

News-Driven Trading: Who Reads the News and When?

Anastassia Fedyk*

October 3, 2018

Abstract

This paper explores the puzzle of increased trading volume around informational releases through the lens of canonical models of gradual information diffusion and differences of opinion. I use a unique dataset of clicks on news by key finance professionals to distinguish between trading among investors who see the news at different times and trading among investors who see the same news but disagree regarding its interpretation. Consistent with gradual information diffusion, dispersion in the timing of investors' attention is strongly predictive of daily trading volume surges around earnings announcements and volume surges within minutes of individual news articles. The differences of opinion channel, measured as heterogeneity of investors reading the news, is generally weaker in explaining trading volume surges, but plays a larger role around more ambiguous news.

Keywords: information diffusion, disagreement, trading volume, price formation

*UC Berkeley, Haas School of Business. Mail: 545 Student Services Building #1900, Berkeley, CA 94720. Email: fedyk@berkeley.edu. I am grateful to Malcolm Baker, Matteo Benetton, John Beshears, John Campbell, Lauren Cohen, Joshua Coval, Kent Daniel, Stefano DellaVigna, Xavier Gabaix, Tom Glocher, Brett Green, Robin Greenwood, Sam Hanson, Marc Kaufmann, David Laibson, Owen Lamont, Andres Liberman, Dmitry Livdan, Chris Malloy, Jeffrey Picel, Martin Schmalz, Daniel Schmidt (discussant), Andrei Shleifer, Jeremy Stein, Adi Sunderam, Carmen Wang, and seminar participants at Harvard University, UC Berkeley Haas, University of Washington Foster, the European Finance Association, the European Summer Symposium on Financial Markets (evening session), and the Society for Economic Dynamics for helpful comments.

1 Introduction

This paper explores drivers of increased trading volume around public information releases. High trading volume around information releases has been a long-standing empirical fact in the literature, and a number of theories of disagreement have been proposed to explain this phenomenon.¹ But empirical understanding of various parties' information sets and disagreement around information releases remains limited.² Does the disagreement occur between individuals who have already seen the news and those who have been inattentive to it, according to gradual information diffusion models? Or is the disagreement driven by different interpretations of the same information by investors with varying beliefs, as in models of differences of opinion?

An ideal setting to answer these questions would be one in which we can observe the information set of each counter-party to every trade. I take a step in this direction by investigating a comprehensive click-level dataset of information consumption by a substantial set of market participants. I find that measures of gradual information diffusion – capturing the dispersion of *timing* of investors' clicks – are strongly predictive of daily trading volume surges around earnings announcements and trading volume surges within minutes of individual news articles. Measures of differences of opinion – capturing the dispersion of the *types* of investors clicking on the news – are also operative albeit somewhat weaker in predicting trading volume around information releases. The relative strengths of these two channels depend on the context of the specific informational release: differences of opinion play a larger role around textually ambiguous news, while gradual information diffusion is more closely tied to trading volume around textually straightforward news.

I structure the empirical investigation of the nature of disagreement around news events using a conceptual framework that nests canonical models of gradual information diffusion and differences of opinion. Each of these models yields testable predictions for the joint dynamics of trading volume and news consumption.³ Gradual information diffusion predicts

¹For example, Kaniel, Liu, Saar, and Titman (2012) and Drake, Roulstone, and Thornock (2012) document heightened trading volume around earnings announcements (this relation is illustrated in Figure 5), while Bali, Bodnaruk, Scherbina, and Tang (2016) find that unusual news flow temporarily increases investor disagreement. For theories of disagreement around information releases, see Karpoff (1986), Harris and Raviv (1993), Kandel and Pearson (1995), Hong and Stein (2007), Banerjee and Kremer (2010), Kondor (2012), and Banerjee, Davis, and Gondhi (2017), among others. For trading volume as a proxy for divergence in investors' opinions, see Garfinkel and Sokobin (2006) and Garfinkel (2009).

²Prior work largely relies on dispersion of sell-side analyst forecasts to capture differences in opinion; see Ajinkya, Atiase, and Gift (1991), Atiase and Bamber (1994), Diether, Malloy, and Scherbina (2002), Anderson, Ghysels, and Juergens (2005), and Bamber, Barron, and Stevens (2011). Carlin, Longstaff, and Matoba (2014) look at differences of opinion more directly, but focus on prepayment in the mortgage-backed security market. Giannini, Irvine, and Shu (2015) and Cookson and Niessner (2016) take a complementary approach to mine; they analyze expressed opinions on StockTwits, but do not capture information acquisition.

³For classic models of gradual information diffusion, see Hong and Stein (1999) and Hirshleifer and Teoh

that trading volume is maximized when the investors are evenly split between those who see the news early and those who read it later. The differences of opinion model predicts that trading volume is highest when the group of investors reading the news is most heterogeneous. Gradual information diffusion makes additional predictions for price formation – that the speed of news consumption is positively related to the speed of price adjustment. Differences of opinion generate additional predictions on the effect of news ambiguity – that investor heterogeneity is more instrumental in spurring trading volume around more ambiguous news events, which admit a wider range of interpretations.

These predictions are tested using a comprehensive anonymized dataset of clicks by finance professionals on 3.5 million news articles between March 2014 and March 2015. The data aggregate news articles from a variety of sources and offer a uniquely comprehensive view of news consumption. The click dataset represents details of 80 million clicks by hundreds of thousands of de-identified financial professionals, comprised predominantly of institutional investors.⁴

The advantages of this dataset over news consumption data used in prior work are threefold.⁵ First, the data represent individual clicks, allowing me to observe the dynamics of investor attention at high frequency. Second, although the data are fully anonymized, clicks by the same reader are linked to each other, allowing me to classify readers into types based on their news consumption patterns. Third, the clicks are linked to article-level characteristics such as novelty, sentiment, and textual ambiguity.

In order to estimate gradual information diffusion from the detailed news consumption data, I tabulate the clicks across time after a given piece of news – for example, across hours after an earnings announcement or across seconds after an individual news release. I use a measure of dispersion, normalized Shannon entropy, to assess the extent to which the clicks are evenly distributed across the time buckets. The higher the value of this proxy – the more dispersed the attention across time – the more scope there is for disagreement between investors who have already seen the news and those who have not.

(2003), among others. For models of differences of opinion, see, for example, Harris and Raviv (1993) and Kandel and Pearson (1995).

⁴Several steps were taken to protect the confidentiality of the underlying reader information. For example, the original identifiers were replaced with stochastically generated numbers assigned randomly over the population – removing the possibility of personal details being inferred from the identification schema. Due to the confidentiality protections utilized in the analysis, the Institutional Review Board at Harvard University made a “not human subject research” determination for this project.

⁵See, for example, Da, Engelberg, and Gao (2011), Drake, Roulstone, and Thornock (2012), and Madsen and Niessner (2016) for the use of Google search volume as a measure of attention; Bauguess, Cooney, and Hanley (2013) and Drake, Roulstone, and Thornock (2015, 2016) for downloads of EDGAR filings; Lawrence, Ryans, Sun, and Laptev (2016) for searches on Yahoo!; and Lumsdaine (2010) and Ben-Rephael, Da, and Israelsen (2017) for the use of Bloomberg’s aggregate daily proxy of institutional investors’ attention.

To capture differences of opinion, I take an approach motivated by the extensive literature in sociology on the notion of homophily: that individuals with ex-ante similar characteristics are more likely to agree with each other.⁶ Applied to a network concept, Golub and Jackson (2012) show that the presence of (different) homophilies slows convergence to consensus, leading to persistent disagreement. Empirically, Chang, Hong, Tiedens, and Wang (2015) find that investors from different linguistic backgrounds are more likely to disagree, while Cookson and Niessner (2016) document that individuals who associate with different investment styles express more diverging opinions on StockTwits.

I apply the notion of homophily to the context of investors attending to financial news, and proxy for differences of opinion using the heterogeneity of the investors who read a given piece of news. In order to measure reader heterogeneity, I employ techniques from machine learning to derive distinct news reading styles directly from the news consumption data, and to classify the readers into distinct styles. Readers in different news consumption styles are likely to have different information sets, different approaches to procuring and processing new information, different priors, and different constraints. As a result, they are likely to take different actions in response to the same news release. The higher the dispersion of the readers who see a given piece of news, the more scope there is for trading between investors who have all seen the same news but disagree regarding its impact.

Both gradual information diffusion and differences of opinion are predictive of trading volume around news, although the effect of the former is stronger. I perform the analysis at two horizons: within days around earnings announcements and within minutes around individual news events. Around earnings announcements, the difference between having all reads concentrated in a single hourly bucket and having the reads perfectly evenly distributed across the 48 post-announcement hourly buckets translates to volume surging by an additional 160% relative to its pre-announcement baseline. This effect is strongly statistically significant, and substantially larger than the effects of firm size, book-to-market ratio, or earnings surprise. By contrast, taking differences of opinion from purely concentrated in one reader type to perfectly split across the types corresponds to a 60% larger surge in trading volume, significant only at the 5% level. Similarly, at the high-frequency resolution around individual news articles, going from attention that is perfectly concentrated in time to perfectly dispersed corresponds to a fourfold increase in ten-minute trading volume following the news, compared to a more modest and less significant two-fold increase accompanying dispersion in types of attending investors.

However, the relative strengths of the two channels of disagreement in predicting trading volume around news depend on the characteristics of the underlying information. In par-

⁶See McPherson, Smith-Lovin, and Cook (2001).

ticular, when a piece of news is more ambiguous, lending itself more easily to differential interpretations, the dispersion of attention across reader types is just as predictive of trading volume surges as dispersion of attention over time. To gauge a news story’s ambiguity, I use machine learning classifiers, trained on data tagged by experts, to characterize the strength of the story’s sentiment (positive, negative, or neutral) and the type of information conveyed (factual versus opinion). I take a combination of the two classifications; thus, a news story labeled as having a strong sentiment in any direction and containing factual information is classified as straightforward, whereas a news story with weak sentiment and opinion-based information is deemed ambiguous. For textually ambiguous news, going from minimal to maximal dispersion in reader types corresponds to volume surging by an additional 350% relative to its pre-news baseline, while for textually clear news the effect is only a 200% increase. The estimated effect of dispersion in timing, on the other hand, is a 370% increase in trading volume around ambiguous news and a 440% increase in volume following more straightforward news.

The present paper contributes to the discourse on disagreement in financial markets by simultaneously capturing the two key channels: differences in timing of information acquisition and heterogeneity of attending investors. Prior work has largely investigated these two channels separately. Empirical evidence on gradual information diffusion and inattention relies on indirect attention proxies such as strategic release of information during times when investors are less likely to be attentive,⁷ as well as more direct measures using aggregate search volumes on platforms such as Google, Yahoo, and Bloomberg.⁸ By considering individual clicks, I am able to capture precise timestamps of attention and to see who is clicking, allowing me to gauge how likely the disagreement is to stem from differences in these investors’ reactions to the same news. In terms of measuring differences of opinion, existing proxies rely predominantly on analyst forecasts and opinions expressed on social media.⁹ This line of work is complementary to my paper: they offer more direct measures of opinion, but do not tie these measures to particular informational content. By contrast, I use an implicit proxy for disagreement based on who is reading the news, but do so in a way that allows me to tie this proxy to specific news events and analyze it side by side with

⁷See DellaVigna and Pollet (2009), Hirshleifer, Lim, and Teoh (2009), deHaan, Shevlin, and Thornock (2015), and Niessner (2015).

⁸See Da, Engelberg, and Gao (2011), Drake, Roulstone, and Thornock (2012, 2015, 2016), Bauguess, Cooney, and Hanley (2013), Ben-Rephael, Da, and Israelsen (2017), Madsen and Niessner (2016), and Lawrence, Ryans, Syn, and Laptev (2016), among others.

⁹For proxies of disagreement based on analyst forecasts, see Ajinkya, Atiase, and Gift (1991), Atiase and Bamber (1994), Diether, Malloy, and Scherbina (2002), Anderson, Ghysels, and Juergens (2005), Bamber, Barron, and Stevens (2011), among others. For proxies based on social media, see, for example, Giannini, Irvine, and Shu (2015), Cookson and Niessner (2016).

the timing of attention. I use the individual click data to effectively bring the “who” and the “when” of information consumption into the same setting and explore both channels of disagreement simultaneously.

The remainder of the paper proceeds as follows. Section 2 outlines the conceptual framework for my empirical tests. Section 3 describes the data. Section 4 details the methodology for constructing proxies of gradual information diffusion and differences of opinion. Section 5 presents the tests on predictability of trading volume from the two forms of disagreement. Section 6 considers the strengths of the two channels of disagreement for news events with varying levels of ambiguity. Section 7 concludes.

2 Conceptual Framework

Disagreement about new information can occur in two fundamentally distinct forms: between those who have seen the information and those who have not (gradual information diffusion), or between those who have all seen the same information but react to it differently (differences of opinion). To structure the empirical tests investigating these channels of disagreement, I present a simple theoretical framework that nests canonical models of gradual information diffusion and differences of opinion.

The conceptual framework is standard in the literature, and closely follows the setups in Kandel and Pearson (1995), Hirshleifer and Teoh (2003), and DellaVigna and Pollet (2009). There is a riskfree asset with zero rate of return and a single risky security with a stochastic payoff $R \sim \mathcal{N}(\bar{R}, \sigma_R^2)$ realized in the final period. In the relatively short-term settings that I consider, the realized value R can be taken to denote the end of day price for day-traders trading on individual news, or the price on which an asset settles in the days following an earnings announcement. The risky asset is in fixed supply X . There are potentially heterogeneous agents, with types indexed by i . At any point in time t , each agent of type i maximizes expected utility of his final wealth $W^{(i)}$ upon realization of R , with respect to the current holdings. The agents have mean-variance utility of the form $\mathbb{E}_{i,t}\{W^{(i)}\} - \frac{A^{(i)}}{2} \text{Var}_{i,t}\{W^{(i)}\}$; for simplicity, I take the risk-aversion coefficient to be identical across agents: $\forall i, A^{(i)} = A$. Each agent of type i is initially endowed with wealth $W_0^{(i)}$. There are no liquidity constraints.

Information in this framework is modeled as a signal arriving during an intermediate period. In particular, there are three periods in the model: in period 0, agents form prior expectations regarding the distribution of R ; in period 1, a noisy signal (news) is released, and agents update their expectations accordingly; in period 2, the value of R is realized and the agents consume their wealth. I assume the following form for the news signal: $N = R + \epsilon$,

where ϵ is a normally distributed noise term, independent of R , with mean μ and variance σ_ϵ^2 . The timeline is depicted in Figure 1.

A key to both gradual information diffusion and differences of opinion is that the agents do not form rational expectations regarding the information sets and actions of others. Instead, each agent acts in accordance only with his own information. In particular, if an agent lacks some piece of information, he fails to recognize that others may be better informed; and if agents hold differing beliefs from each other, they do not factor in others' beliefs. This form of overconfidence is a common modeling device across models of gradual information diffusion (see Hong and Stein (1999) or Hirshleifer and Teoh (2003)) and differences of opinion (see Harris and Raviv (1993) or Kandel and Pearson (1995)).

To fix ideas, I begin with the rational benchmark of all agents receiving information immediately and holding identical beliefs in Section 2.1. In Section 2.2, I incorporate gradual information diffusion as the news signal being observed only by a fraction of attentive investors. Differences of opinion are modeled as all investors having access to the same information, but holding different beliefs regarding the prior distribution of R and the distribution of the signal noise ϵ (Section 2.3).

2.1 Benchmark: Identical Information and Beliefs

I briefly characterize price formation and trading in absence of both gradual information diffusion and differences of opinion. In the benchmark, all agents are privy to all information, and hold identical, correct beliefs.

In period 0, all agents perceive the distribution of the final payoff R to be normal with mean \bar{R} and variance σ_R^2 . Hence, each agent i 's demand for the risky security is $x_0^{(i)}(P_0) = \frac{\bar{R} - P_0}{A\sigma_R^2}$. Imposing the market clearing condition that the net supply of the risky asset is X , the price in period 0 is:

$$P_0 = \bar{R} - A\sigma_R^2 X,$$

where the risk premium $A\sigma_R^2 X$ is zero if the asset is in zero net supply.

Similarly, at $t = 1$, the agents optimize their holdings with update beliefs that $\mathbb{E}_{i,1}\{R\} = \frac{\sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2} \bar{R} + \frac{\sigma_R^2}{\sigma_R^2 + \sigma_\epsilon^2} (N - \mu)$ and $Var_{i,1}\{R\} = \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2}$. The first period price is thus:

$$P_1 = \frac{\sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2} \bar{R} + \frac{\sigma_R^2}{\sigma_R^2 + \sigma_\epsilon^2} (N - \mu) - A \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2} X.$$

In period 2, all uncertainty is resolved, and $P_2 = R$. Hence, the returns in the two periods

are:

$$\Delta P_1 = \frac{\sigma_R^2}{\sigma_R^2 + \sigma_\epsilon^2}(R - \bar{R} + \epsilon - \mu) + C_1; \quad \Delta P_2 = \frac{\sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2}(R - \bar{R}) - \frac{\sigma_R^2}{\sigma_R^2 + \sigma_\epsilon^2}(\epsilon - \mu) + C_2,$$

where $C_1 = A \frac{(\sigma_R)^2}{\sigma_R^2 + \sigma_\epsilon^2} X$ and $C_2 = A \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2} X$ are the constant risk premia.

First, note that, by construction, the news signal enters the price dynamics and holdings identically regardless of clicks on news, since it is assumed that the news signal is observed by all investors and interpreted identically by them. As a result, all investors hold identical positions and there is no trading volume in this baseline model. Trading volume in the benchmark model can be generated by incorporating differential risk-aversion parameters or liquidity shocks to some investors. In neither of these cases, however, does trading volume depend on consumption of information.

Second, note that the correlation between ΔP_1 and ΔP_2 is zero. In the benchmark model, there is no serial correlation in returns, and, trivially, no predictability for return continuation from news consumption dynamics.

The basic benchmark predictions are summarized below.

Prediction 1 (Identical Information and Beliefs Benchmark):

H0.a: Clicks on news stories are not predictive of trading volume around the news.

H0.b: There is no relationship between clicks on news and price dynamics.

Prediction H0.0 is the null hypothesis throughout the empirical analysis in Sections 5 and 6, where I estimate the relationships between clicks on news and trading volume. Prediction H0.b provides the null for additional analyses on return predictability in Appendix A.

2.2 Gradual Information Diffusion

In this subsection, I model the implications of gradual information diffusion, where only a subset of investors immediately attend to the news.¹⁰ Gradual information diffusion predicts that trading volume is highest when investors are evenly split between those who see the news early and those who read it with a delay. The model also predicts that the price adjustment is faster when attention to news is more immediate, and that serial correlation in returns is higher when the split between immediate and delayed attention is more even.

¹⁰The importance of gradual information diffusion is suggested by studies of variation in investor attention and firms' strategic releases of information during periods of distraction. See, for example, DellaVigna and Pollet (2009) on earnings announcements released on Fridays, Hirshleifer, Lim, and Teoh (2009) on earnings announcements released contemporaneously with other announcements, and deHaan, Shevlin, and Thornock (2015) and Niessner (2015) for evidence that firms strategically respond to investors' limited attention by timing their releases.

Formally, gradual information diffusion is modeled as a fraction γ of investors (type $i = 1$) observing the news signal N in period 1, and the remaining $1 - \gamma$ of investors (type $i = 2$) not seeing the signal. In the empirical analysis of news consumption in this paper, the attentive investors are proxied by those who click on the news immediately, while the inattentive investors are modeled by the delayed clicks.

Prior expectations in period 0 are the same as in the benchmark, so prices and holdings in period 0 remain:

$$P_0 = \bar{R} - A\sigma_R^2 X; \quad \forall i, x_0^{(i)} = X$$

In period 1, investors of type $i = 1$ observe the news signal, and update their beliefs accordingly, while investors of type $i = 2$, who are not attentive to the news signal, continue to hold the same beliefs as in period 0. Thus, investors of type 1 perceive $\mathbb{E}_{1,1}\{R\} = \frac{\sigma_i^2}{\sigma_R^2 + \sigma_\epsilon^2} \bar{R} + \frac{\sigma_R^2}{\sigma_R^2 + \sigma_\epsilon^2} (N - \mu)$ and $Var_{1,1}\{R\} = \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2}$, while investors of type 2 perceive $\mathbb{E}_{2,1}\{R\} = \bar{R}$ and $Var_{2,1}\{R\} = \sigma_R^2$. As a result, the investors' demand functions for the asset in period 1 are given by:

$$x_1^{(1)}(P_1) = \frac{\sigma_\epsilon^2 \bar{R} + \sigma_R^2 (N - \mu) - (\sigma_R^2 + \sigma_\epsilon^2) P_1}{A\sigma_R^2 \sigma_\epsilon^2}; \quad x_1^{(2)}(P_1) = \frac{\bar{R} - P_1}{A\sigma_R^2}$$

Imposing the market clearing condition that $\gamma x_1^{(1)}(P_1) + (1 - \gamma) x_1^{(2)}(P_1) = X$ yields:

$$P_1 = \frac{\sigma_\epsilon^2 \bar{R}}{\sigma_\epsilon^2 + \gamma \sigma_R^2} + \frac{\gamma \sigma_R^2}{\sigma_\epsilon^2 + \gamma \sigma_R^2} (N - \mu) - A \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_\epsilon^2 + \gamma \sigma_R^2} X$$

The total trading volume associated with the news event, given by the absolute difference between $\gamma x_0^{(1)}$ and $\gamma x_1^{(1)}$ is:

$$Volume = \frac{\gamma(1 - \gamma)|N - R - \mu + A\sigma_R^2 X|}{A(\sigma_\epsilon^2 + \gamma \sigma_R^2)}$$

Under gradual information diffusion, the relationship between trading volume around the news announcement and the percentage of immediately attentive investors is non-monotonic. There is little disagreement and trading volume when either all or none of the investors see the news immediately, and trading volume is maximized when the split between immediately attentive and inattentive investors is roughly even.

While the key empirical analyses in this paper concern predictions for trading volume, gradual information diffusion also yields predictions regarding the relationship between news

consumption and price formation. The price changes across the periods are given by:

$$\Delta P_1 = \frac{\gamma\sigma_R^2}{\sigma_R^2 + \gamma\sigma_\epsilon^2}(R - \bar{R} + \epsilon - \mu) + C_1; \quad \Delta P_2 = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \gamma\sigma_R^2}(R - \bar{R}) - \frac{\gamma\sigma_R^2}{\sigma_\epsilon^2 + \gamma\sigma_R^2}(\epsilon - \mu) + C_2$$

First, note that the magnitude of the immediate price move, ΔP_1 , is increasing in γ , the percentage of the investing public who observe the news signal in period 1.

Second, correlation between ΔP_1 and ΔP_2 is given by:

$$\text{corr}(\Delta P_1, \Delta P_2) = \frac{\gamma(1 - \gamma)(\sigma_R^2)^2\sigma_\epsilon^2}{(\sigma_\epsilon^2 + \gamma\sigma_R^2)^2},$$

which is maximized at $\gamma^* = \frac{\sigma_\epsilon^2}{2\sigma_\epsilon^2 + \sigma_R^2}$. As a result, serial correlation in returns is largest when the investors are somewhat evenly distributed between those who see the news early and those who do not. The exact correlation-maximizing split depends on the variance of the priors and the noisy signal, where a higher share of informed agents is required to achieve maximal serial correlation when the signal is noisier.

Overall, the predictions of gradual information diffusion can be summarized as follows:

Prediction 2 (Gradual Information Diffusion):

H1.a: Highest trading volume occurs when the clicks on news are dispersed between immediate and delayed.

H1.b: Percentage of clicks on a news event that are immediate is predictive of the fraction of price move that is immediate.

H1.c: Highest serial correlation (continuation) in returns occurs when the split between immediate and delayed clicks is most even.

Prediction H1.a is the primary prediction of the gradual information diffusion channel for disagreement around news. I test this prediction empirically in Section 5 by estimating the relationship between trading volume surges around informational releases and the extent to which attention to those releases is dispersed over time. I do this in two settings: over hours after earnings announcements and during the ten minutes after individual news articles. Gradual information diffusion also generates additional predictions regarding prices, H1.b and H1.c. Empirical support for these predictions is documented in Appendix A.

2.3 Differences of Opinion

This subsection investigates the effects of differences of opinion by considering the case of investors who hold different beliefs regarding the distribution of the payoff and the news

signal. Differences of opinion predict that trading volume around news is driven by the diversity of the investors reading the news.

I model differences of opinion as two types of investors observing the same signal, but interpreting it differently. In particular, suppose that investors of type i hold priors that $R \sim \mathcal{N}(\bar{R}^{(i)}, \sigma_R^2)$ and believe that the noise in the news is distributed according to $\epsilon \sim \mathcal{N}(\mu^{(i)}, \sigma_\epsilon^2)$. Let γ denote the portion of investors who are of type $i = 1$.

In period 0, the demand $x_0^{(i)}$ of investors of type i and the price of the risky asset are determined by the investors' priors and the market clearing condition:

$$P_0 = \gamma \bar{R}^{(1)} + (1 - \gamma) \bar{R}^{(2)} - A \sigma_R^2 X$$

$$x_0^{(1)} = X + \frac{1 - \gamma}{A \sigma_R^2} (R^{(1)} - R^{(2)}); \quad x_0^{(2)} = X + \frac{\gamma}{A \sigma_R^2} (R^{(2)} - R^{(1)})$$

In period 1, prices and holdings depend not only on the investors' priors, but also on their interpretations of the news signal N . Imposing the market clearing condition on the agents' demands gives the following solution for the period 1 price and holdings:

$$P_1 = \frac{\sigma_\epsilon^2 (\gamma R^{(1)} + (1 - \gamma) R^{(2)})}{\sigma_R^2 + \sigma_\epsilon^2} + \frac{\sigma_R^2 (N - \gamma \mu^{(1)} - (1 - \gamma) \mu^{(2)})}{\sigma_R^2 + \sigma_\epsilon^2} - A \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2} X$$

$$x_1^{(1)} = x_0^{(1)} + \frac{1 - \gamma}{A \sigma_\epsilon^2} (\mu^{(2)} - \mu^{(1)}); \quad x_1^{(2)} = x_0^{(2)} + \frac{\gamma}{A \sigma_\epsilon^2} (\mu^{(1)} - \mu^{(2)})$$

Combining the changes in holdings from period 0 to period 1 gives an expression for the trading volume around news:

$$Volume = \frac{\gamma(1 - \gamma)}{A \sigma_\epsilon^2} |\mu^{(1)} - \mu^{(2)}| \quad (2.1)$$

First, note that the trading volume is highest when the population of investors is most evenly distributed between type $i = 1$ and type $i = 2$. Thus, differences of opinion predicts that the trading volume around news is highest when the population of investors reading the news is most diverse.

Second, note that volume in (2.1) is increasing in the difference between the two opinions, $\mu^{(1)}$ and $\mu^{(2)}$. In the news consumption data, the greatest dispersion in possible interpretations of the signal is likely to correspond to the greatest ambiguity of the underlying news story, as more ambiguous news admits a wider range of interpretations. I test this prediction using data on the textual ambiguity of individual news articles.

Third, note that the interaction between ambiguity ($|\mu^{(1)} - \mu^{(2)}|$) and investor diversity ($\gamma(1 - \gamma)$) in predicting trading volume is multiplicative. The effect of investor diversity is

highest when news is most ambiguous (i.e., $|\mu^{(1)} - \mu^{(2)}|$ is largest) and reduces to zero for completely unambiguous news (when $|\mu^{(1)} - \mu^{(2)}| = 0$).

Overall, the predictions of the differences of opinion model are summarized below.

Prediction 3 (Differences of Opinion):

H2.a: Highest trading volume occurs when the population of investors consuming a piece of news is most diverse.

H2.b: Ambiguity of the news article is positively predictive of the trading volume.

H2.c: Diversity of investors reading the news play a larger role in predicting trading volume when the news is more ambiguous.

I test these predictions empirically in Sections 5 and 6. To estimate heterogeneity of investors attending to a piece of news, I classify readers into types using their overall news consumption patterns and techniques from machine learning. I then tabulate the extent to which attention to a particular piece of news is concentrated within a limited set of reader types or dispersed across types. For Predictions H2.b and H2.c, I use machine learning to identify news stories whose text is more subjective and has less polarized sentiment – these are the more ambiguous news. Stories with clear sentiment and fact-based language constitute the sample of less ambiguous news.

3 Data

In order to estimate the extent to which gradual information diffusion and differences of opinion drive trading volume around new information, I need to observe exactly *who* attends to relevant financial information, and *when*. I do so using a unique dataset of clicks on individual news articles by several hundred thousand key finance professionals. These news consumption data are merged with market data to relate trading volume and price formation to attention.

3.1 News Consumption Data

The data on news consumption come from a large financial news database. The database aggregates stories from a variety of sources in real-time, providing a comprehensive landscape of media coverage. The sources of the news include key national and international news wires from major news organizations, company filings, press releases, and content from web sources, including blogs and social media.

The present paper analyzes clicks on 3.5 million financial news articles tagged with U.S. securities over the course of March 22, 2014 to March 2, 2015. The news articles are tagged with individual tickers; there are 12.5 thousand unique tickers represented in the news sample. This consists of all U.S. equities securities, including individual names, indices, open-end funds, and ETFs. There are, on average, 6 thousand new stories tagged with each ticker over the course of the 344 days in the sample. An average article is tagged with 2-3 tickers. Each story receives an average (median) of about 25 (3-4) clicks.

Since timing of reads is integral to the analysis in this paper, I provide summary statistics on the timing of reads relative to the publication of each article in Figure 2. From Panel 1, we can see that the vast majority of reads – 80% – occur within a day of news publication. Frequency of reads decays over the following week, with 4% of reads occurring on the second day after publication, 2% occurring on the third day, etc. A residual 10% of reads captures readers looking at stories more than a week after their publication. Panel 2 displays readership of articles within the first day by hour. 44% of these reads occur within the first hour of the day, with fast decay over the next hours. Similarly, out of the clicks within the first hour of publication, 35% occur within the first 5 minutes, as can be seen from Panel 3. Panel 4 zooms in on the first minute after publication. Since the clicks reflect human readers, very few articles are read immediately in the first five seconds after publication. 39% of the first-minute reads, which is also 2.4% of all reads, occur within 5-15 seconds of when the news becomes available. All in all, the finance professionals in my sample attend to news in a fairly timely manner; however, there is still a meaningful lag between when a piece of information becomes available in the news and when this information disseminates across the landscape of financial market participants.

One caveat is that my dataset of clicks on financial news does not feature consumption of news by algorithmic traders. Some high-frequency traders and quantitative hedge funds consume the news through direct text feeds, and without knowledge of these funds’ individual trading strategies, it is impossible to observe which news they pay “attention” to. However, the current dataset offers a representative view of *human* consumption of financial news by finance professionals.

3.2 Market Data

The news consumption data are merged with market data from several sources. Tests around earnings announcements are conducted using daily trading and return data from the Center for Research in Security Prices (CRSP), and accounting data from Compustat. High frequency tests use trading and return data from QuantQuote.

The earnings announcement tests include all firms for which there are return data in CRSP, earnings numbers in Compustat, and click data in the news consumption dataset. Due to the sample period of the click data, the merged data cover earnings announcements between March 22, 2014 and March 2, 2015. The sample consists of 9,989 earnings announcements by 2,774 firms.

The high frequency tests are run using news tagged with all firms for which there are pricing data in QuantQuote, and shares outstanding and NAICS industry codes in Compustat. The second resolution QuantQuote data include all tickers listed on NYSE and NASDAQ exchanges, and provide prices and numbers of shares traded for each second during the market open. The data are adjusted for splits, dividends, and symbol changes. The merged sample for the high frequency tests covers news releases tagged with 6,134 firms.

4 Methodology

In this section, I discuss the methodology for using the detailed news consumption dataset to construct measures of gradual information diffusion and differences of opinion around individual news events. I capture gradual information diffusion using the precise timestamps of when investors read the news. Differences of opinion are measured using the characteristics of the different investors attending to the news.

4.1 Measuring Gradual Information Diffusion

As a proxy of gradual information diffusion, I look at the normalized Shannon entropy of read times.¹¹ Entropy has a number of applications in fields ranging from thermodynamics to information theory, and has recently been increasingly applied in economics and finance. Philippatos and Gressis (1975) apply entropy to portfolio selection; Stutzer (1996) use entropy to estimate risk-neutral probabilities for derivative pricing; Sims (2003) applies entropy to learning capacity; and Backus, Chernov, and Zin (2014) use entropy to measure pricing kernel’s dispersion. Entropy of a distribution is a natural measure in my context, as it serves to quantify the extent to which readers of the news are heterogeneous either in their timing of clicks or in their reading types.

I measure gradual information diffusion using entropy as follows. For a news article $s_{i,t}$ about firm i at time t , let $\{t_n\}_{n=1}^N$ be N evenly spaced time intervals after t – for example, these might be the 48 one-hour intervals within two days of an earnings announcement. Let $C(t_n)$ denote the set of all clicks on $s_{i,t}$ that occur during the time interval t_n , and define

¹¹See Shannon (1948).

the attention share $p(t_n)$ of the interval t_n as $p(t_n) = |C(t_n)| / \sum_{n=1}^N |C(t_n)|$. Then I use the following proxy for gradual information diffusion:

$$EntropyTime_{i,t} = -\frac{1}{N} \sum_{n=1}^N p(t_n) \log(p(t_n))$$

4.2 Measuring Differences of Opinion

For differences of opinion, the relevant measure is the heterogeneity of the attending investors.¹² To compute investor heterogeneity, I classify finance professionals in my sample into categories based on their overall click histories, in accordance with the intuition that finance professionals with different news consumption patterns likely have different models of the world. I use machine learning techniques to identify 20 disjoint styles of news consumption and classify each of the hundreds of thousands of readers into one of these styles.

First, an important part of the classification problem lies in encoding the readers' click history in a way that is amenable to identifying patterns in their news consumption. Each reader consumes, on average, under 200 of the 3.5 million articles, and each article receives an average of 24 clicks from across more than 400 thousand readers. As a result, encoding readers by their clicks (or absence thereof) on every news article would result in far too sparse a matrix. Before proceeding, this sparse readership matrix must be condensed into a set of meaningful features that would capture a comprehensive representation of each reader's click history. In order to do so, I define the following 66 binary features, which include information on the readers' preferences for specific firms, industries, news sources, and particular types of news, as well as the readers' overall activeness and sophistication:

- **Reading speed (3 features):** For each reader, I compute the incidence of long periods of inactivity as the percentage of lags between consecutive reads that exceed 3 days. I then construct three indicator variables for: frequent readers (those for whom long inactivity occurs less than 1% of the time), moderate readers (those for whom long inactivity occurs 1-5% of the time), and occasional readers (those for whom long inactivity occurs 5-20% of the time). The remaining readers, who see long periods of inactivity more than 20% of the time, are very infrequent consumers of news.
- **Length of stories read (2 features):** I divide the news stories into long (300 words and longer) and short (shorter than 300 words), and compute the number of clicks on

¹²For studies exploring the origins for investor disagreement, see, for example, Cronqvist, Siegel, and Yu (2015) and Chang, Hong, Tiedens, and Wang (2015).

the two types of stories for each reader. The two length features are indicators for readers who prefer long stories (at least 70% of their clicks occur on long stories) and for readers who prefer short stories (at least 70% of their clicks occur on short stories).

- Reading of stale and duplicate stories (10 features):** These features capture the extent to which a reader is prone to consuming old news (stale stories), and in particular reprints of news (duplicate stories). I measure staleness of each story as its textual similarity to preceding stories about the same firm, and duplication as intersection with a single previous story (see Section 5.2 for a detailed discussion of staleness and duplication metrics). Each story is classified into one of five buckets of staleness: stories with staleness $\in [0\%, 20\%]$, $(20\%, 40\%]$, $(40\%, 60\%]$, $(60\%, 80\%]$, and $(80\%, 100\%]$; analogously for duplication. The features denote high (more than one standard deviation above the mean) propensity to read each kind of story. Thus, there are ten features in total: 2 metrics (staleness and duplication) \times 5 buckets each.
- Industry concentration (23 features):** for each industry j of the 23 two-digit NAICS codes, I set $Ind_{i,j}$ equal to 1 if more than 5% of the news stories read by reader i are tagged with firms in industry j , and to 0 otherwise. These 23 features capture the extent to which a reader’s news consumption is concentrated on certain industries.
- Ticker concentration (3 features):** For each reader i , I compute F_i as the number of unique tickers followed by i , scaled by i ’s total number of reads. The readers are then compared against each other: the broad firm focus feature is set to 1 if F_i is more than one standard deviation above the mean, while the narrow firm focus feature is set to 1 if F_i is more than one standard deviation below the mean. The third feature captures whether a reader has a strong preference for a particular firm: it is set to one for any reader who clicks on news about some firm at least twice as often as on news about any single other firm.
- News source concentration (3 features):** For each reader, I compute the number of different news sources from which the reader consumes at least one piece of news, normalized by the reader’s total number of reads. Each reader is then compared against the others, and readers who are at least one standard deviation above the mean in terms of the number of sources are labeled as having a wide news-source focus, while readers who are at least one standard deviation below the mean are labeled as having a narrow focus. Comparing the frequency of the top two sources for each reader, I construct a third feature: readers who read from some source at least twice as frequently as from any one other source are labeled as single-source focused.

- **News source types (16 features):** The news sources are classified into six categories based on type – e.g., one type of sources is press releases, – five categories based on importance, and five categories based on overall attention. For each reader i , feature $S_{i,c}$ is set to 1 if more than 10% of i ’s reads are on news stories published by a source from category c , and to 0 otherwise.
- **Activity level (6 features):** The readers are also classified into six categories based on their historical levels of activity in using the news service.

After representing the readers as points in the 66-dimensional feature space, I sort the readers into types using a randomly selected set of 4,000 readers. This allows me to use a sufficiently representative subset of the dataset to capture its structure, yet keep the problem computationally tractable. For the clustering algorithm, I use affinity propagation, an unsupervised learning technique proposed by Frey and Dueck (2007). Affinity propagation is well suited to the present problem for two reasons. First, this approach forms clusters around datapoints chosen as exemplars, thus identifying a “representative” point for each cluster and facilitating interpretability. Second, the procedure treats all points as potential exemplars, so that every reader is *ex ante* equally likely to be an exemplar, and the most representative readers are chosen. Third, the affinity propagation approach does not rely on a predefined number of clusters, instead identifying the most appropriate number of clusters by iteratively partitioning the dataset. For a novel dataset with relatively unknown structure, the less restrictive approach of leaving the number of clusters flexible is more appealing than pre-specifying an exact number of clusters. Technical details of the affinity propagation algorithm can be found in Appendix B.1.

The resulting clusters can be visualized by projecting the 66-dimensional feature space onto 2 dimensions. For the projection, I use the t-distributed stochastic neighbor embedding technique, introduced by van der Maaten and Hinton (2008).¹³ The results are displayed in Figure 3, with the 21 clusters marked in different colors. The clusters are fairly balanced, with 100-300 points in each of the 21 clusters. To fix ideas, some examples of the cluster exemplars are:

- A reader disproportionately following a single news source, who prefers short stories, follows a single industry, and has historically been moderately active;
- A reader with broad source focus, who has very few long lags between reads, prefers short stories, and has a broad firm focus;

¹³Please refer to Appendix B.2 for detail

- A reader who prefers reading blogs, has a large incidence of long lags between reads, focuses on four industries, and is likely to read stale stories;
- A moderately frequent reader who prefers research reports and short stories, focuses on five industries, is likely to read stale stories, and has historically been quite active.

Having formed the clusters on a subset of the data, I next classify the remaining readers. Recall that the affinity propagation algorithm learns the relative importance of each feature and interactions between them iteratively when forming the clusters. The ensuing classification problem of readers into clusters is best suited to non-linear methods that allow for sufficient flexibility in factoring in interactions between the features.

An intuitive method for visualizing the data and classifying the readers according to a variety of feature combinations is a decision tree. A decision tree repeatedly partitions the data according to one feature per node, until the datapoints at each end-node belong to a single cluster. At each node, the algorithm chooses to partition according to the most informative feature, according to a metric such as Gini impurity or entropy reduction. Figure 4 displays the top few partitions of the decision tree fit to the 4,000 readers sorted into the 21 clusters. Some of the most informative features, chosen as the top nodes, are historical levels of activity, propensity to read blogs, and diversity of news sources that the reader follows. The decision tree classifier performs relatively well on this training dataset. Running the decision tree algorithm on subsets consisting of 90% of the data and testing on the remaining 10%, a technique called cross-validation in the machine learning literature, yields a cross-validation score of 68%, meaning that 68% of the points are classified correctly.

While a decision tree achieves a high degree of accuracy in classifying readers, its performance suffers from the problem of overfitting to the training dataset. Since a decision tree chooses a single feature along which to partition at each node, the method is highly sensitive to small perturbations in the dataset. A more robust approach is using a random forest classifier, which effectively combines a number of decision trees trained on bootstrapped samples from the data and selects from a random subset of candidate features at each node. This approach follows Breiman (2001), and is detailed in Appendix B.3. The random forest classifier achieves a cross-validation score of 80%. The resulting classification of all readers into 21 clusters is used to construct a measure of differences of opinion.

The readers in different clusters represent different styles of attention and investing: they follow a different landscape of industries, have varying amounts of focus, and differ in their levels of activity and sophistication. These differences in the approach to gathering information likely translate to different world-views, leading to differential interpretations of the same news.

My measure of differences of opinion takes advantage of the different information consumption patterns of the identified reader clusters. Let the clusters be indexed by $m \in \{1, \dots, M\}$, and let $c_m(C_{i,s})$ denote the percentage of clicks $C_{i,s}$ on news s about firm i that come from readers classified into cluster m . Then the measure of differences of opinion is:

$$EntropyType_{i,s} = -\frac{1}{\log(M)} \sum_{k=1}^M c_m(C_{i,s}) \log(c_m(C_{i,s})) \quad (4.1)$$

5 Disagreement and Trading Volume

This section estimates the importance of the two models of disagreement in explaining trading volume around news, at two horizons: days around earnings announcements and minutes around individual news articles. Gradual information diffusion is the key driver: the difference between perfect coincidence and perfect dispersion of readership corresponds to a 160% larger increase in trading volume during the two days after earnings announcements, and 400% during the ten minutes after individual news articles. Measures of differences of opinion are substantially less significant in explaining trading volume at both resolutions.

5.1 Trading Volume around Earnings Announcements

In this section, I test the extent to which gradual information diffusion and differences of opinion explain the surge in trading volume around earnings announcements. Measures of gradual information diffusion (dispersion in the timing of attention) and differences of opinion (dispersion in the type of readers) are both predictive of trading volume around the announcement, with the former having a substantially stronger effect.

Trading volume is consistently higher around earnings announcements than in absence of news. Figure 5 plots the daily percentage of shares turned over for the CRSP universe in my sample period of 2014 to 2015. I look between twenty days before and twenty days after each earnings announcement, and aggregate the trading volumes in event time across announcements. In the baseline, approximately 0.6% of shares turn over each day. The turnover is nearly three times higher around the announcement: On the day of an earnings announcement, 1.5% of shares turn over, and this increases further over the next trading day, reaching almost 2% of shares turned over. Trading volume stays elevated for two to three days, after which the market activity comes back to its normal level.

In order to evaluate the extent to which this trading volume spike is related to the two channels of disagreement, I construct the following trading volume and attention variables. For each firm i on announcement date t , let $Volume_{i,t}$ denote the trading volume,

expressed as a percentage of shares turned over, during the day of the announcement, and let $Volume_{i,t+s}$ denote the volume on trading day s after the announcement. For the information set, consider all articles $S_{i,t}$ published about firm i on the date of earnings announcement t . Then let $Clicks_{i,t}$ and $Clicks_{i,t+s}$ denote the number of clicks on articles $S_{i,t}$ during the day t and s trading days later, respectively. For example, $Clicks_{i,t+1}$ includes all clicks by investors who read the earnings news on the next business day after the earnings announcement. All trading volume and click variables are winzorized at the top and bottom 1%.

The tests focus on trading volume and attention on the day of the announcement and the day immediately after, since the spike in trading volume around earnings news occurs on these two dates. In order to capture abnormal trading volume spurred by the news, I take the percentage increase in trading volume from the 20 days preceding the announcement to the announcement window. Namely, I define the trading volume variable as:

$$ImmVolume_{i,t} = \frac{\frac{1}{2}(Volume_{i,t} + Volume_{i,t+1})}{\frac{1}{20} \sum_{s=-20}^{-1} Volume_{i,t+s}} - 1$$

To test whether the abnormal trading volume around earnings announcement is driven mostly by gradual diffusion of the earnings news or differences in its interpretation, I take advantage of two key features of the news click data: the precise timing of the clicks and the knowledge of the clickers' behaviors. Using the measure constructed in Section 4, I estimate the following regression:

$$Volume_{i,t} = \alpha + \beta_1 EntropyTime_{i,t} + \beta_2 EntropyType_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}, \quad (5.1)$$

where the controls X include $Size_{i,t}$, $B/M_{i,t}$, $SUE_{i,t}$, the number of clicks during the two-day announcement window ($|C_{i,t}|$), and year and industry fixed effects. In these low-frequency tests, the measure $EntropyTime_{i,t}$ uses the 48 hours after news publication as the time intervals over which dispersion is computed. Similarly, $EntropyType$ is computed using the clicks within 48 hours of publication, and observing the extent to which these come from different types of readers. The results are presented in Table 1.

Gradual information diffusion, $EntropyTime_{i,t}$ is strongly predictive of trading volume. Going from an entropy value of 0 (corresponding to all clicks falling within the same hour during the two-day post-publication window) to an entropy value of 1 (corresponding to the clicks being evenly split across the 48 hours) corresponds to a 160% increase in abnormal announcement-period trading volume relative to the baseline during the preceding 20 days. The result is robust to the inclusion of a variety of controls, and statistically significant at the 1% level.

Differences of opinion have a milder effect, with a change from entropy value of 0 (clicks only by investors of a single type) to 1 (clicks evenly split between the 21 types of investors) translating to a 50-70% increase in the abnormal trading volume. Furthermore, the effect is significant only at the 5% or 10% level, depending on the exact specification of controls.

5.2 Individual News Articles: Identifying Relevant News Events

The previous section investigates the drivers of daily trading volume around earnings announcements; next, I investigate minute-level trading around precise news articles. To set up the stage for this higher-frequency analysis, I begin by defining a sample of relevant news events. I select news stories that are sufficiently textually novel relative to preceding articles and that receive at least a minimal threshold of attention.

The first screen for relevant news is based on textual novelty. Since the news in my sample is aggregated from a variety of news providers, there are a number of instances of repeated articles published by different sources with varying delays. In these instances, identifying all of the articles as separate news events independently driving trading volume would be misleading. Instead, the goal is to identify the *earliest* dissemination instance of a particular piece of news. To this end, I condition on the individual news articles' textual novelty.

I use a measure of novelty computed following the methodology in Fedyk and Hodson (2015). For each article s tagged with firm i on date t , textual similarity to a preceding article s' tagged with i is computed as the percentage of s 's unique words that appear also in s' :

$$Sim(s, s') = \frac{||s \cap s'||}{||s||},$$

where $|| \cdot ||$ denotes the number of unique words in a set of articles.¹⁴ Then, for each article s , textual novelty is defined as the percentage of unique words in s that are *not* spanned by the closest five preceding articles tagged with the same firm:

$$Novel(s) = 1 - \frac{||s \cap (\cup_{j=1}^5 s'_j(s))||}{||s||}, \quad (5.2)$$

where $\{s'_1(s), \dots, s'_5(s)\}$ are the five most textually similar articles to s .

I limit the sample of relevant news events to news articles that are at least 20% novel, meaning that at least 20% of the words in these articles have not appeared in the closest preceding articles about the same firm. Figure 6 displays the distribution of textual novelty

¹⁴This excludes common stop words such as “a”, “the”, “for”, “where”, etc., and stems all words using the standard stemming algorithm from Porter (1980) (so that words such as “prediction” and “predicted” are represented with the same token, “predict-”).

across the full set of 3.5M articles in the sample. The novelty screen reduces the sample to 1.6M articles.

The second screen for relevant news is based on attention. Since pieces of news that receive little to no attention are unlikely to be relevant for financial markets, I limit the analysis to the set of news article that receive at least one hundred clicks, in total, by the readers in the fifteen relevant industries, and that receive at least ten clicks within the first five minutes of publication. This reduces the news sample to 131.5K relevant articles tagged with 4,078 firms.

5.3 Trading around Individual News Events

In this subsection, I describe the joint dynamics of clicks and trading volume around the individual news articles identified as relevant, and attribute variation in trading volume to measures of gradual information diffusion and differences of opinion.

Trading volume around specific news articles is measured over a ten-minute interval using QuantQuote second-level pre-processed market data. Let $Trading_{i,[t_1,t_2]}$ denote the total trading volume for firm i during the time period from t_1 to t_2 . For a news article s tagged with firm i published at time t , I compute abnormal trading volume as the percent increase in ten-minute trading volume immediately following the publication of the news article relative to the average trading volume over the preceding six non-overlapping ten-minute intervals (i.e. one hour):

$$AbnVolume_{i,s,t} = \frac{Volume_{i,[t,t+10min]}}{\frac{1}{6} \sum_{n=1}^6 Volume_{[t-10n\ min, t-10(n-1)\ min]}} - 1$$

Measures of gradual information diffusion and differences of opinion are constructed following the methodology of Section 4, but now using higher-frequency windows. For gradual information diffusion around article s , I compute $EntropyTime_s$ as entropy of news timing across the 50 twenty-second buckets during ten minutes after news publication. Similarly, for the measure of differences of opinion around article s , $EntropyType_s$, I look at heterogeneity in the types of readers during this ten minute interval post-publication.

In order to measure the extent to which trading volume around individual news articles is driven by gradual information diffusion and differences of opinion, I estimate the following linear regression:

$$AbnVolume_{i,s,t} = \alpha + \beta_1 EntropyTime_s + \beta_2 EntropyType_s + \gamma X_{i,s,t} + \epsilon_{i,s,t} \quad (5.3)$$

where the controls $X_{i,s,t}$ include the total number of clicks on article s within the first ten

minutes of publication, year and hour fixed effects, and firm fixed effects.

Consistent with the evidence from earnings announcements, results at the higher frequency indicate that both gradual information diffusion and differences of opinion are predictive of increased trading volume around individual news events, with gradual information diffusion playing a larger role. Going from completely concentrated to maximally dispersed timing of clicks corresponds to an additional 400% increase in trading volume relative to the pre-news baseline, as can be seen in the first row of Table 2. The result is highly statistically significant, and robust to the inclusion of date, hour, and firm fixed effects. The second row shows the estimates of the effect of differences of opinion: going from fully concentrated to fully dispersed types of readers attending to a piece of news corresponds to an additional 250% increase in short-term trading volume. The effect of differences of opinion, while substantial, is both economically and statistically weaker than that of gradual information diffusion.

5.4 Complementarity Analysis

I confirm that the two channels of disagreement capture distinct aspects of news consumption by considering the interactions between them. I find that the two channels complement each other. On the one hand, timing plays a relatively stronger role when the news is read by a heterogeneous group of investors. On the other hand, diversity of investors reading the news has a larger effect on trading volume when these investors are all looking at the news at the same time.

I begin by considering the differential impact of timing of attention, conditional on the level of investor heterogeneity. To do so, I slice the individual news events into quintiles based on the heterogeneity of attending investors, *EntropyType*. Let $Type_q$ denote the set of all news events whose value of *EntropyType* falls within the q^{th} quintile. I run the following predictive regression within each subsample $Type_q$:

$$\forall s \in Type_q : AbnVolume_{i,s,t} = \alpha^{(q)} + \beta^{(q)} EntropyTime_s + \gamma^{(q)} X_{i,s,t} + \epsilon_{i,s,t}^{(q)}, \quad (5.4)$$

with the full set of controls $X_{i,s,t}$ including the total number of clicks on article s within the first ten minutes of publication, firm size and book-to-market ratio, year and hour fixed effects, and firm fixed effects.

The results, tabulated in Panel 1 of Table 3, show that the differences in the timing of clicks play a stronger role when the population clicking on the news is relatively more heterogeneous. The table displays the estimated coefficients $\beta^{(q)}$, which indicate the effect

of *EntropyTime* within each subsample. The effect of *EntropyTime* is economically sizable and statistically significant at the 1% within the three highest quintiles of *EntropyType*, and much weaker (significant only at the 10% level) in the two lowest quintiles. These results suggest that the relationship between trading volume and gradual information diffusion is stronger when the attention to information comes from a more heterogeneous group of investors. This points to a complementarity between the two channels: differentially informed investors are more likely to trade with each other when they have fundamentally different worldviews.

I consider the complementarity between the two channels further by conditioning on the timing aspect, *EntropyTime*, and observing the effect of investor heterogeneity within each subsample. In particular, in this portion of the analysis, I split the individual news sample into quintiles based on the level of *EntropyTime*, and let $Time_q$ denote the set of all news events in quintile q . I then run the following predictive regression within each subsample:

$$\forall s \in Time_q : AbnVolume_{i,s,t} = \alpha^{(q)} + b^{(q)} EntropyType_s + \gamma^{(q)} X_{i,s,t} + \epsilon_{i,s,t}^{(q)}, \quad (5.5)$$

where the set of controls $X_{i,s,t}$ again includes the total number of clicks on article s within the first ten minutes of publication, firm size and book-to-market ratio, year and hour fixed effects, and firm fixed effects.

The estimates of the coefficients $b^{(q)}$, displayed in Panel 2 of Table 3, indicate that heterogeneity of investors reading the news contributes to disagreement more when these investors all get the information at the same time. The estimated effect of *EntropyType* is large and highly statistically significant within the bottom two quintiles of *EntropyTime*, where the clicks are minimally dispersed across time. By contrast, the coefficients on *EntropyType* from equation (5.5) are generally smaller and not statistically significant across all three of the top *EntropyType* quintiles (with the exception of quintile 5, where the coefficient is marginally significant only at the 10% level). Conceptually, this indicates that disagreement between investors from diverse worldviews is stronger when these individuals are reading the same information at the same time. Consuming the news signal at the same time prompts contemporaneous reactions, which are more likely to vary and spur trade when the reacting individuals come from different backgrounds.

Overall, the conditional analysis reveals that the effect of each of the two channels of disagreement depends in part on the other channel. This highlights complementarity between the two channels: gradual information diffusion plays a stronger role for more diverse investors, and differences of opinion matter more when the investors see the news at the

same time. The two channels of disagreement are both operative in driving trading around informational releases; they capture different – and complementary – aspects of news consumption.

6 Trading Volume and News Ambiguity

I investigate how the relationship between clicks on news and trading volume changes with textual characteristics of the news, testing whether heterogeneity of opinions matters more when the news is less straightforward. I introduce the methodology for measuring ambiguity of news, and then present evidence that differences of opinion are more important in driving trading volume around relatively more ambiguous news events.

6.1 Measuring News Ambiguity

In order to classify news events as textually clear versus textually ambiguous, I characterize news articles along two dimensions. The first is the extent to which each article’s positive or negative sentiment is conveyed in clear language, and the second is the article’s concentration on hard (factual) versus soft (opinion) information. Overall ambiguity is computed as the average of these two proxies.

For the sentiment-based measure, I use a sentiment analyzer trained on a dataset of approximately 10,000 articles tagged by human experts as positive, negative, or neutral. The training data are selected to be representative of the full sample of news articles across sources, topics, and tagged tickers. Each article is annotated by multiple experts and classified according to the majority vote when at least 75% of the annotators agree; articles where no agreement can be reached are dropped from the training set. The experts are provided with an annotation rubric and examples of positive, negative, and neutral articles. The experts’ annotations are checked for speed and answer patterns, and data from experts who answer exceptionally quickly or display patterns of identical answers are dropped from the calculations.

In order to learn the attributes that are associated with particular sentiment, articles are represented as vectors of features, and a binary classification model is built on the feature vectors. The features representing the articles include the following: story length; number of topics covered; indicators for particular unigrams, bigrams, and trigrams in the text; the similarity of the article’s text to the distribution of text in the full sample of financial news; the complexity of the article’s syntactic structure; the density of the article’s semantic concept graph; and indicators for particular patterns of syntactic structure and semantic

relationships. The sentiment question is then posed as a binary classification problem which is solved with a Support Vector Machine (a maximum-margin, Gaussian kernel-based classifier; see Cortes and Vapnik (1995)). The resulting classification of articles into sentiment classes achieves a cross-validation score of 86.3% on the training set. The estimated model is then used to classify any incoming articles.

Sentiment-based ambiguity is computed from the sentiment classifier as the certainty with which the procedure determines the article’s sentiment. Effectively, ambiguity is the inverse of the distance of a given article from the separating hyperplane for its class, normalized to be between 0 and 1. For example, a positive article that is very far in the positive space would have lower sentiment-based ambiguity than a positive article that is very close to the decision line.

Analogous methodology is used for estimating the extent to which the article’s content consists of hard versus soft information. The same training set is tagged as either hard factual information or soft opinion. Then, a classifier is built to predict the type of information from article features. The model’s cross validation score is similar to the sentiment classifier, at 84.6%. The information-type ambiguity is then computed as the distance to the separation between the two classes interacted with an indicator for the classes (1 for soft and -1 for hard information), normalized to be between 0 and 1.

Overall ambiguity of the articles is computed as the average of the two ambiguity metrics. The distribution of the ambiguity scores is right-skewed, so I take a threshold of 75% or more to label an article as ambiguous. A total of 58,000 articles are classified in this way: 25,000 of them labeled as textually ambiguous and 33,000 labeled as textually clear.

Examples of clear news include the following headlines:

- “Deutsche Bank is still recovering from 2015 fines, CEO says after it posts third consecutive annual loss”
- “AT&T earnings: 78 cents per share, vs expected EPS of 65 cents”
- “Qualcomm fined \$1.2 billion for paying Apple to use its mobile chips”

By comparison, below are some examples of ambiguous news:

- “JPMorgan Holds Law Firms’ Feet to the Fire on Diversity”
- “Fuji film announces X-A5 mirrorless camera and first X-series power zoom”
- “The Amazon, Berkshire and JP Morgan Chase Health Care Company Might Be the Perfect Industry Disruption”

6.2 Trading Volume around Ambiguous News

I repeat the primary tests linking trading volume surges around individual news articles to the two measures of disagreement across two samples: for textually clear news articles and for textually ambiguous news articles. The results indicate that differences of opinion plays a stronger role for ambiguous news, but only gradual information diffusion is predictive of trading volume surges around clear news.

To begin with, I look at average trading volume surges across the two samples. Consistent with prediction H2.b, trading volume is higher around more ambiguous news. The increase in the ten-minute trading volume immediately after the news is 22% after textually clear news, and 25% after textually ambiguous news. The difference is significant at the 5% level, with a t-statistic of 2.03.

In order to evaluate the extent to which the effect of investor heterogeneity differs between clear and ambiguous news (prediction H2.c), I estimate (5.3) separately on the sample of textually clear news and the sample of textually ambiguous news. The results are reported in Table 4.

The relative performance of the two channels of news consumption differs across the news samples. In the sample of textually straightforward news, displayed in Panel 1 of Table 4, the point estimate of the effect of *EntropyTime* is substantially higher than that of *EntropyType*, and much more statistically significant. However, in the sample of ambiguous news, presented in Panel 2 of Table 4, *EntropyType* is just as predictive of trading volume as *EntropyTime*, both economically and statistically.

Looking across the samples, the point estimates of the effect of gradual information diffusion (measured by *EntropyTime*) are larger in the sample of straightforward news than in the sample of ambiguous news (an effect size of 440% as compared to 370%). In contrast, results point to the differences of opinion channel (captured by *EntropyType*) being more operative for textually ambiguous news (an effect of a 350% increase in volume for ambiguous news, as compared to only 200% for straightforward news). These results are consistent with prediction H2.c: the dispersion of opinions is more predictive of market activity when the underlying information admits a wider range of interpretations.

7 Conclusion

This paper uses a uniquely detailed dataset of news consumption by key finance professionals to evaluate the extent to which increased trading volume around news events is driven by gradual information diffusion and differences of opinion. I find that disagreement

induced by differential *timing* of news consumption is strongly predictive of trading volume at both daily and minutely horizons. Disagreement regarding the meaning of a piece of news read by a variety of investors is less significant in explaining the surge in trading volume around news.

The results of this paper highlight the importance of attention in the increasingly prolific modern news environment. Despite the push for transparency bringing more and more information to the public domain, informational advantages persist – only here, they take the form of speedy attention to public news rather than possession of private news. As a result, even when we restrict our attention to public information, trading volume in the markets is largely driven by some investors getting the information before others.

References

- [1] Ajinkya, B. B., R. K. Atiase, and M. J. Gift (1991). “Volume of trading and the dispersion in financial analysts’ earnings forecasts.” *Accounting Review*, 66(2): 389-401.
- [2] Anderson, E. W., E. Ghysels, and J. L. Juergens (2005). “Do heterogeneous beliefs matter for asset pricing?” *Review of Financial Studies*, 18(3): 875-924.
- [3] Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013). “Value and momentum everywhere.” *Journal of Finance*, 68(3): 929-985.
- [4] Atiase, R. K. and L. S. Bamber (1994). “Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry.” *Journal of Accounting and Economics*, 17(3): 309-329.
- [5] Backus, D., M. Chernov, and S. Zin (2014). “Sources of entropy in representative agent models.” *Journal of Finance*, 69(1): 51-99.
- [6] Bali, T.G., A. Bodnaruk, A. Scherbina, and Y. Tang (2017). “Unusual News Flow and the Cross-Section of Stock Returns.” *Management Science*, forthcoming.
- [7] Ball, R. and P. Brown (1968). “An empirical evaluation of accounting income numbers.” *Journal of Accounting Research*, 6(2): 159-178.
- [8] Bamber, L. S., O. E. Barron, and D. E. Stevens (2011). “Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research.” *Contemporary Accounting Research*, 28(2): 431-471.

- [9] Banerjee, S., Davis, J., and Gondhi, N. (2017). “When transparency improves, must prices reflect fundamentals better?” Working paper.
- [10] Banerjee, S. and I. Kremer (2010). “Disagreement and learning: Dynamic patterns of trade.” *Journal of Finance*, 65.4: 1269-1302.
- [11] Barberis, N., A. Shleifer, and R. Vishny (1998). “A model of investor sentiment.” *Journal of Financial Economics*, 49(3): 307-343.
- [12] Bauguess, S. W., J. Cooney, and K. W. Hanley (2013). “Investor demand for information in newly issued securities.” Working paper.
- [13] Ben-Rephael, A., Z. Da, and R. D. Israelsen (2017). “It depends on where you search: A comparison of institutional and retail attention.” *Review of Financial Studies*, 30(9): 3009-3047.
- [14] Bernard, V. L. and J. K. Thomas (1989). “Post-earnings-announcement drift: Delayed price response or risk premium?” *Journal of Accounting Research*, 27: 1-36.
- [15] Breiman, L. (2001). “Random forests.” *Machine Learning*, 45(1): 5-32.
- [16] Carlin, B. I., F. A. Longstaff, and K. Matoba (2014). “Disagreement and asset prices.” *Journal of Financial Economics*, 114(2): 226-238.
- [17] Chang, Y. C., H. Hong, L. Tiedens, N. Wang, and B. Zhao (2015). “Does diversity lead to diverse opinions? Evidence from languages and stock markets.” Working paper.
- [18] Cookson, J. A. and M. Niessner (2016). “Why don’t we agree? Evidence from a social network of investors.” Working paper.
- [19] Cortes, C. and V. Vapnik (1995). “Support-vector networks.” *Machine Learning*, 20(3): 273-297.
- [20] Cronqvist, H., S. Siegel, and F. Yu (2015). “Value versus growth investing: Why do different investors have different styles?” *Journal of Financial Economics*, 117(2): 333-349.
- [21] Da, Z., J. Engelberg, and P. Gao (2011). “In search of attention.” *Journal of Finance*, 66(5): 1461-1499.
- [22] Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997), “Measuring mutual fund performance with characteristic-based benchmarks.” *Journal of Finance*, 52(3): 1035-1058.

- [23] deHaan, E., T. Shevlin, and J. Thornock. (2015). "Market (in)attention and the strategic scheduling and timing of earnings announcements." *Journal of Accounting and Economics*, 60(1): 36-55.
- [24] DellaVigna, S. and J. M. Pollet (2009). "Investor inattention and Friday earnings announcements." *Journal of Finance*, 64(2): 709-749.
- [25] Diether, K. B., C. J. Malloy, and A. Scherbina (2002). "Differences of opinion and the cross section of stock returns." *Journal of Finance*, 57(2): 2113-2141.
- [26] Drake, M. S., D. T. Roulstone, and J. A. Thornock (2012). "Investor information demand: Evidence from Google searches around earnings announcements." *Journal of Accounting Research*, 50(4): 1001-1040.
- [27] Drake, M. S., D. T. Roulstone, and J. R. Thornock (2015). "The determinants and consequences of information acquisition via EDGAR." *Contemporary Accounting Research*, 32(3): 1128-1161.
- [28] 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.
- [29] Fama, E. F. and K. R. French (2012). "Size, value, and momentum in international stock returns." *Journal of Financial Economics*, 105(3): 457-472.
- [30] Fedyk, A. and J. Hodson (2015). "When can the market identify stale news?" Working paper.
- [31] Foster, G., C. Olsen, and T. Shevlin (1984). "Earnings releases, anomalies, and the behavior of security returns." *Accounting Review*, 59(4): 574-603.
- [32] Frazzini, A. (2006). "The disposition effect and underreaction to news." *Journal of Finance*, 61(4): 2017-2046.
- [33] Frey, B. J. and D. Dueck (2007). "Clustering by passing messages between data points." *Science*, 315.5814: 972-976.
- [34] Garfinkel, J. A. (2009). "Measuring investors' opinion divergence." *Journal of Accounting Research*, 47(5): 1317-1348.
- [35] Garfinkel, J. A. and J. Sokobin (2006). "Volume, opinion divergence, and returns: A study of post-earnings announcement drift." *Journal of Accounting Research*, 44(1): 85-112.

- [36] Giannini, R. C., P. J. Irvine, and T. Shu (2015). “The convergence and divergence of investors’ opinions around earnings news: Evidence from a social network.” Working paper.
- [37] Grinblatt, M. and T. J. Moskowitz (2004). “Predicting stock price movements from past returns: The role of consistency and tax-loss selling.” *Journal of Financial Economics*, 71(3): 541-579.
- [38] Golub, B. and M. P. Jackson (2012). “How homophily affects the speed of learning and best-response dynamics.” *Quarterly Journal of Economics*, 127(3): 1287-1338.
- [39] Harris, M. and A. Raviv (1993). “Differences of opinion make a horse race.” *Review of Financial Studies*, 6(3): 473-506.
- [40] Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). “Driven to distraction: Extraneous events and underreaction to earnings news.” *Journal of Finance*, 64(5): 2289-2325.
- [41] Hirshleifer, D., and S. H. Teoh (2003). “Limited attention, information disclosure, and financial reporting.” *Journal of Accounting and Economics*, 36(1): 337-386.
- [42] Hong, H., T. Lim, and J. C. Stein (2000). “Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies.” *Journal of Finance*, 55(1): 265-295.
- [43] Hong, H. and J. C. Stein (1999). “A unified theory of underreaction, momentum trading, and overreaction in asset markets.” *Journal of Finance*, 54(6): 2143-2184.
- [44] Hong, H. and J. C. Stein (2007). “Disagreement and the stock market.” *Journal of Economic Perspectives*, 21(2): 109-128.
- [45] Kandel, E., and N. D. Pearson (1995). “Differential interpretation of public signals and trade in speculative markets.” *Journal of Political Economy*, 103(4): 831-872.
- [46] Karpoff, J. M. (1986). “A theory of trading volume.” *Journal of Finance*, 41(5): 1069-1087.
- [47] Kaniel, R., S. Liu, G. Saar, and S. Titman (2012). “Individual investor trading and return patterns around earnings announcements.” *Journal of Finance*, 67(2): 639-680.
- [48] Kondor, P. (2012). “The more we know about the fundamental, the less we agree on the price.” *Review of Economic Studies*, 79(3): 1175-1207.

- [49] Lawrence, A., J. P. Ryans, E. Sun, and N. Laptev, (2016). “Yahoo Finance search and earnings announcements.” Working paper.
- [50] Lumsdaine, R. L. (2010). “What the market watched: Bloomberg news stories and bank returns as the financial crisis unfolded.” Working paper.
- [51] Madsen, J. M. and M. Niessner (2016). “Is investor attention for sale? The role of advertising in financial markets.” Working paper.
- [52] McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). “Birds of a feather: Homophily in social networks.” *Annual Review of Sociology*, 27(1): 415-444.
- [53] Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012). “Time series momentum.” *Journal of Financial Economics*, 104(2): 228-250.
- [54] Niessner, M. (2015). “Strategic disclosure timing and insider trading.” Working paper.
- [55] Philippatos, G.C. and N. Gressis (1975). “Conditions of equivalence among EV, SSD, and EH portfolio selection criteria: the case for uniform, normal and lognormal distributions.” *Management Science*, 21(6): 617-625.
- [56] Porter, M. F. (1980). “An algorithm for suffix stripping.” *Program*, 14(3): 130-137.
- [57] Shannon, C. E. (1948). “A Mathematical theory of communication.” *Bell System Technical Journal*, 27: 656-715.
- [58] Rouwenhorst, K. G. (1998). “International momentum strategies.” *Journal of Finance*, 53(1): 267-284.
- [59] van der Maaten, L. and G. Hinton (2008). “Visualizing data using t-SNE.” *Journal of Machine Learning Research*, 9: 2579-2605.

Appendix A Additional Analyses: Attention and Prices

This section explores price dynamics to provide additional evidence for gradual information diffusion driving disagreement around news. In particular, consistent with gradual information diffusion, I find that delayed attention is predictive of delayed price adjustment at a variety of horizons: within minutes of a news release, within days of an earnings announcement, and even at the level of traditional monthly return momentum.

A.1 Price Dynamics around Individual News Articles

This subsection documents a high-frequency price dynamics result consistent with gradual information diffusion. Looking at prices within minutes of publication of individual news articles, I estimate the extent to which price variance is concentrated immediately after a piece of news, and how this relates to the immediacy of investors' attention to the news. I find that price variance is more immediate when a larger fraction of attention is immediate.

I measure immediacy of the price variance as follows. For a news article s about firm i published during second t , take the ratio of the variance in second-level prices of i during the first minute following t to the variance in second-level prices during the five minutes following t . In particular, let $p_{i,t+t'}$ denote the closing price of firm i 's stock during second $t + t'$. Then the share of immediate price variance is defined for two immediacy windows – the first 60 second and the first 120 seconds:

$$\text{For } \tau \in \{60, 120\} : \text{ImmVar}_{s,i,\tau} = \frac{\text{Variance}\{p_{i,t}, \dots, p_{i,t+\tau}\}}{\text{Variance}\{p_{i,t}, \dots, p_{i,t+300}\}}$$

Immediate attention is defined analogously, as the ratio between the number of clicks on s that occur immediately (within the first 30 seconds or within the first 60 seconds) to the number of clicks that occur anywhere in the five minute interval following s . In particular, for article s tagged with firm i released during second t , let $C_{s,[t,t+r]}$ denote the set of clicks on s that occur between the release second t and r seconds later. Then the immediate attention proxy is measured using two windows, $r \in \{30, 60\}$:

$$\text{ImmClicks}_{s,r} = \frac{|C_{s,[t,t+r]}|}{|C_{s,[t,t+300]}|}$$

In order to estimate the relationship between the immediacy of price variance and the immediacy of attention, I regress $\text{ImmVar}_{s,\tau}$ on $\text{ImmClicks}_{s,r}$ for the different windows τ

and r :

$$ImmVar_{s,i,\tau} = \alpha + \beta ImmClicks_{s,r} + \gamma X + \epsilon_s, \text{ for } (\tau, r) \in \{(30, 30), (30, 60), (60, 60), (60, 120)\},$$

where controls X include article length, total number of clicks, and firm, date, and hour fixed effects.

The results display a consistent relationship between the immediacy of attention and the immediacy of price variance, as displayed in Table 5. A 10% increase in the percentage of clicks occurring within the first minute after article publication corresponds to a 3% larger share of immediacy price variance within the first one to two minutes. This supports prediction H1.b of the gradual information diffusion model: that the size of the immediate price move increases with the share of immediate attention.

A.2 Post-Earnings-Announcement Drift

I demonstrate that the post-earnings-announcement drift is strongest when attention to the earnings news is most delayed. The effect of delayed attention on price formation around earnings announcements supports the findings that trading around the announcements is largely driven by disagreement between early-informed and late-informed investors.

A sizable literature beginning with Ball and Brown (1968) discusses the post-earnings-announcement drift: an upward (downward) drift in abnormal returns following positive (negative) earnings surprises. Bernard and Thomas (1989) investigate whether the drift is driven by a risk premium or a delay in the response to earnings news, and find evidence consistent with the latter. Below, I provide evidence that, consistent with gradual information diffusion driving disagreement and trading volume around earnings news, the post-earnings-announcement drift is greater when attention to news is slower.

The precise prediction of gradual information diffusion for the earnings announcement drift is that the serial correlation in returns is maximized at an interior point, where there is an even distribution of clicks across immediate and delayed (prediction H1.c). However, the distribution of clicks by day after news publication, displayed in Panel 1 of Figure 3, indicates that the vast majority of clicks – 80% – occur on the first day. It is relatively rare to observe attention to news with a delay of a full day or more. As a result, a simplified version of H1.c applicable to the earnings announcement drift is as follows: the drift is stronger when a smaller percentage of attention to the news is immediate.

Throughout the analysis, I compute abnormal (characteristic-adjusted) return for firm i on date s is defined as:

$$AbnRet_{i,s} = Ret_{i,s} - DGTWRet_{i,s},$$

where $Ret_{i,s}$ is the raw return for firm i on date s , and $DGTWRet_{i,s}$ is the value-weighted return of a portfolio of stocks in the same size, value, and momentum quintiles as i (see Daniel, Grinblatt, Titman, and Wermers (1997)). For each earnings announcement by firm i on date t , let $CAR_{i,[t+2,t+20]}$ denote the additive cumulative abnormal return from the second to the twentieth day after the announcement:

$$CAR_{i,[t+2,t+20]} = \sum_{s=t+2}^{t+20} AbnRet_{i,s}$$

I follow the methodology originally introduced by Foster, Olsen, and Shevlin (1984) for measuring the post-earnings-announcement drift. For each earnings announcement t of firm i in fiscal quarter q , I rank $SUE_{i,t}$ against the distribution of SUE in the preceding fiscal quarter $q - 1$. Ranking earnings surprises relative to those from the preceding quarter rather than the current fiscal quarter avoids the look-ahead bias stemming from some firm reporting earnings later than others. Each announcement (i, t) is then placed into a quintile bin according to its ranking relative to the prior quarter earnings surprises.

I also sort announcements based on attention. For each announcement (i, t) , I compare the share of immediate attention around that announcement, $ImmClicks_{i,t}$, against the distribution of immediate attention shares in the preceding fiscal quarter. The announcements are thus sorted into quintiles based on attention, analogously to the sort on SUE .

I measure the post-earnings-announcement drift within each quintile of attention. For each of the twenty five double-sorted attention and earnings surprise portfolios, I take an equal-weighted average of $CAR_{i,[t+2,t+20]}$ over the earnings announcements (i, t) within the portfolio. The rows of Table 6 display the relationship between SUE and the abnormal returns within a particular attention quintile.

Following Foster, Olsen, and Shevlin (1984), statistical significance is determined by comparing the observed average CAR ($ACAR$) for each portfolio against an empirical distribution of $ACAR$ for portfolios drawn from the same $ImmClicks$ quintile. Since each portfolio consists of approximately 300 observations, I draw 300 firm-announcement combinations from the same quintile of $ImmClicks$ and compute the corresponding $ACAR$; I repeat this process 1,000 times and compute the percentage of times when the simulated $ACAR$ is as extreme as the observed value. The estimated difference between the highest and lowest SUE quintiles, displayed in the last column of Table 6, is analogously compared against simulated differences.

The results indicate that there is a significant post-earnings-announcement drift only when the attention to the firm's news is relatively less immediate (i.e., relatively more delayed). This pattern supports prediction H1.c of the gradual information diffusion model,

indicating that the return continuation following earnings announcements is driven by delayed attention of some investors to the earnings news.

A.3 Attention and Momentum

This subsection looks at monthly frequency, and investigates the cross-sectional relationship between return momentum and the speed of attention to news. I find that return momentum is highest when attention to news is most delayed. Analogously to the previous subsections, this finding supports the gradual information diffusion model of disagreement around earnings news.

Return momentum is the widely documented empirical finding that securities that have performed well over the prior 6-12 months continue to outperform relative to those that did poorly, for the next 6-12 months. This result has been documented to hold across geography (see Rouwenhorst (1998) and Fama and French (2012)) and asset class (see Moskowitz, Ooi, and Pedersen (2012) and Asness, Moskowitz, and Pedersen (2013)). A number of explanations have been proposed for return momentum, including gradual information diffusion (see Hong and Stein (1999)), investors holding erroneous beliefs in trending or reversing regimes (Barberis, Shleifer, and Vishny (1998)), and a disposition effect induced by loss aversion (Frazzini (2006)). In this section, I test whether gradual information diffusion is related to momentum by estimating the cross-sectional relationship between a firm’s return momentum and the speed of attention to the firm’s news.

Following a common methodology in the literature (see, e.g., Grinblatt and Moskowitz (2004) or Asness, Moskowitz, and Pedersen (2013)), I measure momentum for each firm in the sample as the serial correlation in that firm’s abnormal monthly returns and the cumulative abnormal returns over the preceding 12-months, skipping the most recent month. In particular, for each firm i and month t , let $AbnRet_{i,[t1,t2]}$ denote firm i ’s cumulative return over months $t1$ to $t2$, adjusted for the equal-weighted market return over the same time period. Then I define momentum for firm i , $Momentum_i$, as the correlation between the series $AbnRet_{i,[t,t]}$ and the lagged series $AbnRet_{i,[t-2,t-12]}$. Since the attention data span the period of March 2014 through March 2015, I construct $Momentum_i$ using $t \in \{March\ 2014, \dots, December\ 2015\}$.

To measure the relationship between return momentum and attention, I define for each firm the following attention proxies, computed over the full sample from March 2014 to March 2015:

- $MeanTimeLag_i$ ($MedTimeLag_i$): average (median) time lag, in hundreds of seconds, from publication to click, across all clicks on articles tagged with firm i ;

- *PercentDay_i* (*PercentWeek_i*): the percentage of clicks on articles tagged with firm i that occur within a day (a week) of publication.

Since momentum varies with firm size (see Hong, Lim, and Stein (2000)), and smaller firms receive attention with a larger delay, I compute the adjusted proxies as residuals from regressions on log market capitalization and NAICS industry dummies, normalized to mean zero and standard deviation one for comparability across proxies.

The results indicate that slower attention to news corresponds to higher return momentum. Table 7 reports the coefficients from linear regressions of momentum against the raw and adjusted attention proxies. For all proxies except for median lag to read, the relationship is strongly significant, regardless of using raw or size- and industry- adjusted proxies. The results are also economically significant, indicating that a hundred second increase in the average (median) time from publication to click corresponds to an increase in the serial correlation in monthly returns of 7% (17%), and a 10% increase in the percentage of clicks occurring more than a day (a week) after article publication predicts a 25% (35%) increase in return momentum. These findings are consistent with the evidence on the post-earnings-announcement drift in the previous subsection, and further support hypothesis H1.c of gradual information diffusion: gradual diffusion of information across news readers generates serial correlations in returns.

Appendix B Technical Details

B.1 Clustering Readers by News Consumption Patterns: Affinity Propagation

In this section, I briefly present the affinity propagation method for clustering readers according to their news consumption patterns. For further detail on this methodology, please refer to Frey and Dueck (2007).

Let I denote the set of datapoints to be clustered, and let $s(i, k)$ denote the similarity between points $i, k \in I$. In this paper, $s(i, k)$ is the negative Euclidean distance between the readers in the 66-dimensional feature space. Hence, the range of $s(i, k)$ is between -66 and 0.

Affinity propagation chooses exemplars and associated clusters through an iterative procedure that updates pairwise measures of representability (the extent to which point k is suitable as an exemplar for point i , relative to all other available exemplars) and availability (the extent to which point k is available as an exemplar given accumulated support from

other points' preference for k as exemplar). Availability $a(i, k)$ is initialized at 0 for all pairs of datapoints (i, k) . The iterative updating process then proceeds as follows.

$$r(i, k)^{(t)} = \lambda r(i, k)^{(t-1)} + (1 - \lambda) \left[s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\} \right] \quad (\text{B.1})$$

$$a(i, k)^{(t)} = \lambda a(i, k)^{(t-1)} + (1 - \lambda) \left[\min\{0, r(k, k) + \sum_{i' \notin \{i, k\}} \max(0, r(i', k))\} \right] \quad (\text{B.2})$$

Effectively, representability $r(i, k)$ increases in the similarity of candidate exemplar k to point i and decreases in the similarity of i to other points and their availability as potential exemplars. Availability $a(i, k)$ of k as an exemplar increases in $r(k, k)$ – the extent to which k wants to be its own exemplar – and decreases in the suitability of other points as exemplars for k . The array of parameters $r(k, k)$ is set by the researcher to indicate a preference for a large number of finer clusters versus a small number of larger clusters. In the main analysis, I set $r(k, k) = -200, \forall k \in I$, which produces 20 clusters.

The other free parameter is the dampening factor λ , included to avoid large oscillations in the optimization problem. The analysis is conducted setting $\lambda = 0.9$.

B.2 Reader Type Visualization: t-Distributed Stochastic Neighbor Embedding

In this section, I describe the t-distributed stochastic neighbor embedding technique for nonlinear dimensionality reduction, which is used for visualizing the high-dimensional space of readers in two dimensions. For further details on this methodology, please consult van der Maaten and Hinton (2008).

First, we represent the readers as points in a 66-dimensional space of features, X . For any two points $x_i, x_j \in X$, let $\|x_i - x_j\|^2$ denote the Euclidean distance between x_i and x_j . Then define $p_{j|i}$ as:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}; \quad p_{j|j} = 0 \quad (\text{B.3})$$

The interpretation of $p_{j|i}$ is the probability of point x_j being chosen as the closest neighbor to x_i , when the neighbors are picked in proportion to their probability density under a Gaussian centered at x_i . The variance σ_i^2 is chosen such that perplexity is the same around each i :

$$\forall i, k : \text{Perp}(P_i) = 2^{-\sum_j p_{j|i} \log_2(p_{j|i})} = \text{Perp}(P_k) = 2^{-\sum_j p_{j|k} \log_2(p_{j|k})}$$

Perplexity can be interpreted as the effective number of neighbors, so that roughly the same number of neighbors is considered around each point, by setting higher variance σ_i^2 in less dense regions. In the representation I produce, perplexity is set to a default value of 30.

The conditional probabilities defined in (B.3) are converted into symmetric total probabilities as follows:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}, \quad (\text{B.4})$$

where n is the number of readers.

The target low-dimensional space, which in my case is two-dimensional, is likewise represented by probabilities proportional to similarities between the points. But in this case, the tSNE uses the Student t-distribution, rather than the Gaussian distribution, as the heavier tails of the Student t-distribution help to fit distant points into the lower-dimensional space without inducing excessive crowding among the nearer points. Thus, for two points y_i, y_k in the two-dimensional space Y , define:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \quad (\text{B.5})$$

In order to represent the high-dimensional points $\{x_i, \dots, x_n\}$ in the low-dimensional space Y , the tSNE procedure chooses the points $\{y_1, \dots, y_n\}$ so as to minimize the Kullback-Leibler divergence of the induced probability distribution Q from the distribution P :

$$\{y_1^*, \dots, y_n^*\} = \arg \min_{y_1, \dots, y_n} KL(P||Q) = \arg \min_{y_1, \dots, y_n} \sum_{i=1}^n \sum_{j=1}^n p_{ik} \log\left(\frac{p_{ij}}{q_{ij}}\right) \quad (\text{B.6})$$

The optimization is performed using the gradient descend method with default parameters for the maximum number of iterations, learning rate (the rate at which new gradient values are incorporated at each iteration), momentum (the extent to which previous updates are incorporated at each iteration), and initial exaggeration (inflation of early values of p_{ij} for tighter, widely separated clusters).

B.3 Reader Classification into Clusters: Random Forest

In this subsection, I briefly describe the random forest classification algorithm for classifying the remaining readers.¹⁵ The reader-type categories are constructed using affinity propagation clustering on a subset of the sample. This reduces computational complexity

¹⁵For more details on random forests and their convergence properties, see Breiman (2001).

of the clustering step, but leaves the problem of classifying the remaining readers into the newly defined clusters.

The most intuitive classification method, which highlights the relative importance of the various features in partitioning the space into clusters, is a decision tree. A decision tree sequentially splits the space on the features, at each node choosing the feature that is most informative for the classification, according to the selected criterion (e.g., according to minimizing entropy or Gini impurity). For example, in the top of the decision tree for the reader classification problem, displayed in Figure 4, the first node splits the data according to historical level of activity, indicating that the readers' propensity to be quite active is most informative in partitioning the data into clusters.

While decision trees are appealing in their simplicity and interpretability, they have the drawback of high variance, meaning that they are highly sensitive to small perturbations in the training data, leading to a tendency to overfit. To mitigate this, the technique of tree bagging averages over predictions from multiple trees. In particular, tree bagging repeatedly bootstraps, with replacement, a random training set from the available data, and builds a decision tree classifier. Then, for each data point x , the overall prediction is taken as the majority vote from the trees whose training sets do not include x .

While tree bagging reduces the overfitting problem relative to a single decision tree, the trees built on subsets of the training data are likely to be highly correlated if the same features are chosen in the early nodes of every tree. In order to minimize correlation between the trees, random forest classifiers incorporate random split selection: when building each tree, at every node, instead of choosing among all features, the algorithm chooses among a randomly selected subset of the features. This methodology further reduces the sensitivity of the algorithm to the particular training data used.

The classification of readers into clusters is performed by a random forest classifier built with 250 trees, using Gini impurity to choose among 8 randomly selected features at every node.

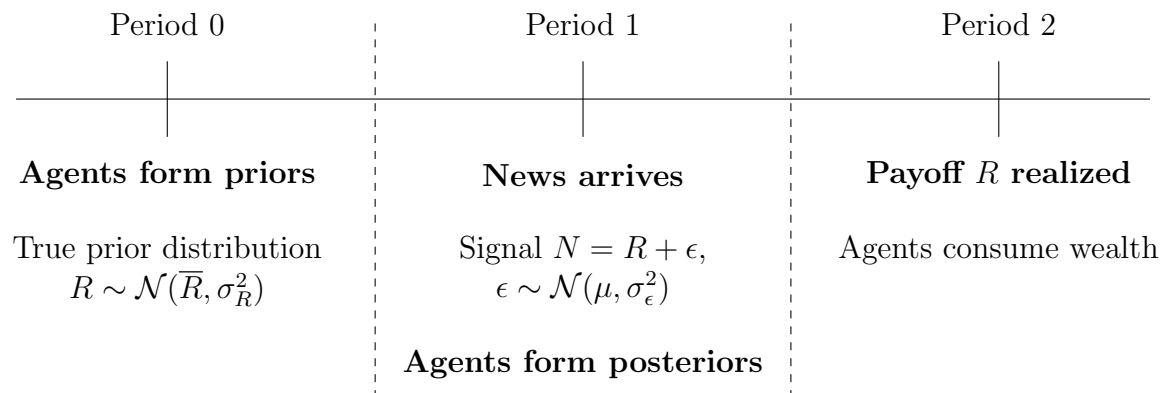


Figure 1: Model timeline.

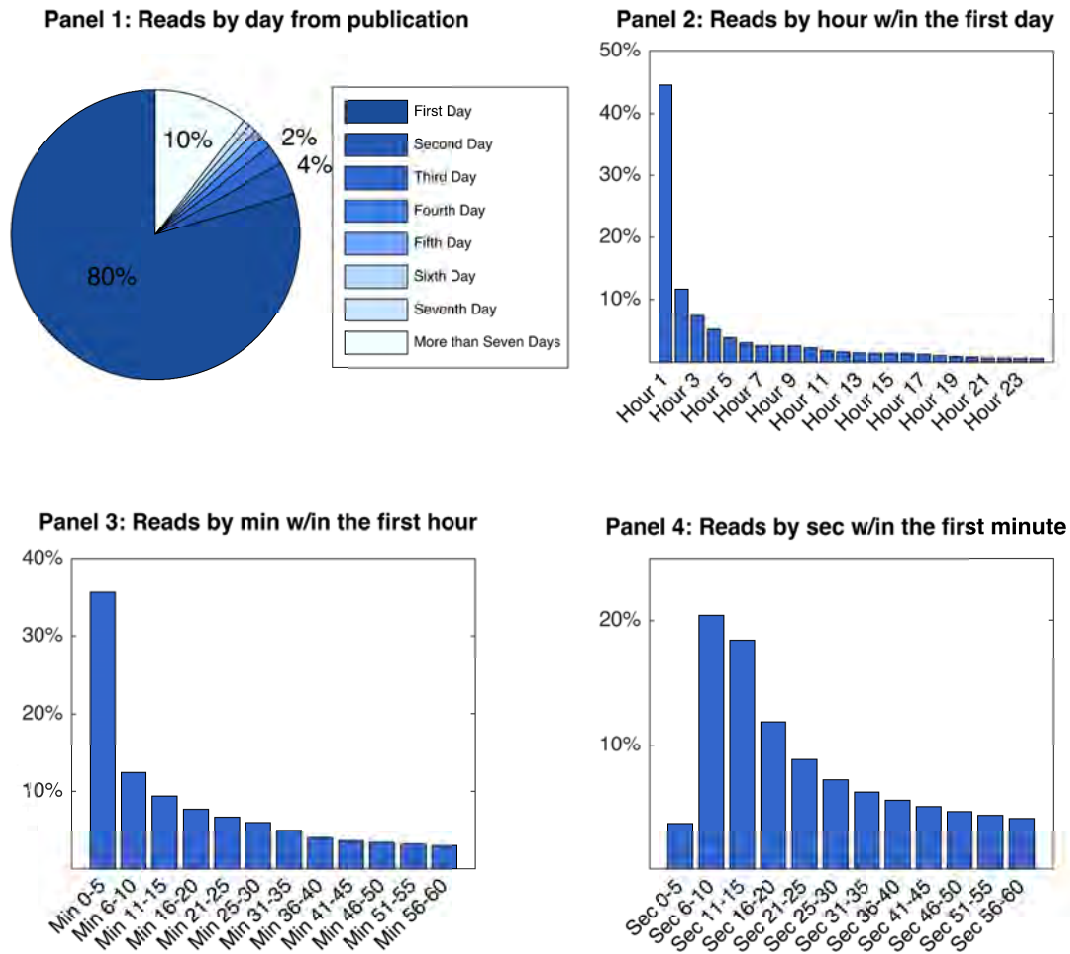


Figure 2: Time lag from publication of a news article to the read. Panel 1 displays the distribution of reads across days from publication. Panel 2 zooms into the first day, and displays the distribution of reads by hour from publication. Panel 3 shows the distribution of reads within the first hour, while Panel 4 presents the distribution of reads within the first minute of news publication.

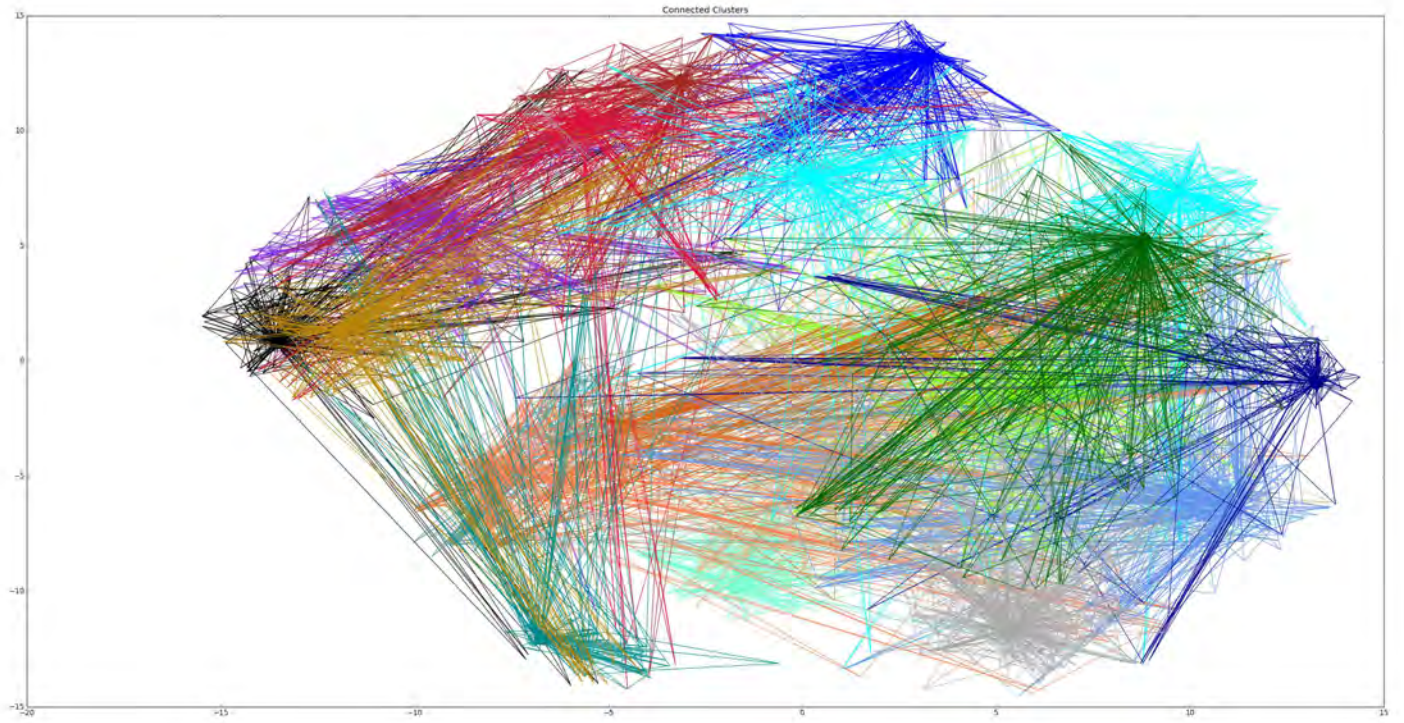


Figure 3: News readers clustered by their news consumption patterns. The figure displays the results of the affinity propagation clustering algorithm run on 4,000 randomly selected readers, encoded as vectors of 66 binary features. The clusters are projected into two dimensions for visualization using t-distributed stochastic neighbor embedding.

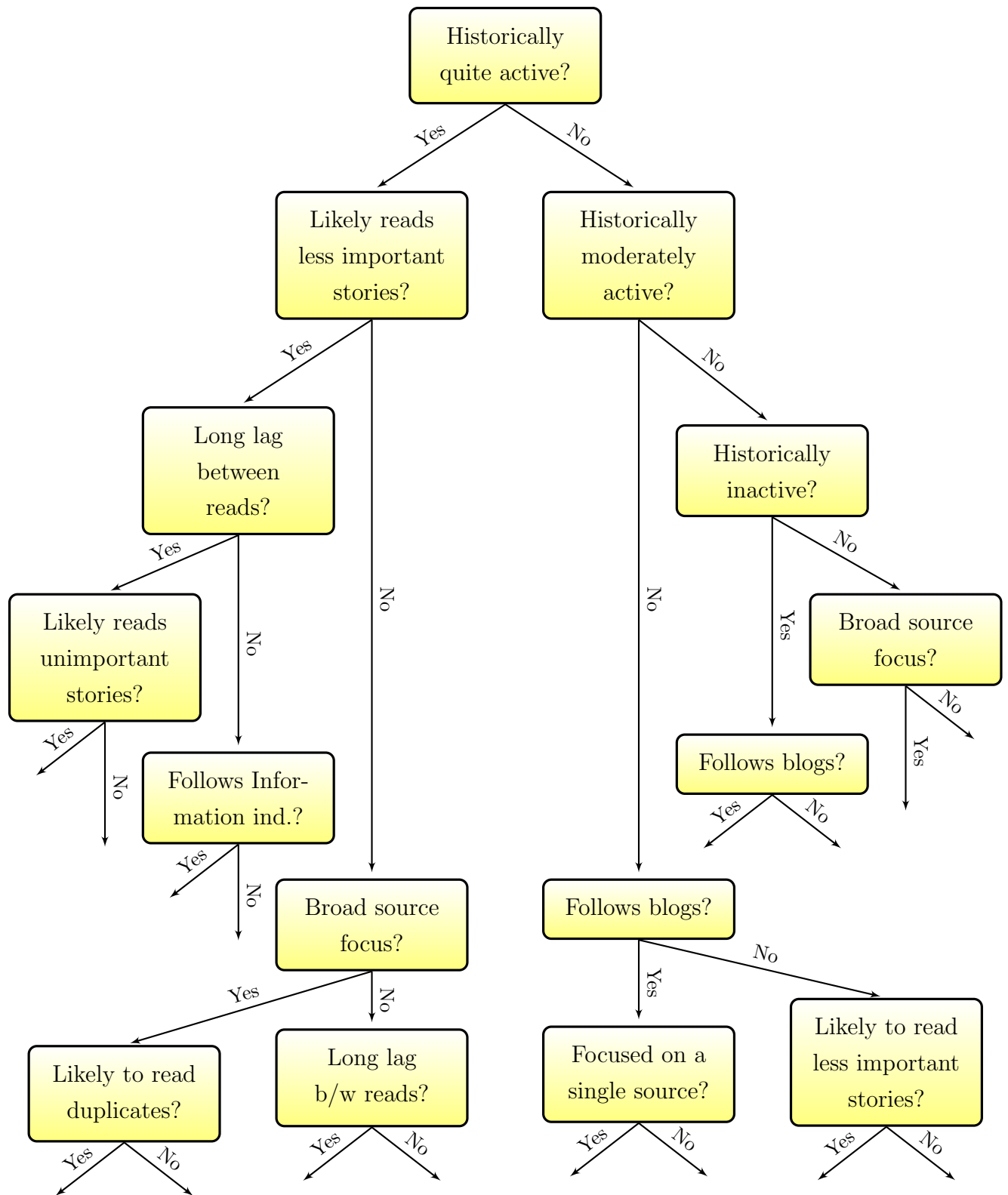


Figure 4: Top several splits of the decision tree classifying readers into clusters. Each split is done using the feature that is most informative at that point, according to Gini impurity.

Daily Trading Volume around Earnings Announcements

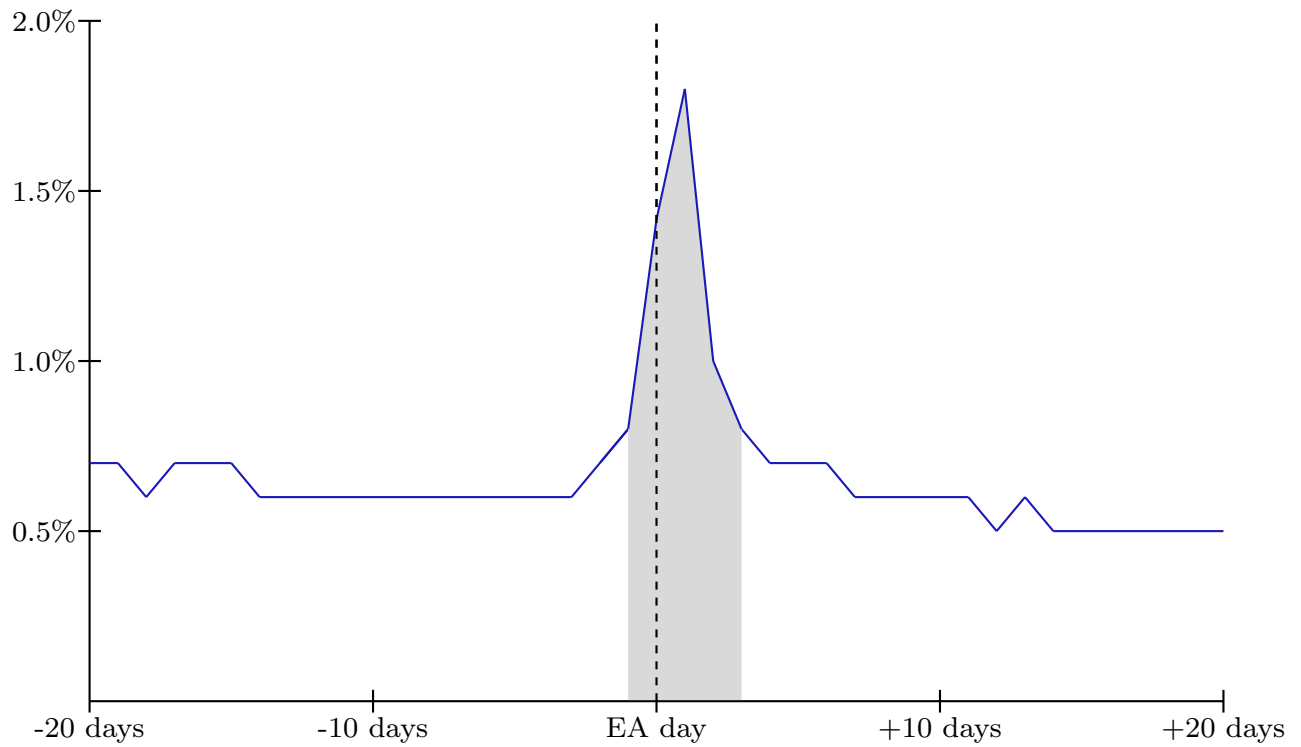


Figure 5: Daily trading volume around earnings announcements in the 2014-2015 sample. Trading volume is computed as the percentage of shares that are turned over on each day, from 20 days preceding the announcement to 20 days after the announcement.

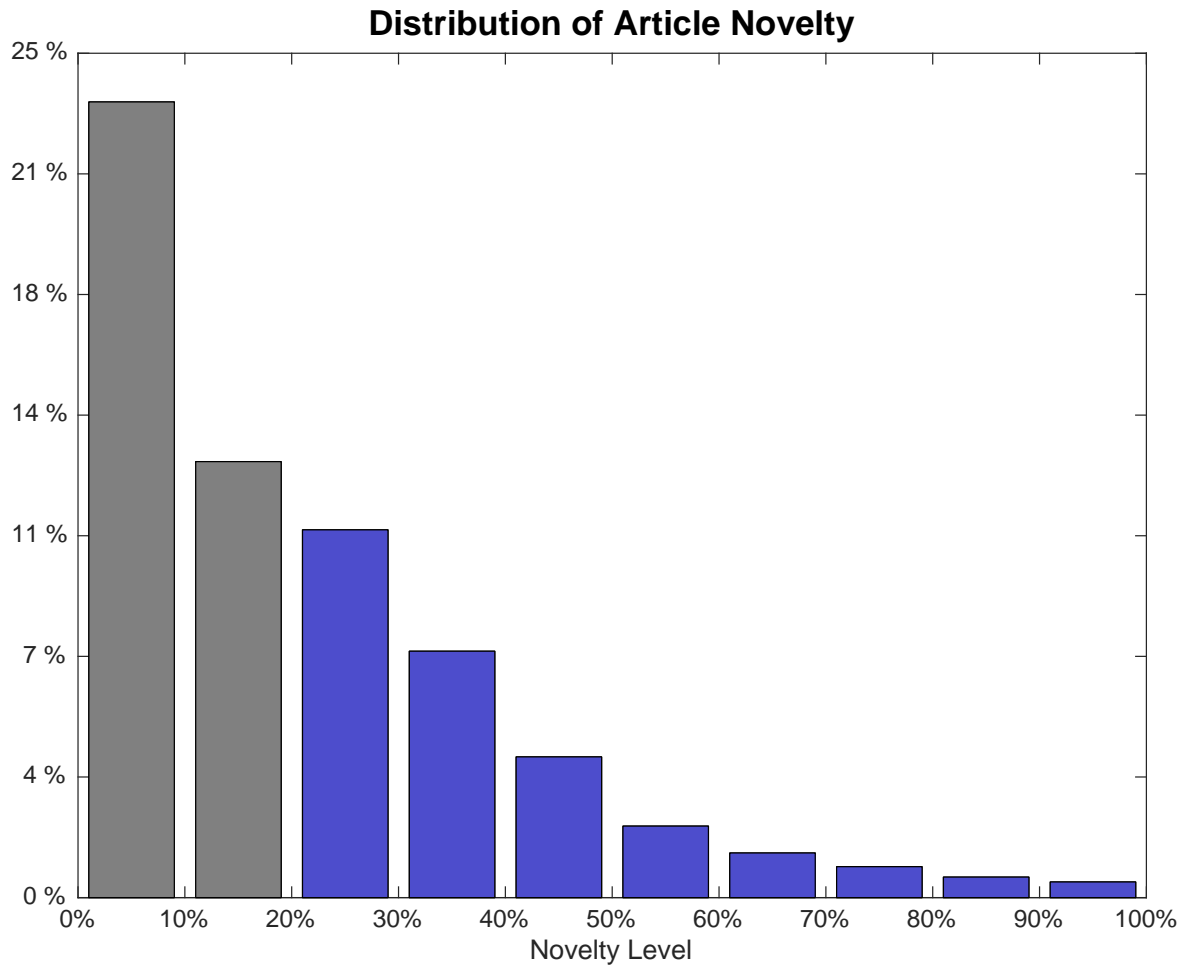


Figure 6: The distribution of textual novelty in the news sample. Novelty is computed as the percentage of unique words in a story that are not spanned by the closest five preceding stories tagged with the same firm. Blue bars indicate the portion of the distribution that is included in the analysis as relevant news (novelty above the 20% threshold).

Table 1: Trading volume tests around earnings announcements. The table presents estimates from a regression of the surge in trading volume on the day of and the day after an earning announcement, $AbnVolume_{i,t}$ on measures of gradual information diffusion ($EntropyTime_{i,t}$) and differences of opinion ($EntropyType_{i,t}$).

$$AbnVolume_{i,t} = \alpha + \beta_1 EntropyTime_{i,t} + \beta_2 EntropyType_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$$

The vector of controls $X_{i,t}$ includes the total number of clicks during the two-day announcement window, size, book to market, and earnings surprise (in all columns); year fixed effects (in columns 2 and 3); and industry fixed effects (in column 3).

	(1)	(2)	(3)
EntropyTime	1.62**	1.61**	1.58**
Standard error	(0.31)	(0.31)	(0.32)
EntropyType	0.57*	0.53†	0.70*
Standard error	(0.29)	(0.29)	(0.32)
Controls			
TotalReads	X	X	X
Size	X	X	X
B/M	X	X	X
SUE	X	X	X
Year FE		X	X
Industry FE			X
**, *, and † denote significance at the 1%, 5%, and 10% levels, respectively.			

Table 2: Trading volume tests around individual news releases. The table presents estimates from a regression of the surge in trading volume within ten minutes of a news release, $AbnVolume_{i,t}$, on measures of gradual information diffusion ($EntropyTime_{i,t}$) and differences of opinion ($EntropyType_{i,t}$):

$$AbnVolume_{i,t} = \alpha + \beta_1 EntropyTime_{i,t} + \beta_2 EntropyType_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$$

The vector of controls $X_{i,t}$ includes the total number of clicks during the ten-minute post-news window, firm size, and book to market ratio (in all columns); day and hour fixed effects (in columns 2 and 3); and firm fixed effects (in column 3).

	(1)	(2)	(3)
EntropyTime	4.10**	4.32**	4.23**
Standard error	(1.10)	(0.87)	(1.33)
EntropyType	2.85**	2.26*	2.58**
Standard error	(0.82)	(0.86)	(0.88)
Controls			
TotalReads	X	X	X
Size	X	X	X
B/M	X	X	X
Year FE		X	X
Hour FE		X	X
Firm FE			X
**, *, and † denote significance at the 1%, 5%, and 10% levels, respectively.			

Table 3: Complementarity analysis around individual news releases. The table presents estimates from regressing the surge in trading volume within ten minutes of a news release, $AbnVolume_{i,t}$, on each of the two measures of disagreement – gradual information diffusion ($EntropyTime_{i,t}$) and differences of opinion ($EntropyType_{i,t}$) – while conditioning on the other.

The controls includes total clicks during the ten-minute post-news window, firm size, and book to market ratio (in all columns); day and hour fixed effects (in columns 2 and 3); and firm fixed effects (in column 3).

Panel 1 presents the results from estimating the following regression within each quintile of $EntropyType$: $AbnVolume_{i,t} = \alpha + \beta_1 EntropyTime_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$

Panel 2 reports the results from estimating the following regression within each quintile of $EntropyTime$: $AbnVolume_{i,t} = \alpha + \beta_2 EntropyType_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$

In both panels, the controls includes total clicks during the ten-minute post-news window, firm size, book to market ratio, day and hour fixed effects, and firm fixed effects.

Panel 1: Conditioning on $EntropyType$					
	Quintile of $EntropyType$				
	lowest – (1)	(2)	(3)	(4)	(5) – highest
EntropyTime	2.66†	1.53†	5.07**	4.43**	2.97**
Standard error	(1.41)	(0.87)	(1.27)	(1.18)	(0.93)
Controls					
TotalReads	X	X	X	X	X
Size, B/M	X	X	X	X	X
Year, Hour FE	X	X	X	X	X
Firm FE	X	X	X	X	X

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

Panel 2: Conditioning on $EntropyTime$					
	Quintile of $EntropyTime$				
	lowest – (1)	(2)	(3)	(4)	(5) – highest
EntropyTime	1.67**	3.55**	1.04	0.68	1.23†
Standard error	(0.67)	(0.93)	(1.20)	(0.88)	(0.66)
Controls					
TotalReads	X	X	X	X	X
Size, B/M	X	X	X	X	X
Year, Hour FE	X	X	X	X	X
Firm FE	X	X	X	X	X

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

Table 4: Trading volume tests around individual news releases, partitioned by news ambiguity. The table presents estimates from regressing the surge in trading volume within ten minutes of a news release, $AbnVolume_{i,t}$, on measures of gradual information diffusion ($EntropyTime_{i,t}$) and differences of opinion ($EntropyType_{i,t}$):

$$AbnVolume_{i,t} = \alpha + \beta_1 EntropyTime_{i,t} + \beta_2 EntropyType_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$$

The controls includes total clicks during the ten-minute post-news window, firm size, and book to market ratio (in all columns); day and hour fixed effects (in columns 2 and 3); and firm fixed effects (in column 3).

Panel 1 presents the results from estimating the regression on the sample of ambiguous news.

Panel 2 reports the results from estimating the regression on the sample of news articles labeled as textually straightforward.

Panel 2: Straightforward News			
	(1)	(2)	(3)
EntropyTime	4.33**	4.37**	4.36**
Standard error	(1.39)	(1.29)	(1.43)
EntropyType	2.12**	1.78	2.03†
Standard error	(0.98)	(1.11)	(1.06)
Controls			
TotalReads	X	X	X
Size, B/M	X	X	X
Year, Hour FE		X	X
Firm FE			X

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

Panel 2: Ambiguous News			
	(1)	(2)	(3)
EntropyTime	3.61**	3.70**	3.66**
Standard error	(1.14)	(1.08)	(1.21)
EntropyType	3.58**	3.34**	3.47**
Standard error	(1.10)	(1.27)	(1.21)
Controls			
TotalReads	X	X	X
Size, B/M	X	X	X
Year, Hour FE		X	X
Firm FE			X

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

Table 5: Relationship between immediacy of attention and immediacy of price variance. For each news event s tagged with firm i , $ImmVar_{s,i,\tau}$ is the ratio of the price variance within an immediate window (first one or two minutes after s) to the price variance over five minutes after the publication of s . Immediate attention $ImmClick_{s,r}$ is the percentage of the first five-minute clicks that occur within the first r seconds. The table reports the coefficients from regressions of $ImmVar_{i,\tau}$ on $ImmClick_{s,r}$ over different timing specifications:
 $ImmVar_{s,i,\tau} = \alpha + \beta ImmClick_{s,r} + \gamma X + \epsilon_s$, for $(\tau, r) \in \{(30, 30), (30, 60), (60, 60), (60, 120)\}$,
The vector of controls X includes article length, total number of clicks within the first five minute window, and firm, date, and hour fixed effects.

		Immediate price variance window (τ)		
		30 seconds	1 minute	2 minutes
Click window (r)	30 seconds	0.15†	0.21*	
	Standard error	(0.09)	(0.10)	
	1 minute		0.33*	0.37*
	Standard error		(0.14)	(0.17)
		*, † denote significance at the 5%, 10% level.		

Table 6: Cumulative abnormal returns two to twenty days following earnings announcements, by earnings surprise and attention speed. The announcement observations are double-sorted into quintiles based on *SUE* and *ImmClicks*, the percentage of clicks on the announcement-day news that occur within 48 hours after the news publication. Average cumulative abnormal returns are reported for each portfolio, as well as for the difference between top-SUE and bottom-SUE portfolios within each attention quintile.

		SUE quintile					
		1 (bottom)	2	3	4	5(top)	Diff (5-1)
ImmClicks quintile	1 (bottom)	-1.24%	-1.31%	-1.04%	0.00%	0.55%*	1.79%*
	2	-1.30%†	-0.06%	-0.89%	0.09%	0.64%†	1.94%*
	3	-1.28%†	-0.39%	-0.59%	0.16%	-0.98%	0.30%
	4	-0.49%	-0.48%	0.10%	-0.30%	-0.44%	0.05%
	5	0.25%	0.39%	0.32%	0.04%	-0.17%	-0.42%
* and † denote significance at the 5% and 10% levels, respectively.							

Table 7: Cross-sectional regression of firm-level momentum against proxies of attention:

$$Momentum_i = \alpha + \beta AttentionProxy_i + \epsilon_i,$$

where $Momentum_i$ is the correlation between monthly abnormal returns for firm i and lagged monthly returns over the prior twelve months, skipping the most recent month; and $AttentionProxy_i$ is one of the four proxies of investor attention, in raw or adjusted form: average time from publication to click on articles about firm i ($MeanTimeLag$), median time from publication to click ($MedTimeLag$), the percentage of clicks occurring within a day of publication ($PercentDay$), and the percentage of clicks within a week of publication ($PercentWeek$). Adjusted proxies are computed as normalized (to mean zero and standard deviation one) residuals from a regression of the raw proxies against log market capitalization and NAICS industry dummies.

	Lag to read		Percentage of quick reads	
	<i>MeanTimeLag</i>	<i>MedTimeLag</i>	<i>PercentDay</i>	<i>PercentWeek</i>
Raw				
Coefficient	0.07**	0.17*	-0.23**	-0.35**
Standard error	(0.01)	(0.08)	(0.04)	(0.06)
Adj.				
Coefficient	0.01**	0.003	-0.01**	-0.01**
Standard erro	(0.004)	(0.005)	(0.004)	(0.004)
* and † denote significance at the 5% and 10% levels, respectively.				