When Can the Market Identify Old News?*

Anastassia Fedyk[†] and James Hodson[‡]

August 30, 2019

Abstract

Why do investors react to old information? We provide survey evidence to experimentally document that active finance professionals are more susceptible to old information when it comes as a *recombination* of content from multiple sources. To evaluate the market implications of this mechanism, we exploit a comprehensive dataset of news passing through the Bloomberg terminal. Recombination of old information prompts larger price moves and subsequent reversals than direct reprints of old stories. Furthermore, while overall reactions to old information have declined over time, differential reactions to recombination stories have risen. Altogether, we document investors' increased sophistication in discarding simple reprints but continuing susceptibility to recombination of old information.

Keywords: news, information processing, investor inattention

1   Introduction

News, as the word implies, refers to information that is believed to be new. Yet not all financial "news" contains new information. Due to the proliferation of information transmission mechanisms that has occurred over the past two decades, finance professionals encounter increasing volumes of reprints and recombinations of old news. On the Bloomberg terminal alone, readers are faced with over a million articles per day from thousands of sources, most of them containing few or no new facts. Market participants' task of sifting through this large volume of news and identifying novel content is nontrivial, and prior studies document that financial markets sometimes react to old information.[1]

Why would stale information prompt market reactions? We argue that it is the *recombination* of old information from multiple sources that prompts market reactions to stale news. First, we show experimentally that finance professionals perceive higher novelty in articles that recombine facts from several sources, compared to direct reprints of single previous articles. Second, consistent with this mechanism, we find that, controlling for overall news staleness, 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. The time trends of our coefficients indicate that the initial price impact of overall stale news has declined over time, and yet the *differential* reaction to recombinations versus reprints has risen. These findings suggest that investors are paying progressively less attention to mere duplication of pre-existing stories, but continue to react to recombinations.

Our analysis exploits a novel dataset of all relevant news articles appearing on the Bloomberg terminal, which is uniquely suited to studying 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 specific equity securities and marked as highly relevant by the editorial staff, our sample covers over 17 million news articles published between January 2000 and December 2014. 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 provided by other organizations, and the remainder gathered from the web. As a result, for any news article, we can observe whether similar information has already been published in a variety of sources. This enables us to construct comprehensive measures of staleness and aggregation for any given piece of news.

We define the staleness of a given news article by assessing the percentage of unique terms

---

[1]See Huberman and Regev (2001), Tetlock (2011), Gilbert, Kogan, Lochstoer, and Ozyildirim (2012), and Drake, Roulstone, and Thornock (2016).

in the article that have already appeared in similar recent articles about the same firm. Our comprehensive dataset allows us to effectively identify large amounts of stale news – 48% of the articles in our sample are at least 80% stale (i.e., at least 80% of the unique words in these articles have appeared in similar recent stories about the same firm). Most of these stale news stories are direct repeats of individual preceding articles. For example, consider the following two headlines from November 2014:

- *Bloomberg News* on November 10: "Apple Said Developing Curved iPhone Screens"
- *New York Post* on November 11: "Apple developing larger, curved-screen iPhones"

Unsurprisingly, approximately 90% of the unique words in the text of the *New York Post* article had appeared in the *Bloomberg News* article. As such, we label this *New York Post* article as a "reprint" – an article whose text is a direct duplication of a *single* textually closest previous story about the same firm.

We are interested in distinguishing reprints from what we term "recombinations": articles that do not contain novel information but that draw upon *several* preceding sources. For example, consider the following three pieces of news from October 3, 2012:

- *The Fly On The Wall*: "VentiRx announces collaboration with Celgene for VTX-233"
- *Bloomberg News*: "Celgene to retain exclusive option to acquire VentiRx"
- *Briefing.com*: "VentiRx Pharmaceuticals announced the formation of an exclusive, world-wide collaboration with Celgene [...] Celgene will retain the exclusive option to acquire VentiRx"

The first story announces Celgene's collaboration with VentiRx for the development of a new drug. The second story tells us another piece of information: that Celgene now holds an exclusive option to acquire VentiRx. It is likely that both of these can be viewed as positive news for Celgene. The *Briefing.com* report captures both of these pieces of news in a single article. It does not provide any new information, but neither does it directly reprint a single previous story.

We argue that from an information processing standpoint, recombination is fundamentally different from direct reprinting. While it is relatively straightforward to recognize that the two headlines in our reprint example contain the same information, more attentive processing is required to identify when an article carries information from several sources. Experimental evidence by Enke and Zimmermann (2017) finds that human subjects tend to ignore correlations when analyzing a stream of reports representing different combinations of the same underlying signals. Such correlation neglect is stronger for more complex structures, and extreme reduction in the complexity of information structures eliminates this

cognitive error altogether. In the domain of financial news, this cognitive limitation would manifest itself as failure to identify recombinations of facts as non-novel, despite a greater ability to screen out simple reprints.

Our experimental investigation of finance professionals' responses to news confirms this conjecture. We survey a sample of active finance professionals from a variety of institutions ranging from large broker dealers such as Goldman Sachs and Morgan Stanley to investment management firms such as State Street and hedge funds such as Two Sigma. These finance professionals are asked to assess novelty of a stream of headlines. The headlines include two samples, of twenty headlines each, regarding two hypothetical firms, Argosy Logistics Inc. and Laker Pharmaceuticals LLC, specifically designed to capture novel news, reprints, and recombinations. All three categories of news are similar in length (an average of roughly 7.6 words each), and the two categories of old news – reprints and recombinations – are constructed to be equally old, with an average of 81.1% of their terms appearing in previous headlines about the same firm. Each participant sees one of the two sets of experimental headlines (about Argosy or Laker), randomly interspersed with a selection of twenty filler headlines. The filler headlines are randomly drawn from a filler sample consisting of thirty actual Reuters headlines from August 2017. The filler headlines serve to simulate the information overload that market participants likely face when reading real-world news.

We document a cognitive limitation in processing recombination of old information. Overall, finance professionals in our sample correctly perceive headlines containing new information to be more novel than headlines containing old news: on average, novel headlines are marked with novelty ratings of 4.52 on a 7-point scale, as compared to 2.82 for stale news headlines. However, finance professionals are susceptible to recombination of old information. In particular, our experimental participants rate recombinations, on average, with novelty scores of 3.03 on a 7-point scale, statistically significantly higher than similarly worded, same length, and equally stale reprints, which are assessed to be have an average novelty rating of 2.61. On an individual level, 72% of participants perceive recombinations to be, on average, more novel than their reprint counterparts. By contrast, only 19% perceive reprints to be more novel than recombination news.

We consider the asset pricing implications of the documented susceptibility to information recombination. First, given that finance professionals generally perceive old news – especially reprints – as less novel than actual new news, we should observe smaller price reactions and less trading accompanying stale news articles. Second, since recombinations are perceived as more novel than simple reprints, we expect to observe larger price moves and more trading activity associated with recombination news. Third, since the larger price reactions to recombination news arise due to a cognitive limitation on the investors' part,

these reactions should be followed by larger subsequent reversals.

To test these predictions empirically, we begin by characterizing each news article in the Bloomberg news sample as either a novel piece of news, a reprint, or a recombination. We mark an article as novel if at least 40% of the terms in its text have not appeared in the closest five preceding articles about the same firm. All other news stories, 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 stale 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. For example, consider two articles, A and B, each containing text that is entirely covered by the previous five most similar stories. Suppose that 80% of the unique words in article A have already appeared in the single most similar previous story about the same firm, but only 50% of the unique words in article B have appeared in its single most similar neighbor. Then both of the articles would be considered old news, but article A would be a reprint, while article B would be a recombination.

Our empirical tests are run at daily horizons, aggregating all articles published about a given firm on a given day. On each day, we compute the proportion of news articles about each firm that are old (either reprints or recombinations), as well as the proportion of news articles that are recombinations. Consistent with our first prediction and prior research, higher proportions of old news are associated with smaller absolute abnormal price changes and trading volumes: there is less market activity when more of the firm's news is stale. On the other hand, 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 stories from reprints to recombinations corresponds to a 13 basis point larger absolute price move on the following trading day. Consistent with our third conceptual prediction, the larger price responses following higher proportions of recombination news are more likely to reverse over the following week. The market evidence aligns with the idea that the relatively higher abnormal returns following recombination news articles reflect the cognitive limitations in processing these news articles that we document in our experiment.

We exploit the time series of our large dataset to investigate how the documented effects vary over time. 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 old news; this suggests that investors have become less responsive to overall stale information. The *differential* response to recombination news, however, has increased in recent years. Altogether, the results point to increased investor sophistication in screening out simple reprints, but continuing suscep-

5

tibility to articles that recombine previously available information from multiple sources.

Additional analyses suggest that our main results are robust to the choices we make in the construction of the news categories, including: (a) the number of days over which we look for previous articles; (b) the number of closest preceding articles about the same firm that are included in the comparisons; (c) the thresholds for novelty, reprints, and recombinations; and (d) the use of continuous staleness and recombination measures instead of discrete classification. We also confirm that our market results are not driven by any systematic differences in the residual novel content in reprints versus recombinations. In particular, we repeat the analysis considering only the stories that have close to no novel content – articles that are at least 90% stale – and show that our results are robust to this specification.

This paper contributes to the literature on cognitive biases in financial markets by proposing a channel leading to market failure to identify stale content in news.[2] Our paper relates to the growing number of studies examining the effects of limited investor attention in processing financial information on asset prices.[3] In particular, 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. We document this cognitive limitation experimentally and empirically confirm its implications for asset prices. Our "recombination effect" helps explain the puzzle of overreactions to old news, which has been documented 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. In his investigation of reactions to stale news, Tetlock (2011) observes that absolute abnormal returns are generally lower when the news about a firm is more stale, but finds evidence of some overreaction to stale news. Evidence from Gilbert, Kogan, Lochstoer, and Ozyildirim (2012) more closely hints at the mechanism we develop in our paper: investors overreact to a recombination of previously released inputs into summary statistics in form of the Leading Economic Indicator. Our paper sheds light on a specific type of inattention that contributes to such reactions to old news, namely the failure to identify recombination of previously available information across multiple sources.

The remainder of the paper proceeds as follows. Section 2 presents the experimental investigation of finance professionals' perception of news novelty, showing that market par-

---

[2]For studies on the effects of financial news on security prices see, among others: Barber and Loeffler (1993), Busse and Green (2002), Chan (2003), Tetlock (2007), Fang and Peress (2009), Engelberg, and Parsons (2011), Carvalho, Klagge, and Moench (2011), Dougal, Engelberg, Garcia, and Parsons (2012), Solomon (2012), Hillert, Jacobs, and Müller (2014), and Liu, Sherman, and Zhang (2014)).

[3]See, for example, Klibanoff, Lamont, and Wizman (1998), Engelberg (2008), DellaVigna and Pollet (2009), Hirshleifer, Lim, and Teoh (2009), Da, Engelberg, and Gao (2011), Drake, Roulstone, and Thornock (2012), Boulland, Degeorge, and Ginglinger (2017), Ben-Rephael, Da, and Israelsen (2017), and Fedyk (2017).

ticipants are more susceptible to recombination of old information than to direct reprints. Section 3 describes the data and methodology for evaluating 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 conducts robustness tests using different methods for computing the key proxies, and Section 6 concludes.

2 Experimental Evidence

In this section, we document a cognitive limitation in processing of old information. Individuals are less likely to identify stale information in news articles that recombine old information from multiple sources than in direct reprints of single previous articles. After documenting this limitation in an experimental setting involving active finance professionals, we draw the empirical implications for trading volumes and asset price dynamics, which are tested in Section 4.

2.1 Design

We propose difficulty in identifying recombination of old information as a potential mechanism for the market's reactions to stale news. In order to document this cognitive limitation, we recruit 41 active finance professionals to participate in an online experiment eliciting their beliefs regarding the novelty of news headlines.

The participants span the entire landscape of finance professionals. The breakdown of participants across industry affiliations is presented in Table 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, PIMCO, and State Street, private equity firms such as Bain Capital and Lindsay Goldberg, and hedge funds including Two Sigma and Point 72. The remaining one third of the sample consists of individuals from investment banks (e.g., Barclays), financial news organizations (Bloomberg and Financial Times), the Federal Reserve Board, private investors, and finance-oriented employees of universities or tech companies such as Facebook. These individuals include key decision makers such as partners and managing director of relevant investment firms, as well as active younger employees such as analysts and traders.

Each participant faces a stream of forty news headlines, including twenty headlines about one of two fictitious companies, Argosy Logistics Inc and Laker Pharmaceuticals LLC. We exploit two samples of experimental headlines, about two different hypothetical firms, in order to ensure that our findings are qualitatively similar across separate samples of news

7

headlines. Each participant is independently randomly assigned to one of two headline samples; out of the 41 experimental participants, 20 see the Argosy news sample and 21 see the Laker news sample. The remaining twenty headlines are filler headlines about other firms, aimed to simulate the type of informational overload market participants are likely to face while consuming real-world information. The filler headlines are drawn from a sample of thirty actual news headlines published by Reuters during August 2017. Each participant sees 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 set and the filler headlines, are presented in Table 2.

The individual headlines about Argosy and Laker represent different informational content. Some are novel news, for which approximately 20% of the terms have appeared in previous headlines about the same firm. The rest are old news, for which approximately 80% of the terms are covered by preceding headlines about the same firm. The old news are further split into two categories: reprints, for which at least 80% of the previously seen terms are spanned by a *single* preceding article about the same firm, and recombinations, for which the single closest previous neighbor covers less than 80% of the overall old content.

Importantly, the experimental setting allows us to test the proposed mechanism in a fully controlled environment, where we design 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 the desired information structures. The two sets of experimental headlines (about Argosy and Laker) are specifically designed to be: (1) equal in average length across the three groups (novel news, reprints, and recombinations); (2) equal in average location across reprints and recombination; and (3) indistinguishable in average level of stale content across reprints and recombinations. These balance checks are presented in Table 3. Overall across the two samples (Argosy and Laker), both reprints and recombinations have exactly 81.1% of their terms spanned by preceding headlines about the same firm.

For concreteness, the following are examples of novel headlines about Argosy, a fictitious trucking company:

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"

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

- "Argosy beats expectations: Q3 trucking results eps up 0.1 on 1.1"

75% of the terms in this headline already appear in preceding headlines about Argosy, all of them from the novel headline (2).

By contrast, below is a headline that combines information from both headline (1) and headline (2):

- "Argosy Q3 earnings beat expectations, but design business down"

This headline mentions both the positive earnings surprise and the letdown of the design business. 82.5% of its terms appear in preceding headlines, but no single preceding headline covers more than 50%. a The headlines are presented sequentially, in the manner depicted in Figure 1. Each headline is presented in the middle of the screen, along with a tag indicating whether the headline pertains to a specific firm, such as Toyota, or topic, such as World News. In the instructions, the participants are 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 rating is elicited via the novelty scale below the headline. This is a 7-point scale, ranging from "Nothing New" to "Completely New".

To simulate the type of cognitive overload that the participating finance professionals are likely to face in real-world news environments, such as the scrolling news on the Bloomberg terminal screen, we place a time constraint on each question. Each participant is allotted exactly 10 seconds to mark the novelty of each headline. There is a timer above the headline, which shows the amount of time left. Once the 10 seconds elapse, the experiment progresses to the next question. If the participant has not marked anything in that time, (s)he receives a pop-up notification that (s)he missed that question; missing more than 3 questions results in being disqualified.

The participants are incentivized to detect novelty as well as they can: the five participants whose answers most correctly match actual novelty of the stories receive $90 bonuses. The entirety of the survey takes 7 minutes to complete, and all participants are offered a $10 gift card as thanks for their participation.

2.2 Results

The results of the experimental survey confirm the conjectured effect: finance professionals are better able to distinguish reprint stories as old news than equally stale recombination stories.

A total of 41 finance professionals took part in the survey; 1 individual was disqualified after partial completion due to missing more than three questions, and another 4 stopped the survey partway. The remaining 36 participants completed the entirety of the survey. The results from the full set of responses are presented in Table 4. Individual-level results for the 36 participants who successfully completed the entirety of the survey are displayed

in Figure 2.

The pooled results show that, in general, participants correctly identify novel stories as containing more new information than either recombinations or reprints, but are more susceptible to recombination of information than to direct reprinting. As can be seen in Panel 1 of Table 4, on average, new stories receive novelty scores of 4.52 out of 7 (4.16 in the Argosy headlines subsample and 4.84 in the Laker headlines subsample). Both recombinations and reprints are identified as containing substantially less new information than their novel counterparts; on average, reprints receive novelty scores of 2.61 out of 7 (2.65 in the Argosy subsample and 2.57 in the Laker subsample), while recombinations are rated as 3.03 out of 7 (3.23 in the Argosy subsample and 2.85 in the Laker subsample).

The differences in novelty rankings across these three groups of news are significant and qualitatively consistent across the two samples of experimental news headlines. As displayed in Panel 2 of Table 4, novel stories are ranked, on average, 1.70 points out of the 7-point novelty scale higher than their non-novel counterparts (pooled across both reprints and recombinations). This difference is highly statistically significant. More importantly, the difference in average rankings between recombinations and reprints is economically meaningful at 0.42 points out of the 7-point novelty score and statistically significant at the 5% level with a t-statistic of 2.05. This difference is also positive, albeit less significant, in the smaller samples restricted to either the Argosy news headlines (where the difference is 0.58) or the Laker news sample (where the difference is 0.27).

The individual-level results confirm the findings from the pooled analysis. We compute the difference in average responses regarding recombination vs. reprint news separately for each experimental participant who completed the entirety of the survey. Figure 3 displays the distribution of these individual-level differences. The median difference is 0.40 points out of the 7-point scale (the average difference is 0.48). The difference is positive for 72.2% of the finance professionals who participated in the survey (19.4% of participants rate reprints higher, and 8.3% of participants perceive the two sets of news equally). Overall, the results point to active finance professionals perceiving recombination stories as more new than their reprint counterparts, even in a setting where these two sets of stories are specifically designed to be equally "old news."

2.3   Asset Pricing Implications

In this subsection, we briefly trace out the empirical implications of the evidence documented in our experiment. 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 reverse in the future. We formally test three empirical predictions.

First, as a baseline, the results in the previous subsection indicate that active finance professionals do perceive new information to be more novel than old news. In particular, novel headlines receive an average novelty score of 4.32 out of 7, significantly higher than the average score of 2.69 that experimental participants assign to stale news (pooling both reprints and recombinations). This yields our first empirical prediction:

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

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

Second, central to the recombination mechanism, market participants perceive recombination headlines to be more novel than reprint headlines of matching length, relative location, and actual staleness level. The average novelty ranking that the active finance professionals assign to recombination stories is 2.91 out of 7, significantly higher than the 2.47 out of 7 assigned to reprints. 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 stories are associated with larger trading volumes and absolute price changes immediately following publication than reprint stories.*

Third, to the extent that reactions to old news reflect a failure to recognize previously released information, we should expect these initial reactions to correct with time. And particularly, the stronger initial reactions to recombination news, reflecting a cognitive limitation in processing of more complex information structures, should be followed by larger subsequent price reversals. We summarize this in the following empirical prediction:

Prediction 3 *Initial reactions to old news are subject to subsequent reversal. In particular, during the 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 stories see more reversal than the initial price moves after reprint stories.*

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

3   Data and Textual Analysis

To test the asset pricing implications of the recombination channel documented in the previous section, we use news data from the Bloomberg terminal. Bloomberg boasts one of the most comprehensive news databases in the world and provides a comprehensive picture of the full landscape of financial media. Descriptive statistics of our sample and the various screens imposed on the data are tabulated in Table 5.

Most news data sources, including the widely used Dow Jones and Reuters newswires,[4] 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 premium 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 stories passing through the Bloomberg terminal on a typical business day in our sample (excluding social media) is on the order of 1 million stories per day, several times larger than other similar services. The Bloomberg news database, by virtue of including a wide breadth of articles practically instantaneously, offers a way to analyze reactions to the media in a much more comprehensive manner than has previously been possible.

We impose a number of conditions on the news used in the analysis, in order to benefit from the breadth of coverage while limiting noise. We start with a sample of approximately 60,000 articles per day retrieved from the Bloomberg news database over the period ranging from January 1, 2000 to December 31, 2014, including non-financial news such as crime, sports, and entertainment.[5] Financial news 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 stories 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 includes roughly 29,500 news articles per day.

---

[4]Although Reuters' and Bloomberg's news operations are similarly sized, Bloomberg dominates the US company news landscape, whereas Reuters is more popular in Europe.

[5]The volume of news stories decays with time. As of January 1, 2014, Bloomberg publishes 0.9-1.2 million stories per day, but does not store the content over time. Web content tends to be deleted first, within 3-6 months, and certain other content is deleted from the system within a few years due to contractual obligations. Table 5 displays the average numbers of stories per day by year.

We limit our attention to articles whose tagged securities are especially relevant, using Bloomberg's relevance tags. The majority of the relevant articles are tagged with more than one security code. Prior work has used indirect proxies for relevance, such as limiting the sample to articles tagged with one or two securities.[6] Bloomberg's explicit relevance markers offer a more direct way to screen relevant stories. 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 example, they may talk about the company's main competitor. Articles with relevance scores around 50% are only tangentially related to the security in question. For example, consider the following three headlines about Apple:

- 95% relevance for AAPL: "Apple announces event on 3/17 to unveil new iPad" (Piper Jaffray, 2/28/2012)

- 70% relevance for AAPL: "Android Grows U.S. Smartphone Market Shr to 50.1%" (*Bloomberg First Word*, 4/3/2012)

- 50% relevance for AAPL: "JCPenney lowered to BB at Standard & Poor's on new strategy" (*Bloomberg Newswire*, 3/7/2012)

Our sample is comprised of all news stories 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 4,000 news articles per day. Each story is linked to an average of 1.30 securities.

3.1 *Old News*

Throughout our empirical analysis, we proxy for old information in news using the extent to which the textual content of each individual news article is spanned by the text of preceding articles about the same security.[7]

To analyze the textual content of the news, for each article s in the sample, we first extract the unique words in the text of the article. We exclude stop words (common words such as "a", "the", "in", "when", etc.) and stem all words using the standard stemming

---

[6]See, for example, Tetlock (2011).

[7]For prior studies of reactions to news events with methodology based on textual analysis or machine learning, see Antweiler and Frank (2006), Das and Chen (2007), Schumaker and Chen (2009), Loughran and McDonald (2011), and Bollen, Mao, and Zeng (2011).

algorithm first developed by Porter (1980) (so that words such as "earned" and "earning" are represented with the same term, "earn-"). We denote by the norm $||\cdot||$ the number of unique terms in a set of articles. Thus, for example, $||s||$ denotes the number of unique terms in article s, while $||s_1 \cap s_2||$ is the number of unique terms appearing in both s_1 and s_2.

We estimate the staleness of a news article's content by its textual similarity to previously published articles. For each story s in the database tagged with company i, we look at all stories s' tagged with company i published no more than three days (72 hours) before the publication of s.[8] We then compute the textual similarity of s to each s', denoted by $Sim(s, s')$, as the percentage of s's unique words that appear also in s'.

The large breadth of the data makes it highly likely that the terms in any given article have appeared somewhere in the preceding articles about the same firm. To screen out this effect, we consider only the five preceding articles that are most textually similar to s. That is, we identify the five articles $\{s'_1(s), ..., s'_5(s)\}$ that span s the most:

$$s'_i(s) = \arg \max_{s' \notin \{s'_j(s)\}_{j<i}} Sim(s, s')$$

The extent to which a news story s contains old information is then defined as:

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

or the percentage of s's unique terms that also appear in the five articles published in the preceding 72 hours that are most similar to s.

Our measure of stale content is very similar to the methodology introduced by Tetlock (2011), with one key innovation. Instead of defining staleness as the average intersection of an incoming article s with the closest previous articles about the same firm, we consider the percentage of s that is spanned by at least one of its closest previous articles. This is meant to differentiate between cases exemplified by the following hypothetical scenarios: (1) 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; (2) 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 stale (50%) in these two cases, despite the fact that it contains new information in

---

[8]We compute our measures only for weekdays to tie the news to market reactions, and because weekend news activity is very low (see Table 5). However, we do not exclude articles published on weekends altogether. Instead, we group all articles published on weekends with those occurring on the following Monday. Thus, on Monday, we compare against all articles published since the previous Wednesday, on Tuesday – all articles since the preceding Thursday; and on Wednesday – all articles since the preceding Friday.

case (1) and nothing new whatsoever in case (2). The metric based on the percentage of terms spanned *overall* seeks to get around this issue.

The distribution of the old content measure is presented in Panel 1 of Figure 3. A large percentage of articles, nearly 40%, are almost entirely (90% or more) spanned by preceding news about the same firm. Another 30% of the sample is evenly distributed between being 60% and 90% spanned. Note that very few stories have staleness levels below 10%, as even unrelated news articles likely share some common words.

We define old news as any article for which at least 60% of its content has appeared in the closest previous articles about the same firm. In particular, we define the following dummy variable:

$$OldNews(s) = \mathbb{1}_{Old(s) \geq 60\%} \tag{2}$$

From Panel 1 of Figure 3, which displays a histogram of $Old(s)$, we can see that roughly 70% of the articles in our sample are classified as old news according to the criterion in (2); the remaining 30% are novel news.

Robustness checks in Section 5.1 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 in (2). We also show that our results are robust to using the continuous measure of old content defined in (1) instead of the binary indicator variable.

3.2   *Reprints and Recombinations*

The experimental evidence in Section 2 identifies a difference in cognitive processing of two types of old information: reprints of previous articles and recombinations that do not contain much novel content but combine information from two or more previous sources.

Figure 4 illustrates this distinction. Each panel of the figure displays a news story s (in grey), and the content of s that was already present in each of the five nearest stories $s'_1(s), ..., s'_5(s)$ (marked in dark blue or red). The bottom row displays $\cup_{i=1}^{5} s'_i(s)$, and likewise marks the content intersecting with s in dark blue. 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, we observe situations where the measure of $Old(s)$ is at 90%. However, the two cases are very different. The top panel of Figure 4 illustrates a reprint: s is almost an exact copy of $s'_1(s)$, marked in red 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 red).

To differentiate reprints from recombinations, we look at the extent to which the content of each story s is spanned by its *single* closest previous neighbor, $s'_1(s)$. In particular, we define the following measure:

$$ClosestNeighbor(s) = \frac{\max_{s'} ||s \cap s'||}{||s||} = \frac{||s \cap s'_1(s)||}{||s||} \tag{3}$$

The bimodal distribution of the $ClosestNeighbor$ measure is tabulated in Panel 2 of Figure 3. There is a large number of stories almost all of whose words have appeared in the single closest neighbor (capturing exact reprints), and a large share of stories approximately half of whose words appeared in the single closest neighbor (potential recombinations). The distribution of the residual content spanned not by the closest neighbor but by the next four closest articles is presented Panel 3. Most stories tend to have relatively little content coming from sources other than the single closest neighbor, but there are some stories with as much as a third of the content incrementally spanned by these secondary sources.

The intersection with the closest previous neighbor is only meaningful when contextualized with the article's overall level of old information. In particular, for two articles with the same value of $OldNews(s)$ – say, at 90% – the one with the higher value of $ClosestNeighbor(s)$ is more likely to be a reprint, and the one with the lower value of $ClosestNeighbor(s)$ is more likely to be a recombination.

Hence, 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 Figure 5. First, following (2), 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. Articles with less than 60% of their text spanned by prior articles are classified as novel. Second, for each article in the old news category, we consider the share of that article's stale content that is spanned by the single closest neighbor. If this share is higher than 80%, then we classify the article as a reprint; if that share is smaller than 80%, then the article is considered to be a recombination. In particular, we define the following two dummy variables to capture reprints and recombinations:

$$Reprint(s) = OldNews(s) \times \mathbb{1}_{ClosestNeighbor(s)/Old(s) \geq 0.8} \tag{4}$$

$$Recombination(s) = OldNews(s) \times \mathbb{1}_{ClosestNeighbor(s)/Old(s) < 0.8} \tag{5}$$

Overall, in our sample, the majority of stories see almost all of their stale content spanned by a single previous story – these are the reprints. But there are also a number of articles

that have as much as a third of their stale content coming from sources other than the single closest neighbor – these recombine information from multiple sources. As can be seen from Panel 4 of Figure 3, among stories classified as old news, approximately 75% are reprints and approximately 25% are recombinations.

Summary statistics of the article characteristics are presented in the first three columns of Table 6. We tabulate, for the full sample and each year in the sample, the average length of the articles in terms of unique terms, the percentage of the article that are classified as old news, and the percentage of the articles that are classified as recombinations. In general, the average length of articles has decreased over the fourteen-year sample. The prevalence of old news, especially direct reprints, has steadily gone up along with the overall increase in news volume documented in Table 5, 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 the individual article-level classifications into firm-level metrics, which are then tied to trading volumes and returns. In particular, for each firm on each date, we consider the percentage of the news articles that are old news and the percentage of the news articles that are recombinations.

Consider any firm i, and let $S_{i,t}$ denote the set of articles published on date t that are tagged with firm i. Then we construct the following firm-date-level measures:

$$PrcOld_{i,t} = \frac{\sum_{s \in S_{i,t}} OldNews(s)}{|S_{i,t}|} \tag{6}$$

$$PrcRecombination_{i,t} = \frac{\sum_{s \in S_{i,t}} Recombination(s)}{|S_{i,t}|} \tag{7}$$

Average values of these two firm-level measures over the full sample and separately for each year in the sample are presented in the last two columns of Table 6.

We take the following approach to screen out the effect of firm-specific news dynamics on firm-level measures of old information and recombinations. For each of the two measures defined in (16)-(17), we take the residuals from daily cross-sectional regressions of the measure on the log of the number of stories for firm i on date t, the log of the average number of unique terms per story, and the square of the log average number of unique terms per story. 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, higher volumes of old news are associated with smaller trading volumes and price responses, but larger shares of recombination stories garner stronger reactions. These reactions to recombination of old information largely reverse in subsequent weeks.

4.1   *Market Reactions to Old News*

We analyze the relationship between market activity – abnormal trading volumes and returns – and news content. First, we confirm that, in general, higher volumes of old news are associated with lower abnormal trading volumes and smaller absolute abnormal returns over the next trading day. Second, holding the overall share of old information constant, a larger proportion of recombination stories is followed by stronger market reactions.

We begin by estimating the association between old news and market reactions. We estimate the following next-day abnormal return and volume regressions:

$$|AbnRet|_{i,t+1} = a + bAbnPrcOld_{i,t} + gX_{i,t} + e_{i,t} \tag{8}$$

$$AbnVol_{i,t+1} = \alpha + \beta AbnPrcOld_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \tag{9}$$

where a, b, α and β are single coefficients, and g and γ denote 1×9 vectors of coefficients. $AbnRet_{i,t+1}$ denotes the abnormal return for firm i on date $t+1$, calculated as the difference between firm i's return on date $t+1$ and the return on a value-weighted index of all firms in our universe on date $t+1$. $AbnVol_{i,t+1}$ designates abnormal trading volume for firm i on date $t+1$, defined as the difference between the fraction of shares turned over for firm i on date $t+1$ and the value-weighted average of the fraction of shares turned over for all firms in our sample. $X_{i,t}$ is a set of controls, which includes the following variables:

- $Stories_{i,t}$ is the number of news articles published on date t that are tagged with firm i (with a relevance of at least 70%).

- $AbnStories_{i,[t-5,t-1]}$ is abnormal volume of news for firm i over the previous week, defined as the difference between the average number of stories over $[t-5, t-1]$ and the average number of stories over the preceding three months, $[t-60, t-6]$.

- $Terms_{i,t}$ is the average number of unique terms in stories published on date t and tagged with firm i.

- $MCap_{i,t}$ is log market capitalization of firm i as of market open on date t.

- $BM_{i,t}$ is the ratio of firm i's book value as of the latest quarterly earnings report preceding date t to its market capitalization as of market open on date t.

- $AbnRet_{i,[t-5,t-1]}$ denotes cumulative abnormal returns for firm i over the preceding five business days.

- $AbnVol_{i,[t-5,t-1]}$ is the average abnormal trading volume for firm i over the preceding five business days.

- $AbnVolatility_{i,[t-5,t-1]}$ is abnormal volatility computed as the difference between firm i's volatility and the value-weighted average volatility of all firms in our sample on date t, averaged over the preceding five business days.

- $Illiq_{i,[t-5,t-1]}$ is the log of the illiquidity measure from Amihud (2002), computed as the prior-week average of $10^6 \times |Ret_{i,t}|/Volume_{i,t}$, where $Volume_{i,t}$ is firm i's dollar volume on date t.

We estimate (8) and (9) using daily cross-sectional Fama and MacBeth (1973) regressions. Standard errors robust to heteroskedasticity and five days of autocorrelation are computed following Newey and West (1987). Note that the computation of $AbnStories_{i,[t-5,t-1]}$ requires three months of preceding data. As a result, we begin computing our dependent variables, key measures, and control variables on April 1, 2000, leaving January 1 - March 31, 2000 as a buffer for the computation.

The estimates of coefficients b and β from specifications (8) and (9), scaled to correspond to the effect of a 10% increase in $AbnPrcOld_{i,t}$, are tabulated in Table 7, columns (1) and (3). The estimated coefficients indicate that an additional 10% of the news about firm i being stale corresponds to an 8 basis points smaller absolute abnormal return on the next trading day and 0.02% lower abnormal trading volume. This implies that, on average, going from a piece of news in the lowest quartile based on old content to a piece of news in the top quartile based on old content corresponds to approximately 17 basis points lower absolute abnormal return and 0.04% lower trading volume on the next trading day. The effects are both economically meaningful and statistically significant at the 1% level, confirming that the market reacts less strongly when news is more stale (Prediction 1 stemming from the experimental evidence).

Having established the expected relationship between overall old news and market reactions, we now turn to the key prediction of our proposed mechanism: that the market reacts to old news more strongly when the old news recombines information from multiple sources than when the old news consists of direct reprints. In order to test for the differential response to recombinations and reprints, we estimate the relationship between absolute abnormal trading volumes and returns and the share of recombination news, controlling for

the overall share of old news. To this effect, we estimate the following two specifications:

$$|AbnRet|_{i,t+1} = a + b_1 AbnPrcOld_{i,t} + b_2 AbnPrcRecombination_{i,t} + gX_{i,t} + e_{i,t} \qquad (10)$$

$$AbnVol_{i,t+1} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \qquad (11)$$

where $X_{i,t}$ is the same set of controls as we included in (8) and (9), $a, b_1, b_2, \alpha, \beta_1$, and β_2 are single coefficients, and g and γ denote 1×9 vectors of coefficients.

The key Prediction 2 states that, holding $AbnPrcOld_{i,t}$ constant, there is a larger market reaction for firms whose news stories score higher on the $AbnPrcRecombination_{i,t}$ measure – i.e., firms whose old news consists mostly of recombination stories, rather than reprints that directly duplicate single pre-existing articles. As a result, we expect to see a positive coefficient on $AbnPrcRecombination_{i,t}$ in specifications (10) and (11). The effect of old news in general, as before, is posited to be negative – firms with more stale content are expected to see smaller trading volumes and absolute price changes. Hence, the coefficient on $AbnPrcOld_{i,t}$ should remain negative.

The results from estimating specifications (10) and (11) using daily cross-sectional Fama-MacBeth regressions are reported in Columns (2) and (4) of Table 7. The estimated coefficients on $AbnPrcOld_{i,t}$, b_1 and β_1, are similar to the estimates of coefficients b and β that we observed in specifications (8) and (9), in terms of both magnitude and statistical significance, indicating that firms with higher shares of old news tend to have smaller absolute abnormal returns and abnormal trading volumes.

Importantly, the share of recombination stories, captured by $AbnPrcRecombination_{i,t}$, has a positive relationship with absolute abnormal returns and abnormal trading volumes, indicating larger market reactions to non-novel news content that recombines previously available information rather than directly reprinting a single source. Holding all else – including the overall share of old news, $AbnPrcOld_{i,t}$ – equal, an additional 10% of a firm's old news consisting of recombination stories (as opposed to reprints) corresponds to an additional 13 basis points absolute abnormal return and additional 0.03% abnormal trading volume on the following business day. This translates to firms in the top quartile in terms of the share of recombination news experiencing 14 basis points larger price moves and 0.03% larger trading volumes on the next day. The estimates of the coefficients on $AbnPrcRecombination_{i,t}$ are highly statistically significant for both outcome variables.

The results thus far confirm the empirical predictions of the cognitive limitations documented in the experiment in Section 2. In general, readers identify old news – both reprints and recombination – as less novel than new news. However, this is not done perfectly. In particular, finance professionals are susceptible to mistaking recombination news articles for

novel, more so than the less complex reprints. As a result, although market participants generally react less to old news than to new news, the coefficients on $AbnPrcRecombination_{i,t}$ indicate that the market does much better at identifying reprints than recombinations.

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 constitute overreactions? The mechanism documented in Section 2 predicts that the observed reactions are a consequence of imperfect processing of the more complex information structures in recombination news: since these news stories combine old information from several sources, it is more difficult for investors to identify that the reported information is not novel. In this subsection, we rule out the following alternative: finance professionals have no cognitive limitation in identifying that recombinations contain old news, and recombination stories prompt larger market reactions purely by serving the valuable function of facilitating information processing through the juxtaposition of pre-existing but inadequately noted information. Under this alternative, reactions to recombination news would constitute corrections of prior underreactions to the constituent information, rather than overreaction to previously seen old news.

In order to differentiate between these two narratives, we consider the dynamics of return reversals. If the observed reactions to recombination news reflect overreactions driven by finance professionals mistaking recombinations to be more new than they actually are, then we expect these reactions to subsequently reverse (Prediction 3). If, however, the reactions to recombination news come purely from corrections of prior underreactions, then we should observe no subsequent reversal. To this effect, we estimate the following specification:

$$AbnRet_{i,[t+t_1,t+t_2]} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcOld_{i,t} \times AbnRet_{i,t+1} + \beta_3 AbnRet_{i,t+1} +$$

$$\delta_1 AbnPrcRecombination_{i,t} + \delta_2 AbnPrcRecombination_{i,t} \times AbnRet_{i,t+1} + \gamma X_{i,t} + \epsilon_{i,t}, \quad (12)$$

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 $AbnRet_{i,[t,t+1]}$ is the signed abnormal return for firm i over the trading day immediately following the news measures. We consider several delayed windows: $(t_1, t_2) \in \{(2, 4), (2, 6), (2, 11)\}$. These windows capture subsequent returns during the three days, one week, and two weeks following the initial $t + 1$ return. We include the same set of controls as in Section 4.1: $Stories_{i,t}$, $AbnStories_{i,[t-5,t-1]}$, $Terms_{i,t}$, $MCap_{i,t}$, $BM_{i,t}$, $AbnRet_{i,[t-5,t-1]}$, $AbnVol_{i,[t-5,t-1]}$, $AbnVolitility_{i,[t-5,t-1]}$, and $Illiq_{i,[t-5,t-1]}$.

Specification (12) assesses the two portions of Prediction 3: (a) the extent to which market reactions following stale content in general tend to reverse in subsequent trading; and (b) 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 following larger shares of old news relative to the returns following mostly novel news. 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 3(b) of our proposed mechanism 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 coefficient estimates, tabulated in Table 8, support our empirical predictions. The coefficients on both interaction terms, $AbnPrcOld_{it,} \times AbnRet_{i,t+1}$ and $AbnPrcRecombination_{i,t} \times AbnRet_{i,t+1}$, are negative and statistically significant in all of the considered specifications. For example, a comparison of the coefficients on $AbnPrcOld_{it,} \times AbnRet_{i,t+1}$ and $AbnRet_{i,t+1}$ in column (2) suggests that when an additional 10% of a firm's news on a given day is non-novel, then that firm's next business day return is subject to a twice larger reversal over the following week. Thus, in general, market reactions to old information are subject to more reversal than reactions driven by novel news.

Across all specifications, the estimated coefficients on the incremental reversal following recombination news, $AbnPrcrecombination_{i,t} \times AbnRet_{i,t+1}$, are consistently negative. In fact, they are both economically larger and more statistically significant than those on $AbnPrcOld_{it,} \times AbnRet_{i,t+1}$. Combining these results with the findings in Section 4.1, we see that reactions to recombination news are not only stronger than those to direct reprints, but also more likely to reverse over subsequent weeks. This traces out the empirical implications of the initial reactions to recombination news reflecting frictions in information processing. The results support Prediction 3 over alternative narratives regarding corrections of prior underreactions.

4.3 Time Series of Effects

In this subsection, we address the following question: has the presence of increasingly more sophisticated arbitrageurs attenuated reactions to recombination news? If so, then the observed effects should be driven by the earlier data in our sample, and we should observe a weakening of the results over time. Evidence in Section 2 pushes against this conjecture: in a controlled experimental setting, active finance professionals remain susceptible to old

news in recombination form. We now investigate whether this experimentally documented susceptibility translates to continued market effects.

Availing ourselves of a long time period spanning 2000-2014, we investigate the dynamics of the estimated effects of the $AbnPrcOld_{i,t}$ and $AbnPrcRecombination_{i,t}$ variables by computing a time series of the estimated coefficients. To do so, we re-estimate regression specification (10) separately for each full year in our sample. Thus, we run the following daily cross-sectional Fama-MacBeth regression separately for each $T \in [2001, 2014]$[9]:

$$\forall T \in [2001, 2014], t \in T : |AbnRet|_{i,t+1} = \alpha_T + \beta_{1,T} AbnPrcOld_{i,t}+ \quad\quad (13)$$

$$\beta_{2,T} AbnPrcRecombination_{i,t} + \gamma_T X_{i,t} + \epsilon_{i,t}$$

where $X_{i,t}$ denotes the same vector of controls as in the full-sample specification (10); $\alpha_T, \beta_{1,T}$, and $\beta_{2,T}$ are individual coefficients estimated for year T, and each γ_T is a 1×9 vector of coefficients on the controls.

We plot the estimated coefficients $\beta_{1,T}$ on $AbnPrcOld_{i,t}$ and $\beta_{2,T}$ on $AbnPrcRecombination_{i,t}$ for each year $T \in [2001, 2014]$ in Figure 6. The estimated coefficients are displayed in solid line and accompanied by the 95% confidence intervals in dashed lines.

Market reactions to old news in general, compared to new news, decline from 2001 to 2014, as can be seen in Panel 1 of Figure 6. The annual estimates of $\{\hat{\beta}_{1,T}\}_{T \in [2001,2014]}$ are negative throughout the sample except for 2002, and statistically significant at the 5% level for nine out of the fourteen years. There is a downward trend in the time series of the estimated coefficients through our fourteen-year period. The negative coefficients increase in magnitude and become more statistically significant towards the end of the sample period. For example, in 2001, an additional 10% of a firm's news on a given day consisting predominantly of stale content corresponds to a statistically insignificant 6 basis points smaller absolute abnormal return over the next trading day. By contrast, in 2014, the effect is a precisely estimated 9 basis points.

While overall market reactions to old news lessen over time, the differential reaction to recombination of information, as compared to reprints, increases. Panel 2 of Figure 6 plots the time series of the estimated coefficients $\{\hat{\beta}_{2,T}\}_{T \in [2001,2014]}$. As can be seen from the graph, the relationship between the $AbnPrcRecombination$ measure and the following business day's absolute abnormal return (controlling for overall old news, $AbnPrcOld$) is positive and significant at the 5% level in eleven out of the fourteen years in the sample. The exception years (2001, 2003, and 2004) all fall at the beginning of the sample. The

___

[9]We exclude the year 2000, since we have only partial data in that year, computing the relevant metrics and control variables starting on April 1, 2000, due to the fact that the control variable $AbnStories[t-5, t-1]$ requires three months of preceding data.

positive relationship is substantially stronger in the second half of the sample. While the effect is statistically indistinguishable from 0 in 2001, in 2014, an additional 10% of a firm's news featuring recombination stories (as opposed to simple reprints) translates to 18 basis points larger absolute abnormal returns on the following trading day, strongly statistically significant.

These results suggest that the differential market reaction to recombination of old information is not a vestige of the earlier years in the sample. On the contrary, the recombination effect has strengthened over time. The larger market reactions to recombinations compared to reprints are strong and significant through the fourteen years in our sample, and especially prominent in the most recent five years. The mechanism documented in Section 2 is as active in driving market reactions now as a decade ago.

Taken together, the time series results paint the following picture. The overall effect of old news on stock returns has declined over time. Market participants appear to be getting progressively better at identifying old content and not reacting to it. However, this is driven predominantly by decreased reactions to simple reprints: the differential reactions to recombinations, as compared to reprints, have risen markedly in the second half of the sample. Thus, 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   Robustness Checks

In this section, we confirm that our results are robust to alternative definitions of old news and recombination of information. We vary discretionary thresholds in the construction of *AbnPrcOld* and *AbnPrcRecombination* variables, consider continuous measures of old and recombination content instead of discrete classification of each article, and restrict our sample in order to screen out any differences in residual novel content between reprints and recombinations.

5.1   *Alternative Constructions*

When constructing our measures of old news and recombination of information in Section 3, for any given article, we consider the five most similar articles about the same firm over the preceding three business days. We now test the robustness of our results to the discretionary choices in this specification. In particular, we consider the following three dimenstions:

- **Look-back window.** We alter the number of days over which we search for previous articles. For each article s about firm i, we try looking for the textually most similar n articles about firm i in the past τ days, where $\tau \in \{3\,days, 5\,days, 10\,days\}$.

- **Number of closest neighbors.** When searching for the closest articles about the same firm in the previous τ days, we allow the number n of the most similar articles considered in the computation of $Old(s)$ to be either 5 stories or 10 stories: $n \in \{5, 10\}$.

- **Impose Minimum Limit.** We also check that our results are robust to limiting the sample of firms on each date to only those firms that actually have at least n stories in the preceding τ days. To fix ideas, consider the case with $\tau = 3$ days and $n = 5$. Without the minimum limit, if at the time of publication of article s about firm i, there are only 4 preceding articles tagged with i during the prior 3 days, then the measure of old content in s, $Old(s)$, is computed using those 4 stories. With the minimum limit, the article s is simply dropped from the sample. Letting L denote whether or not we impose the limit (Y or "Yes" for imposing the limit, and N or "No" otherwise), we repeat the analysis for $L \in \{Y, N\}$.

Allowing τ to vary over $\{3\,days, 5\,days, 10\,days\}$, n to be in $\{5\,stories, 10\,stories\}$, and L to be either Y or N yields 12 distinct constructions of the $Old(s)$ measure. The $ClosestNeighbor(s)$ and $Recombination(s)$ metrics are then computed as before, from (3) and (5). Firm-level shares $PrcOld(s)$ and $PrcRecombination(s)$ are recalculated according to (6) and (7). Abnormal measures are obtained by taking residuals from cross-sectional daily regressions against firm-level article volume, log of the firm-level average story length, and the square of the log of the average story length.

We reestimate the full-sample regressions of the next business day absolute abnormal returns and abnormal trading volumes against the measures of old and recombination news from each of the 12 constructions. That is, we estimate the following specifications, which correspond to (10) and (11), respectively, for each of the 12 versions of $AbnPrcOld_{i,t}^k$ and $AbnPrcRecombination_{i,t}^k$ (where $k \in \{1, ..., 12\}$ indexes the variable constructions):

$$|AbnRet|_{i,t+1} = a^k + b_1^k AbnPrcOld_{i,t}^k + b_2^k AbnPrcRecombination_{i,t}^k + g^k X_{i,t} + e_{i,t}^k, \quad (14)$$

$$AbnVol_{i,t+1} = \alpha^k + \beta_1^k AbnPrcOld_{i,t}^k + \beta_2^k AbnPrcRecombination_{i,t}^k + \gamma^k X_{i,t} + \epsilon_{i,t}^k, \quad (15)$$

where the vector $X_{i,t}$ includes the standard controls: $Stories_{i,t}$, $AbnStories_{i,[t-5,t-1]}$, $Terms_{i,t}$, $MCap_{i,t}$, $BM_{i,t}$, $AbnRet_{i,[t-5,t-1]}$, $AbnVol_{i,[t-5,t-1]}$, $AbnVolatility_{i,[t-5,t-1]}$, and $Illiq_{i,[t-5,t-1]}$.

The results are consistent across the various constructions, as can be seen from Table 9, which displays the estimated coefficients $\hat{b}_1^k, \hat{b}_2^k$ (in Panel 1) and $\hat{\beta}_1^k$ and $\hat{\beta}_2^k$ (in Panel 2).

The estimates of the effect of overall old news, captured by the coefficients $\hat{b}_1^k$ and $\hat{\beta}_1^k$ on $AbnPrcOld_{i,t}^k$, are negative across construction methods and statistically significant at the 1% level in all but four instances. The differential reactions to recombination of information, captured by the estimated coefficients on $AbnPrcRecombination_{i,t}^k$, are significant at the 1% level for every single construction method. The economic estimates are also robust to specification, with an additional 10% of the news about a given firm on a given day being recombinations rather than reprints – holding all else, including the overall share of old news, equal – corresponding to between 6 basis points and 21 basis points higher absolute abnormal return on the following trading day. The results using the baseline specification of $\tau = 3\ days$, $n = 5\ articles$ and exclusion flag L set to N (marked in blue) are roughly in the middle of the range of results observed in Table 6.

Next, we repeat the main return reversal analysis using the alternative constructions of $AbnPrcOld_{i,t}^k$ and $AbnPrcRecombination_{i,t}^k$. In particular, for each $k \in \{1, ..., 12\}$, we estimate:

$$AbnRet_{i,[t+3,t+6]} = \alpha^k + \beta_1^k AbnPrcOld_{i,t}^k + \beta_2^k AbnPrcOld_{i,t}^k \times AbnRet_{i,t+1} + \beta_3^k AbnRet_{i,t+1} +$$

$$\delta_1^k AbnPrcRecombination_{i,t}^k + \delta_2^k AbnPrcRecombination_{i,t}^k \times AbnRet_{i,t+1} + \gamma^k X_{i,t} + \epsilon_{i,t}^k,$$

where the vector of controls $X_{i,t}$ again includes the standard set of controls.

The results are displayed in Table 10. Here, just as in Table 9, the coefficient estimates are consistent across specifications, and the baseline estimates (highlighted) are roughly in the middle of the observed range of results. Overall, the results from the main specifications considered in Section 4 are robust to different methods of measuring old content and recombination of information in news.

5.2 *Continuous Measures of Old News and Reprints*

The methodology thus far classifies each article as novel news, a reprint, or a recombination. In this subsection, we confirm that our results are robust to an alternative approach that assigns each article s continuous measures of the oldness and recombination of its textual content.

For each firm i on date t, we compute the firm-level oldness and recombination measures, $ExtentOld_{i,t}$ and $ExtentRecombination_{i,t}$, as follows. We average the respective measures for each article in our sample that is published on date t and tagged with firm i:

$$ExtentOld_{i,t} = \frac{1}{|S_{i,t}|} \sum_{s \in S_{i,t}} Old(s) \tag{16}$$

$$ExtentRecombination_{i,t} = \frac{1}{|S_{i,t}|} \sum_{s \in S_{i,t}} (Old(s) - ClosestNeighbor(s)), \qquad (17)$$

where $S_{i,t}$ denotes the set of stories on date t tagged with firm i, and $|S_{i,t}|$ is the number of elements in $S_{i,t}$. The firm-level measure of news oldness is simply the average, across articles about that firm on the given date, of the fraction of content that has already been seen in prior news about the same firm. The firm-level measure of news recombination is the average of article-level residual old content, excluding the content spanned by the single closest neighbor.

For each article, the recombination measure the continuous percentage of the content in that article that is old but not covered by the single closest neighbor. In Figure 5, Panel 1 displays an article with a closest neighbor measure of 90% and a recombination measure of 0%, while the article portrayed in Panel 2 has a closest neighbor measure of 50% and a recombination measure of 40%. The identification in the empirical tests in this subsection comes from observing differential reactions to variations in the continuous firm-level recombination variable, conditional on the overall level of oldness of news.

As before, we adjust the firm-level proxies to screen out the effect of news volume and length by taking residuals from daily cross-sectional regressions of $ExtentOld_{i,t}$ and $ExtentRecombination_{i,t}$ 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 be mean zero and standard deviation one. Thus, all quoted effects correspond to a one standard deviation change in the explanatory variables. We denote these orthogonalized variables by $AbnExtentOld_{i,t}$ and $AbnExtentRecombination_{i,t}$.

We repeat our analysis using $AbnExtentOld_{i,t}$ and $AbnExtentRecombination_{i,t}$ instead of $AbnPrcOld_{i,t}$ and $AbnPrcRecombination_{i,t}$, respectively. To assess initial reactions, we estimate:

$$|AbnRet|_{i,t+1} = a + b_1 AbnExtentOld_{i,t} + b_2 AbnExtentRecombination_{i,t} + gX_{i,t} + e_{i,t} \quad (18)$$

$$AbnVol_{i,t+1} = \alpha + \beta_1 AbnExtentOld_{i,t} + \beta_2 AbnExtentRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \quad (19)$$

The results, displayed in Panel 1 of Table 11, are consistent with those in Table 7. A one standard deviation increase in the continuous measure of abnormal old content corresponds to 6 basis points smaller absolute abnormal returns and 0.02% lower trading volumes over the next trading day. However, a one standard deviation increase in the continuous measure of abnormal recombination of information increases the next-day absolute abnormal return by 11 basis points and the next-day trading volume by 0.02%.

27

To evaluate return reversals, we estimate:

$$AbnRet_{i,[t+3,t+6]} = \alpha + \beta_1 AbnExtentOld_{i,t} + \beta_2 AbnExtentOld_{i,t} \times AbnRet_{i,t+1} + \beta_3 AbnRet_{i,t+1} +$$

$$\delta_1 AbnExtentRecombination_{i,t} + \delta_2 AbnExtentRecombination_{i,t} \times AbnRet_{i,t+1} + \gamma X_{i,t} + \epsilon_{i,t}$$

$$(20)$$

The results, which are tabulated in Panel 2 of Table 11, are qualitatively similar to those in Table 8. Overall, when firm i's news on date t contains relatively more old content, immediate returns on $t+1$ are smaller in magnitude and more likely to reverse over the following one to two weeks. When firm i's old news on date t is less spanned by single preceding neighbors (i.e., includes more recombination of information), the absolute abnormal return on date $t+1$ is larger and even more likely to reverse over the coming weeks.

5.3   *Isolating Stale Content*

Our last robustness check is aimed to rule out 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, rather than the cognitive limitations identified in Section 2.

Recall that 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 stories with equal amounts of old content, one a recombination and the other a reprint, it could be the case that the recombination story is simply more likely to contain important information in the portion of its text that is novel.

To fix ideas, consider the following two stories, where blue denotes preciously seen text:

Story 1: | 70% | ? |

Story 2: | 40% | 30% | ? |
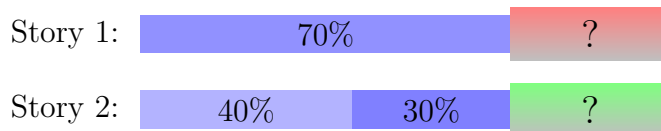
For both stories, 70% of their text has been seen in the closest prior stories about the same firm. Story 1 is a reprint: all 70% of its old content comes from a single previous source. Story 2 is a recombination: 40% of its old content comes from one previous source, and 30% – from another. Both of these articles also include 30% novel content. The quality of this novel content differs from article to article, and we wish to rule out the possibility that the content differs systematically between reprints and recombinations. For example, suppose that there is a selection bias in reporters, where the more competent reporters choose to

publish recombination stories akin to Story 2, while the less competent reporters produce reprints akin to Story 1. Then the novel content in Story 2 (marked in green) will contain more valuable information than the novel content in Story 1 (marked in red), prompting a stronger market reaction to Story 2 (the recombination) than to Story 1 (the reprint).

We address this concern in two ways. First, the return reversal results in Section 4.2 are inconsistent with the narrative of the reactions to recombinations arising purely from higher quality of novel content. If the larger reactions to recombination news were driven by the superior novel content in such news, then we would expect to see no reversal of these reactions. The fact that we do observe larger reversals of returns following larger amounts of recombination news indicates that our results reflect overreaction and therefore cannot be driven purely by recombination stories' superior novel content.

Second, as an additional check that the differences between recombinations and reprints are not due to differences in the stories' residual novel content, we repeat our main tests using only those stale news stories that have next to no novel content. In particular, we reclassify articles into novel news, reprints, and recombinations using an old content threshold of 90%. That is, we only classify a story as old news (either a reprint or a recombination) if at least 90% of its text is spanned by the most similar preceding stories about the same firm. This ensures that both reprints and recombinations contain practically no novel content.

Each article s is classified as follows:

$$OldNews^{90\%}(s) = \mathbb{1}_{Old(s) \geq 90\%} \tag{21}$$

$$Recombination^{90\%}(s) = OldNews^{90\%}(s) \times \mathbb{1}_{ClosestNeighbor(s)/Old(s) < 0.8}, \tag{22}$$

where $Old(s)$ and $ClosestNeighbor(s)$ are computed according to (1) and (5), respectively.

The firm-level shares are aggregated as before, according to (6) and (7), but with the $OldNews^{90\%}(s)$ and $Recombination^{90\%}(s)$ dummies instead of $OldNews(s)$ and $Recombination(s)$, respectively. The firm-level variables are orthogonalized to firm-level measures of news volume, log average news length, and the square of log average news length. We denote the resulting abnormal percentages of overall old news and recombination news for firm i on date t by $AbnPrcOld_{i,t}^{90\%}$ and $AbnPrcRecombination_{i,t}^{90\%}$, respectively.

We then estimate the following specifications:

$$|AbnRet|_{i,t+1} = a + b_1 AbnPrcOld_{i,t}^{90\%} + b_2 AbnPrcRecombination_{i,t}^{90\%} + gX_{i,t} + e_{i,t} \tag{23}$$

$$AbnVol_{i,t+1} = \alpha + \beta_1 AbnPrcOld_{i,t}^{90\%} + \beta_2 AbnPrcRecombination_{i,t}^{90\%} + \gamma X_{i,t} + \epsilon_{i,t} \tag{24}$$

where $X_{i,t}$ is the usual set of controls.

The results are presented in Panel 1 of Table 12. The estimated coefficients are similar in size and significance to their counterparts in Columns (2) and (4) of Table 7. The estimate of the coefficient on old content is somewhat larger when we use $AbnPrcOld_{i,t}^{90\%}$ instead of $AbnPrcOld_{i,t}^{90\%}$, and the estimate of the recombination effect is marginally smaller when we use $AbnPrcRecombination_{i,t}^{90\%}$ instead of $AbnPrcRecombination_{i,t}$. Still, even in this substantially restricted set of old news, an additional 10% of stories being recombinations rather than reprints translates into an additional 12 basis points larger absolute abnormal returns and 0.03% larger abnormal trading volumes over the next trading day.

We also reestimate the return reversal regression using the $AbnPrcOld_{i,t}^{90\%}$ and $AbnPrcRe$-$combination_{i,t}^{90\%}$ proxies:

$$AbnRet_{i,[t+3,t+6]} = \alpha + \beta_1 AbnPrcOld_{i,t}^{90\%} + \beta_2 AbnPrcOld_{i,t}^{90\%} \times AbnRet_{i,t+1} + \beta_3 AbnRet_{i,t+1} +$$

$$\delta_1 AbnPrcRecombination_{i,t}^{90\%} + \delta_2 AbnPrcRecombination_{i,t}^{90\%} \times AbnRet_{i,t+1} + \gamma X_{i,t} + \epsilon_{i,t} \quad (25)$$

The estimated effects are similar to those catalogued in Section 4.2, as can be seen from Panel 2 of Table 12. Overall, the magnitudes and statistical significance of the coefficients in both the initial reactions and the subsequent reversals are consistent with our prior tests. The findings remain qualitatively similar when we restrict our attention to news articles that have practically no novel content, indicating that differences in residual novel content are not driving our empirical results.

6   Conclusion

The present paper investigates why market participants fall prey to old news. We conjecture that limited attention and neglect of correlations across media sources make readers susceptible to certain forms of repetition of stale content in financial news. While news stories that directly reprint information from previous articles are easily identifiable as old news, stories that combine information from several different sources are more difficult to distinguish from novel information. We document this mechanism experimentally. Finance professionals employed at a variety of institutions ranging from large banks such as Goldman Sachs to hedge funds such as Two Sigma systematically misperceive recombination of old information from multiple sources as more novel than reprints of information from a single source.

We test the asset pricing implications of this "recombination effect" using a uniquely comprehensive database of news passing through the Bloomberg terminal. While a higher share of news about a firm on a given date being old corresponds to lower absolute abnormal

returns and trading volumes on the following trading day, market reactions are significantly larger when a higher percentage of the old news consists of recombination stories rather than reprints. Holding all else equal, including the overall amount of old news, an additional 10% of a firm's news on a given day being recombinations rather than reprints corresponds to roughly 13 basis points higher absolute abnormal returns on the next business day. Consistent with the idea that market reactions to old news in general, and to recombination of information in particular, reflect imperfect information processing, we observe subsequent return reversals. Larger shares of recombination news correspond not only to larger next day returns, but also larger reversals of those returns over the following one to two weeks. The time series of the estimated coefficients indicates that the recombination effect has strengthened over time.

We interpret these results as evidence 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 of cognitive biases and limitations and how they contribute to market stability and efficiency would constitute a productive direction for future work.

References

[1] Ahern, K. R., and D. Sosyura (2014). "Who writes the news? Corporate press releases during merger negotiations." *The Journal of Finance*, 69.1: 241-291.

[2] Amihud, Y. (2002). "Illiquidity and stock returns cross-section and time-series effects." *Journal of Financial Markets*, 5.1: 31-56.

[3] Antweiler, W., and M. Z. Frank (2006). "Do US stock markets typically overreact to corporate news stories," Working paper, University of British Columbia.

[4] Barber, B. M., and T. Odean (2008). "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors." *Review of Financial Studies*, 21.2: 785-818.

[5] Barber, B. M., and D. Loeffler (1993). "The 'Dartboard' column: Second-hand information and price pressure." *Journal of Financial and Quantitative Analysis*, 28.2: 273-284.

[6] Ben-Rephael, A., Z. Da, and R. D. Israelsen (2017). "Depends on where you search: A comparison of institutional and retail attention." *Review of Financial Studies*, forthcoming.

[7] Bollen, J., H. Mao, and X. Zeng (2011). "Twitter mood predicts the stock market." *Journal of Computational Science*, 2.1: 1-8.

[8] Busse, J. A., and T. C. Green (2002). "Market efficiency in real time." *Journal of Financial Economics*, 65.3: 415-437.

[9] Carvalho, C., N. Klagge, and E. Moench (2011). "The persistent effects of a false news shock." *Journal of Empirical Finance*, 18(4): 597-615.

[10] Chan, W. (2003). "Stock price reaction to news and no-news: drift and reversal after head- lines." *Journal of Financial Economics*, 70: 223-260.

[11] Da, Z., J. Engelberg, and P. Gao (2011). "In search of attention." *The Journal of Finance*, 66.5: 1461-1499.

[12] Das, S. R., and M. Y. Chen (2007). "Yahoo! for Amazon: Sentiment extraction from small talk on the web." *Management Science*, 53.9: 1375-1388.

[13] DellaVigna, S., and J. M. Pollet (2009). "Investor inattention and Friday earnings announcements." *The Journal of Finance*, 64.2: 709-749.

[14] Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons (2012). "Journalists and the stock market." *Review of Financial Studies*, 25.3: 639-679.

[15] Drake, M.S., D. T. Roulstone, and J. R. Thornock (2016). "The usefulness of historical accounting reports." *Journal of Accounting and Economics,* 61.2: 448-464.

[16] Engelberg, J. E. (2008). "Costly information processing: Evidence from earnings announcements." Working paper, University of California San Diego.

[17] Engelberg, J. E., and C. A. Parsons (2011). "The causal impact of media in financial markets." *The Journal of Finance,* 66.1: 67-97.

[18] Enke, B., and F. Zimmermann (2017). "Correlation neglect in belief formation." *Review of Economic Studies,* forthcoming.

[19] Fedyk, A. (2017). "Front Page News: The Effect of News Positioning on Financial Markets." Working paper, Harvard University.

[20] Fang, L., and J. Peress (2009). "Media Coverage and the Cross-section of Stock Returns." *The Journal of Finance*, 64.5: 2023-2052.

[21] Fama, E. F., and J. D. MacBeth (1973). "Risk, return, and equilibrium: Empirical tests." *The Journal of Political Economy*, 81.3: 607-636.

[22] Gilbert, T., S. Kogan, L. Lochstoer, and A. Ozyildirim (2012). "Investor inattention and the market impact of summary statistics." *Management Science*, 58.2: 336-350.

[23] Hillert, A., H. Jacobs, and S. Müller (2014). "Media makes momentum." *Review of Financial Studies*, 27.12: 3467-3501.

[24] Hirshleifer, D., S. S. Lim, and S.H. Teoh (2009). "Driven to distraction: Extraneous events and underreaction to earnings news." *The Journal of Finance*, 64.5: 2289-2325.

[25] Huberman, G., and T. Regev (2001). "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar." *The Journal of Finance*, 56.1: 387-396.

[26] Klibanoff, P., O. Lamont, and T. A. Wizman (1998). "Investor Reaction to Salient News in Closed-End Country Funds." *The Journal of Finance*, 53.2: 673-699.

[27] Liu, L., A. Sherman, and Y. Zhang (2014). "The long-run role of the media: Evidence from initial public offerings." *Management Science*, 60.8: 1945-1964.

[28] Loughran, T., and B. McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance*, 66.1: 35-65.

[29] Newey, W. K., and K. D. West (1987). "A simple positive semi-definite, heteroskedasticity and autocorrelatio nconsistent covariance matrix." *Econometrica*, 55.3: 703-708.

[30] Porter, M. F. (1980). "An algorithm for suffix stripping." *Program: Electronic Library and Information Systems*, 14.3: 130-137.

[31] Solomon, D. H. (2012). "Selective publicity and stock prices." *The Journal of Finance,* 67.2: 599-638.

[32] Schumaker, R. P., and H. Chen (2009). "Textual analysis of stock market prediction using breaking financial news: The AZFin text system." *ACM Transactions on Information Systems (TOIS)*, 27.2: 12-19.

[33] Tetlock, P. C. (2007). "Giving content to investor sentiment: The role of media in the stock market." *The Journal of Finance*, 62.3: 1139-1168.

[34] Tetlock, P. C. (2011). "All the news that's fit to reprint: Do investors react to stale information?" *Review of Financial Studies*, 24.5: 1481-1512.

How New is the Informatiom in this Headline?

6 seconds remaining

Climate Deal Threatened by Penguin Couple

World News

Rate this article...
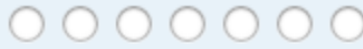
(Nothing New) ○ ○ ○ ○ ○ ○ ○ (Completely New)

Figure 1: Screenshot of the experimental screen, asking participating finance professionals to mark novelty of a news headline.
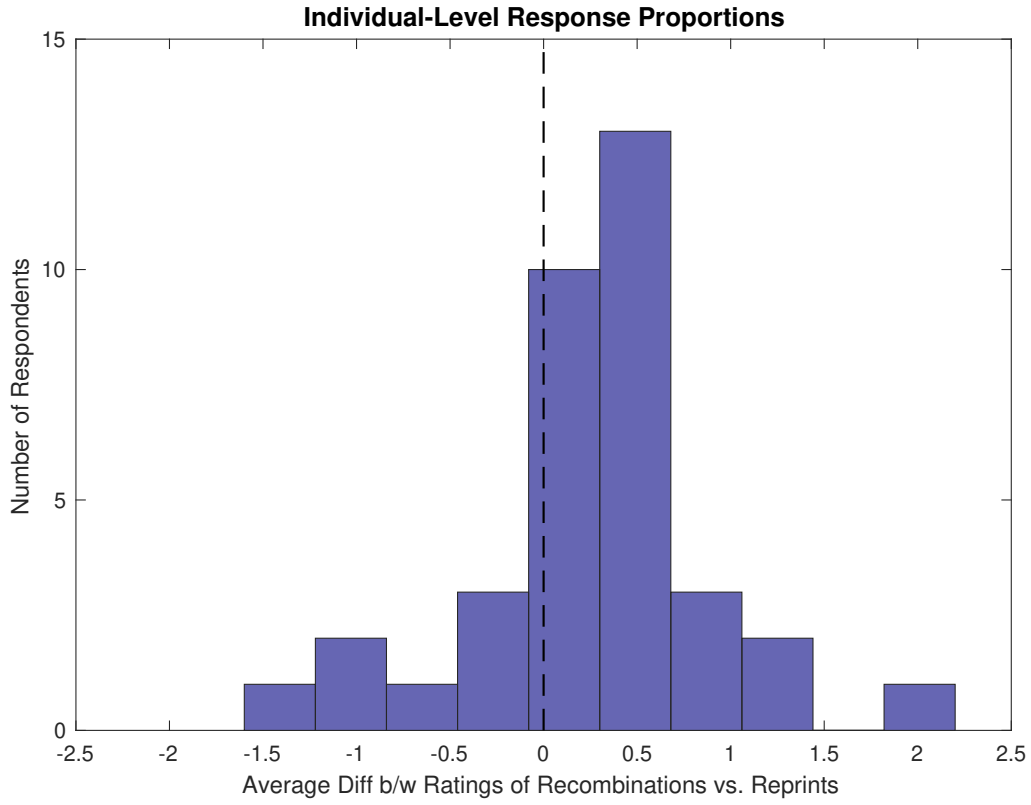
Figure 2: Individual-level experimental responses: for each participant, we compute the difference in average novelty scores assigned to recombination vs. reprint news.
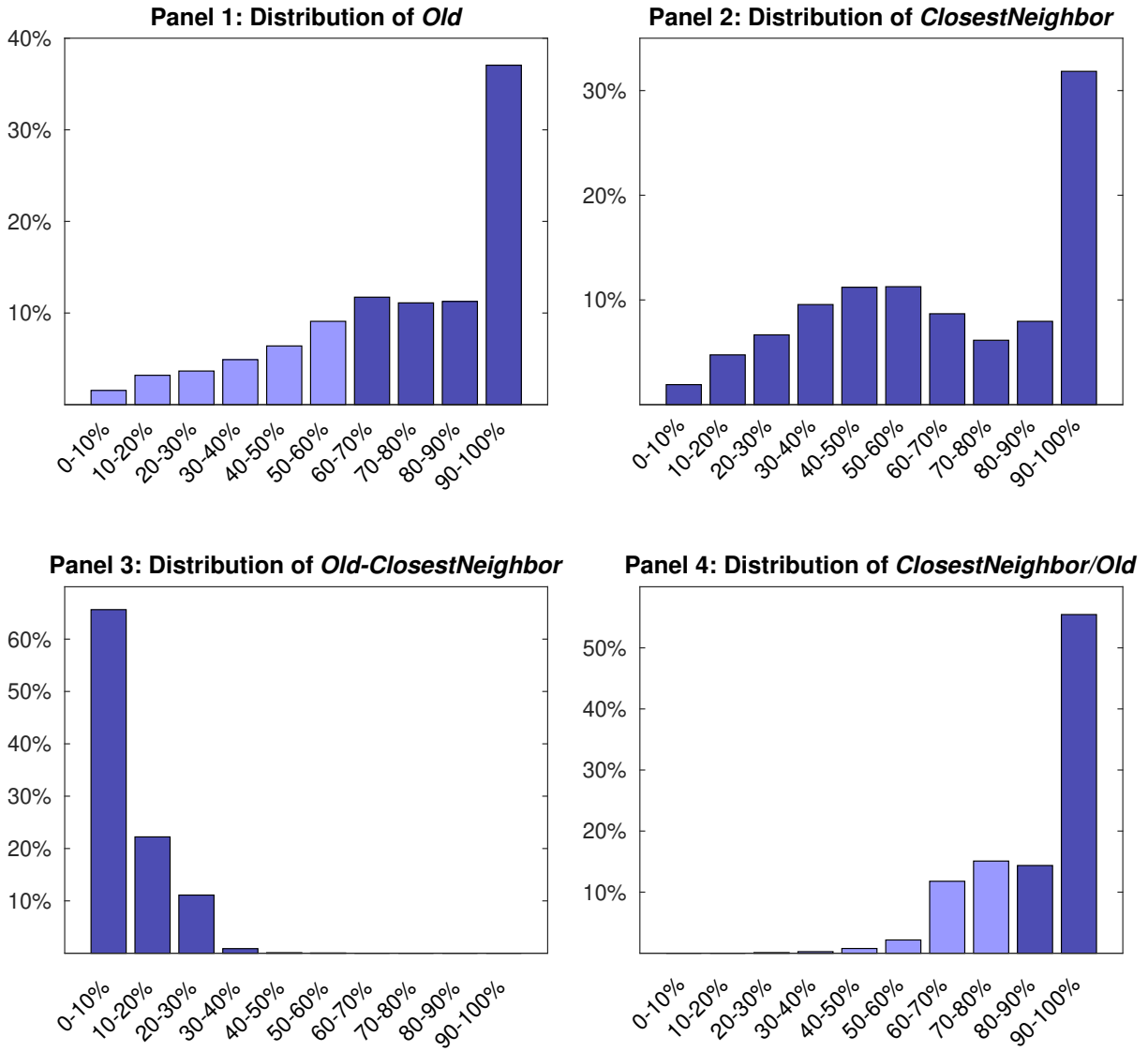
Figure 3: Distributions of key article-level variables. **Panel 1** presents a histogram of the *Old* measure. **Panel 2** displays the distribution of the *ClosestNeighbor* measure, while **Panel 3** shows the distribution of the article-level differences between *Old* and *ClosestNeighbor*. **Panel 4** displays the distribution of *ClosestNeighbor / Old* – a measure of the extent to which an article's stale content comes from the single closest source.
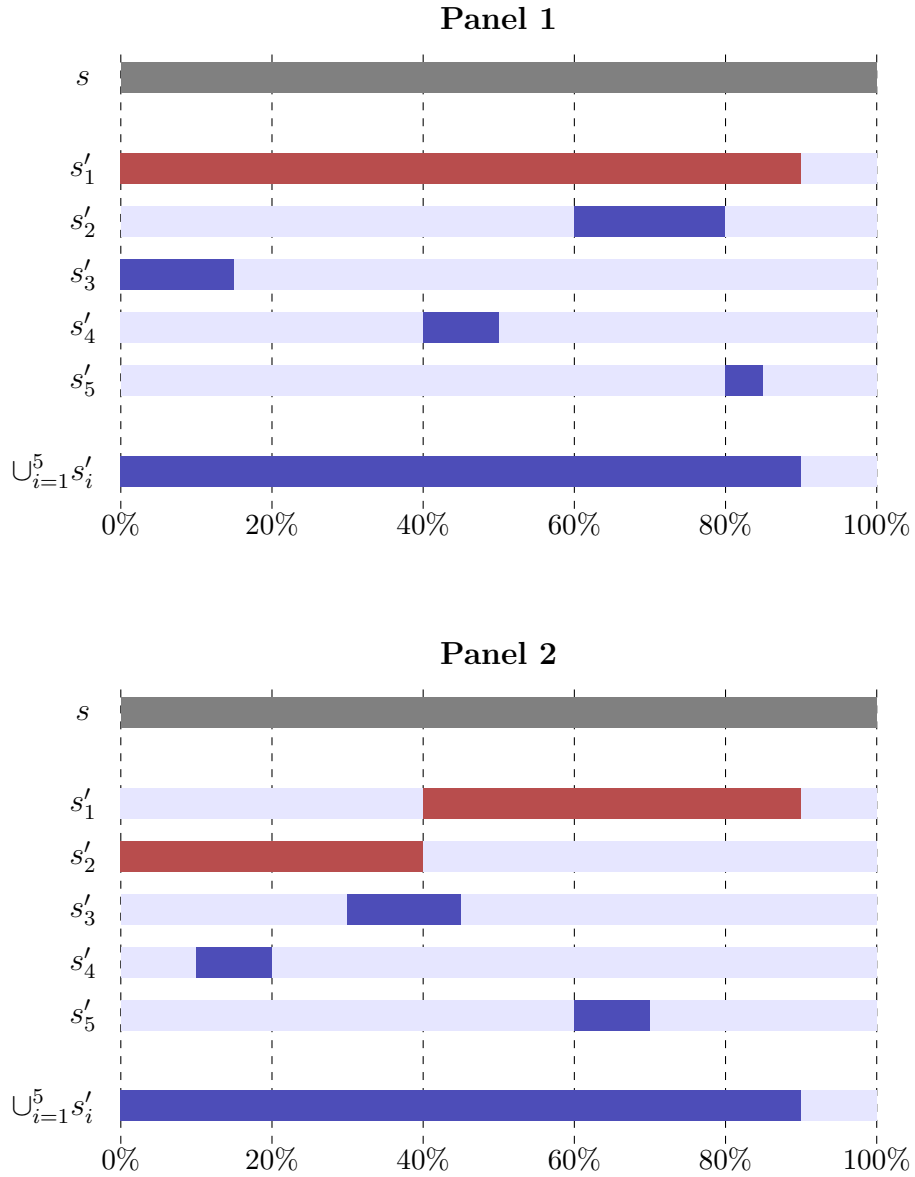
Figure 4: Two news articles, both 90% stale, display drastically different forms of staleness. **Panel 1** features an article, s, 90% of whose content is spanned by the single closest neighbor, $s'_1(s)$; this article is not only 90% stale, but an almost exact reprint of a previous article. **Panel 2**, on the other hand, features an article s that is also 90% stale but that recombines information from two previous articles: $s'_1(s)$ and $s'_2(s)$.

Article s

$Old(s) \geq 0.6$?

No

Yes

Novel News

$ClosestNeighbor(s)/Old(s) \geq 0.8$?

No

Yes

Recombination
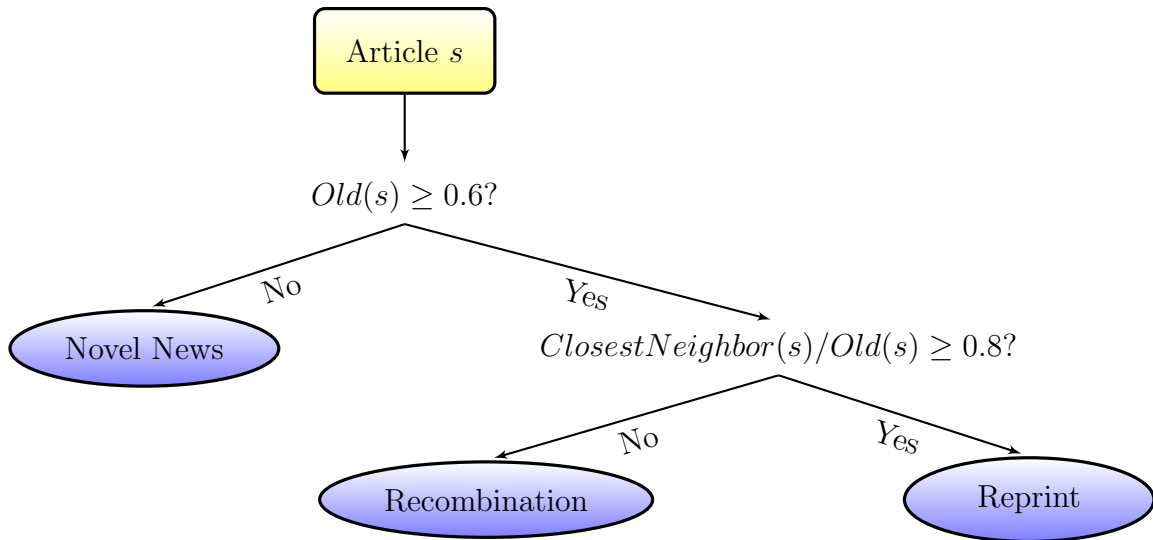
Reprint

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

Panel 1: Time Series of the Coefficient on *AbnPrcOld*

Panel 2: Time Series of the Coefficient on *AbnPrcRecombination*
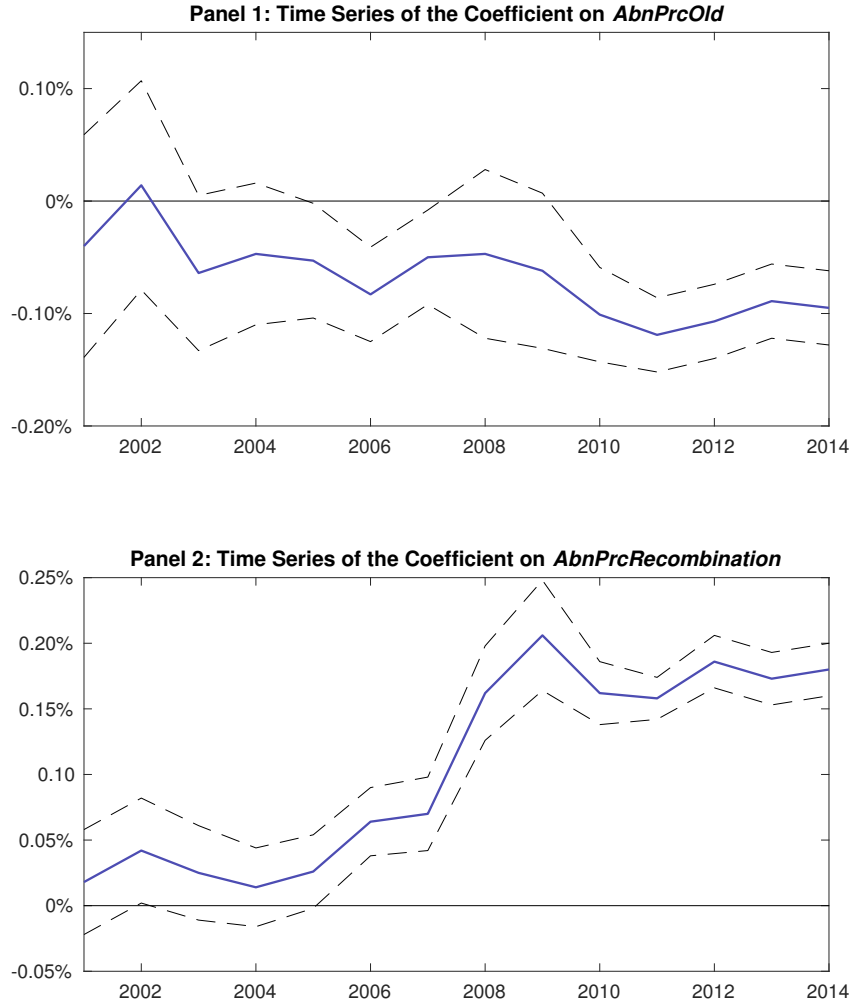
Figure 6: Time series of the coefficients from regressions of the next business day abrnomal return on *AbnPrcOld* and *AbnPrcRecombination*. Daily Fama and MacBeth (1973) cross-sectional regressions are estimated for each year $T \in [2001, 2014]$:

$$|AbnRet|_{i,t+1} = \alpha_T + \beta_{1,T} AbnPrcOld_{i,t} + \beta_{2,T} AbnPrcRecombination_{i,t} + \gamma_T X_{i,t} + \epsilon_{i,t}$$

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 blue line; dashed lines show 95% confidence bounds.

Table 1: Breakdown by affiliation of the finance professionals participating in the experimental survey.

Affiliation Type	Institution	Percentage
	Hedge Funds	9.8%
	Two Sigma, Point 72, Och Ziff Capital Management, Dockyard Capital Management	
	Investment Managers	14.6%
	Fidelity, PIMCO, State Street, Alliance Bernstein, Hayfin Capital Management, Devonshire Holdings	
	Private Investors	2.4%
	Banks and Broker Dealers	24.4%
	Goldman Sachs, Morgan Stanley, Wells Fargo, ABN AMRO, Commonwealth Bank, Royal Bank of Scotland	
	Investment Banks	4.9%
	Barclays, Macquarie	
	Government Agencies	2.4%
	Federal Reserve Board	
	Private Equity & Venture Capital	19.5%
	Bain Capital, Lindsay Goldberg, Bessemer Venture Partners, Hammerhead Equity Partners, Garrison Investment Group, Charlesbank Capital, ACON Investments	
	Media	4.9%
	Bloomberg, Financial Times	
	Technology Companies	4.9%
	Facebook, Uber	
	Academic and Non-Profits	7.3%
	Other Finance Professionals	4.9%

Table 2: Headlines used in the experiment on perception of news novelty. Panels 1 and 2 present headlines regarding fictitious firms Argosy and Laker, respectively. Panel 3 displays filler headlines.

Panel 1: Headlines about Argosy

Headline	Type
Argosy's misfit design business down, some tough questions to answer	Novel
Argosy Trucking Q3 results above expectations eps 1.2 vs 1.1	Novel
Argosy beats expectations: Q3 trucking results eps up .1 on 1.1	Reprint
Argosy Q3 earnings beat expectations, but design business down	Recomb.
Upgrade to BUY from HOLD on Positis, Hansen, Argosy (Casey Watts)	Novel
Argosy upgraded to BUY as earnings top expectations for Q3	Recomb.
Argosy logistics operation boosts jobs, business expands, "more to come"	Novel
Strong results indicative of broader logistics uplift: Star, Argosy	Novel
Strong results indicate broader boost of logistics players	Reprint
Miller, Jensen, Tower choose Argosy as contracts explode	Novel
Argosy IntelligentLogistix play yet to bear fruit, Carlson states	Novel
Argosy touts IntelligentLogistix play as new contracts explode	Recomb.
Fishing for acquisition targets, Bugle eyes direct to consumer Argosy play	Novel
Carlson waiting for Argosy IntelligentLogistix play to bring results	Reprint
Argosy contracts explode as Frank targets Bugle acquisition deal	Recomb.
Retail competition good for parcel carriers as Argosy expands fleet	Novel
TruckingLife: Retailers compete as Argosy expands parcel carrying fleet again	Reprint
DePaul celebrates 125 years of entrepreneurship: Argosy's Frank to speak	Novel
Frank talks Bugle acquisition as results above expectations, business expands (Argosy)	Recomb.
Strong indication of retail competition as Argosy expands fleet	Reprint

Panel 2: Headlines about Laker

Headline	Type
Russell quits amid claims of harassment from ex-employees	Novel
More scandal at Laker as CFO Russell fends off allegations of misconduct	Novel
An improbable source: Laker CEO George ousts Russell on hardline	Novel
Laker CEO George takes hardline to force Russell out, source says	Reprint
Laker's AdventiMed releases major landmark in DP2 cures (PharmaToday)	Novel
Laker CFO leaves on scandal during DP2 release	Recomb.
Laker CFO "not atypical in an industry rife with problems", says Blake	Novel
Amid harassment claims from ex-employees, Russell forced to quit in disgrace	Reprint
Pharma rife with scandals as Russell quits amid claims of harassment	Recomb.
Another week, another indiscretion: Laker's female employees walk out	Novel
David George apologizes, pledges compensation as Laker scandal continues	Novel
Laker's female employees walk out, George issues apology	Recomb.
Employee action at Laker: female employees walk out over indiscretion	Reprint
Laker CEO George pledges compensation for scandal	Reprint
Russell "couldn't keep hands to himself", as OFCO probes deeper	Novel
"Indecent", Potemus questions juniors among harassment claims	Novel
AdgentiMed releases DP2, Laker under OFCO probe	Recomb.
DP2 affects 3/4 of adults in the US by age 55, Harvard study finds	Novel
OFCO probes deeper into CFO Russell's behavior	Reprint
Harvard study finds promise, while Potemus questions juniors	Recomb.

Table 1 (Continued). Headlines used in the experiment on perception of news novelty. Panels 1 and 2 present headlines regarding fictitious firms Argosy and Laker, respectively. Panel 3 displays filler headlines.

Panel 3: Filler Headlines

Headline
Dow boasts eighth record close in a row, banks lead
Toyota plans truck, possibly SUV production in Mexico after Trump threat
U.S. payrolls report boosts dollar, equities, bond yields
Exclusive: China regulators plan to crack down further on overseas deals
Former drug company executive Martin Shkreli is convicted of fraud
Wells Fargo to pay U.S. $108 million over veterans' loans
Elliott discloses NXP stake, to push for higher price in Qualcomm deal
Volkswagen brand CEO says Tesla has abilities Volkswagen lacks
Viacom's weak forecast shows need for merger: analysts
Airbus deliveries remain hampered by engine delays
Payroll firm ADP seeks to fend off activist investor Ackman
U.S. judge dismisses lawsuit over Buffett's See's 'kosher' chocolate
Clariant, Huntsman investor backs merger, fears fight is a distraction
N.Y. Fed lifts U.S. third-quarter GDP growth view near 2 percent
RBS makes first-half profit, may move some jobs to Amsterdam
Toyota eyes possible SUV production at Guanajuato, Mexico plant
Trump's Attorney General Jeff Sessions is bringing the hammer down on leaks
DOJ warns the media could be targeted in crackdown on leaks
Martin Shkreli Is Found Guilty of Fraud
Convicted 'Pharma Bro' has an image problem, lawyer concedes
'Pharma bro' Martin Shkreli convicted of 3 counts of fraud
Special counsel utilizing DC grand jury in Russia investigation
Special Counsel Robert Mueller Impanels Washington Grand Jury in Russia Probe
Murder conviction in Blackwater case thrown out, other sentences overturned
Asian Americans are divided after the Trump administration's move on affirmative action
Justice Department to Investigate Harvard's Admissions Process
Torch tower residents in Dubai wake up to the building engulfed in flames yet again
Torch Tower fire: They walked down 50 floors with two cats
Venezuela ushers in new pro-government chamber as opposition struggles to regroup
More than 1000 dead exotic animals found in California complex

Table 3: Balance checks on the experimental headlines along three dimensions: length (in terms of the average number of words; **Panel 1**), location (in terms of the location in the sequence of 20 headlines; **Panel 2**), and staleness (as the average % of words that have already appeared in previously located headlines; **Panel 3**). 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	8.5	12.1	12.0

Panel 3: Staleness (% 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%

Table 4: Experimental results. **Panel 1** presents the mean participant responses fro novel news, reprints, and recombinations. **Panel 2** shows the differences in responses, with accompanying standard errors, across two comparisons: novel news vs. old news (pooling both reprints and recombinations) and between the two sets of old news (recombinations vs. reprints). The results are computed across three samples: headlines about Argosy Logistics Inc.; headlines about Laker Pharmaceuticals LLC; and the combined set of headlines..

Panel 1: Mean responses

	Novel	Reprint	Recombination
Argosy sample	4.16	2.65	3.23
Laker sample	4.84	2.57	2.85
Overall	4.52	2.61	3.03

Panel 2: Comparisons

		Novel vs. Old	Recombination vs. Reprint
Argosy sample	Diff	1.21**	0.58*
	(SE)	(0.21)	(0.29)
Laker sample	Diff	2.13**	0.27
	(SE)	(0.20)	(0.28)
Overall	Diff	1.70**	0.42*
	(SE)	(0.15)	(0.20)

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

Table 5: Summary statistics of Bloomberg news data, consisting of the volume of stories, computed on a daily basis and separately for weekends, and the density of relevant tags per story. I present the overall volume of news, stories tagged with security codes, and stories tagged with security codes with at least a 70% relevance score. 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 stories	Daily stories w/ sec. codes	Daily stories w/ relevance ≥ 70	Weekend stories w/ relevance ≥ 70	Tags per story
Full Sample					
Mean	70,723	29,570	4,026	494	1.30
Median	60,783	25,212	2,675	351	1.26
25th Percentile	31,668	12,353	1,311	124	1.22
75th Percentile	94,010	41,785	6,032	917	1.37
By year (mean)					
2000	20,940	8,727	940	82	1.60
2001	23,844	9,692	1,062	116	1.40
2002	26,844	10,621	1,159	133	1.39
2003	30,190	12,458	1,309	122	1.32
2004	33,417	14,432	1,577	169	1.27
2005	38,251	16,372	1,582	156	1.26
2006	47,784	20,755	2,194	310	1.21
2007	60,024	24,775	2,596	350	1.22
2008	66,819	27,258	2,673	466	1.20
2009	75,303	28,976	2,777	596	1.18
2010	82,227	36,901	3,774	482	1.22
2011	93,058	42,208	5,966	972	1.25
2012	102,778	45,471	6,096	883	1.25
2013	170,409	69,775	12,582	1,269	1.35
2014	188,964	75,128	14,106	1,411	1.37

Table 6: Summary statistics of extracted news text, including article-level and firm-level numbers of words per story, 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%

Table 7: Results from Fama and MacBeth (1973) regressions of next-day absolute abnormal returns and abnormal trading volumes on abnormal shares of old news and recombination stories:

$Column\ (1): |AbnRet|_{i,t+1} = a + bAbnPrcOld_{i,t} + gX_{i,t} + e_{i,t}$

$Column\ (2): |AbnRet|_{i,t+1} = a + b_1AbnPrcOld_{i,t} + b_2AbnPrcRecombination_{i,t} + gX_{i,t} + e_{i,t}$

$Column\ (3): AbnVol_{i,t+1} = \alpha + \beta AbnPrcOld_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$

$Column\ (4): AbnVol_{i,t+1} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$

The controls $X_{i,t}$ include the number of news stories published about firm i on date t ($Stories_{i,t}$), the abnormal number of stories over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), the average number of unique terms per story ($Terms_{i,t}$), 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 var: $|AbnRet_{i,t+1}|$ | | Dependent var: $AbnVol_{i,t+1}$ | |
	(1) Old News Only	(2) Old News & Recombinations	(3) Old News Only	(4) Old News & Recombinations
$AbnPrcOld_{i,t}$	-0.078%**	-0.082%**	-0.019%**	-0.021%**
	(0.013%)	(0.014%)	(0.005%)	(0.005%)
$AbnPrcRecombination_{i,t}$		0.134%**		0.032%**
		(0.009%)		(0.004%)
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
$AbnVolitility_{i,[t-5,t-1]}$	X	X	X	X
$Illiq_{i,[t-5,t-1]}$	X	X	X	X
R^2	0.221	0.228	0.173	0.181

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

Table 8: Results of Fama-MacBeth regressions of abnormal returns over $[t + t_1, t + t_2]$ on the abnormal returns on $t + 1$, interacted with overall share of old news and share of recombination news:

$Specification\ (12): AbnRet_{i,[t+t_1,t+t_2]} = \alpha + \beta_1 AbnPrcOld_{i,t} + \beta_2 AbnPrcOld_{i,t} \times AbnRet_{i,t+1} + \beta_3 AbnRet_{i,t+1} + \delta_1 AbnPrcRecombination_{i,t} + \delta_2 AbnPrcRecombination_{i,t} \times AbnRet_{i,t+1} + \gamma X_{i,t} + \epsilon_{i,t}$

The regressions are run over three horizons of delayed abnormal returns: three days ($t_1 = 2$ and $t_2 = 4$), one week ($t_1 = 2$ and $t_2 = 6$), and two weeks ($t_1 = 2$ and $t_2 = 11$). Controls include firm-date measures of the number of news stories ($Stories_{i,t}$), the abnormal number of stories over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), the average number of unique terms per story ($Terms_{i,t}$), 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.

The coefficients on the interaction terms are scaled to correspond to the effect sizes from a 10% increase in $AbnPrcOld_{i,t}$ and $AbnPrcRecombination_{i,t}$.

	(1) $AbnRet_{i,[t+2,t+4]}$	(2) $AbnRet_{i,[t+2,t+6]}$	(3) $AbnRet_{i,[t+2,t+11]}$
$AbnRet_{i,t+1}$	-0.021†	-0.024*	-0.026*
	(0.011)	(0.011)	(0.012)
$AbnPrcOld_{i,t} * AbnRet_{i,t+1}$	-0.019*	-0.022*	-0.023**
	(0.008)	(0.009)	(0.009)
$AbnPrcRecombination_{i,t} * AbnRet_{i,t+1}$	-0.026**	-0.028**	-0.033**
	(0.005)	(0.006)	(0.006)
R^2	0.077	0.075	0.070
Controls:			
$\quad Stories_{i,t}$	X	X	X
$\quad AbnStories_{i,[t-5,t-1]}$	X	X	
$\quad Terms_{i,t}$	X	X	X
$\quad MCap_{i,t}$	X	X	X
$\quad BM_{i,t}$	X	X	X
$\quad AbnRet_{i,[t-5,t-1]}$	X	X	X
$\quad AbnVol_{i,[t-5,t-1]}$	X	X	X
$\quad AbnVolitility_{i,[t-5,t-1]}$	X	X	X
$\quad Illiq_{i,[t-5,t-1]}$	X	X	X

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

Table 9: Robustness of market reactions to old and recombination news. *AbnPrcOld* and *AbnPrcRecombination* are computed with varied look-back windows τ, numbers n of considered closest articles, and whether firms with fewer than n articles in the past τ days are included in the analysis.

In **Panel 1**, we estimate the following specification for each variable construction method k:

$|AbnRet|_{i,t+1} = \alpha^k + \beta_1^k AbnPrcOld_{i,t}^k + \beta_2^k AbnPrcRecombination_{i,t}^k + \gamma^k X_{i,t} + \epsilon_{i,t}^k$

Panel 2 considers the following specification for each variable construction method k:

$AbnVol_{i,t+1} = \alpha^k + \beta_1^k AbnPrcOld_{i,t}^k + \beta_2^k AbnPrcRecombination_{i,t}^k + \gamma^k X_{i,t} + \epsilon_{i,t}^k$

Controls include the number of news stories ($Stories_{i,t}$), abnormal volume of stories over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average story length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), illiquidity ($Illiq_{i,[t-5,t-1]}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), and abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$). 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. The coefficients on the interaction terms are scaled to correspond to the effect sizes from a 10% increase in $AbnPrcOld_{i,t}$ and $AbnPrcRecombination_{i,t}$.

Panel 1: $|AbnRet_{i,t}|$

| | Coefficient on: | |
Construction Method	$AbnPrcOld_{i,t}$	$AbnPrcRecombination_{i,t}$
t=3, n= 5, L=N	-0.082%**	0.134%**
	(0.013%)	(0.009)
t = 3, n = 5, L=Y	-0.085%**	0.118%**
	(0.013)	(0.017)
t = 3, n=10, L=N	-0.079%**	0.205%**
	(0.013)	(0.013)
t = 3, n=10, L=Y	-0.024%	0.064%**
	(0.018)	(0.024)
t=5, n=5, L=N	-0.067%**	0.120%**
	(0.012)	(0.013)
t=5, n=5, L=Y	-0.089%**	0.106%**
	(0.013)	(0.012)
t=5, n=10, L=N	-0.042%*	0.182%**
	(0.016)	(0.013)
t=5, n=10, L=Y	-0.037%*	0.112%**
	(0.014)	(0.017)
t=10, n=5, L=N	-0.189%**	0.095%**
	(0.012)	(0.014)
t=10, n=5, L=Y	-0.093%**	0.106%**
	(0.012)	(0.008)
t=10, n=10, L=N	-0.182%**	0.135%**
	(0.013)	(0.012)
t=10, n=10, L=Y	-0.028%	0.072%**
	(0.017)	(0.023)

Table 9 (Continued). Robustness of market reactions to old and recombination news.

Panel 2: $AbnVol_{i,t}$

Construction Method	Coefficient on:	
	$AbnPrcOld_{i,t}$	$AbnPrcRecombination_{i,t}$
t=3, n= 5, L=N	-0.021%**	0.032%**
	(0.005%)	(0.004)
t = 3, n = 5, L=Y	-0.019%**	0.026%**
	(0.005)	(0.004)
t = 3, n=10, L=N	-0.018%**	0.039%**
	(0.006)	(0.005)
t = 3, n=10, L=Y	-0.007%	0.015%**
	(0.005)	(0.004)
t=5, n=5, L=N	-0.015%**	0.028%**
	(0.005)	(0.004)
t=5, n=5, L=Y	-0.023%**	0.024%**
	(0.006)	(0.005)
t=5, n=10, L=N	-0.014%*	0.036%**
	(0.005)	(0.004)
t=5, n=10, L=Y	-0.011%*	0.022%**
	(0.005)	(0.004)
t=10, n=5, L=N	-0.028%**	0.020%**
	(0.005)	(0.004)
t=10, n=5, L=Y	-0.022%**	0.020%**
	(0.005)	(0.004)
t=10, n=10, L=N	-0.027%**	0.027%**
	(0.004)	(0.004)
t=10, n=10, L=Y	-0.006%†	0.015%**
	(0.004)	(0.004)

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

Table 10: Robustness of return reversals following old and recombination news. *AbnPrcOld* and *AbnPrcRecombination* are computed with varied look-back windows τ, numbers n of considered closest articles, and whether firms with fewer than n articles in the past τ days are included in the analysis. We estimate the following specification for each variable construction method k:

$AbnRet_{i,[t+2,t+11]} = \alpha^k + \beta_1^k AbnPrcOld_{i,t}^k + \beta_2^k AbnPrcOld_{i,t}^k \times AbnRet_{i,t+1} + \beta_3^k AbnRet_{i,t+1} + \delta_1^k AbnPrcRecombination_{i,t}^k + \delta_2^k AbnPrcRecombination_{i,t}^k \times AbnRet_{i,t+1} + \gamma^k X_{i,t} + \epsilon_{i,t}^k$

Controls include the number of news stories ($Stories_{i,t}$), abnormal volume of stories over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average story length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), illiquidity ($Illiq_{i,[t-5,t-1]}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), and abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$). 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. The coefficients on the interaction terms are scaled to correspond to the effect sizes from a 10% increase in $AbnPrcOld_{i,t}$ and $AbnPrcRecombination_{i,t}$.

	Coefficient on:	
	$AbnPrcOld_{i,t}^k \times AbnRet_{i,t+1}$	$AbnPrcRecombination_{i,t}^k \times AbnRet_{i,t+1}$
t=3, n= 5, L=N	-0.023**	-0.033**
	(0.009)	(0.006)
t = 3, n = 5, L=Y	-0.008	-0.022**
	(0.010)	(0.006)
t = 3, n=10, L=N	-0.024**	-0.037**
	(0.009)	(0.006)
t = 3, n=10, L=Y	-0.002	-0.009
	(0.010)	(0.007)
t=5, n=5, L=N	-0.021*	-0.030**
	(0.009)	(0.006)
t=5, n=5, L=Y	-0.036**	-0.035**
	(0.009)	(0.006)
t=5, n=10, L=N	-0.020*	-0.016*
	(0.009)	(0.006)
t=5, n=10, L=Y	-0.018†	-0.028**
	(0.010)	(0.007)
t=10, n=5, L=N	-0.022*	-0.026**
	(0.009)	(0.006)
t=10, n=5, L=Y	-0.020*	-0.013*
	(0.009)	(0.006)
t=10, n=10, L=N	-0.015†	-0.014*
	(0.008)	(0.006)
t=10, n=10, L=Y	-0.029**	-0.030**
	(0.009)	(0.006)

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

Table 11: Replication of key results using continuous measures *AbnExtentOld* and *AbnExtentRecombination*.

Panel 1 estimates Fama-MacBeth regressions of the next-day absolute abnormal returns and abnormal trading volumes on the *AbnExtentOld* and *AbnExtentRecombination* measures:

$|AbnRet|_{i,t+1} = a + b_1 AbnExtentOld_{i,t} + b_2 AbnExtentRecombination_{i,t} + gX_{i,t} + e_{i,t}$,

$AbnVol_{i,t+1} = \alpha + \beta_1 AbnExtentOld_{i,t} + \beta_2 AbnExtentRecombination_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$,

Panel 2 estimates the reversal of the abnormal returns over the next weeks, $(t+2, t+11)$:

$AbnRet_{i,[t+2,t+11]} = \alpha + \beta_1 AbnExtentOld_{i,t} + \beta_2 AbnExtentOld_{i,t} \times AbnRet_{i,t+1} + \beta_3 AbnRet_{i,t+1} + \delta_1 AbnExtentRecombination_{i,t} + \delta_2 AbnExtentRecombination_{i,t} \times AbnRet_{i,t+1} + \gamma X_{i,t} + \epsilon_{i,t}$

Controls include the number of news stories ($Stories_{i,t}$), abnormal volume of stories over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average story length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), illiquidity ($Illiq_{i,[t-5,t-1]}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), and abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$). 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: $|AbnRet|_{i,t+1}$	Dependent variable: $AbnVol_{i,t+1}$
$AbnExtentOld_{i,t}$	-0.058%**	-0.017%**
	(0.012)	(0.004)
$AbnExtentRecombination_{i,t}$	0.112%**	0.024%**
	(0.013)	(0.004)

Panel 2: Return Reversal

Explanatory variable	Dependent variable: $AbnRet_{i,[t+2,t+11]}$
$AbnExtentOld_{i,t} \times AbnRet_{i,t+1}$	-0.015**
	(0.006)
$AbnExtentRecombination_{i,t} \times AbnRet_{i,t+1}$	-0.017**
	(0.003)

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

Table 12: Replication of key results restricting attention to news articles whose text contains at least 90% old content.

Panel 1 estimates Fama-MacBeth regressions of the next-day absolute abnormal returns and abnormal trading volumes on the $AbnPrcOld^{90\%}$ and $AbnPrcRecombination^{90\%}$ measures:

$|AbnRet|_{i,t+1} = a + b_1 AbnPrcOld_{i,t}^{90\%} + b_2 AbnPrcRecombination_{i,t}^{90\%} + gX_{i,t} + e_{i,t}$,

$AbnVol_{i,t+1} = \alpha + \beta_1 AbnPrcOld_{i,t}^{90\%} + \beta_2 AbnPrcRecombination_{i,t}^{90\%} + \gamma X_{i,t} + \epsilon_{i,t}$,

Panel 2 estimates the reversal of the abnormal returns over the next two weeks, $(t+2, t+11)$:

$AbnRet_{i,[t+2,t+11]} = \alpha + \beta_1 AbnPrcOld_{i,t}^{90\%} + \beta_2 AbnPrcOld_{i,t}^{90\%} \times AbnRet_{i,t+1} + \beta_3 AbnRet_{i,t+1} +$

$\delta_1 AbnPrcRecombination_{i,t}^{90\%} + \delta_2 AbnPrcRecombination_{i,t}^{90\%} \times AbnRet_{i,t+1} + \gamma X_{i,t} + \epsilon_{i,t}$

Controls include the number of news stories ($Stories_{i,t}$), abnormal volume of stories over the past week relative to preceding three months ($AbnStories_{i,[t-5,t-1]}$), average story length ($Terms_{i,t}$), log market capitalization ($MCap_{i,t}$), book-to-market ratio ($BM_{i,t}$), illiquidity ($Illiq_{i,[t-5,t-1]}$), and prior-week abnormal returns ($AbnRet_{i,[t-5,t-1]}$), abnormal volume ($AbnVol_{i,[t-5,t-1]}$), and abnormal volatility ($AbnVolatility_{i,[t-5,t-1]}$). 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 $AbnPrcOld^{90\%}$ and $AbnPrcRecombination^{90\%}$ measures.

Panel 1: Market Reaction

Explanatory variable	Dependent variable:	
	$|AbnRet|_{i,t+1}$	$AbnVol_{i,t+1}$
$AbnPrcOld_{i,t}^{90\%}$	-0.093%**	-0.025%**
	(0.031)	(0.007)
$AbnPrcRecombination_{i,t}^{90\%}$	0.123%**	0.028%**
	(0.028)	(0.006)

Panel 2: Return Reversal

Explanatory variable	Dependent variable:
	$AbnRet_{i,[t+2,t+11]}$
$AbnPrcOld_{i,t}^{90\%} \times AbnRet_{i,t+1}$	-0.028*
	(0.014)
$AbnPrcRecombination_{i,t}^{90\%} \times AbnRet_{i,t+1}$	-0.030*
	(0.012)

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