feedback and guidance, as well as for the excellent suggestions from the anonymous referee. The authors are also grateful to John Beshears, John Campbell, Lauren Cohen, Felipe Cortes (discussant), Stefano DellaVigna, Joseph Engelberg (discussant), Benjamin Enke, Erik Eyster, Robin Greenwood, Samuel Hanson, Alan Huang (discussant), Victoria Ivashina, Shimon Kogan (discussant), Alan Kwan (discussant), David Laibson, Owen Lamont, Christopher Malloy, Ulrike Malmendier, Sendhil Mullanaithan, Christopher Parsons, Andrei Shleifer, Jeremy Stein, Adi Sunderam, Paul Tetlock, Florian Zimmermann, and seminar participants at Aalto University, BI Norwegian Business School, Fundação Getúlio Vargas, Harvard University, Teradata Corporation, UC Berkeley Haas, UC Santa Barbara, University of São Paulo, the News and Finance Conference, the Rome Junior Finance Conference, AFA, EFA, SITE Workshop on Experimental Economics, EBEM, FIRS, ESSFM Gerzensee, NFA, and ECWFC at the WFA for insightful comments. Hodson acknowledges funding from the EU Research and Innovation programme Horizon 2020 under grant agreement No. 675044. The experimental component of this paper was deemed "not human subject research" by the UC Berkeley IRB. Jonathan Bodine, Richard Gong, and Willy Wu provided valuable research assistance.

Supplementary material

Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.jfineco.2023.04.008

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financ. Market. 5 (1), 31–56.
Barber, B.M., Huang, X., Jorion, P., Odean, T., Schwarz, C., 2022. A (sub) penny for your thoughts: tracking retail investor activity in taq. Working paper.
Barber, B.M., Loeffler, D., 1993. The ædartboardg column: second-hand information and price pressure. J. Financ. Quant. Anal. 28 (2), 273–284.
Ben-Rephael, A., Da, Z., Israelsen, R.D., 2017. It depends on where you search: a comparison of institutional and retail attention. Rev. Financ. Stud. 30 (9), 3009–3047.
Boehmer, E., Jones, C.M., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. J. Finance 76 (5), 2249–2305.
Boulland, R., Degeorge, F., Ginglinger, E., 2017. News dissemination and investor attention. Rev. Financ. 21 (2), 761–791.
Bushee, B., Cedergren, M., Michels, J., 2020. Does the media help or hurt retail investors during the IPO quiet period? J. Account. Econ. 69 (1), 101261.
Busse, J.A., Green, T.C., 2002. Market efficiency in real time. J. Financ. Econ. 65 (3), 415–437.
Carvalho, C., Klagge, N., Moench, E., 2011. The persistent effects of a false news shock. J. Empir. Finance 18 (4), 597–615.
Cohen, L., Malloy, C., Nguyen, Q., 2020. Lazy prices. J. Finance 75 (3), 1371–1415.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.
Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. J. Finance 66 (5), 1461–1499.
Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. J. Finance 52 (3), 1035–1058.
DellaVigna, S., Pollet, J.M., 2009. Investor inattention and friday earnings announcements. J. Finance 64 (2), 709–749.
DeMarzo, P.M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. Q. J. Econ. 118 (3), 909–968.
Dougal, C., Engelberg, J., Garcia, D., Parsons, C.A., 2012. Journalists and the stock market. Rev. Financ. Stud. 25 (3), 639–679.
Drake, M.S., Roulstone, D.T., Thornock, J.R., 2016. The usefulness of historical accounting reports. J. Account. Econ. 61 (2–3), 448–464.
Engelberg, J., 2008. Costly information processing: evidence from earnings announcements. Working paper.
Engelberg, J.E., Parsons, C.A., 2011. The causal impact of media in financial markets. J. Finance 66 (1), 67–97.

Enke, B., Schwerter, F., Zimmermann, F., 2020. Associative memory and belief formation. Working paper.
Enke, B., Zimmermann, F., 2019. Correlation neglect in belief formation. Rev. Econ. Stud. 86 (1), 313–332.
Eyster, E., Weizsacker, G., 2010. Correlation neglect in financial decision–making. Working paper.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81 (3), 607–636.
Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. J. Finance 64 (5), 2023–2052.
Fedyk, A., 2021. Disagreement after news: gradual information diffusion or differences of opinion? Rev. Asset Pric. Stud. 11 (3), 465–501.
Fedyk, A., 2022. Front page news: the effect of news positioning on financial markets. J. Finance Forthcoming.
Gilbert, T., Kogan, S., Lochstoer, L., Ozyildirim, A., 2012. Investor inattention and the market impact of summary statistics. Manage. Sci. 58 (2), 336–350.
Hirshleifer, D., Hsu, P.-H., Li, D., 2018. Innovative originality, profitability, and stock returns. Rev. Financ. Stud. 31 (7), 2553–2605.
Hirshleifer, D., Lim, S.S., Teoh, S.H., 2009. Driven to distraction: extraneous events and underreaction to earnings news. J. Finance 64 (5), 2289–2325.
Hirshleifer, D., Sheng, J., 2022. Macro news and micro news: complements or substitutes? J. Financ. Econ. 145 (3), 1006–1024.
Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36 (1–3), 337–386.
Huberman, G., Regev, T., 2001. Contagious speculation and a cure for cancer: anonevent that made stock prices soar. J. Finance 56 (1), 387–396.
Kaniel, R., Parham, R., 2017. Wsj category kings–the impact of media attention on consumer and mutual fund investment decisions. J. Financ. Econ. 123 (2), 337–356.
Klibanoff, P., Lamont, O., Wizman, T.A., 1998. Investor reaction to salient news in closed-end country funds. J. Finance 53 (2), 673–699.
Murfin, J., Pratt, R., 2019. Comparables pricing. Rev. Financ. Stud. 32 (2), 688–737.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. Econometrica 55 (3), 703–708.
Ortoleva, P., Snowberg, E., 2015. Overconfidence in political behavior. Am. Econ. Rev. 105 (2), 504–535.
Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. J. Financ. Econ. 80 (3), 563–602.
Petrović, S., Osborne, M., Lavrenko, V., 2010. Streaming first story detection with application to twitter. In: Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics, pp. 181–189.
Porter, M.F., 1980. An algorithm for suffix stripping. Program 14 (3), 130–137.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. J. Finance 62 (3), 1139–1168.
Tetlock, P.C., 2011. All the news that's fit to reprint: do investors react to stale information? Rev. Financ. Stud. 24 (5), 1481–1512.