



Essays on Information in Financial Markets

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050135>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Essays on Information in Financial Markets

A dissertation presented

by

Anastassia Fedyk

to

The Committee for the PhD in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2018

© 2018 Anastassia Fedyk
All rights reserved.

Dissertation Advisors:
Professor David Laibson
Professor Andrei Shleifer

Author:
Anastassia Fedyk

Essays on Information in Financial Markets

Abstract

In efficient markets, information, once public, should be instantaneously and correctly reflected in prices. But do investors consume information efficiently? Or do they allocate attention based on presentation, vary in their reading speed and sophistication, and hold overconfident beliefs? My dissertation addresses frictions in information processing and belief formation that may prevent even publicly available information from being correctly and immediately priced.

I find that positioning of financial news plays a large role in determining how the information gets reflected in prices. When a piece of news is saliently highlighted to investors, the price response can be very efficient, taking under an hour. The incorporation of less saliently presented information takes much longer, however: although the price paths eventually converge, this process can take multiple days. I also address the puzzle of increased trading volume around news events. I find that differences in *when* investors see the news are just as instrumental in explaining trading volume as the diversity of *who* is reading the news, especially for more straightforward news. Lastly, I consider biased belief formation and overconfidence. Focusing specifically on the domain of present bias, I document that individuals are aware of this bias in others but remain overoptimistic specifically about themselves. This wedge in beliefs is relevant not only for trading in financial markets, but also in a variety of other settings including teams in the workplace.

Contents

Abstract	iii
Acknowledgments	xi
Introduction	1
1 Front Page News: The Effect of News Positioning on Financial Markets	4
1.1 Introduction	4
1.2 Data Sources and Empirical Strategy	11
1.2.1 Natural Experiment in News Positioning	11
1.2.2 Market Data	19
1.3 Conceptual Framework	20
1.3.1 Evolution of Prices and Trading Volumes	24
1.3.2 Empirical Predictions	26
1.4 News Positioning and Market Dynamics	28
1.4.1 News Positioning and Short-Term Market Dynamics	28
1.4.2 News Positioning and Longer-Term Price Dynamics	35
1.5 News Positioning versus News Importance	41
1.5.1 Discussion: Attention versus Inference	44
1.6 Additional Analyses	45
1.6.1 Quiet Times vs. Busy Times	46
1.6.2 Distributions of Topics	47
1.6.3 Market Participants' Perceptions of News	54
1.7 Conclusion	61
2 News-Driven Trading: Who Reads the News and When?	62
2.1 Introduction	62
2.2 Conceptual Framework	67
2.2.1 Benchmark: Identical Information and Beliefs	69
2.2.2 Gradual Information Diffusion	71
2.2.3 Differences of Opinion	73
2.3 Data	75

2.3.1	News Consumption Data	75
2.3.2	Market Data	77
2.4	Methodology	78
2.4.1	Measuring Gradual Information Diffusion	78
2.4.2	Measuring Differences of Opinion	79
2.5	Disagreement and Trading Volume	86
2.5.1	Trading Volume around Earnings Announcements	86
2.5.2	Individual News Articles: Identifying Relevant News Events	89
2.5.3	Trading around Individual News Events	90
2.5.4	Complementarity Analysis	92
2.6	Trading Volume and News Ambiguity	96
2.6.1	Measuring News Ambiguity	96
2.6.2	Trading Volume around Ambiguous News	98
2.7	Conclusion	100
3	Asymmetric Naïveté: Beliefs about Self-Control	101
3.1	Introduction	101
3.2	Experimental Design	107
3.2.1	Experimental task	108
3.2.2	Experimental Timeline and Payments	110
3.2.3	Work Decisions	112
3.2.4	Predictions	115
3.2.5	Sample	120
3.3	Reduced-Form Results	121
3.3.1	Present Bias and Beliefs	122
3.3.2	Incentivizing Predictions	125
3.3.3	Juxtaposing Predictions about Self and Others	127
3.4	Structural Estimation	129
3.4.1	Model	129
3.4.2	Empirical Estimation	132
3.4.3	Individual Level Estimates	136
3.5	External Evidence: Classroom Experiment	139
3.5.1	Design	139
3.5.2	Results	140
3.6	Applications	144
3.6.1	Competitive Environments	144
3.6.2	Group Decision Making	145
3.6.3	Principal-Agent Settings	147

3.7 Conclusion	148
References	150
Appendix A Appendix to Chapter 1	162
A.1 Technical Details	162
A.1.1 Latent Dirichlet Allocation	162
A.1.2 Text Preprocessing	164
A.1.3 Topic Model Estimation	165
Appendix B Appendix to Chapter 2	167
B.1 Additional Analyses: Attention and Prices	167
B.1.1 Price Dynamics around Individual News Articles	167
B.1.2 Post-Earnings-Announcement Drift	169
B.1.3 Attention and Momentum	171
B.2 Technical Details	174
B.2.1 Clustering Readers by News Consumption Patterns: Affinity Propagation	174
B.2.2 Reader Type Visualization: t-Distributed Stochastic Neighbor Embedding	175
B.2.3 Reader Classification into Clusters: Random Forest	176
Appendix C Appendix to Chapter 3	178
C.1 Additional Reduced-Form Results	178
C.1.1 Results Including Attrited Participants	178
C.1.2 Pilot Study	178
C.1.3 Warm-up Amounts and Projection Bias	181
C.2 Additional Structural Estimates	184
C.2.1 Different Baseline Levels of Effort for Self and Others	184
C.2.2 Different Utility Function Parameters for Self and Others	185
C.3 Experimental Instructions	187
C.3.1 Study Sign-Up	187
C.3.2 Instructions and Questions: First Day	188
C.3.3 Instructions and Questions: Dates 2, 3, and 4	207
C.4 Informed Consent	213

List of Tables

1.1	Examples of articles in the “primary important” and “secondary important” categories.	14
1.2	Summary statistics of the hand-collected news sample.	17
1.3	Comparison of trading volumes and absolute price changes immediately following SI news articles that are pinned to the front page and those that are not.	31
1.4	Short-term continuation in returns after front page and non-front page SI news articles.	34
1.5	Continuation in returns over longer horizons following front page and non-front page SI news.	36
1.6	Differences in trading volumes and absolute price changes following front page and non-front page SI news articles over longer horizons.	38
1.7	Trading volumes and absolute price changes following PI and front page SI news articles.	42
1.8	Serial correlation in returns following PI and SI front page news articles. Each column estimates the following specification:	44
1.9	Comparison of trading volumes and absolute price changes within ten minutes of non-front page SI articles published during quiet versus busy times.	48
1.10	Topics identified from a large and representative dataset of financial news from Reuters.. . . .	51
1.11	Results from pairwise comparisons between sets of news articles from different positions and different levels of importance.	53
1.12	Summary statistics of the financial experts surveyed regarding the news.	55
1.13	Aggregated responses of financial experts to the news survey.	59
2.1	Trading volume tests around earnings announcements.	89
2.2	Trading volume tests around individual news releases.	93
2.3	Complementarity analysis around individual news releases.	94
2.4	Trading volume tests around individual news releases, partitioned by news ambiguity.	99

3.1	Decision and prediction questions posed to participants in Groups 1, 2, and 3.	119
3.2	Numbers of recruited participants and attrition rates across experimental sessions.	121
3.3	Reduced-form results from participants who complete the entirety of the experiment.	124
3.4	Predicted differences in one's own and others' decisions, sliced by incentive.	126
3.5	Predicted differences in one's own and others' decisions, sliced by method of presenting the questions.	128
3.6	Parameter estimates from the structural model.	135
3.7	Comparison of students' predictions of their own and their classmates' completion dates.	142
B.1	Relationship between immediacy of attention and immediacy of price variance.	169
B.2	Cumulative abnormal returns two to twenty days following earnings announcements, by earnings surprise and attention speed.	171
B.3	Cross-sectional regression of firm-level momentum against proxies of attention.	173
C.1	Pooled results from all experimental participants, including participants who attrited between Week 1 and Week 4.	179
C.2	Results from the pilot run of the online experiment.	180
C.3	Pooled decisions, actual differences, and predicted differences, sliced by warm-up amount.	182
C.4	Parameter estimates from the structural model, allowing for different perceived baseline levels for self versus others when the decision is made for future dates.	185
C.5	Parameter estimates from the structural model, allowing for different model parameters for responses about self versus other.	186

List of Figures

1.1	Bloomberg terminal screen displaying company news as of 3:01PM EST on December 15, 2016.	13
1.2	Process illustrating how Bloomberg news articles are pinned to the prominent front page positions at the top of the news screen.	16
1.3	Distribution of daily volumes of “primary important” (PI) and “secondary important” (SI) news articles across the days in the sample.	19
1.4	Model timeline, illustrating the diffusion of information for front page and non-front page news articles, as well as corresponding shares of informed investors in each period.	23
1.5	Market dynamics following front page and non-front page SI news articles. .	30
1.6	Price paths after front page and non-front page SI news, sliced by the direction of the initial 5-minute moves.	40
1.7	Distributions of topics across news articles from different categories.	52
1.8	Example question from the survey administered to active finance professionals and MBA students.	58
1.9	Individual-level responses from the survey of financial experts.	60
2.1	Model timeline.	68
2.2	Time lag from publication of a news article to the read.	77
2.3	News readers clustered by their news consumption patterns.	83
2.4	Top several splits of the decision tree classifying readers into clusters.	85
2.5	Daily trading volume around earnings announcements in the 2014-2015 sample.	87
2.6	The distribution of textual novelty in the news sample.	91
3.1	Screenshot of the experimental task.	109
3.2	Breakdown of demographics and task enjoyment responses from the pilot study.	110
3.3	Experimental timeline. Stars indicate questions regarding decisions for immediate work.	111
3.4	An example of a work decision screen.	113
3.5	An example screen aggregating a participant’s work decisions for a given date.	114

3.6	Examples of screens eliciting participants' predictions of their own and others' work decisions.	116
3.7	A comparison of work decisions for the current date (immediate decisions) and for future dates (ahead-of-time decisions).	123
3.8	Distribution of the structural parameter estimates across individual experimental participants.	137
3.9	Results from the classroom experiment.	141
A.1	Log likelihood for different estimations of the Latent Dirichlet Allocation model.	166
C.1	Payment schedule.	190

Acknowledgments

I am deeply indebted to David Laibson, Andrei Shleifer, John Campbell, Christopher Malloy, Lauren Cohen, and Jeremy Stein for their kind support and excellent guidance. For financial support, I am thankful to the Harvard Business School, Hillcrest Asset Management, the Lab for Economic Applications and Policy, the Pershing Square Venture Fund for Research on the Foundations in Human Behavior, and the Russell Sage Foundation.

I am very grateful to Philippe Aghion, Daniel Andrei, Joshua Coval, Stefano DellaVigna, Benjamin Edelman, Xavier Gabaix, Robin Greenwood, Samuel Hanson, Oliver Hart, Owen Lamont, Andres Liberman, Kristof Madarasz, Adi Sunderam, Dmitri Taubinski, Alex Rees-Jones, and especially John Beshears and Matthew Rabin for many enlightening discussions. I have likewise benefitted from conversations with Ned Augenblick, Linda Babcock, Malcolm Baker, Max Bazerman, Claire Celerier, Katie Coffman, Lucas Coffman, Kent Daniel, Benjamin Enke, Uri Gneezy, Jerry Green, Emir Kamenica, Patrick Kelly, Botond Kőszegi, Peter Koudijs, George Loewenstein, Michael Luca, William Mann, Dmitry Muravyev, Muriel Niederle, Marina Niessner, Emily Oster, Daniel Paravisini, Julien Penasse, Devin Pope, Todd Rogers, Martin Schmalz, Kelly Shue, Charles Sprenger, Tomasz Strzalecki, and Florian Zimmermann. I am also thankful for invaluable feedback from my friends and officemates, Chris Andersen, Vitaly Bord, Kirill Borusyak, William Diamond, Xavier Jaravel, Seunghyup Lee, Jonathan Libgober, Gordon Liao, Weiling Liu, Jeffrey Picel, Marc Kaufmann, Thomas Powers, Gianluca Rinaldi, Stanislav Sokolinski, Carmen Wang, Wentao Xiong, and Yao Zeng.

My family, especially my sisters Valeria and Maria Fedyk, have provided a wonderful sense of balance. And most of all, I am indebted my parents, Tatiana and Yurii Fedyk, and my husband, James Hodson, for their boundless love and support.

To my parents, my grandmother, and my beloved husband

Introduction

How is information incorporated into asset prices? In the traditional semi-strong form efficient view of financial markets, information, once public, should be instantaneously and correctly reflected in asset prices. However, financial professionals may not consume information efficiently – they may allocate their attention based on the way in which information is presented, vary in their reading speed and sophistication, and hold biased beliefs regarding their ability. These frictions form the basis for my dissertation research.

The first chapter of my dissertation exploits a natural experiment in “front page” positioning of news articles on the Bloomberg terminal to trace out how presentation of information affects its incorporation into asset prices. The front page and non-front page articles are indistinguishable by either algorithmic analysis or by the target audience of active finance professionals. I find that pinning a news article to the front page leads to 280% higher trading volumes and 180% larger price changes within the first ten minutes after publication, and induces a stronger short-term price drift for 30-45 minutes. At longer horizons, once the front page articles are removed from their prominent positions, the reactions to non-front page news begin to catch up. However, the incorporation of non-front page information is much slower. Although the price paths after front page and non-front page news eventually converge, this process takes multiple days. A comparison against differential reactions following news articles of varying levels of editorial importance indicates that news positioning can even play a stronger role in driving short-term market activity than news importance.

In the second chapter, I use a detailed dataset of individual clicks on news to dig deeper

into the mechanisms through which attention to news impacts market activity. Conceptually, there are two channels through which attention could drive market activity: trade can happen either between two investors one of whom has seen the news and the other has not, or between two investors who have both seen the news but disagree regarding its interpretation. I use a unique dataset of over 80 million clicks on news tagged with U.S. equity securities by hundreds of thousands of finance professionals to directly observe attention to public information releases. These granular data allow me to simultaneously compute for each piece of news: (1) how dispersed attention is over time; (2) how dispersed attention is over different types of investors; and (3) how strongly these two variables relate to abnormal trading volume following the news. To characterize different types of investors, I use machine learning techniques and classify investors based on their news consumption patterns. I find that both, differences in timing of *when* investors see the news and differences in *who* sees the news, are strongly predictive of trading volume. Although gradual information diffusion is stronger in explaining trading volume around news overall, differences of opinion play a larger role when the news is more textually ambiguous.

The third chapter of my dissertation takes a different approach and investigates a more general tendency towards biased beliefs. In particular, I use a classroom survey and a laboratory setup to experimentally investigate individuals' awareness of their own and others' susceptibility to a particular bias: present bias. In the classroom, I ask students to predict when they or their peers would turn in an assignment. In the laboratory experiment, participants engage in a real effort task and are asked to predict how much work they and others would choose to do for future dates, as well as how much work they would choose to do immediately. In both settings, I find that individuals are quite accurate in predicting others' present bias, but display virtually no awareness of their own present bias, despite having plentiful information regarding their own past behavior. This wedge in beliefs regarding self versus others is relevant not only for trading in financial markets, but also for a variety of other settings including teachers' beliefs regarding their students guiding classroom assignments, spouses' beliefs regarding each other shaping their

consumption decisions, and employees' beliefs regarding each other affecting performance in the workplace.

Overall, my research suggests that even when information is plentifully available, we still see that (1) how the information is presented plays a large role in determining how quickly it is incorporated into asset prices; (2) news spurs increased disagreement and hence trading; and (3) individuals are susceptible to biases and hold overconfident beliefs regarding themselves compared to others. Especially with the modern-day proliferation of information, timely and effective processing of complex streams of information is an important factor for financial markets' stability and efficiency.

Chapter 1

Front Page News: The Effect of News Positioning on Financial Markets

1.1 Introduction

How does information get incorporated into asset prices? A number of theoretical models propose potential frictions that may prevent even publicly available information from being instantaneously reflected in prices.¹ Multiple empirical studies lend suggestive evidence to this view.² However, tracing out incorporation of information in real time remains difficult, and requires a detailed understanding of the variation across individual pieces of information.

In this paper, I capture the causal effect of prominence of news positioning on the way

¹See, for example, Peng and Xiong (2006), DellaVigna and Pollet (2009), and Andrei and Hasler (2014) on limited attention to publicly available information; and Harris and Raviv (1993), Kandel and Pearson (1995), Cao and Ou-Yang (2009), and Banerjee and Kremer (2010) on differential interpretations of public information. A related large literature including Kyle (1985), Holden and Subrahmanyam (1992), Wang (1994), Hirshleifer *et al.* (2009), Cao *et al.* (2002), Foucault *et al.* (2016), and Andrei and Cujean (2017), among others, considers incorporation of private information.

²See, for example, Foster *et al.* (1984), Bernard and Thomas (1989), and Peress (2008) on the post-earnings announcement drift; Loh (2010), Da *et al.* (2011, 2015), Drake *et al.* (2012, 2015, 2016), Schmidt (2013), Curtis *et al.* (2014), and Ben-Rephael *et al.* (2017) on predictability of market dynamics from proxies of attention; Huberman and Regev (2001), Tetlock (2011), Gilbert *et al.* (2012), and Fedyk and Hodson (2015) on reactions to stale news; and Carvalho *et al.* (2011) and Marshall *et al.* (2014) on reactions to false news.

the underlying information is incorporated into asset pricing, using a natural experiment in the way news articles are pinned to the top of the Bloomberg terminal news screen, in the “front page” positions. I show that positioning a news article about a security on the front page has a fast and substantial effect on the security’s market dynamics. News articles that get pinned to the front page induce 280% higher trading volumes and 180% larger price changes within ten minutes of publication, and 17% more continuation in returns over consecutive five-minute intervals. These articles are accompanied by a strong price drift for approximately 30-45 minutes after publication, consistent with the fact that the front page articles remain prominently positioned for half an hour to an hour. After that, the information in front page news appears to be fully incorporated, and the reactions to non-front page articles begin to gradually catch up. However, the incorporation of non-front page information is much slower. Although the price paths after front page and non-front page news eventually converge, this process takes multiple days. Interestingly, differences in news positioning have an even stronger effect on market dynamics than differences between news articles marked with distinct importance labels by the editorial staff.

My empirical design exploits a natural experiment based around a category of Bloomberg news articles whose placement depends on the contemporaneous volume of other articles, rather than on their own content. I focus on news articles about individual U.S. equity securities, and hand-collect a sample of news between March 2014 and December 2015. The news articles in my hand-collected sample fall into three categories: “primary important,” “secondary important,” and “all other” news. News articles marked as “primary important” are always pinned to the prominent front page positions, displacing the previous front page news and remaining on the front page for, on average, half an hour to an hour. News articles marked as “all other” are never placed in the front page positions. News articles marked as “secondary important” constitute the category of interesting variation. Any particular news article in this category is given a front page slot if and only if, at the precise moment when the article is released, there is at least one such slot remaining from the “primary important” news. As a result, “secondary important” news articles that make it to the front

page position and those that do not are marked as equally significant. Their positions vary due to contemporaneous numbers of “primary important” articles, rather than their own underlying content.

I structure the empirical analysis of the market dynamics following front page versus non-front page “secondary important” news articles using a theoretical framework that reflects standard models of limited attention and gradual information diffusion.³ The three-period model considers a news signal published by the main news source of interest and also reported by alternative news sources. The framework incorporates two standard features from models of gradual information diffusion: (1) only a fraction of investors are attentive to the news signal from each source in each period; and (2) investors update their beliefs in a naïve Bayesian manner, incorporating their own information but not rational expectations of the information that may have been obtained by other investors. Front page positioning is represented by more prominent and longer-lasting reporting by the main news source. A larger incidence of investors are attentive to the news signal from the main source when it is published on the front page, and this persists into the second period.

This framework generates several predictions. First, the front page news articles are accompanied by larger immediate trading volumes and absolute price changes. Second, the initial returns accompanying front page news articles are more likely to continue in the short-term. Third, the front page news articles are followed by *lower* longer-term price continuation.

My empirical results confirm these predictions. There are significant differences in market dynamics following “secondary important” news articles that are pinned to the front page and those that are not. Consistent with the first prediction, front page news articles are accompanied by substantially higher trading volumes and absolute price changes for the tagged securities immediately after publication. For example, these articles are, on average, accompanied by 280% larger trading volumes and 180% larger absolute price

³For models of limited attention and gradual information diffusion, see, for example, Hong and Stein (1999), DellaVigna and Pollet (2009), and Andrei and Hasler (2014).

changes during the ten minutes following article publication.

Since pinning a news article to a front page position makes it visible for a longer period of time, the front page positioning also induces more persistent short-term market reactions, confirming the second prediction of the gradual information diffusion framework. Front page news articles are accompanied by significantly higher serial correlation in price changes over a variety of short-term horizons. For example, these articles are, on average, followed by 17% larger serial correlations in price changes across consecutive five-minute intervals.

I also find empirical support for the third prediction of the model: after the initial period of about forty-five minutes, the price drift is significantly stronger for *non*-front page news articles. Most front page articles are displaced from their prominent positions after half an hour to an hour, and the incorporation of front page information is virtually complete during this period. The incorporation of non-front page information takes much longer. We begin to see some convergence during the hours after the news. For example, the initial returns from the first 30-45 minutes after publication of non-front page news are accompanied by a drift of 14-29% over the subsequent hour. However, the convergence is quite gradual, and differences in price effects persist even days after news publication. Securities mentioned in front page news articles see 34 basis points larger absolute cumulative returns measured from the moment of publication to two days later, relative to securities mentioned in non-front page news; the difference is statistically significant at the 10% level. Five days out, the difference declines to a statistically insignificant 25 basis points, and fifteen days after the news, the difference is a statistically indiscernible 8 basis points.

I compare the market effects of news positioning against the effects of news importance. In particular, I estimate market dynamics following two sets of news articles that receive equally prominent positions but that differ in their importance, as marked by the editorial staff. These are: (1) "secondary important" articles that make it to the front page; and (2) "primary important" articles, all of which make it to the front page by default. Articles in both of these categories are prominently positioned, but the articles in the second category are marked by the editorial staff, *ex ante*, to be more important than those in the first

category.

I find that news importance is not as significant in driving short-term market activity as positioning. Trading volumes following news publication are not statistically different for securities mentioned in more (“primary important”) versus less important (“secondary important”) news articles pinned to the front page. Absolute price changes are 80% (66%) larger during the first five (ten) minutes following the more important news articles, but the relative difference is smaller than that induced by the news positioning. The short term price drift is statistically indistinguishable for more and less important news articles, holding front page position constant. Overall, the results indicate that news positioning plays an even larger role for short-term market dynamics than editorial markings of the importance of the underlying news.

I perform a number of additional analyses to confirm that the results are not driven by systematic differences between “secondary important” articles that receive a front page slot and those that do not. First, I consider the possibility that, due to market participants’ distraction during periods with high volumes of news, articles published during quieter times garner larger reactions. To address this possibility, I hold position constant and compare non-front page “secondary important” articles released during times with different amounts of contemporaneous news activity. I document that the non-front page articles released during quiet times are, if anything, accompanied by *less* substantial reactions than the non-front page articles released during busy times.

Second, using techniques from machine learning and a representative corpus of financial news from Reuters, I learn the mixtures of topics generally discussed in financial news, such as earnings announcements, technology, and litigation. I then use the trained model to compare the distributions of identified topics appearing in the text of the individual Bloomberg news articles in my hand-collected samples. I find no systematic differences between the distributions of topics discussed in the front page versus non-front page “secondary important” news articles. The distribution of topics covered by the “primary important” news articles, by comparison, does differ slightly from the distribution of topics

appearing in “secondary important” news.

Third, a survey of 150 active finance professionals indicates that absent salient positioning, market participants find front page “secondary important” headlines to be indistinguishable from non-front page ones. The survey participants consist of key decision makers at a broad range of financial institutions, including broker dealers such as Bank of America and Goldman Sachs, investment management firms such as BlackRock and PIMCO, hedge funds such as Bridgewater and AQR, and private equity firms such as Blackstone and Warburg Pincus. The finance professionals confirm Bloomberg editorial staff’s judgment of news importance. They consistently identify the “primary important” news articles as, on average, more impactful than “secondary important” news articles (“primary important” headlines are chosen as more impactful 61% of the time, significantly different from 50%). By contrast, these finance professionals identify the front page “secondary important” news articles as more impactful than their non-front page counterparts only 48% of the time, not significantly different from 50%. I also repeat this analysis using a smaller survey of 27 MBA students from top business schools and find qualitatively similar results.

My findings build on the growing literature evaluating the impact of media on financial markets.⁴ Prior empirical strategies for estimating the causal impact of media use exogenous variation in news arrival through weather-related disruptions (see Engelberg and Parsons (2011)), newspaper strikes (see Peress (2014)), disruptions to boat routes (see Koudijs (2016)), and staggered implementation of robo-journalism (see Blankespoor *et al.* (2017)), as well as variation in security relevance tags (see von Beschwitz *et al.* (2015)) and headline complexity and degree of quantification (see Umar (2017) and Huang *et al.* (2017)). Klibanoff *et al.* (1998) find that for closed-end country funds, the incidence of news on the front page of the *New York Times* is correlated with a higher elasticity of price with respect to asset value. Huberman and Regev (2001) further highlight the importance of prominent news

⁴See Busse and Green (2002), Barber and Loeffler (1993), Chan (2003), Fehle *et al.* (2005), Antweiler and Frank (2006) Barber and Odean (2008), Fang and Peress (2009), Engelberg *et al.* (2012), Solomon (2012), Dougal *et al.* (2012), Rogers *et al.* (2013), Hillert *et al.* (2014), Ahern and Sosyura (2014), Liu *et al.* (2014), Yuan (2015), and Boulland *et al.* (2017).

positioning by analyzing a case of (mostly) stale information, initially reported in *Nature* in November 1997, getting reprinted on the front page of the *New York Times* in May 1998. Furthermore, Lawrence *et al.* (2017) present compelling evidence that promotion of earnings announcement news on Yahoo! Finance to a subset of website visitors increases the abnormal return on the announcement date.⁵

The present paper contributes to the literature by providing a clear counterfactual. The natural experiment in positioning of news on the Bloomberg terminal offers clean variation in institutional investor attention in an important setting that represents the main source of information for a large set of finance professionals. This allows me to document two things. On the one hand, when information is especially saliently highlighted, the market response is quite efficient: prices respond within an hour (and largely within the first minutes) of news publication. These highlighted news events, for which limited attention and cognitive processing limitations play a minor role, illustrate a best-case scenario. On the other hand, the price formation process is this efficient *only* for especially highlighted news. In other cases – even with public, easily accessible news consumed by sophisticated institutional investors – attention is more gradual and the price formation process takes substantially longer, on the order of days or even weeks.

These findings provide systematic evidence that it is not enough to make financially-relevant information easily accessible: how saliently the information is presented plays an important role in determining whether the information is immediately reflected in asset prices. The price impact induced by front page positioning occurs quickly, but the comparable non-front page information takes surprisingly long to converge, given that these are all easily accessible news articles available on the Bloomberg terminal. For more obscure or private information, similar mechanisms are likely to apply at longer horizons, generating phenomena such as months-level momentum.

⁵The importance of prominent positioning and alphabetical ordering has also been documented in other contexts; see, for example, Ho and Imai (2008), Jacobs and Hillert (2015), and Feenberg *et al.* (2017). In financial markets, presentation has been shown to drive mutual fund flows (Kaniel and Parham (2017)) and attract attention to securities independent of information flows (Wang. (2017)).

The remainder of the paper proceeds as follows. Section 1.2 describes the data and the natural experiment in news positioning. Section 1.3 outlines the conceptual framework of market dynamics following more and less prominently positioned news. Section 1.4 presents the key empirical findings on the differential market dynamics following front page and non-front page news articles. Section 1.5 explores the effect of news importance, holding position constant, by comparing “secondary important” news articles that are positioned on the front page against “primary important” front page news articles. Section 1.6 presents additional analyses of news content, confirming that the front page “secondary important” articles in the sample are indistinguishable from their non-front page counterparts by both algorithmic analysis and the target audience of market participants. Section 1.7 concludes.

1.2 Data Sources and Empirical Strategy

In order to capture the casual effect of news presentation on trading volumes and returns, I use quasi-random variation in positioning of news articles on the Bloomberg terminal. Two key features of these data make them especially well-suited to the current analysis. First, Bloomberg is one of the largest financial news providers and a main source of news for finance professionals, making it an ideal setting to estimate the effect of attention to news on financial markets. Second, the data include a natural experiment of quasi-random positioning for a subset of news articles. The news data are merged with market data to relate news presentation to trading volumes and price formation.

1.2.1 Natural Experiment in News Positioning

In this subsection, I describe the quasi-random variation in news positioning that I use in my research design. In particular, I concentrate on a subset of news articles that are sometimes prominently positioned, and sometimes not, depending on the volume of other articles released around the same time and not on the characteristics of the news articles themselves.

The full sample of news passing through the Bloomberg terminal is aggregated from a

variety of sources in real-time. The sources of news include key national and international news wires from a comprehensive set of news organizations, company filings, press releases, and content from web sources, including blogs and social media. The news articles are disseminated electronically to over 300,000 finance professionals through the subscription-based terminal. Overall, there are millions of articles tagged with U.S. equity securities during the sample period of March 22, 2014 - December 31, 2015.

There are differences in how Bloomberg presents individual news articles on the terminal. Generally, the news screen features a scrolling list of news articles, where newly published articles replace the older ones at the top of the screen. However, some of the news articles written directly by Bloomberg News get pinned to the top of the screen. At any given point in time, there are at most three such pinned articles. Figure 1.1 shows a screenshot of a default Bloomberg news screen covering all company-relevant news. The top three articles are pinned and remain at the top, while the articles below continually move down as new publications arrive. It is these positions, highlighted in yellow font at the top of the default company news screen in Figure 1.1, that I term "front page" throughout this paper.

Effectively, there are three broad categories of news articles passing through the Bloomberg terminal: the primary "primary important" (PI) articles; the "secondary important" (SI) articles, and "all other" articles. The assignment of individual articles to these categories reflects the journalistic and editorial opinion regarding the importance of a given piece of news. Each of the two important categories, PI and SI, comprises roughly 0.1-0.5% of all news, so both of these categories of articles capture news of fairly rare perceived importance. I exclude market wrap articles, in order to focus on new information relevant to individual securities, and hand-collect all articles that are tagged with at least one publicly traded U.S. equity security, that are published between 8AM and 5PM EST during the sample period, and that are either in the PI category (1,419 unique PI articles) or the SI category (4,887 unique SI articles).

For the most part, PI articles represent significant company news, such as earnings reports and M&A decisions. A few representative examples of the sample of PI news are



Figure 1.1: Bloomberg terminal screen displaying company news as of 3:01PM EST on December 15, 2016.

provided in Panel 1 of Table 1.1.

SI articles likewise include significant events, such as changes in regulation and drug approvals. However, this set of news also features articles that are likely to capture the readers' curiosity, but that are less immediately relevant to financial markets, such as moves of top well-known traders and perks in financial firms. A few representative examples of SI articles are presented in Panel 2 of Table 1.1.

The classification of articles into categories of relative importance plays a role in how prominently the articles are positioned. When an article from the PI category is released, it is immediately placed in a prominent front page position, displacing whichever news article was in that position previously. Once on the front page, a news article remains there until the earlier of two things occurs: either a new PI article comes out and displaces the old article, or a predefined amount of time (on the order of hours) elapses. Occasionally, there are not enough PI articles at a given point in time to fill all of the front page slots. In

Table 1.1: Examples of articles in the “primary important” and “secondary important” categories.

Panel 1: “Primary important” news articles

Date	Headline
3/25/2014	Walgreen 2Q Adj. EPS Misses Est.
4/7/2014	Tekmira Says FDA Modifies TKM-Ebola Drug to Partial Hold
4/25/2014	United Technologies Reports SEC Formal Investigation, Subpoena
8/14/2014	Icahn Reports 6.63% Stake in Gannett, Urges Splitting Co.
10/28/2014	Amgen Restarts Buyback, Boosting Dividend; 2015 View Tops Ests.
12/23/2014	Stryker Said to Plan Smith & Nephew Takeover Bid Within Weeks
1/27/2015	Amgen 4Q Adj. EPS, Rev. Top Ests.; Ivabradine, T-vec Delayed
5/13/2015	Nissan Forecasts 6% Gain in Profit on U.S. Demand, Weak Yen
5/19/2015	Computer Sciences Corp. to Split Into Two Companies
7/30/2015	Sanofi Profit Beats Estimates as Multiple Sclerosis Drugs Gain
9/10/2015	Morrison Earnings Miss Analysts’ Estimate as Grocer Cut Prices
9/14/2015	Standard Chartered Said to Plan Cutting 250 Managing Directors
11/24/2015	Fed Says It’s Overhauling Standards for Large-Bank Examiners
1/15/2016	Wal-Mart to Close 269 Stores in U.S., Globally
2/25/2016	Apple Says U.S. Can’t Force It to Unlock Terrorist’s iPhone

Panel 2: “Secondary important” news articles

Date	Headline	FP
4/7/2014	Honeywell CEO Makes Biggest Executive Shift Naming Vice Chairmen	N
5/19/2014	AstraZeneca Chairman “Surprised” Pfizer Took Last Offer Public	N
6/3/2014	Robertson’s Stock Picker Singh Said to Become Newest Tiger Cub	Y
6/24/2014	Morgan Stanley Gets 90,000 Applications for Summer Program	N
7/10/2014	TRW Said to Receive Takeover Approach From ZF Friedrichshafen	Y
8/1/2014	Judge Grants Preliminary Approval to Apple e-Book Settlement	N
9/26/2014	Pimco Said to Have Discussed Firing Gross Before Exit to Janus	Y
12/5/2014	CNN’s Candy Crowley to Leave Cable News Network After 27 Years	N
1/20/2015	FXCM Plunges as Bailout Lets Leucadia Force Sale of Brokerage	Y
3/12/2015	Viacom Says Chairman Redstone Will Miss Monday’s Annual Meeting	N
4/28/2015	McDonald’s Axes Seven Sandwiches in Push to Get Its Menu Right	Y
6/3/2015	Pandora Internet Radio Wins U.S. Nod to Buy South Dakota Station	N
6/11/2015	Biotech Led by 29-Year-Old CEO Now Worth Billions With No Sales	Y
7/29/2015	High-Density Drone Flights Possible Within Decade, Google Says	N
9/21/2015	Clinton’s Tweet on High Drug Prices Sends Biotech Stocks Down	Y
10/22/2015	Amazon Sales Top Estimates on Prime Day Event, Cloud Computing	Y
12/21/2015	Chipotle Probed for New Outbreak of Different E. Coli Strain	Y
1/14/2015	Apple, Ericsson Sue Each Other Over Phone Patent Royalties	N
2/27/2016	Lenovo to Purge Adware From New PCs After Superfish Controversy	N

this case, the next SI article to be published, upon its release, takes the available front page position. The process of article positioning is depicted in Figure 1.2.

As a result, there are two categories of news articles deemed equally important but having different positions: the SI articles that come out at a time when there are available front page slots and the SI articles that come out at a time when front page slots are unavailable. I hand-collect the positions of the SI articles in my sample. This subset of the news sample – SI articles in various positions – forms the basis for my causal analysis.

Screening of the articles confirms that there are no systematic differences in content between SI news articles that are placed on the front page and those that are not. Both include significant events, such as:

- “T-Mobile Said to Plan to Turn Down Iliad’s \$15 Billion Offer” (not front page)
- “Chipotle Probed for New Outbreak of Different E. Coli Strain” (front page)

But both front page and non-front page SI news articles also feature news events that carry less immediately relevant impact for financial markets. For example:

- “Morgan Stanley Gets 90,000 Applications for Summer Program” (not front page)
- “Pimco Said to Have Discussed Firing Gross Before Exit to Janus” (front page)

In Section 1.6.2, I compare the texts of the front page and non-front page SI news articles formally using machine learning techniques. I find no systematic differences between the two categories of news. Similarly, a survey of active finance professionals and MBA students from top business school programs indicates that human financial experts do not perceive the front page SI news articles to be any more significant than their non-front page counterparts.

Table 1.2 presents the distribution over time of PI and SI news articles published between the hours of 8AM and 5PM. All numbers are cited in ticker-articles, so that articles tagged with more than one U.S. equity security ticker are included one time for each tagged U.S. security. Overall, there are 2,362 PI article-tickers in the sample and 8,233 SI article-tickers,

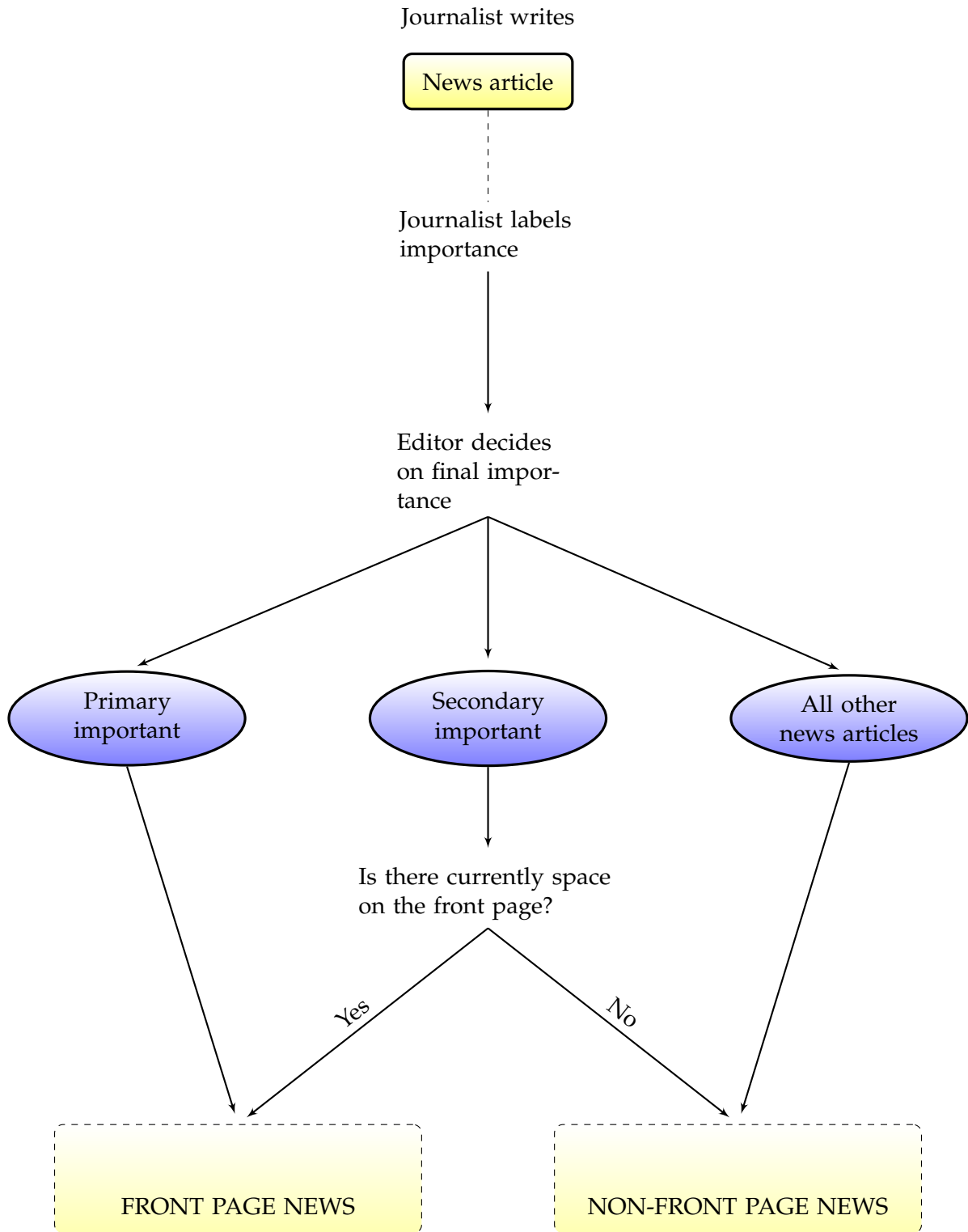


Figure 1.2: Process illustrating how Bloomberg news articles are pinned to the prominent front page positions at the top of the news screen.

Table 1.2: *Summary statistics of the hand-collected news sample.*

Panel 1: News Articles By Month

Hour of Day	PI articles	SI articles	FP SI articles
January	106	346	78
February	125	305	63
March	208	461	85
April	284	891	123
May	222	830	104
June	232	776	90
July	245	1,009	132
August	152	640	97
September	238	495	134
October	239	757	125
November	157	854	104
December	154	869	139
Total	2,362	8,233	1,274

Panel 2: News Articles By Hour

Hour of Day	PI articles	SI articles	FP SI articles	% SI articles on FP
8AM - 9AM	370	745	134	18%
9AM - 10AM	285	1,054	135	13%
10AM - 11AM	189	1,090	174	16%
11AM - 12PM	173	942	155	16%
12PM - 1PM	147	935	142	15%
1PM - 2PM	171	896	147	16%
2PM - 3PM	213	819	158	19%
3PM - 4PM	147	808	134	17%
4PM - 5PM	667	944	95	10%
Total	2,362	8,233	1,274	15%

of which 1,274 are given a front page position. The articles are roughly evenly distributed across the months of the year, with a lower volume of articles in January and February (and, to a lesser extent, March), since the sample begins on March 22, 2014 and hence does not cover these months in 2014. Over hours of the day, PI news articles peak at the start and end of the business day, during 8-10AM and especially during 4-5PM, while the SI news articles are more evenly distributed during the day. Consistent with the SI articles' positioning being determined by the concurrent volume of PI news, a lower percentage of SI articles makes it to the front page during the hours that see a higher volume of PI articles. The correlation between the hourly numbers of PI articles and the hourly likelihoods of SI articles receiving front page positions is -81%.

Examining the timing of news releases in the sample, I find no evidence of strategic release timing of the SI news articles. Of all the front page SI news articles, only 1.4% have a non-front page article released up to one minute before or after the front page article's publication. Similarly, only 0.7% of the front page SI articles are accompanied by non-front page SI articles within 30 seconds before or after. A mere 0.2% of the articles in the front page SI sample have a non-front page article released within 10 seconds of their publication. This low volume of SI news articles leaves little scope for influencing article position by strategically timing the exact seconds of when the articles are released. As a result, the process is unlikely to be contaminated by editorial staff being faced with multiple SI articles to be released at the same time and strategically releasing the more important ones first.

I also find that the article volume does not appear to be driven by editorial targets. I observe the distribution of articles across days and find that the volumes of PI and SI news articles vary dramatically from day to day. The number of PI news articles ranges from 0 to 40 per day, while SI articles can number anywhere between 0 and 67 per day. There is also little relationship between the numbers of PI and SI articles on any given day. The daily numbers of the two types of articles display a low correlation of 25%. As shown in Figure 1.3, any given day can see a large number of PI articles accompanied by few SI articles, and vice versa. Overall, the distribution of PI and SI articles across days indicates that the

Daily Numbers of "Primary Important" and "Secondary Important" Articles

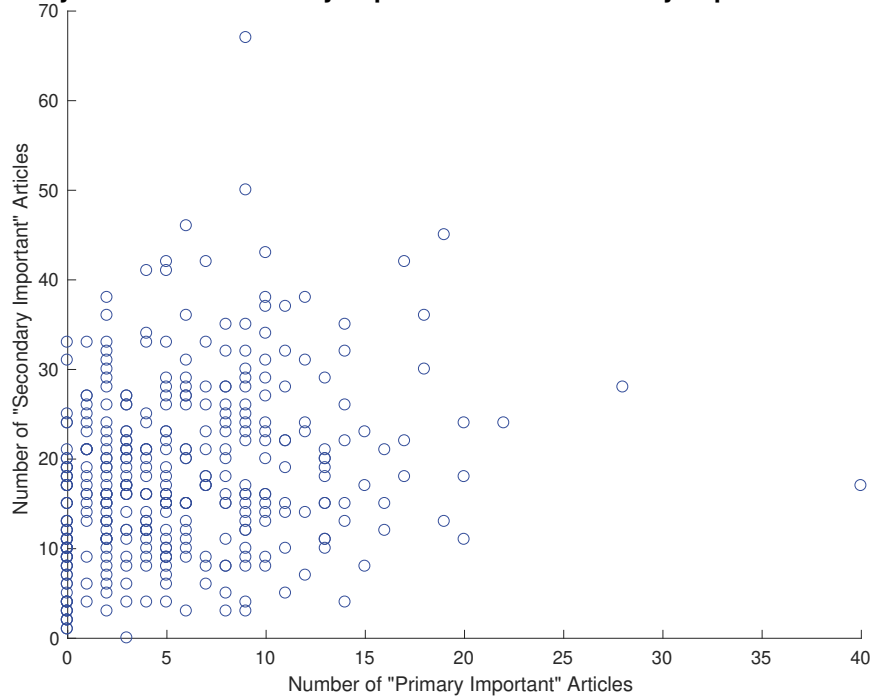


Figure 1.3: Distribution of daily volumes of "primary important" (PI) and "secondary important" (SI) news articles across the days in the sample.

editorial staff is not targeting particular numbers of high-importance articles. Instead, the patterns are more consistent with the evaluation of each article's importance being based on its own merit, independently of the volume of other news.

1.2.2 Market Data

I use the security ticker tags to merge the news position data with market data from several sources. Industry classification, market capitalization, and shares outstanding come from Compustat. High frequency price and trading data come from QuantQuote. 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 high frequency tests are run using news articles tagged with all firms for which

there are pricing data in QuantQuote, and shares outstanding and NAICS industry codes in Compustat. The merged sample includes 948 front page SI article-ticker pairs, 4,930 non-front page SI article-ticker pairs, and 1,650 PI article-ticker pairs. All of these article-ticker pairs have at least one price data point in QuantQuote on the day of publication, but not necessarily within shorter windows. Recall that PI news articles are more likely to come out during the hours of 8-9AM EST and especially 4-5PM EST. As a result, the empirical tests, which require market data within short windows of publication, reduce the PI news sample more substantially than the two SI news samples.

1.3 Conceptual Framework

In this section, I present a conceptual framework formalizing the intuition regarding the differences between front page and non-front page news articles. I outline two key aspects in the way investors are likely to pay differential attention to news articles in different positions, and then trace out the implications of these aspects for the process of incorporation of information into asset prices.

The conceptual framework follows the setups in Hirshleifer and Teoh (2003) and DellaVigna and Pollet (2009). There is a risk-free asset with a zero rate of return and a single risky security with a stochastic payoff R normally distributed with mean \bar{R} and variance σ_R^2 , realized in an unmodeled final period T . In the relatively short-term empirical settings that I consider, the realized value R can be taken to denote, for example, the price on which an asset settles in the days following an earnings announcement or the price of the combined enterprise following an acquisition. The risky asset is in fixed supply X . For expositional simplicity, I fix $X = 0$, so that the asset is in zero net supply; this simplifies the notation without affecting the results.

There is a continuum of investors with total mass equal to 1, who maximize mean-variance utility. In particular, let $W^{(i)}$ denote investor i 's final wealth at the end of the game

at time T . Then at any point in time t , investor i maximizes expected utility of the form

$$\mathbb{E}_{i,t}\{W^{(i)}\} - \frac{A^{(i)}}{2} \text{Var}_{i,t}\{W^{(i)}\} \quad (1.1)$$

with respect to his current holdings. For expositional simplicity, I take the risk-aversion coefficient to be identical across investors and normalize it to one: $\forall i, A^{(i)} = 1$. Each investor i is initially endowed with wealth $W_0^{(i)}$. There are no liquidity constraints.

Information in this framework is modeled as a signal arriving at a particular point in time and gradually diffusing across the population of investors. In particular, there are four periods in the model. In period 0, investors form prior expectations regarding the distribution of R . In period 1, a noisy signal (news) is released, and investors update their expectations accordingly. In periods 2 and 3, investors continue to update their beliefs following the news signal. At the end of the game, in the unmodeled period T , the true value of R is realized and the investors consume their final 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 0 and variance σ_ϵ^2 .

The news signal is not immediately observed by all investors. Instead, the main news source, S , reports the news signal N for some number of periods. Mass $\gamma > 0$ of investors are attentive to the main source S in each period t . Thus, in each period t that S reports the news signal N , a fraction γ of investors who had not observed the news signal prior to t now become aware of N .

I model the difference between front page and non-front page news with two key features. First, front page news articles induce more attention overall, so that the fraction of investors attentive to the news signal is higher: $\gamma = \bar{\gamma}$ in the case of front page news and $\gamma = \underline{\gamma} < \bar{\gamma}$ in the case of non-front page news. Second, front page news corresponds to the signal being reported by S for longer. Thus, for non-front page news, investors can observe the signal N from the main source S only in period 1. For front page news, by contrast, investors can also observe the signal from the main source S in period 2.

Investors may also learn the news from alternative sources, albeit at a lower rate. In

particular, in any period when the news is not being reported by S , a fraction $\xi > 0$ of uninformed investors still observe the news signal. This additional information channel can be interpreted as investors finding the news through filters or active searches once it scrolls off the top of the Bloomberg terminal screen, or reading the news from other providers. This channel is a minor one in the model, and I assume that most investors who receive the news do so from the main source S . In particular, I assume that:

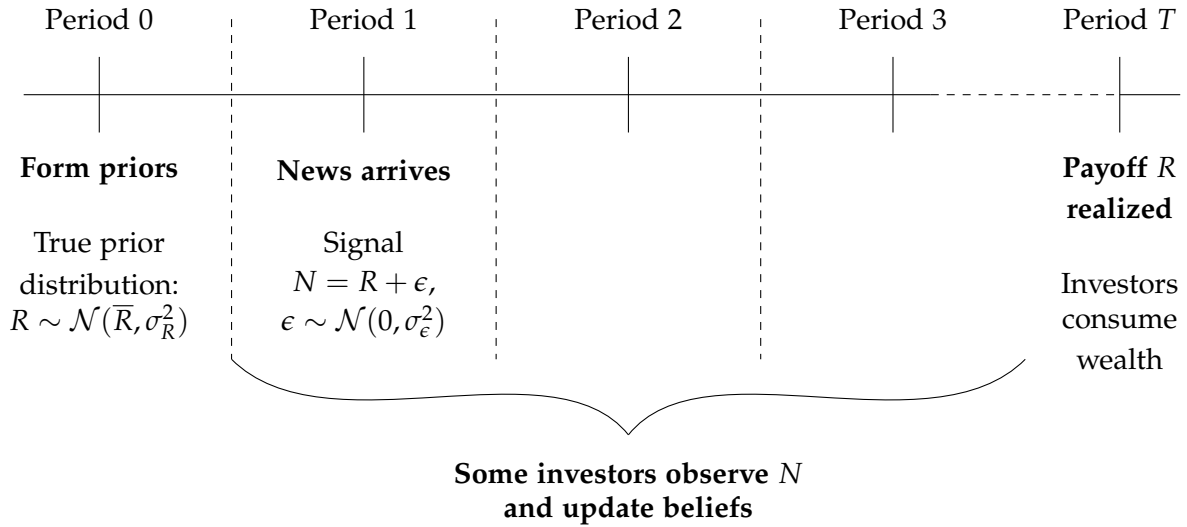
$$\xi < \frac{1 - \bar{\gamma}}{1 - \underline{\gamma}} \gamma \quad (1.2)$$

This condition ensures that once the main news source stops actively reporting the news (i.e., when the news is not on the front page), the fraction of informed investors does not increase faster than when the source continues to report (front page news). Consistent with the information disseminating relatively slowly over the short horizons considered in my empirical analysis, I also assume that both γ and ξ are small: $\gamma, \xi \ll 1/2$.

The model timeline is depicted in Figure 1.4. In each period t , let I_t denote the set of informed investors, who observe the news signal either during or prior to t , and let $F_t = |I_t|$ be the share of informed investors. I denote the remaining uninformed investors by U_t . Let F_t^{FP} and F_t^{NFP} denote the values of F_t in the cases of front page and non-front page news, respectively. Figure 1.4 illustrates the arrival of information and the evolution of the share of informed investors for both front page and non-front page news.

The key frictions in the model are that (1) some investors are inattentive; and (2) investors update their beliefs in a naïve Bayesian manner. Namely, some of the investors do not observe the public signal, and all investors update their beliefs with respect to only their own information, without taking into account the information sets and actions of others. In particular, while all investors observe equilibrium prices in all periods, they do not use the information contained in the price history to update their beliefs. These assumptions are standard modeling devices in models of gradual information diffusion (see Hong and Stein (1999), Hirshleifer and Teoh (2003), or Peng and Xiong (2006)).

I characterize the price path and trading volume following a news signal as a function



Non-front page news:

	Hazard rate $\underline{\gamma}$ of observing N	Hazard rate $\underline{\zeta}$ of observing N	Hazard rate $\underline{\zeta}$ of observing N
$F_0 = 0$	$F_1 = \underline{\gamma}$	$F_2 = \underline{\gamma} + (1 - \underline{\gamma})\underline{\zeta}$	$F_3 = \underline{\zeta} + (1 - \underline{\zeta}) \times [\underline{\gamma} + (1 - \underline{\gamma})\underline{\zeta}]$

Front page news:

	Hazard rate $\bar{\gamma}$ of observing N	Hazard rate $\bar{\gamma}$ of observing N	Hazard rate $\bar{\zeta}$ of observing N
$F_0 = 0$	$F_1 = \bar{\gamma}$	$F_2 = \bar{\gamma} + (1 - \bar{\gamma})\bar{\gamma}$	$F_3 = \bar{\zeta} + (1 - \bar{\zeta}) \times [\bar{\gamma} + (1 - \bar{\gamma})\bar{\gamma}]$

Figure 1.4: Model timeline, illustrating the diffusion of information for front page and non-front page news articles, as well as corresponding shares of informed investors in each period.

of the fraction of attentive investors F_t (Section 1.4.1). The empirical predictions for the differences in market dynamics following front page and non-front page news are then derived in Section 1.4.2.

1.3.1 Evolution of Prices and Trading Volumes

I begin by characterizing the price levels and trading volumes in terms of the fraction of attentive investors F_t , without distinguishing whether the news signal is reported on the front page or not.

Price levels. First, note that the uninformed investors hold the prior beliefs that the return R is normally distributed with mean \bar{R} and variance σ_R^2 . The informed investors attend to the signal and update their beliefs in a naïve Bayesian manner. Hence, their beliefs are given by:

$$\forall t \in \{0, 1, 2, 3\}, i \in I_t : \mathbb{E}_t^{(i)}\{R\} = \frac{\sigma_\epsilon^2 \bar{R} + \sigma_R^2 N}{\sigma_R^2 + \sigma_\epsilon^2}; \text{Var}_t^{(i)}\{R\} = \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2} \quad (1.3)$$

Next, note that optimization of the mean-variance preferences given by (1.1) with the above beliefs results in the following demand functions by the two groups of investors during any period t :

$$\forall t \in \{0, 1, 2, 3\}, i \in I_t : x_t^{(i)} = \frac{\sigma_\epsilon^2 (\bar{R} - P_t) + \sigma_R^2 (N - P_t)}{\sigma_R^2 \sigma_\epsilon^2} \quad (1.4)$$

$$\forall t \in \{0, 1, 2, 3\}, i \in U_t : x_t^{(i)} = \frac{\bar{R} - P_t}{\sigma_R^2} \quad (1.5)$$

where P_t denotes the price of the risky asset in period t .

The market clearing condition each period is that the total demand from the informed and uninformed investors must equal the zero net supply. Hence, in each period t , the price of the asset P_t must satisfy:

$$\forall t \in \{0, 1, 2, 3\} : F_t \frac{\sigma_\epsilon^2 (\bar{R} - P_t) + \sigma_R^2 (N - P_t)}{\sigma_R^2 \sigma_\epsilon^2} + (1 - F_t) \frac{\bar{R} - P_t}{\sigma_R^2} = 0 \quad (1.6)$$

Solving this equation gives the following expression for the price of the asset during

each period t :

$$\forall t \in \{0, 1, 2, 3\} : P_t = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + F_t \sigma_R^2} \bar{R} + \frac{F_t \sigma_R^2}{\sigma_\epsilon^2 + F_t \sigma_R^2} N \quad (1.7)$$

Absolute price changes. Taking the first differences yields the absolute price change between any two consecutive periods:

$$\forall t \in \{1, 2, 3\} : |\Delta P_t| = |P_t - P_{t-1}| = \frac{(F_t - F_{t-1}) \sigma_R^2 \sigma_\epsilon^2 |N - \bar{R}|}{(\sigma_\epsilon^2 + F_t \sigma_R^2)(\sigma_\epsilon^2 + F_{t-1} \sigma_R^2)} \quad (1.8)$$

Price continuation. In order to calculate the continuation in the price path, recall that the news signal has the form $N = R + \epsilon$, where R and ϵ are independent normal variables with $R \sim \mathcal{N}(\bar{R}, \sigma_R^2)$ and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Hence, price continuation, measured as the slope in a regression predicting the price change in period $t + 1$ from the price change in period t , is given by:

$$\forall t \in \{1, 2\} : Cont(t, t + 1) = \frac{Cov(\Delta P_t, \Delta P_{t+1})}{Var(\Delta P_t)} = \left(\frac{F_{t+1} - F_t}{F_t - F_{t-1}} \right) \left(\frac{\sigma_\epsilon^2 + F_{t-1} \sigma_R^2}{\sigma_\epsilon^2 + F_{t+1} \sigma_R^2} \right) \quad (1.9)$$

Note that this expression is defined for any setting where a non-trivial set of investors learns the news during the earlier period t . This holds for both the front page and the non-front page news in my setting, since even in absence of reporting by the main source S , news diffuses at the low but nonzero hazard rate ζ .

Trading volumes. Trading volume in each period t consists of all holdings that exchange hands between periods $t - 1$ and t . In each period, the newly informed investors, i.e. investors $i \in I_t \cap U_{t-1}$ change their demand following receipt of the news signal, inducing a change in the equilibrium price and the other investors' equilibrium holdings. Let $x_i^{(I)}$ denote the equilibrium holdings, in period t , of an investor $i \in I_t$; similarly, let $x_i^{(U)}$ denote the equilibrium holdings of an investor $u \in U_t$. Trading volume in each period can be expressed as a function of the newly informed investors' holdings as follows:

$$\forall t \in \{1, 2, 3\} : TV_t = (F_t - F_{t-1}) |x_i^{(I)} - x_{i-1}^{(U)}| \quad (1.10)$$

Taking the holdings from (1.4)-(1.5) and the equilibrium price levels from (1.7) then gives

the following expression for each period's trading volume:

$$\forall t \in \{1, 2, 3\} : TV_t = (F_t - F_{t-1}) \frac{(1 - F_t)\sigma_\epsilon^2 + F_{t-1}(\sigma_\epsilon^2 + \sigma_R^2)}{(\sigma_\epsilon^2 + F_t\sigma_R^2)(\sigma_\epsilon^2 + F_{t-1}\sigma_R^2)} |N - \bar{R}| \quad (1.11)$$

1.3.2 Empirical Predictions

I now compare the expressions for price changes, trading volumes, and price continuation for front page and non-front page news, and derive empirical predictions for differential market dynamics following different article positions.

Before proceeding, I note the evolution of the share of informed investors, F_t , in the cases of front page and non-front page news. In the first period, $F_0^{NFP} = F_0^{FP} = 0$. After that, the share of informed investors following non-front page news evolves as follows:

$$F_t^{NFP} = \begin{cases} \underline{\gamma} & \text{for } t = 1 \\ \underline{\gamma} + (1 - \underline{\gamma})\underline{\xi} & \text{for } t = 2 \\ \underline{\xi} + (1 - \underline{\xi})(\underline{\gamma} + (1 - \underline{\gamma})\underline{\xi}) & \text{for } t = 3 \end{cases} \quad (1.12)$$

Following front page news, meanwhile, the share of informed investors evolves as follows:

$$F_t^{FP} = \begin{cases} \bar{\gamma} & \text{for } t = 1 \\ \bar{\gamma} + (1 - \bar{\gamma})\bar{\gamma} & \text{for } t = 2 \\ \bar{\xi} + (1 - \bar{\xi})(\bar{\gamma} + (1 - \bar{\gamma})\bar{\gamma}) & \text{for } t = 3 \end{cases} \quad (1.13)$$

Combining the shares of informed investors in (1.12)-(1.13) with the price changes in (1.8) gives the immediate absolute price changes after non-front page and front page news:

$$|\Delta P_1^{NFP}| = \frac{\underline{\gamma}\sigma_R^2}{\sigma_\epsilon^2 + \underline{\gamma}\sigma_R^2} |N - \bar{R}|; \quad |\Delta P_1^{FP}| = \frac{\bar{\gamma}\sigma_R^2}{\sigma_\epsilon^2 + \bar{\gamma}\sigma_R^2} |N - \bar{R}| \quad (1.14)$$

Given that $\bar{\gamma} > \underline{\gamma}$, the first-period absolute price change is larger following front page news than following non-front page news.

Similarly, trading volumes at the news release in the first period are given by:

$$TV_1^{NFP} = \frac{\underline{\gamma}(1 - \underline{\gamma})}{(\sigma_\epsilon^2 + \underline{\gamma}\sigma_R^2)} |N - \bar{R}|; \quad TV_1^{FP} = \frac{\bar{\gamma}(1 - \bar{\gamma})}{(\sigma_\epsilon^2 + \bar{\gamma}\sigma_R^2)} |N - \bar{R}| \quad (1.15)$$

The relationship between immediate trading volume around the news signal and the percentage of immediately informed investors is non-monotonic. Trading volume is low if 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. Recall that $\gamma, \epsilon \ll 1/2$, reflecting the empirical setting I consider, where the proportion of the population who see any news article immediately (within the first few minutes of publication) is relatively low even for front page news. As a result, the split of attentive versus inattentive investors is more equal and the immediate trading volume is higher when the news is pinned to the front page.

Together, the price and volume expressions give the first empirical prediction regarding the immediate market response to front page and non-front page news.

Prediction 1 (Immediate Market Response) *Front page news articles are followed by larger trading volumes and absolute price moves immediately (within minutes) after the news.*

How does the price response play out outside of the immediate window? To see this, I turn to the continuation in the price path. I begin with the short-term continuation:

$$Cont^{NFP}(\Delta P_1, \Delta P_2) = \frac{(1 - \underline{\gamma})\underline{\xi}}{\underline{\gamma}} \times \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + [\underline{\gamma} + (1 - \underline{\gamma})\underline{\xi}]\sigma_R^2} \quad (1.16)$$

$$Cont^{FP}(\Delta P_1, \Delta P_2) = (1 - \bar{\gamma}) \times \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + [\bar{\gamma} + (1 - \bar{\gamma})\bar{\gamma}]\sigma_R^2} \quad (1.17)$$

Note that from condition (1.2), the first term of $Cont^{FP}(\Delta P_1, \Delta P_2)$ is larger than the first term of $Cont^{NFP}(\Delta P_1, \Delta P_2)$. The second term is larger in $Cont^{NFP}(\Delta P_1, \Delta P_2)$, since $\bar{\gamma} > \underline{\gamma} > \underline{\xi}$. However, for sufficiently low levels of immediate attention $\bar{\gamma}$ and $\underline{\gamma}$, the former effect dominates. This results in the following empirical prediction.

Prediction 2 (Immediate Return Continuation) *Front page news articles are accompanied by higher continuation in the short-term price changes.*

While front page news articles are followed by a larger immediate reaction that continues in the , the longer term dynamics are quite different. To see this, note that the continuation

in returns from the second to the third period for front page and non-front page news is given by:

$$Cont^{NFP}(\Delta P_2, \Delta P_3) = (1 - \xi) \times \frac{\sigma_\epsilon^2 + \underline{\gamma}\sigma_R^2}{\sigma_\epsilon^2 + [\xi + (1 - \xi)(\underline{\gamma} + (1 - \underline{\gamma})\xi)]\sigma_R^2} \quad (1.18)$$

$$Cont^{FP}(\Delta P_2, \Delta P_3) = \frac{(1 - \bar{\gamma})\xi}{\bar{\gamma}} \times \frac{\sigma_\epsilon^2 + \bar{\gamma}\sigma_R^2}{\sigma_\epsilon^2 + [\xi + (1 - \xi)(\bar{\gamma} + (1 - \bar{\gamma})\bar{\gamma})]\sigma_R^2} \quad (1.19)$$

Note that with $\xi < \underline{\gamma} < \bar{\gamma} \ll 1/2$, expressions (1.18)-(1.19) imply that the continuation from the second period to the third is actually lower for front page news compared to non-front page news. This yields the third empirical prediction of the gradual information diffusion framework.

Prediction 3 (Delayed Return Continuation) *Front page news articles induce lower continuation in the long-term price changes.*

In the next section, I test Predictions 1, 2, and 3 by observing the market dynamics following front page and non-front page Bloomberg news articles in my hand-collected sample. For the immediate news release window, $t = 1$, I look at the 5-10 minutes following publication of each individual news article. As the short-term subsequent window, $t = 2$, I consider 30-45 minutes following the news, as the front page news articles tend to remain prominently positioned for approximately half an hour to an hour. For the longer horizon, $t = 3$, I consider windows of 60, 90, and 120 minutes following the news release.

1.4 News Positioning and Market Dynamics

Using the natural experiment in news positioning, in this section, I empirically estimate the causal effect of front page news positioning on financial markets.

1.4.1 News Positioning and Short-Term Market Dynamics

I begin the analysis of differential activity following comparable front page and non-front page news articles by observing the short-term trading volume surges and price dynamics

following the two types of SI news. Placing a piece of news on the front page is associated with substantially larger trading volumes and absolute price changes within minutes of publication, as well as with higher continuation in the short-term price paths.

Consistent with Prediction 1, the more saliently positioned front page news articles induce significantly higher trading volumes. The median 15-second trading volume, computed as the percentage of shares turned over during the ten minutes before and after SI news articles, is displayed in Panel 1 of Figure 1.5. The median non-front page SI news article is accompanied by virtually no increase in trading volume (plotted in light blue in the figure) relative to the pre-news baseline. There is, however, a pronounced increase in the trading volumes following SI news articles that appear on the front page (displayed in dark blue). The difference in averages is even starker. Over the ten minutes after a news release, the average non-front page SI news article is accompanied by a total of 0.05% turnover. The average ten-minute trading volume after front page news is almost four-fold larger, at 0.19%. The difference is statistically significant at the 1% level, with a t-statistic of 4.52, as reported in Panel 1 of Table 1.3. The estimated difference remains identical when controlling for month and hour fixed effects, log market capitalization, and industry fixed effects.

Does the increased market activity reflected in trading volume correspond to larger price changes? Panel 2 of Figure 1.5 presents the average absolute percentage price changes following front page and non-front page SI news articles. The absolute price changes are calculated separately for each firm over every five-second interval. The graph averages the price changes in event time over the cross-section of firms. As a reference, the graph also plots, in dashed lines, the baseline price changes computed over the same time period for the same securities 24 hour prior to the publication of the news articles. Confirming the comparability of the two sets of articles, the pre-news baselines are statistically indistinguishable for the two samples of news articles. After publication, both front page news articles and non-front page news articles are accompanied by larger absolute price changes than their respective baselines.

Two patterns emerge from a visual inspection of the absolute price changes. First,

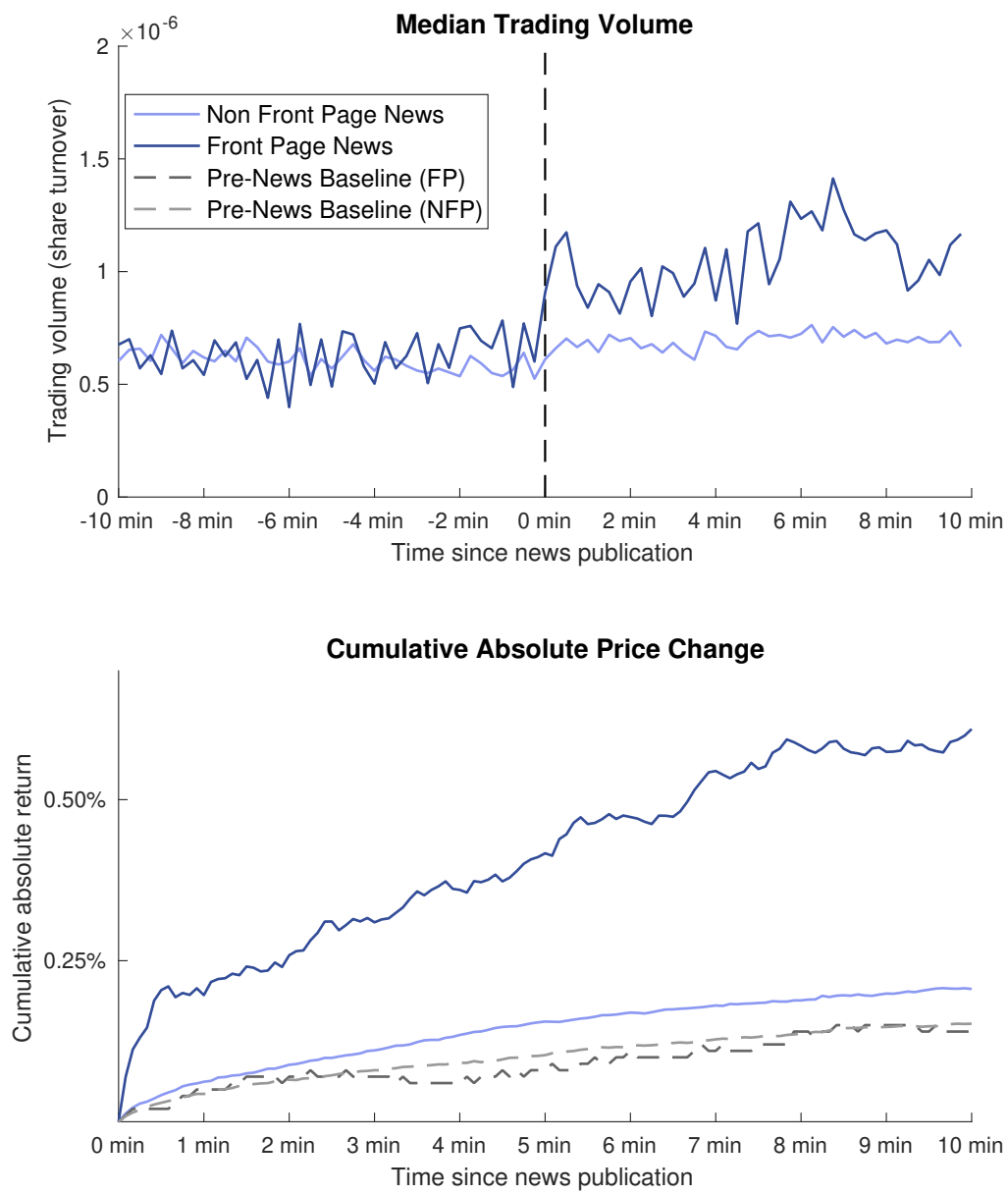


Figure 1.5: Market dynamics following front page and non-front page SI news articles.

Table 1.3: Comparison of trading volumes and absolute price changes immediately following SI news articles that are pinned to the front page and those that are not.

Panel 1: Trading Volume

	Front Page SI News	Non-Front Page SI News	Difference (FP – NFP)
First 5 min	0.10%	0.02%	0.07%**
Standard Error	(0.012%)	(0.001%)	(0.013%)
# Observations	847	4,095	–
First 10 min	0.19%	0.05%	0.14%**
Standard Error	(0.030%)	(0.002%)	(0.031%)
# Observations	858	4,233	–
First 1 hour	0.58%	0.26%	0.32%**
Standard Error	(0.143%)	(0.012%)	(0.143%)
# Observations	897	4,459	–

Panel 2: Absolute Price Changes

	Front Page SI News	Non-Front Page SI News	Difference (FP – NFP)
First 5 min	0.42%	0.16%	0.26%**
Standard Error	(0.041%)	(0.006%)	(0.042%)
# Observations	847	4,095	–
First 10 min	0.60%	0.21%	0.39%**
Standard Error	(0.065%)	(0.006%)	(0.066%)
# Observations	858	4,233	–
First 1 hour	0.98%	0.51%	0.47%**
Standard Error	(0.091%)	(0.020%)	(0.094%)
# Observations	897	4,459	–

** denotes significance at the 1% level.

the overall price change from the time of news publication to ten minutes later is much larger for SI news articles that are positioned on the front page than for those that are not. Second, corresponding to the more persistent attention garnered by the front page news articles being saliently positioned for longer, price changes after these news articles are more persistent. I consider these two effects in greater detail below.

I begin the statistical analysis of price effects by looking at the differential immediate price reactions to front page and non-front page SI news articles. Lending further support to Prediction 1, the average absolute price change within the first ten minutes after front page SI news articles is 60 basis points, compared to 21 basis points for non-front page SI news. The difference of 39 basis points is statistically significant at the 1% level, with a t-statistic of 5.91, as can be seen from Panel 2 of Table 1.3. The result is robust to the inclusion of controls: the estimated difference is 40 basis points when accounting for month and hour fixed effects, and 36 basis points when also controlling for log market capitalization and industry fixed effects. The results are similar at a shorter horizon of five minutes following the news, with an average absolute price change of 42 basis points accompanying front page news articles, compared to 16 basis points for non-front page news articles (t-statistic on the difference is 6.19). The contrast is less stark, but still significant when the window is extended to one hour following the news. The average absolute price change over the hour following front page SI news articles is 0.98%, whereas the average absolute price change over the hour following non-front page SI news articles is 0.51% (t-statistic on the difference is 5.01).

Having established empirical support for the first prediction of my conceptual framework, I now turn to Prediction 2. The theoretical prediction states that price paths following the front page SI news articles should display more short-term continuation, reflecting the more persistent attention garnered by news articles that stay at the top of the terminal screen for longer. I test the extent to which front page positioning induces higher short-term return continuation formally by estimating the following specification:

$$Ret_{s,i,[t+t_1,t+t_2]} = \alpha + \beta_1 Ret_{s,i,[t,t+t_1]} + \beta_2 FP_s + \beta_3 Ret_{s,i,[t,t+t_1]} \times FP_s \quad (1.20)$$

$$+\gamma X_{i,t} + \epsilon_{s,i,[t+t_1,t+t_2]},$$

where $Ret_{s,i,[t,t+t_1]}$ denotes the return on security i during the immediate period $[t, t + t_1]$ after publication of news article s , and $Ret_{s,i,[t+t_1,t+t_2]}$ is the return during the delayed period $[t + t_1, t + t_2]$. FP_s is an indicator variable equal to one for SI news articles that are pinned to the front page and zero for SI news articles not on the front page. The controls $X_{i,t}$ include month and hour of day fixed effects, as well as log firm size and industry fixed effects. The tests are run over the following time windows: $(t_1, t_2) \in \{(3 \text{ min}, 5 \text{ min}), (5 \text{ min}, 10 \text{ min}), (5 \text{ min}, 15 \text{ min}), (5 \text{ min}, 20 \text{ min}), (10 \text{ min}, 20 \text{ min}), (10 \text{ min}, 30 \text{ min})\}$.

Confirming Prediction 2, front page news articles are followed by higher serial correlation in returns at all considered short-term horizons, except for the shortest horizon of $(t_1, t_2) = (3 \text{ min}, 5 \text{ min})$. The coefficient of interest, β_3 , is positive and statistically significant across the other time specifications, as displayed in Table 1.4. For example, relative to non-front page SI news articles, front page SI news articles induce 17% more continuation in returns from the first five minutes after publication to the next five minutes. This result is economically sizable. For every 1% price move within the first five minutes after a front page SI news articles, there is an additional 17 basis points move in the same direction during the following 5 minutes, compared to non-front page SI news articles. The effect is also precisely estimated, with a t-statistic of 5.70 without controls, 5.67 with month and hour fixed effects, and 5.57 with the full set of controls including log firm size and industry fixed effects. Results over other windows are qualitatively similar, with the coefficient β_3 falling between 0.17 and 0.32, depending on the considered time windows.

Interestingly, the non-front page SI news articles are actually followed by short-term return reversal from the first five minutes to the next five to ten minutes, consistent with the literature on short-term price reversals.⁶ The coefficient on $Ret_{s,i,[t,t+t_1]}$ not interacted with the front page indicator is negative and statistically significant for $(t_1, t_2) \in \{(5 \text{ min}, 10 \text{ min}), (5 \text{ min}, 15 \text{ min}), (5 \text{ min}, 10 \text{ min})\}$. Effectively, these news articles, which are

⁶See, for example, Atkins and Dyl (1990) Ederington and Lee (1995), Fung *et al.* (2000), Chordia *et al.* (2002), Zawadowski *et al.* (2006), and Heston *et al.* (2010).

Table 1.4: Short-term continuation in returns after front page and non-front page SI news articles.

	$t_1 = 5 \text{ min}, t_2 = 10 \text{ min}$			$t_1 = 3 \text{ min}, t_2 = 5 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	-0.076**	-0.077**	-0.079**	0.040*	0.035†	0.034†
Standard error	(0.025)	(0.028)	(0.028)	(0.018)	(0.018)	(0.018)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	0.171**	0.170**	0.167**	0.033	0.041†	0.038†
Standard error	(0.030)	(0.030)	(0.030)	(0.022)	(0.022)	(0.022)
# FP SI articles	859	859	858	848	848	847
# Non-FP SI articles	4,235	4,235	4,233	4,097	4,097	4,095

	$t_1 = 5 \text{ min}, t_2 = 15 \text{ min}$			$t_1 = 5 \text{ min}, t_2 = 20 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	-0.120**	-0.121**	-0.117**	-0.170**	-0.168**	-0.168**
Standard error	(0.025)	(0.025)	(0.025)	(0.031)	(0.031)	(0.031)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	0.261**	0.258**	0.254**	0.313**	0.313**	0.317**
Standard error	(0.031)	(0.031)	(0.031)	(0.037)	(0.038)	(0.038)
# FP SI articles	864	864	863	871	871	869
# Non-FP SI articles	4,267	4,267	4,265	4,273	4,273	4,271

	$t_1 = 10 \text{ min}, t_2 = 20 \text{ min}$			$t_1 = 10 \text{ min}, t_2 = 30 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.043†	0.044†	0.046†	-0.011	-0.010	-0.014
Standard error	(0.026)	(0.026)	(0.027)	(0.034)	(0.034)	(0.034)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	0.173**	0.174**	0.170**	0.208**	0.208**	0.212**
Standard error	(0.028)	(0.028)	(0.029)	(0.037)	(0.038)	(0.038)
# FP SI articles	871	871	869	892	892	890
# Non-FP SI articles	4,273	4,273	4,271	4,310	4,310	4,308

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

prominently positioned at the top of the terminal screen only for short periods of time, see the initial five-minute price reactions partially reverse within the following minutes. On the other hand, front page SI news articles, which are prominently positioned for longer, are followed by a strong price drift over the short term.

1.4.2 News Positioning and Longer-Term Price Dynamics

Placing a piece of news on the front page induces sizable short-term price effects; do we see the non-front page information eventually catch up? As the front page news articles get removed from their prominent positions, the differences in diffusion of information contained in these articles and the non-front page articles gradually diminish. The conceptual framework predicts that at longer horizons, front page news articles should see *less* continuation in returns. I find evidence in support of this prediction: over longer horizons of one to two hours after the news, *non*-front page information induces substantially more price drift than front page news. The incorporation of non-front page information is much slower, however, and full convergence does not occur for days after the news.

I begin evaluating longer-term price continuation by estimating specification (1.20) over the following windows: $t_1 \in \{5 \text{ min}, 10 \text{ min}\}$ and $t_2 \in \{45 \text{ min}, 60 \text{ min}, 90 \text{ min}\}$. The results are reported in Panel 1 of Table 1.5.

The results reveal an interesting pattern of dynamics: the immediate returns over the first five minutes after news publication are more positively predictive of subsequent returns following front page news than following non-front page news, up to approximately forty-five minutes. But over longer horizons of sixty or ninety minutes, the effect is no longer present. Continuation in returns from the first five minutes to the remainder of the first hour is statistically indistinguishable for front page versus non-front page SI news articles. From the first five minutes to the remainder of the first hour and a half, there is slightly less continuation following front page news (significant at the 10% level).

Similarly, the initial ten-minute returns induced by front page news are followed by a stronger drift for about forty-five minutes. During the first forty-five minutes, front page

Table 1.5: Continuation in returns over longer horizons following front page and non-front page SI news.

Panel 1: Return continuation from the first 5-10 minutes up to 45-90 minutes

	$t_1 = 5 \text{ min}, t_2 = 45 \text{ min}$			$t_1 = 10 \text{ min}, t_2 = 45 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.084**	0.083*	0.080*	0.121**	0.120**	0.123**
Standard error	(0.033)	(0.033)	(0.033)	(0.040)	(0.040)	(0.040)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	0.338**	0.342**	0.344**	0.315**	0.314**	0.317**
Standard error	(0.058)	(0.058)	(0.059)	(0.048)	(0.048)	(0.048)
# Obs (FP; Non-FP)	894; 4,421	894; 4,421	892; 4,418	894; 4,421	894; 4,421	892; 4,418

	$t_1 = 5 \text{ min}, t_2 = 60 \text{ min}$			$t_1 = 10 \text{ min}, t_2 = 60 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.284**	0.283**	0.281**	0.154**	0.155**	0.157**
Standard error	(0.048)	(0.048)	(0.049)	(0.050)	(0.050)	(0.051)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	-0.123	-0.122	-0.126	0.081	0.083	0.080
Standard error	(0.079)	(0.081)	(0.081)	(0.059)	(0.060)	(0.060)
# Obs (FP; Non-FP)	899; 4,462	899; 4,462	897; 4,459	899; 4,462	899; 4,462	897; 4,459

	$t_1 = 5 \text{ min}, t_2 = 90 \text{ min}$			$t_1 = 10 \text{ min}, t_2 = 90 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.225**	0.228**	0.229**	0.124*	0.120*	0.122*
Standard error	(0.048)	(0.049)	(0.050)	(0.053)	(0.053)	(0.054)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	-0.115†	-0.113*	-0.109†	-0.103†	-0.106†	-0.107†
Standard error	(0.071)	(0.071)	(0.071)	(0.060)	(0.060)	(0.061)
# Obs (FP; Non-FP)	901; 4,475	901; 4,475	899; 4,472	901; 4,475	901; 4,475	899; 4,472

Panel 2: Return continuation from the first 30-45 minutes up to 90-120 minutes

	$t_1 = 30 \text{ min}, t_2 = 90 \text{ min}$			$t_1 = 45 \text{ min}, t_2 = 90 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.254**	0.248**	0.249**	0.191**	0.187**	0.186**
Standard error	(0.029)	(0.029)	(0.030)	(0.016)	(0.016)	(0.016)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	-0.143**	-0.142**	-0.145**	-0.215**	-0.215**	-0.214**
Standard error	(0.032)	(0.032)	(0.032)	(0.020)	(0.020)	(0.020)
# Obs (FP; Non-FP)	901; 4,475	901; 4,475	899; 4,472	901; 4,475	901; 4,475	899; 4,472

	$t_1 = 30 \text{ min}, t_2 = 120 \text{ min}$			$t_1 = 45 \text{ min}, t_2 = 120 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.266**	0.267**	0.255**	0.226**	0.224**	0.221**
Standard error	(0.035)	(0.035)	(0.036)	(0.017)	(0.017)	(0.017)
$Ret_{s,i,[t,t+5 \text{ min}]} \times FP_s$	-0.185**	-0.183**	-0.188**	-0.264**	-0.268**	-0.273**
Standard error	(0.032)	(0.032)	(0.033)	(0.021)	(0.021)	(0.021)
# Obs (FP; Non-FP)	903; 4,491	903; 4,491	901; 4,488	903; 4,491	903; 4,491	901; 4,488

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

news articles induce an additional drift of approximately 31% of the initial ten-minute return. Expanding the window to sixty minutes, the continuation becomes statistically indistinguishable between front page and non-front page news, and at ninety minutes there is weakly more continuation for article that are *not* pinned to the front page.

As I shift the window even further, the results lend empirical support to Prediction 3. Panel 2 of Table 1.5 reports estimates of specification (1.20) over the following windows: $t_1 \in \{30 \text{ min}, 45 \text{ min}\}$ and $t_2 \in \{90 \text{ min}, 120 \text{ min}\}$. The non-front page news articles are followed, on average, by 25-27% continuation in returns from the first half-hour to the remainder of the 90-120 minutes. Front page news articles, however, see 14-19% less continuation. The differences are highly statistically significant. Similarly, the returns from the first forty-five minutes are substantially less likely to continue if the news article is pinned to the front page. Non-front page news articles see a continuation of 19-22% from the first 45 minutes to the remainder of the first 90-120 minutes. By contrast, front page news articles actually see no return continuation over the same time windows.

Coupled with the results in Table 1.4, the longer-term price dynamics highlight the differences in the speed of incorporation of front page and non-front page information. Pinning a piece of news on the front page induces a stronger drift in returns up to forty-five minutes, and the reactions to non-front page articles begin to catch up over the remainder of the first couple of hours after news publication. Theoretically, these patterns are fully consistent with the gradual information diffusion framework outlined in Section 1.3. Practically, the results indicate that for news articles consumed by sophisticated finance professionals through a subscription-based platform such as Bloomberg, the market dynamics track the discretionary positioning in real time.

While the price impact of front page information occurs quickly, it takes substantially longer for non-front page information to be fully reflected in asset prices. Table 1.6 presents the average differences in trading volumes and absolute price changes one, two, five, ten, and fifteen days after front page and non-front page SI news articles. The differences are estimated controlling for month and hour fixed effects, log market capitalization, and

industry fixed effects. For the trading volume tests, I look at the total trading volume over a 10-minute window d days after news publication, where $d \in \{1, 2, 5, 10, 15\}$. Similarly, the absolute price changes are calculated as the absolute percentage difference in price from the time of news publication to exactly d days later.

Table 1.6: Differences in trading volumes and absolute price changes following front page and non-front page SI news articles over longer horizons.

Number of Days after News	Difference in:	
	Trading Volume	Absolute Price Change
$d = 1$	0.02%**	0.38%**
Standard Error	(0.01%)	(0.14%)
# Obs – SI FP	892	892
# Obs – SI NFP	4,432	4,432
$d = 2$	0.03%†	0.34%†
Standard Error	(0.02%)	(0.19%)
# Obs – SI FP	888	888
# Obs – SI NFP	4,415	4,415
$d = 5$	-0.01%	0.25%
Standard Error	(0.02%)	(0.20%)
# Obs – SI FP	890	890
# Obs – SI NFP	4,422	4,422
$d = 10$	-0.01%	0.18%
Standard Error	(0.03%)	(0.22%)
# Obs – SI FP	878	878
# Obs – SI NFP	4,403	4,403
$d = 15$	0.01%	0.08%
Standard Error	(0.03%)	(0.25%)
# Obs – SI FP	885	855
# Obs – SI NFP	4,411	4,411

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

The results indicate that some of the non-front page information is reflected in prices within the first few days, but a portion of the gap in market reactions induced by positioning remains even days after the news. The elevated level of trading volume accompanying front page news lasts for at most two days (the difference is statistically significant at the 1% level one day out, and at the 10% level two days out). The difference in absolute price changes one day after the news is highly statistically significant, but milder than the difference from just the first hour, tabulated in Table 1.3. This difference declines slightly but remains economically similar two days after the news, significant at the 5% level. Five days after

news publication, the price impact of front page positioning declines substantially and is no longer statistically significantly different from the price impact of non-front page news, although the difference remains economically visible at around 25 basis points. The gap is milder still, at 18 basis points, ten days out after the news, and converges to a statistically indiscernible 8 basis points fifteen days out after the news.

These patterns are reinforced in a graphical evaluation of the directional price paths following front page and non-front page news, displayed in Figure 1.6. I group the SI news articles in my sample along two dimensions: (1) their position (front page versus non-front page) and (2) the direction of the initial five-minute price move (positive versus negative). I take the average cumulative price paths across news articles in each category, in event time from the time of publication to various time windows. These price paths are plotted in solid lines for front page news and in dashed lines for non-front page news. The price paths for articles accompanied by positive initial five-minute price changes are shown in blue, while the price paths for articles accompanied by negative initial price changes are shown in red. In each case, the price change accompanying a given piece of news is computed relative to the market return over the same time period, in order to screen out the directional equity premia at longer horizons. Standard error bars are shaded in gray.

The figure shows a variety of time windows ranging from minutes to days after the news. In the immediate term (0-10 minutes, displayed in the first quarter of the figure), front page news articles are accompanied by larger price changes, in both the positive and the negative domains, consistent with the absolute price change results reported in Panel 2 of Table 1.3. This gap widens for about 45 minutes, and then begins to narrow, as can be seen from the price paths over the first hour after news publication. The narrowing of the gap continues for hours after the news, as front page articles see no additional price moves, while non-front page information continues being incorporated into prices. In the last quarter of the figure, I show the price responses from publication to one, two, five, ten, and fifteen days out after the news. Although the standard errors become very wide at these horizons, the economic magnitudes show no difference in the long-term reactions to

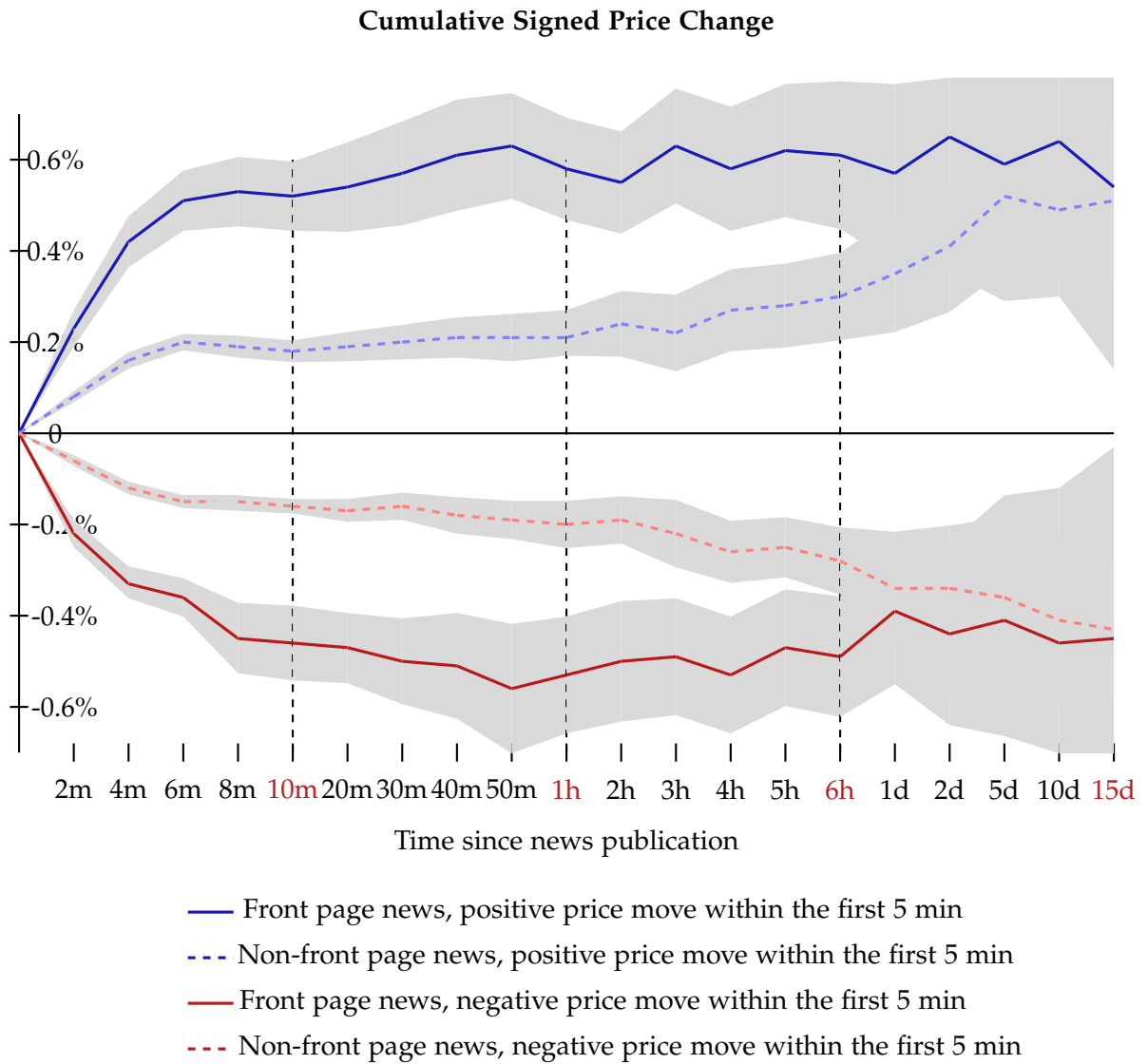


Figure 1.6: Price paths after front page and non-front page SI news, sliced by the direction of the initial 5-minute moves.

front page versus non-front page news, due to the non-front page news articles gradually catching up to their front page counterparts.

Overall, the effect of differential news positioning is stark and quick, and takes a while to converge. The gradual catching up of the reactions to non-front page information begins as early as an hour after publication, but the diffusion of information in non-front page news is quite slow. As a result, the effect of news positioning can be statistically noticeable and economically meaningful even several days after the news.

1.5 News Positioning versus News Importance

In this section, I compare the estimated effects of news positioning against the effects of news importance, as marked by the editorial staff. I estimate the relationship between news importance and market dynamics by concentrating on news articles that are all equally prominently positioned but that vary in importance – i.e., by comparing front page news articles from the PI and SI categories. The difference in market reactions following these two types of news is qualitatively different from and quantitatively weaker than the difference induced by front page positioning.

I limit my attention only to news articles that are pinned to the front page, so that there is no variation in the prominence of the article positions. I include all front page news articles, regardless of their importance markings, and estimate the difference in market reactions following the more (“primary important”) and less important (“secondary important”) news articles.

First, I note that the trading volumes immediately following front page PI news articles are not statistically different from the trading volumes following front page SI news articles. As displayed in Panel 1 of Table 1.7, during the first five minutes after a front page news article, on average, an additional 0.09% of shares turn over when the article is from the PI category, but this difference is not statistically significant. Similarly, during the first ten minutes, front page PI news articles are followed by an additional 0.10% in trading volume compared to front page SI news articles, significant only at the 10% level. The pattern

remains similar over longer horizons, with an average of 0.18% additional shares turned over during the hour following front page PI news articles, with the difference remaining statistically insignificant.

Table 1.7: *Trading volumes and absolute price changes following PI and front page SI news articles.*

Panel 1: Trading Volume

	Front Page SI News	PI News	Difference (PI–SI)
First 5 min	0.10%	0.18%	0.09%
Standard Error	(0.02%)	(0.04%)	(0.04%)
# Observations	847	1,291	–
First 10 min	0.19%	0.29%	0.10%†
Standard Error	(0.02%)	(0.04%)	(0.05%)
# Observations	858	1,306	–
First 60 min	0.57%	0.74%	0.18%
Standard Error	(0.08%)	(0.10%)	(0.13%)
# Observations	897	1,349	–

Panel 2: Absolute Price Changes

	Front Page SI News	PI News	Difference (PI–SI)
First 5 min	0.44%	0.79%	0.35%**
Standard Error	(0.04%)	(0.05%)	(0.05%)
# Observations	847	1,291	–
First 10 min	0.61%	1.01%	0.40%**
Standard Error	(0.07%)	(0.06%)	(0.09%)
# Observations	858	1,306	–
First 60 min	0.99%	1.40%	0.41%**
Standard Error	(0.09%)	(0.07%)	(0.11%)
# Observations	897	1,349	–

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

Second, while PI articles are accompanied by larger price impact than SI articles, the effect is less significant and less persistent than the difference in absolute price changes induced by front page positioning. As can be seen from Panel 2 of Table 1.7, in the first five minutes, front page PI news articles are followed by an additional 0.35% absolute price change, an increase of 80% over the front page SI articles; the difference is significant but statistically weaker than the difference between front page and non-front page SI articles.

The difference in absolute price changes following front page PI news articles versus front page SI news articles remains similar in magnitude and declines in statistical significance as the window is extended to ten and then sixty minutes. Overall, PI news articles are followed by larger price reactions immediately in the first five to ten minutes, but do not see further differences from the front page SI news articles over longer horizons. This contrasts with the difference between front page and non-front page SI news articles documented in Table 1.3, which continues to grow over the hour following the news.

This result is corroborated by a comparison of the continuation in the price paths following PI and front page SI news articles, which I estimate using the following specification:

$$Ret_{s,i,[t+t_1,t+t_2]} = \alpha + \beta_1 Ret_{s,i,[t,t+t_1]} + \beta_2 PI_s + \beta_3 Ret_{s,i,[t,t+t_1]} \times PI_s \quad (1.21)$$

$$+ \gamma X_{i,t} + \epsilon_{s,i,[t+t_1,t+t_2]},$$

where $Ret_{s,i,[t,t+t_1]}$ denotes the return on security i during the immediate period $[t, t + t_1]$ after publication of news article s , and $Ret_{s,i,[t+t_1,t+t_2]}$ is the return during the delayed period $[t + t_1, t + t_2]$. PI_s is an indicator variable equal to one if the front page article comes from the “primary important” category and zero if the article is from the “secondary important” category. The controls $X_{i,t}$ include month and hour fixed effects, log firm size, and industry fixed effects. The considered time windows are $(t_1, t_2) \in \{(5 \text{ min}, 10 \text{ min}), (5 \text{ min}, 15 \text{ min})\}$. Table 1.8 presents the results.

The estimated coefficient on $Ret_{s,i,[t,t+t_1]} \times PI_s$ indicates that front page PI news articles are not accompanied by any more short-term price drift compared to front page SI news articles. The difference is neither economically notable, nor statistically significant. Over the same time horizons, the difference in price drift following front page and non-front page SI news articles is 17% and highly statistically significant (see Table 1.4).

Recall that this analysis considers *only* news articles positioned on the front page, but of both categories: “primary important” and “secondary important.” Whereas the results in Section 1.4.2 keep article importance constant (only SI articles) and vary front page positioning, the analyses in this section keep the positioning constant but vary article

Table 1.8: Serial correlation in returns following PI and SI front page news articles. Each column estimates the following specification:

	$t_1 = 5 \text{ min}, t_2 = 10 \text{ min}$			$t_1 = 5 \text{ min}, t_2 = 15 \text{ min}$		
	(1)	(2)	(3)	(1)	(2)	(3)
$Ret_{s,i,[t,t+5 \text{ min}]}$	0.093*	0.091*	0.089*	0.137**	0.134**	0.136**
Standard error	(0.037)	(0.037)	(0.037)	(0.039)	(0.039)	(0.039)
$Ret_{s,i,[t,t+5 \text{ min}]} \times PIs$	-0.016	-0.020	-0.027	-0.025	-0.020	-0.026
Standard error	(0.043)	(0.044)	(0.044)	(0.045)	(0.045)	(0.045)
# PI articles	1,294	1,294	1,291	1,310	1,310	1,306
# FP SI articles	859	859	858	864	864	863

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

importance. As can be seen from a comparison of Tables 1.7-1.8 against Tables 1.3-1.4, differences in article importance correspond to milder differences in market dynamics than differences in article positioning. These findings suggest that article positioning is even more instrumental in driving market reactions than differences in article importance, as marked by Bloomberg’s journalistic and editorial staff and confirmed by the target audience of finance professionals (see Section 1.6.3 below).

1.5.1 Discussion: Attention versus Inference

The comparison of differential reactions to news position and news importance helps highlight the channel behind the market response to front page positioning. The drift patterns associated with prominent positioning, which are not observed for differential importance, indicate that the positioning effect is driven by attention patterns rather than inference regarding the importance of the underlying news.

Effectively, there are two mechanisms that could induce heightened market activity following front page news articles relative to non-front page articles. First is the attention channel highlighted by the conceptual framework in Section 1.3: front page news articles receive more immediate attention, corresponding to higher trading volumes and absolute price changes. The second mechanism is inference regarding the importance of the underlying

ing news: investors perceive the superior position to signal greater importance of the front page articles.

While both channels produce increases in immediate trading volumes and price changes, only the attention channel predicts the type of subsequent dynamics observed in the data. As captured by Predictions 2 and 3, the attention channel predicts that front page articles should be accompanied by more short-term drift and less continuation at longer horizons. If instead the initial reactions are driven by inference regarding the articles' importance, there is no reason to observe a pattern of higher short-term drift and subsequent gradual convergence.

The results on the differences between PI and SI front page news articles further support the gradual information diffusion interpretation. The differences between reactions to articles of actual varying importance are immediate, inducing no differential drifts. If the effect of positioning were driven primarily by the inference channel, then the timing of the positioning effect should be comparable to the effect of importance. Instead, differential positioning induces differences in incorporation of information that creates predictability in returns at a variety of horizons. This corresponds more closely to the timing of the conceptual model of gradual information diffusion.

Altogether, the effects of news positioning are not only more substantial than the differences between articles of varying editorial importance in the immediate term, but also induce differences in return predictability further out. These results support the importance of gradual information diffusion and highlight news consumption as playing a significant role in causally driving market dynamics around information releases.

1.6 Additional Analyses

I present additional analyses confirming the exclusion restriction of my natural experiment design: that the SI news articles that are pinned to the front page do not systematically differ from those that are not. First, I show that, holding position constant, the news articles published during quiet times (when more front page slots are available) do not generally

induce stronger market reactions than the articles published during busy times. Second, I use machine learning techniques to show that the distributions of topics discussed in the texts of the front page and non-front page SI news articles do not systematically differ. Lastly, I conduct a survey of active finance professionals and MBA students at top business school programs to highlight that, in absence of the differential positioning, the target audience finds the two sets of news to be indistinguishable in terms of importance and expected market impact.

1.6.1 Quiet Times vs. Busy Times

In this subsection, I address the potential concern that the differential reactions to front page and non-front page SI news articles are driven by the fact that the former are released during generally quieter times (when there are fewer PI news articles), rather than by different amounts of attention to the two types of articles. A few points are worth noting here.

First, to the extent that increased market activity during quiet times reflects increased attention dedicated to the security due to few other contemporaneous events, the results would still capture the attention channel. In fact, various indicators of “quiet times” have been used as indirect proxies for attention in prior work⁷ The analysis in Section 1.4 captures the variation in attention more precisely through the salience of news positioning.

Second, in my sample, without the differential positioning, SI news articles published during quiet times, when little goes on in the markets, are likely to be accompanied, if anything, by *less* market activity than the SI news articles published during the busier times. This would push in the direction of finding less market activity after the front page SI news articles (i.e., SI news articles released during quieter times), dampening my results.

I document this finding by comparing the non-front page SI news articles released during relatively quiet times with non-front page SI news articles released during relatively

⁷See DellaVigna and Pollet (2009) on earnings announcements released on Fridays, and Hirshleifer *et al.* (2009) on earnings announcements released contemporaneously with other announcements. Accordingly, deHaan *et al.* (2015) and Niessner (2015) provide evidence that firms strategically respond to investors’ limited attention by timing their releases. In other contexts, distraction has been shown to affect liquidity provision (Corwin and Coughenour (2008)) and corporate actions (Kempf *et al.* (2016))

busy times. Thus, I hold news position (non-front page) and news importance (SI) constant, and vary only the numbers of contemporaneous news releases.

To differentiate busy times from quiet times, I consider the contemporaneous volumes of articles within three time intervals: (1) on the same day as a given non-front page SI article; (2) within five hours of a given non-front page SI news article; and (3) within two hours of a given article. Non-front page SI news articles for which the contemporaneous volumes of other news fall below the median form the "quiet times" sample. Non-front page SI news articles with at or above-median contemporaneous volumes of other news form the "busy times" sample.

As displayed in Table 1.9, holding editorial importance markings and position constant, SI news articles that come out during quieter times are *not* accompanied by larger trading volumes and absolute price changes than the SI news articles published during busier times. If anything, price changes and trading volumes are smaller following non-front page SI news articles published during quiet times. These patterns are qualitatively consistent across definitions of quiet and busy times using the one day, five hour, and two hour windows. Statistically, the differences in absolute price changes and trading volumes after non-front page articles that come out during quiet and busy times are only discernible when the volume of contemporaneous news is measured on a daily level. Economically, the differences are small across the board, within a range of 1-3 basis points.

These results confirm that the differential market reactions following front page and non-front page SI news articles are not driven by the SI news articles that come out during quiet times (and are therefore more likely to take an available front page position) carrying more important content than the articles that come out during busy times.

1.6.2 Distributions of Topics

To rule out systematic differences in the content of front page and non-front page articles, I directly analyze the text of the news articles across different positions and levels of importance. The distribution of topics discussed in front page SI news articles is statistically

Table 1.9: Comparison of trading volumes and absolute price changes within ten minutes of non-front page SI articles published during quiet versus busy times.

Panel 1: 10-Minute Trading Volume

	News in Quiet Times	News in Busy Times	Difference
Window: 1 day	0.04%	0.06%	-0.02%*
Standard Error	(0.003%)	(0.006%)	(0.007%)
# Observations	3,383	3,576	–
Window: 5 hours	0.05%	0.06%	-0.01%
Standard Error	(0.002%)	(0.006%)	(0.008%)
# Observations	2,011	4,948	–
Window: 2 hours	0.05%	0.06%	-0.01%
Standard Error	(0.003%)	(0.006%)	(0.009%)
# Observations	3,224	3,735	–

Panel 2: 10-Minute Absolute Price Changes

	News in Quiet Times	News in Busy Times	Difference
Window: 1 day	0.20%	0.24%	-0.03%*
Standard Error	(0.01%)	(0.01%)	(0.02%)
# Observations	3,383	3,576	–
Window: 5 hours	0.21%	0.22%	-0.01%
Standard Error	(0.01%)	(0.01%)	(0.01%)
# Observations	2,011	4,948	–
Window: day	0.21%	0.21%	-0.01%
Standard Error	(0.01%)	(0.01%)	(0.01%)
# Observations	3,224	3,735	–

* denotes significance at the 5% levels.

indistinguishable from the distribution of topics covered by non-front page SI articles. By contrast, the distribution of topics discussed in PI news articles does differ somewhat from the SI news articles, with a larger focus on company operations and the healthcare industry, and lower coverage of regulations and the financial services industry.

Topic analysis provides an intuitive way to compare the content value of different news articles. The existing literature on the effect of news on financial markets considers

textual characteristics such as sentiment,⁸ grammatical structure,⁹ and complexity.¹⁰ The methodology in this section contributes to the literature by proposing an intuitive approach to identifying common topics in financial news and representing the news articles in terms of these prototypical areas of focus.

The topic analysis proceeds in two steps. First, I use a large corpus of news articles from Reuters to analyze textual patterns in financial news in general by representing the articles in the space of meaningful features and identifying a set of broadly applicable topics. Second, I apply the trained topic model to the news articles in the PI, front page SI, and non-front page SI samples of hand-collected Bloomberg news articles.

For the first step of the process, I use the Latent Dirichlet Allocation algorithm proposed by Blei *et al.* (2003) following similar methods employed in genetics (see, for example, Pritchard *et al.* (2000)). The Latent Dirichlet Allocation approach is particularly well suited to the problem at hand, because it represents all documents as being generated from an underlying set of topics by a latent process. This admits modeling out-of-sample documents as mixtures over the topics identified from the training data – i.e., modeling the news articles from the various Bloomberg categories in terms of topics identified from the larger sample of Reuters news. For a description of the Latent Dirichlet Allocation methodology, please refer to Appendix A.1.1.

In order to train the topic model on a dataset that is similar yet distinct from the Bloomberg news articles that I ultimately classify and evaluate, I use the Thomson Reuters Text Research Collection 2 (TRC2), part of the Thomson Reuters Research Collection described in Lewis *et al.* (2004). This training corpus includes approximately 1.8M news articles spanning the full spectrum of financial news reported by Reuters during the period of 2008-2009, and is available from the National Institute of Standards and Technology.

⁸See Tetlock (2007), Das and Chen (2007), Tetlock *et al.* (2008), Loughran and McDonald (2011), Bollen *et al.* (2011), García (2013), and Uhl (2014).

⁹See Engelberg (2008).

¹⁰See Li (2008), You and Zhang (2009), Miller (2010), Lehavy *et al.* (2011), Loughran and McDonald (2014), and Umar (2017).

Appendix A.1.2 describes the pre-processing of the news articles in order to represent them in terms of meaningful textual features ready to be inputted into the Latent Dirichlet Allocation algorithm.

The output of the Latent Dirichlet Allocation model provides an intuitive conceptualization of the identified topics in terms of the most frequently occurring words conditional on each topic. I estimate the model for $k = \{10, 15, 20, 25, 30, 35, 40\}$ topics, and observe that the specification with 15 topics performs best in terms of model log likelihood (see Appendix A.1.3 for details on the topic model estimation). The topics identified in this specification are presented in Table 10. For each topic, the table displays the fifteen terms in the vocabulary that are most likely to appear conditional on that topic. For each of the topics, the set of common terms forms a single coherent theme; for clarity of reference, each topic is labeled with a concise name capturing its theme. For example, the topic whose most common terms are "court," "case," "judge," "federal," etc. is labeled "Litigation;" while the topic whose most common terms are "deal," "offer," "price," "bid," etc. is labeled "Mergers & Acquisitions."

The topics in Table 1.10 are listed in order of their estimated frequencies, which are presented in the last column. The most common topic to appear in the training corpus of Reuters financial news relates to technology, followed by financial reports such as earnings, and then news regarding financial institutions such as hedge funds and banks. Other common topics include automobile and air transport industries, litigation, and management. Overall, the identified topics are generally applicable and representative of concepts discussed in financial news.

For the second part of the process, I take advantage of the Latent Dirichlet Allocation's ability to represent out-of-sample documents as mixtures over the identified topics. I apply this to characterize the distribution of topics in news articles from three categories: (1) PI news articles; (2) SI news articles that appear on the front page; and (3) non-front page SI news articles.

The results suggest that there are some distinct topic patterns for the select set of news

Table 1.10: Topics identified from a large and representative dataset of financial news from Reuters..

	Topic Label	Most Common Terms	Frequency
#1	Technology	data, technology, companies, security, information, comment, including, according, government, card, software, credit, declined, store, did	23.83%
#2	Earnings & Performance	percent, year, sales, quarter, million, analysts, share, rose, revenue, estimates, profit, earnings, fell, cents, average	9.84%
#3	Financial Services	million, year, bank, today, officer, financial, chief, according, statement, executive, firm, largest, new york, investment, unit	8.29%
#4	Automobile	vehicles, cars, downturn, sales, automaker, deliveries, turnover, air, current, safety, in-house, auto, backlog, switches, parts	6.48%
#5	Air transport	internet, service, search, aircraft, today, flight, plane, contract, engine, carrier, air, airline, satellite, web, traffic	6.22%
#6	Litigation	court, case, judge, federal, workers, law, claims, million, filed, trial, state, lawsuit, ruling, lawyers, attorney	5.96%
#7	Management	ceo, president, job, board, women, chairman, director, vice, named, executive, world, role, according, chief, leave	5.70%
#8	Healthcare	drug, patients, care, percent, flu, health, treatment, disease immunize, today, study, research, treatments, medical, medicines	4.92%
#9	Operations	according, years, got, little, long, later, industry, great, left, good, costs, international, commercial, saying, end	4.69%
#10	Business & Strategy	year, percent, executive, chief, market, officer, brand, today, products, global, world, plans, month, second, sales	4.63%
#11	Mergers & Acquisitions	deal, offer, price, people, bid, shares, comment, buy, companies, analyst, takeover, shareholders, matter, investors, call	4.40%
#12	Advertising	tv, like, food, video, according, subscribers, products, review, pay, media, content, digital, cable, website, advertising	4.15%
#13	Regulations	offer, regulator, today, agency, government, information, review, adjudicate, statement, public, rules, letter, asked, questions, mailed	3.66%
#14	Retail	stores, chain, retailer, sales, retail, years, home, online, customers, holiday, shoppers, foods, black, season, target	3.62%
#15	Employees	companies, time, people, make, week, including, work, interview, want, just, need, way, making, does, spokesman	3.61%

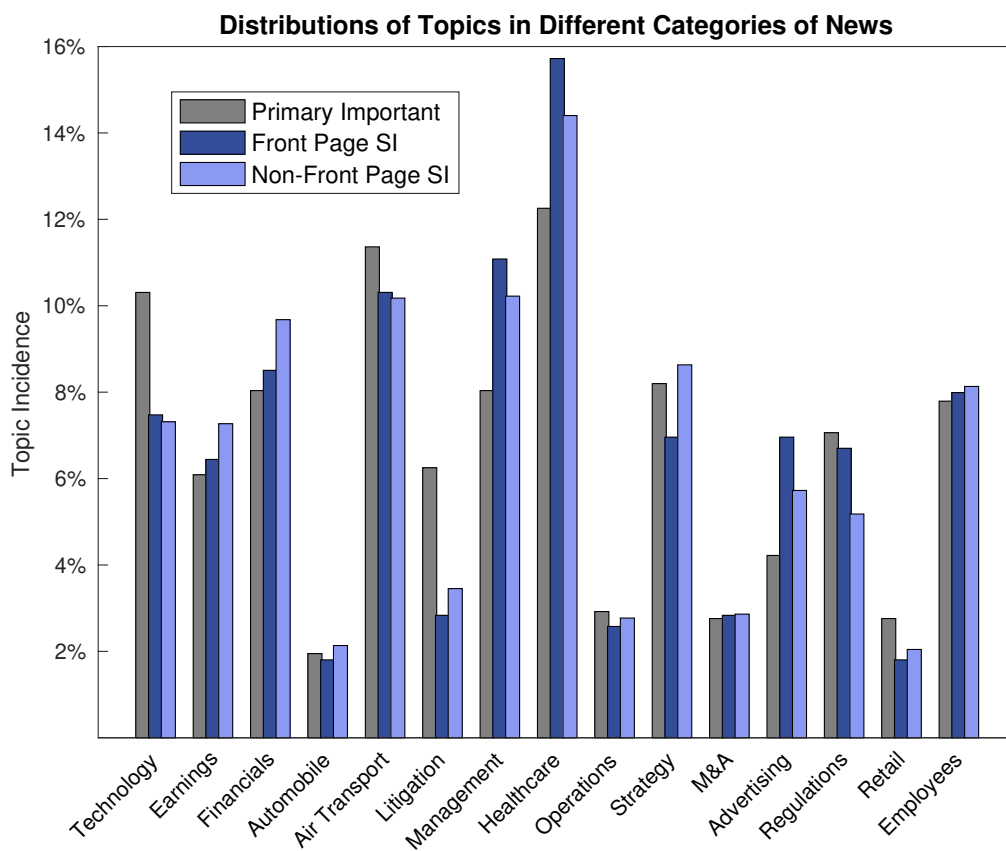


Figure 1.7: Distributions of topics across news articles from different categories.

articles marked as especially important by the editorial staff (PI news), but no differences between the content of SI news articles that make it to the front page and those that do not. The distribution of topics for each category of news is displayed in Figure 1.7. All three categories of news overweight content regarding the financial services industry, regulations, the retail industry, and company employees. Coverage of technology, earnings reports, and the healthcare industry is also common, although technology is far less ubiquitous than in the training corpus. PI news articles are more likely to cover news related to the healthcare industry and company operations; they have a lower focus on the financial services industry and regulations. Front page and non-front page SI news articles are very similar in terms of the distribution of topics, with only minor differences (non-front page SI news articles are more likely to feature news about M&A deals and company employees, whereas front page

articles contain more discussions of litigation and the financial services industry).

For a formal comparison of the distributions of topics across the different categories of news, I perform a Pearson χ -square test of independence pairwise between any two categories (see Rao and Scott (1981)). The results are tabulated in Table 1.11.

Table 1.11: Results from pairwise comparisons between sets of news articles from different positions and different levels of importance.

Panel 1: Front Page SI versus Non-Front Page SI

# Topics in Model	p-value
10 topics	0.8670
15 topics	0.8776
20 topics	0.8731
25 topics	0.7801

Panel 2: PI versus Front Page SI

# Topics in Model	p-value
10 topics	0.1236
15 topics	0.0836†
20 topics	0.0526†
25 topics	0.0417*

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

In the main specification with 15 topics, the distribution of topics in PI news articles is weakly statistically significantly different, at the 10% level, from the distribution of topics covered by the front page SI news articles (see Panel 2 of Table 11). This is robust to varying the number of topics, with the difference becoming significant at the 5% level in the specification with 25 topics but falling short of the 10% statistical significance threshold when the number of topics is reduced to 10. More importantly, the front page and non-front page SI articles are statistically indistinguishable in terms of their textual content, with a p-value of 87.76% in the primary specification with 15 topics. The similarity in the two distributions is robust to varying the topic model specification, with all p-values above 75%.

Overall, the results point to some distinction in the content of “primary important” news

articles from the content of “secondary important” articles. But the distributions of topics are statistically indistinguishable across front page and non-front page SI news articles. This supports the identifying assumption of independence of the prominence of the SI news articles’ positions from their underlying content.

1.6.3 Market Participants’ Perceptions of News

In order to directly assess the market’s perceptions of the underlying news articles in my hand-collected sample, I survey the target audience of the news: active finance professionals and current MBA students at top business schools. Without the differential positioning, these individuals do not perceive front page SI headlines to be any more impactful than non-front page ones. They do, however, perceive the PI news articles to be more impactful, supporting Bloomberg editorial staff’s decisions to mark these articles as more important.

For this part of the analysis, I survey 150 active professionals from a number of financial institutions, as well as 25 current students at top MBA programs. The breakdown of these individuals across affiliations is presented in Table 1.12. The majority of the sample (78.6%) covers active professionals from a representative landscape of financial institutions. The remainder consists of current MBA students at Harvard Business School, the Wharton School, Columbia Graduate School of Business, the University of Chicago Booth School of Business, UVA Darden School of Business, and the McDonough School of Business at Georgetown University.

The sample of active finance professionals is representative of the full landscape of the financial services industry. The bulk (81%) of the active professionals come from large banks and broker dealers such as JP Morgan and Morgan Stanley, investment management firms such as BlackRock and State Street, hedge funds such as Bridgewater Associates and AQR Capital Management, and private equity firms such as the Blackstone Group and Warburg Pincus. The remainder of the sample spans consulting firms such as the Boston Consulting Group, government agencies such as the Federal Reserve Board, financial offices of corporations such as Nike and Walt Disney, pension funds such as North Carolina

Table 1.12: Summary statistics of the financial experts surveyed regarding the news.

Affiliation Type	Institution	Percentage
MBA Students		21.4% of Total
Breakdown:		
	Harvard Business School	42.3%
	The Wharton School	42.3%
	Columbia Graduate School of Business	3.9%
	University of Chicago Booth School of Business	3.9%
	UVA Darden School of Business	3.9%
	McDonough School of Business, Georgetown	3.9%
Professionals		78.6% of total
Breakdown:		
	Hedge Funds	8.0%
	<i>Bridgewater Associates, AQR Capital Management, Tudor Investment Corp, BlueMountain Capital Management, Blue Ridge Capital, QTrade Capital, Bluegrass Capital, One East Partners</i>	
	Investment Managers	20.7%
	<i>BlackRock, The Vanguard Group, State Street, Fidelity, Pacific Investment Management Company, Wellington Management Company, Northern Trust Company, T. Rowe Price, Dodge & Cox Funds, Acadian Asset Management, Eachwin Capital, Crane Asset Management, Wafra Investment Advisory Group, Cambridge Associates, Broadfin Capital</i>	
	Pension Funds	0.7%
	<i>North Carolina Retirement System</i>	
	Private Investors	0.7%
	Banks and Broker Dealers	40.7%
	<i>JP Morgan, Morgan Stanley, Goldman Sachs, Bank of America Merrill Lynch, BNP Paribas, Credit Suisse, Deutsche Bank, Wells Fargo, Royal Bank of Canada, UBS, Standard Chartered Bank, Citizens Bank, The NEX Group, HSBC, Edelweiss, Royal Bank of Scotland, SunTrust Bank, Berliner Volksbank, First Republic Bank</i>	

Table 1.12 (Continued)

Affiliation Type	Institution	Percentage
Professionals		
Breakdown:		
	Investment Banks <i>Barclays Capital, Lazard</i>	2.7%
	Insurance <i>Massachusetts Mutual Life Insurance, Voya Financial, Nippon Life Insurance, Liberty Mutual</i>	3.3%
	Government Agencies & Sovereign Wealth Funds <i>The Federal Reserve Board, Abu Dhabi Investment Authority, World Bank Group</i>	2.0%
	Corporations <i>Nike, Inc., Shaw's Supermarkets, Tiffany & Co., The Walt Disney Company, Philips, ReBio LLC</i>	4.0%
	Private Equity & Venture Capital <i>Blackstone Group, Warburg Pincus, Motive Partners, Garrison Investment Group, ATL Partners, Tamarisc, Pomona Capital, Clearview Capital, Cerberus Capital Management, Madrona Partners</i>	11.3%
	Consulting <i>Boston Consulting Group</i>	0.7%
	Non-Profit <i>Ford Foundation</i>	0.7%
	Financial Advisory, Taxes, and Real Estate <i>Princeton Tax Services, DK Partners, Condor Partners</i>	1.3%
	Media	0.7%
	Other finance professionals	2.0%

Retirement System, insurance companies such as Liberty Mutual, and other areas of the financial services industry.

The respondents largely constitute key decision makers within their respective firms. Many of the professionals from larger corporations such as banks, broker dealers, and large investment management firms are at the principal or managing director levels within their organizations, including heads of regional offices. The sample also includes chairmen, partners, and C-level executives. This sample is broadly reflective of the client base consuming Bloomberg news through the terminal. Approximately 87% of the professionals in my sample report having used a Bloomberg terminal at some point, with 63% actively using the terminal on an ongoing basis.

In the survey, each respondent is asked to answer a series of twenty-five questions about news headlines. The respondent is told that the headlines come from a news provider who chooses how prominently the headlines are displayed based in part on the importance and market impact of the underlying news. Each question presents two headlines, and asks the respondent to specify which headline the respondent thinks had larger market impact and deserves more prominence. A screenshot with an example question is displayed in Figure 1.8.

The survey questions span two sets of comparisons: (1) between front page SI news articles and PI news articles; and (2) between front page SI news articles and non-front page SI news articles. In particular, in each question, one of the two headlines (in random position – either on the left or on the right of the screen) is randomly selected from the front page SI news category. The other headline is randomly selected, with equal likelihoods, from the categories of PI news (approximately 37.5% of the questions) and non-front page SI news (approximately 62.5% of the questions).

The respondents are incentivized to identify the relative news importance as accurately as they can. Each respondent receives a \$10 gift (an Amazon.com gift card or a lunch voucher to a venue of the respondent's choice) for completing the survey. In addition, the five respondents whose answers most closely match actual differences in positioning by the

Which Financial News Headline Is More Important?

(Question 4) For the news headlines below, please select the radio button next to the headline that you think had **larger market impact** and is **more deserving of prominence**.

**ALLSTATE THIRD-QUARTER
PROFIT MORE THAN DOUBLES
ON PREMIUM GAINS**

**EINHORN SAYS BULLISH ON
TECHNOLOGY, SEEKS TO
CLARIFY BUBBLE CALL**

EXIT

NEXT

Figure 1.8: Example question from the survey administered to active finance professionals and MBA students.

news provider receive additional prizes of \$90 each.

The results indicate that the financial experts in the sample do not distinguish between front page and non-front page SI news articles. Panel 1 of Table 1.13 presents the incidence of front page SI news articles being chosen as more important than non-front page SI news articles, with standard errors clustered by participant. The sample of 150 finance professionals identifies the front page articles as more impactful 48.24% of the time, not statistically different from 50%. Similarly, the smaller sample of 26 MBA students choose the front page news 45.05% of the time, which is, if anything, lower than 50% (marginally statistically significant at the 10% level). Pooling across both samples, front page articles are chosen as more impactful 47.78% of the time. The results are very similar when I exclude attritors (participants who do not answer all 25 questions). Thus, absent differential positioning, the target audience of finance professionals does not perceive the front page SI news articles as being any more important than their non-front page counterparts.

The market participants do, however, identify the “primary important” articles as more impactful, validating Bloomberg’s importance markings. As Panel 2 of Table 1.13 reveals, active finance professionals choose PI news articles over front page SI news articles 61.16% of the time, significantly higher than 50% at the 1% level. MBA students are somewhat weaker at identifying “primary important” news, choosing them 57.54% of the time, significant

Table 1.13: *Aggregated responses of financial experts to the news survey.*

Panel 1: Front Page SI versus Non-Front Page SI

Respondent Type	Choosing Front Page	Standard Error	# Respondents
Finance Professionals	48.24%	(1.21%)	150
MBA students	45.05%†	(2.65%)	26
All Respondents	47.78%*	(1.11%)	176
Finance Professionals (excl. attritors)	48.16%	(1.24%)	136
MBA Students (excl. attritors)	44.83%†	(2.69%)	25
All Respondents (excl. attritors)	47.67%*	(1.14%)	161

Panel 2: PI versus Front Page SI

Respondent Type	Choosing PI	Standard Error	# Respondents
Finance Professionals	61.16%**	(2.13%)	150
MBA Students	57.54%*	(3.55%)	26
All Respondents	60.58%**	(1.87%)	176
Finance Professionals (excl. attritors)	61.59%**	(2.20%)	136
MBA Students (excl. attritors)	57.66%*	(3.61%)	25
All Respondents (excl. attritors)	60.95%**	(1.93%)	161

** denotes a proportion differing from 50% with significance at the 1% level.

at the 5% level. Pooling all responses, PI stories are chosen as more impactful 60.58% of the time, significantly higher than 50% at the 1% level. Overall, the results point to the Bloomberg editorial staff correctly identifying, on average, the news most relevant for the target demographic: the higher importance ranking assigned to the PI news articles is corroborated by the surveyed market participants.

Similar patterns hold at the individual level. For each respondent, I calculate the percentage of times that the respondent chooses a front page SI news article over a non-front page one, and the percentage of times that the respondent chooses a PI news article over an SI one. A histogram of these individual-level percentages is displayed in Figure 1.9. The incidence of choosing front page articles over non-front page ones is presented in blue; the distribution is centered around 50%, is symmetric, and resembles a normal distribution. Overall, this distribution is consistent with there being no distinction between the two sets of articles, and the differences between individuals' choices coming from noise and the

variation in the randomly selected questions faced by different individuals. The incidence of choosing PI news articles over SI ones, presented in gray, paints a different picture. Very few respondents choose PI news articles less than 40% of the time, and the distribution is centered around 60%, with a number of respondents choosing the PI news articles as often as 90-100% of the time.

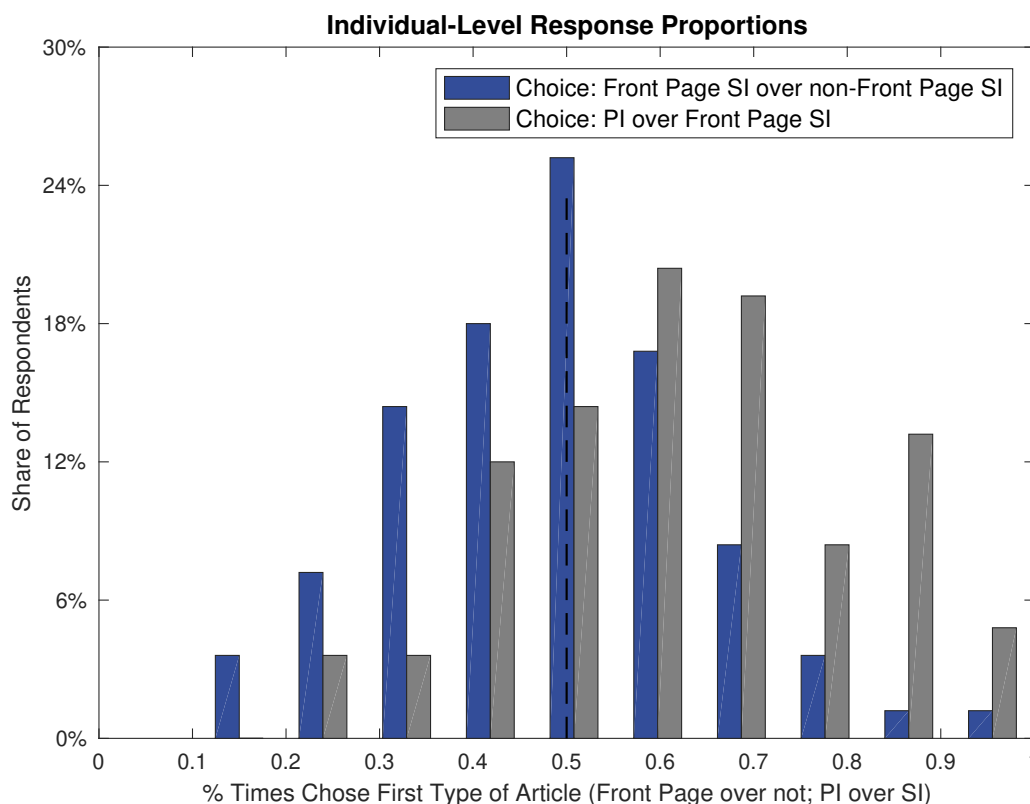


Figure 1.9: Individual-level responses from the survey of financial experts.

Overall, the target audience of the news perceives no systematic differences between the SI news articles that get placed on the front page and those that do not. This is consistent with the quasi-random positioning of these news articles. There is a stark juxtaposition between the significantly different market dynamics following these two sets of news and the market participants' lack of distinction between them in the survey. This juxtaposition highlights the extent to which salient news positioning can induce different reactions to

otherwise identically important content.

1.7 Conclusion

This paper takes advantage of a natural experiment in news positioning to directly estimate the effect of news consumption on financial markets. For two news articles of equal importance, pinning one to a prominent position induces 280% higher trading volume during the ten-minute window after the news, 180% larger absolute price change, and substantially higher short-term return continuation. Interestingly, differences in news positioning play an even larger role for market dynamics than differences in the editorial markings of importance of the underlying news articles' content.

The results in this paper highlight the importance of how information is presented for the way in which the information is incorporated into asset prices. In the modern informational environment, where investors face millions of news articles per day, the distinction between public and private information becomes somewhat blurred, and even public information may not be immediately and efficiently priced.¹¹ My analysis traces out incorporation of information in real time using a natural experiment on a highly relevant platform, the Bloomberg terminal. My results capture momentum in price responses to information, and show that the speed of incorporation depends on the method of dissemination. For more obscure or private information, similar mechanisms are likely to apply at longer horizons, generating phenomena such as month-level momentum.

¹¹Dugast and Foucault (2017) propose a mechanism whereby increased data availability does not necessarily lead to more efficient prices in the long run. Farboodi *et al.* (2017) document declining price informativeness for firms outside of the S&P 500.

Chapter 2

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

2.1 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

¹For example, Kaniel *et al.* (2012) and Drake *et al.* (2012) document heightened trading volume around earnings announcements, while Bali *et al.* (2017) 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 Boulland *et al.* (2017), among others. For trading volume as a proxy for divergence in investors' opinions, see Garfinkel and Sokobin (2006) and Garfinkel (2009).

²Prior work large rely on dispersion of sell-side analyst forecasts to capture differences in opinion; see Ajinkya *et al.* (1991), Atiase and Bamber (1994), Diether *et al.* (2002), Anderson *et al.* (2005), and Bamber *et al.* (2011). Carlin *et al.* (2014) look at differences of opinion more directly, but focus on prepayment in the mortgage-backed security market. Giannini *et al.* (2015) and Cookson and Niessner (2016) take a complementary approach to mine; they analyze expressed opinions on StockTwits, but do not capture information acquisition.

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 of 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 but significantly weaker in predicting trading volume around information releases. However, differences of opinion do play a larger role around textually ambiguous 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 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 heterogenous. 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 generating trading volume around more ambiguous news events that 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

³For classic models of gradual information diffusion, see Hong and Stein (1999) and Hirshleifer and Teoh (2003), among others. For models of differences of opinion, see, for example, Harris and Raviv (1993) and Kandel and Pearson (1995).

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 three-fold.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.

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 *et al.* (2015) find that investors from

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 *et al.* (2011), Drake *et al.* (2012), and Madsen and Niessner (2016) for the use of Google search volume as a measure of attention; Bauguess *et al.* (2013) and Drake *et al.* (2015, 2016) for downloads of EDGAR filings; Lawrence *et al.* (2017) for searches on Yahoo!; and Lumsdaine (2010) and Ben-Rephael *et al.* (2017) for the use of Bloomberg’s aggregate daily proxy of institutional investors’ attention.

⁶See McPherson *et al.* (2001).

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 difference 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 have different information sets and different approaches to procuring and processing new information; hence, they are likely to interpret the news according to different models. 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 market impact.

Both gradual information diffusion and differences of opinion are predictive of trading volume around news, but the effect of the former is stronger. I perform the analysis at two horizons: within days around an earnings announcement 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 hour 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 particular, 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

⁷See DellaVigna and Pollet (2009), Hirshleifer *et al.* (2009), deHaan *et al.* (2015), and Niessner (2015).

⁸See Da *et al.* (2011), Drake *et al.* (2012, 2015, 2016), Bauguess, Cooney, and Hanley (2013), Ben-Rephael, Da, and Israelsen (2015), Madsen and Niessner (2016), and Lawrence *et al.* (2017), among others.

differences in these investors' interpretations of the 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 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 2.2 outlines the conceptual framework for my empirical tests. Section 2.3 describes the data. Section 2.4 details the methodology for constructing proxies of gradual information diffusion and differences of opinion. Section 2.5 presents the key test of the paper on predictability of trading volume from the two forms of disagreement. Section 2.6 considers the strengths of the two channels of disagreement for news events with varying levels of ambiguity. Section 2.7 concludes.

2.2 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

⁹For proxies of disagreement based on analyst forecasts, see Ajinkya *et al.* (1991), Atiase and Bamber (1994), Diether *et al.* (2002), Anderson *et al.* (2005), Bamber, Barron, and Stevens (2011), among others. For proxies based on social media, see, for example, Giannini *et al.* (2015), Cookson and Niessner (2016).

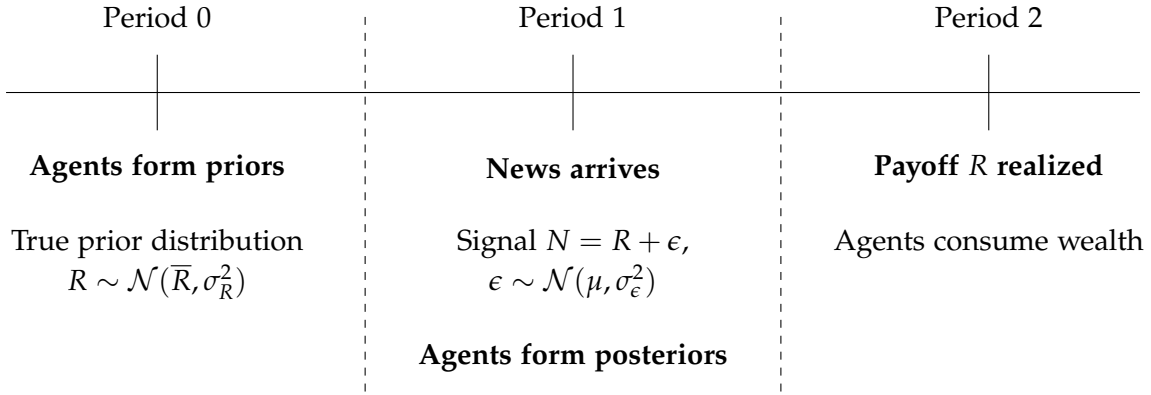


Figure 2.1: Model timeline.

(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 heterogenous 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 2.1.

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 2.2.1. In Section 2.2.2, 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 2.2.3).

2.2.1 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 4 (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 2.5 and 2.6, 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 B.1.

2.2.2 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.

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 $\text{Var}_{1,1}\{R\} = \frac{\sigma_R^2 \sigma_\epsilon^2}{\sigma_R^2 + \sigma_\epsilon^2}$, while investors of type 2 perceive $\mathbb{E}_{2,1}\{R\} = \bar{R}$ and $\text{Var}_{2,1}\{R\} = \sigma_R^2$. As a result, the investors' demand functions for the asset in period 1 are given by:

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

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 *et al.* (2009) on earnings announcements released contemporaneously with other announcements, and deHaan *et al.* (2015) and Niessner (2015) for evidence that firms strategically respond to investors' limited attention by timing their releases.

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

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

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:

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

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 5 (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 2.5 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 B.1.

2.2.3 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 6 (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 2.5 and 2.6. 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.

2.3 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.

2.3.1 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 land-

scape 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 2.2. 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

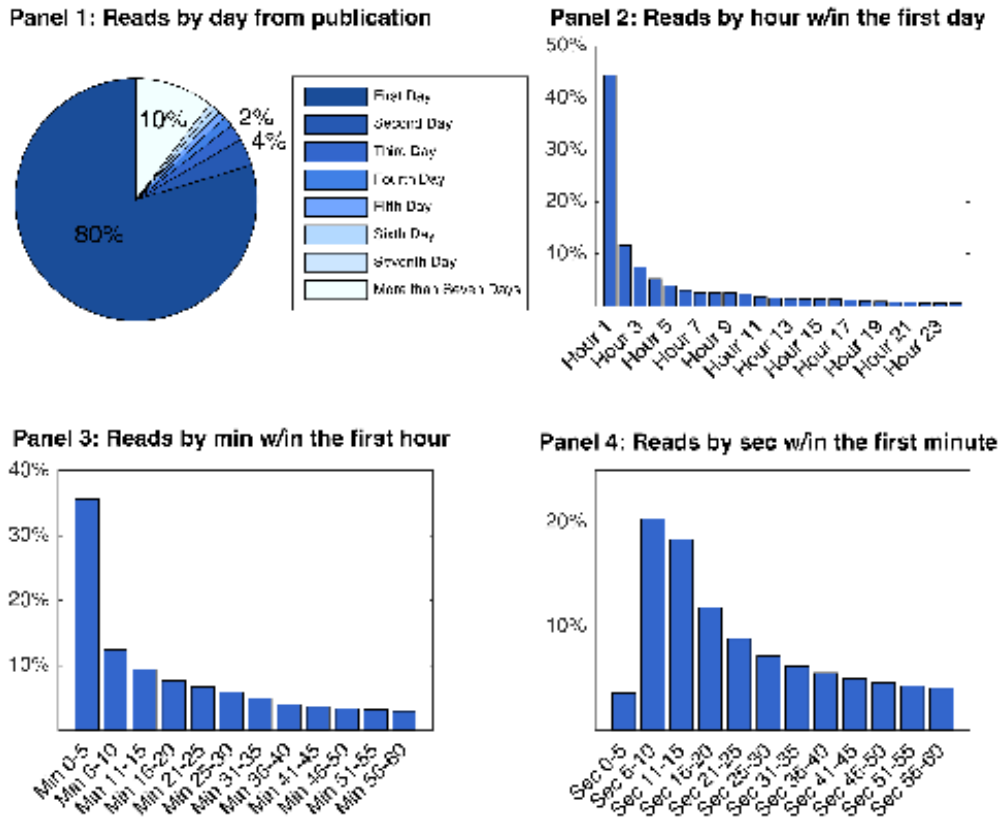


Figure 2.2: Time lag from publication of a news article to the read.

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.

2.3.2 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.

2.4 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.

2.4.1 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 *et al.* (2014) use entropy to measure pricing kernel's dispersion. Entropy of a distribution is a natural measure in my context, as it serves to

¹¹See Shannon (1948).

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 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:

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

2.4.2 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

¹²For studies exploring the origins for investor disagreement, see, for example, Cronqvist *et al.* (2015) and Chang *et al.* (2015).

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 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 2.5.2 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.2.1.

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 2.3, 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;
- 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;

¹³Please refer to Appendix B.2.2 for detail

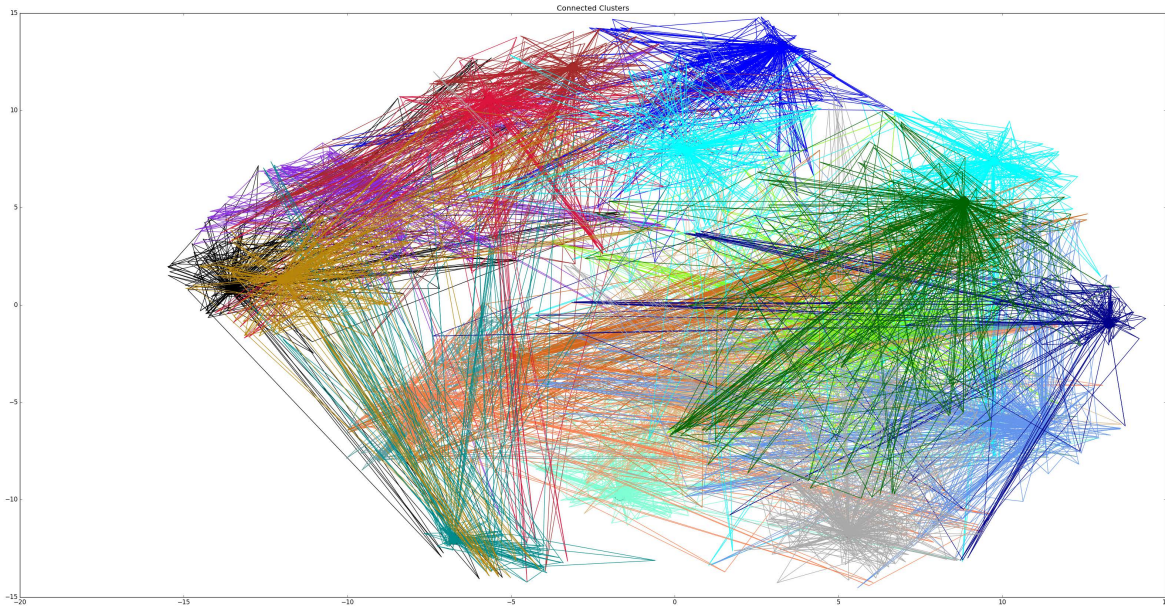


Figure 2.3: *News readers clustered by their news consumption patterns.*

- 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 2.4 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.2.3. 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:

$$EntropyType_{i,s} = -\frac{1}{\log(M)} \sum_{k=1}^M c_k(C_{i,s}) \log(c_k(C_{i,s})) \quad (2.2)$$

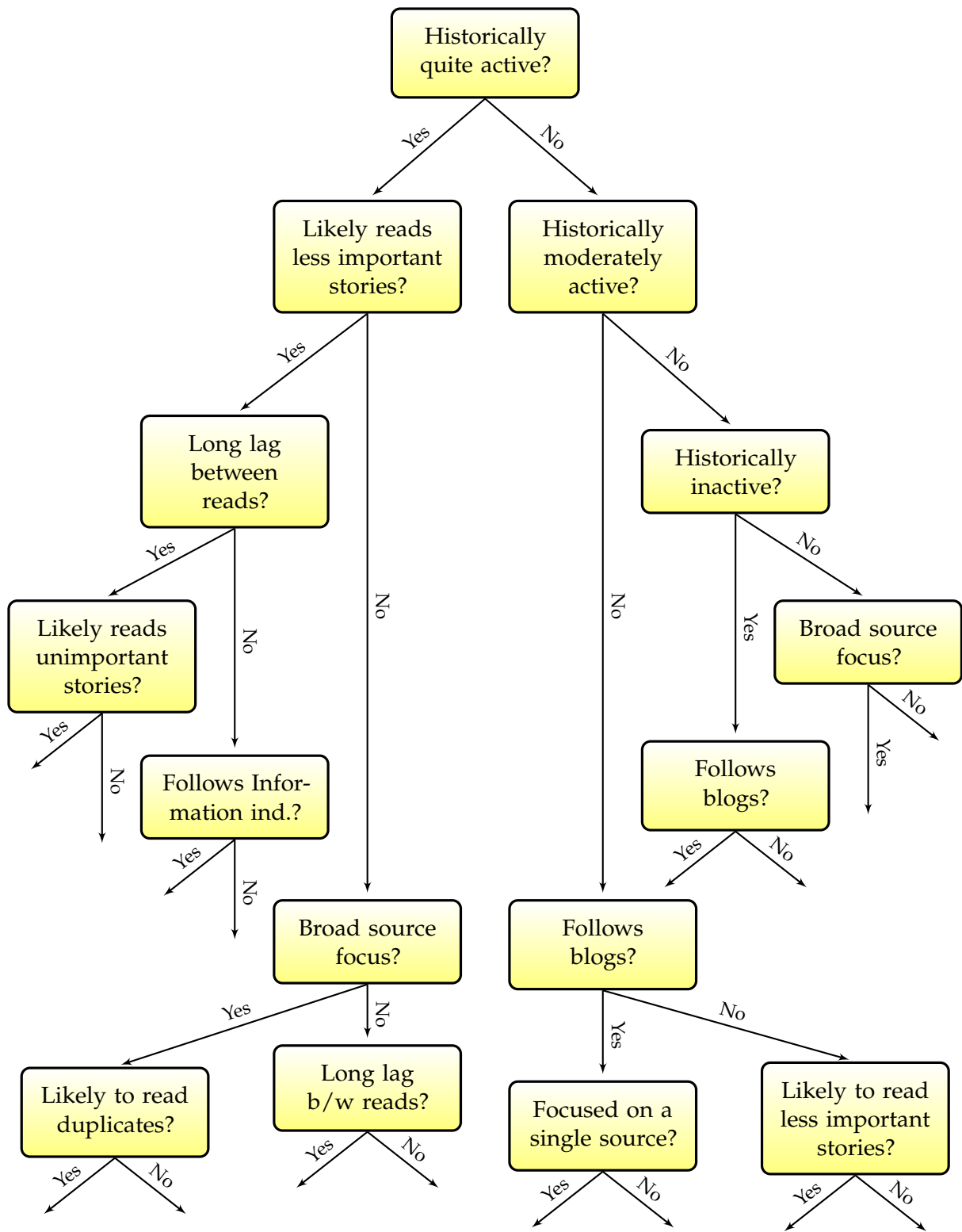


Figure 2.4: Top several splits of the decision tree classifying readers into clusters.

2.5 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.

2.5.1 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 2.5 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,

Daily Trading Volume around Earnings Announcements

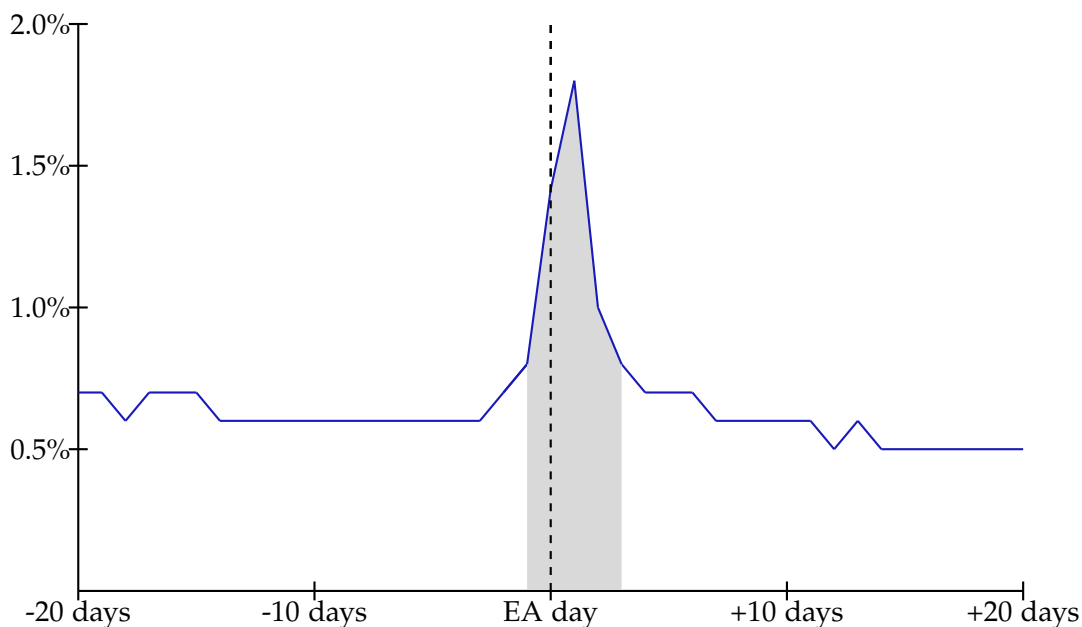


Figure 2.5: Daily trading volume around earnings announcements in the 2014-2015 sample.

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 2.4, I estimate the following regression:

$$Volume_{i,t} = \alpha + \beta_1 EntropyTime_{i,t} + \beta_2 EntropyType_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}, \quad (2.3)$$

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 2.1.

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.

Table 2.1: *Trading volume tests around earnings announcements.*

	(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.

2.5.2 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:

$$Novel(s) = 1 - \frac{||s \cap (\cup_{j=1}^5 s'_j(s))||}{||s||}, \quad (2.4)$$

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 2.6 displays the distribution of textual novelty 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.

2.5.3 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.

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-").

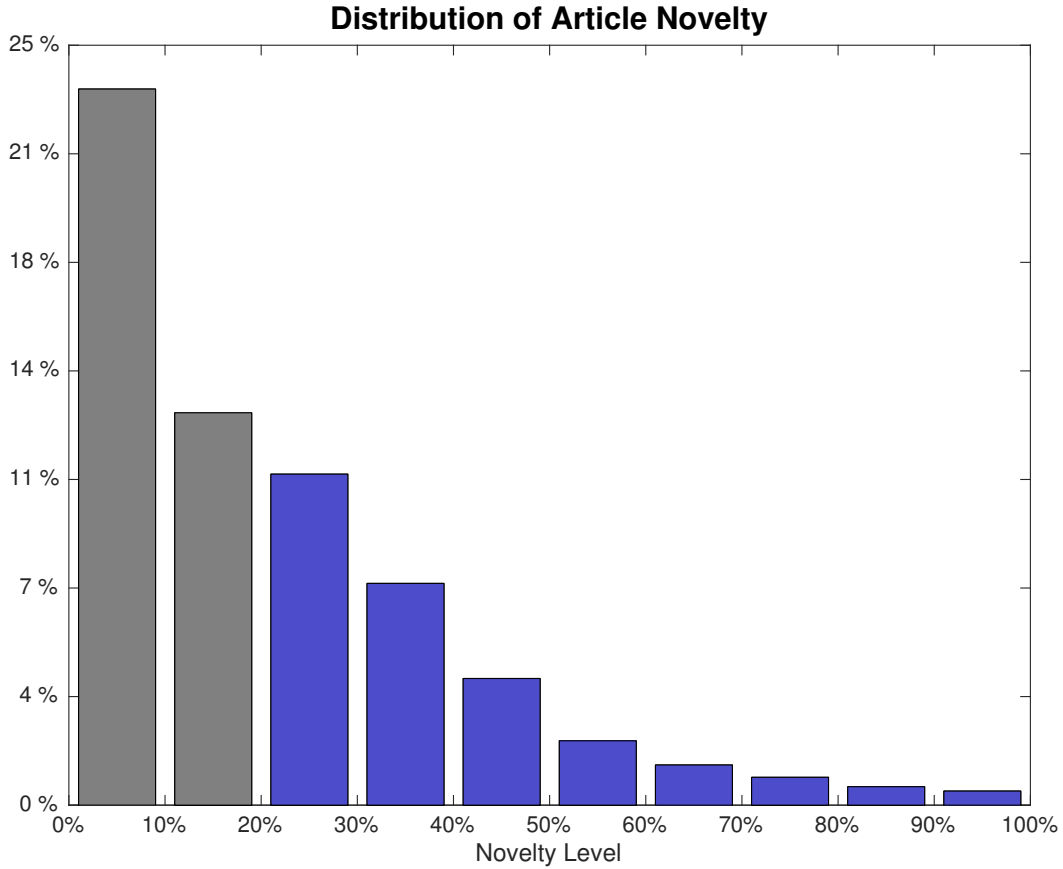


Figure 2.6: *The distribution of textual novelty in the news sample.*

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 2.4, 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:

$$AbnVolume_{i,s,t} = \alpha + \beta_1 EntropyTime_s + \beta_2 EntropyType_s + \gamma X_{i,s,t} + \epsilon_{i,s,t} \quad (2.5)$$

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 2.2. 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.

2.5.4 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

Table 2.2: *Trading volume tests around individual news releases.*

	(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.

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 $Type_q$:

$$\forall s \in Type_q : AbnVolume_{i,s,t} = \alpha^{(q)} + \beta^{(q)} EntropyTime_s + \gamma^{(q)} X_{i,s,t} + \epsilon_{i,s,t}^{(q)} \quad (2.6)$$

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 2.3, 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

Table 2.3: Complementarity analysis around individual news releases.

Panel 1: Conditioning on *EntropyType*

	Quintile of <i>EntropyType</i>				
	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

Panel 2: Conditioning on *EntropyTime*

	Quintile of <i>EntropyTime</i>				
	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.

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:

$$\forall s \in Time_q : AbnVolume_{i,s,t} = a^{(q)} + b^{(q)} EntropyType_s + \gamma^{(q)} X_{i,s,t} + \epsilon_{i,s,t}^{(q)} \quad (2.7)$$

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 2.3, 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 (2.7) 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.

2.6 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.

2.6.1 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"

2.6.2 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 (2.5) separately on the sample of textually clear news and the sample of textually ambiguous news. The results are reported in Table 2.4.

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 2.4, 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

Table 2.4: Trading volume tests around individual news releases, partitioned by news ambiguity.

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

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.

news, presented in Panel 2 of Table 2.4, *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.

2.7 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.

Chapter 3

Asymmetric Naïveté: Beliefs about Self-Control

3.1 Introduction

While time-inconsistent preferences have been gaining prominence in economics, helping explain a variety of individual behaviors ranging from life-time savings to exit rates from unemployment,¹ empirical work has concentrated on individuals' awareness of their own present bias without considering beliefs regarding others.² Yet many situations where time-inconsistent preferences are likely to play a key role (including teams in corporations, households' consumption decisions, educational settings, and political negotiations) involve *interactions* among biased individuals. Households' savings decisions and demand for commitment devices depend on the spouses' expectations regarding each others' future behavior. In the workplace, managers' ability to effectively delegate tasks hinges on their awareness of their subordinates' present bias. And across both educational and corporate environments, efficacy of incentive schemes such as tournaments depends on the individuals'

¹See, for example, Laibson (1997), DellaVigna and Paserman (2005), and Laibson *et al.* (2008).

²See, for example, Ariely and Wertenbroch (2002), DellaVigna and Malmendier (2006), Acland and Levy (2015), Augenblick *et al.* (2015), and Augenblick and Rabin (2018).

relative perception of themselves compared to others. In these situations, understanding of *others'* present bias plays a key role in determining equilibrium outcomes.

In order to lay the foundation for analyzing interactions among present biased individuals, this paper experimentally investigates individuals' beliefs regarding their own and others' present bias within a single unified framework. To what extent are people aware of the self-control problems of others? Are beliefs regarding others more correct than those regarding self? I document that while individuals are largely unaware of their own tendency to procrastinate, they hold much more sophisticated beliefs about others. This wedge in beliefs is consistent with the notion of bias blind spots documented in the social psychology literature,³ and suggests that naïveté regarding one's own present bias is a form of overconfidence rather than a lack of awareness of time-inconsistency in general.

I measure beliefs regarding one's own and others' present bias using both laboratory experiment and field survey evidence. First, I construct a large-scale online laboratory experiment to isolate these beliefs. The experiment addresses the issue of incentive compatibility and allows for structural estimation of parameters reflecting beliefs regarding self and others. The results point to a wedge in beliefs: the participants are naïve about their own present bias, but expect present bias in others. Second, as a test of the external validity of the documented wedge in beliefs, I conduct a field survey in an undergraduate accounting class. The classroom experiment confirms that the wedge in beliefs between self and others is operative and substantial in a real-world setting – the classroom.

The online laboratory experiment runs over the course of four weeks, and recruits participants from the Harvard Decision Sciences Lab. The participants engage in a real-effort task that involves identifying characters on a computer screen, and are asked how much work they would like to perform at different wages. Work decisions are elicited for the current date and for future dates, allowing for an estimate of present bias to be computed by comparing decisions about future work to decisions about immediate work. Some of the participants are also asked, on each date, to predict the choices that they will make on future

³See Pronin *et al.* (2002), Ehrlinger *et al.* (2005), and West *et al.* (2012).

dates about immediate work on those dates. This provides an estimate of the degree to which the participants are aware of the time-inconsistency in their own preferences. Another group of participants is asked to predict the average answers of the other experimental participants – both the others’ current preferences for future dates and the choices that others will make when the future dates actually arrive. The predicted differences capture the beliefs that experimental participants hold about others’ present bias. To investigate the robustness of the results to asking the two sets of prediction questions (about self and about others) separately versus together, a third group of participants receives both sets of questions.

My experimental design introduces three innovations relative to previous experiments on present bias using real effort tasks (Augenblick *et al.* (2015); Augenblick and Rabin (2018)). First, in order to elicit participants’ beliefs regarding others and not just themselves, I take steps to ensure that experimental participants are correctly calibrated regarding the population of “others” for their predictions. I do so through an interactive display of the demographics and self-reported task-enjoyment from the pilot run of the experiment.⁴ The interactive display provides summaries by gender, age, race, marital status, educational attainment, and employment. Second, given that the real effort task is performed online, it is important to ensure that “immediate” work decisions are indeed perceived as imminent. On each date, as soon as a work decision is selected to be implemented, the participant must complete the chosen amount of work immediately, with a total of no more than fifteen minutes of break. A prominently displayed timer on the webpage alerts the participant to the countdown. Third, since laboratory subject pools at universities tend to consist of homogeneous populations largely featuring students, I implement a staggered sessions design to ensure that the participants’ answers are not affected by systematic shocks such as school deadlines, university-wide events, or weather disruptions. I run the experiment in five non-overlapping sessions spread across January - August, 2016.

⁴Importantly, the pilot draws from the same Harvard Decision Sciences Lab participant pool, and the composition of participants does not differ across the pilot and the main experiment.

The results of the experiment indicate that participants display significant present bias, are quite naïve about their own present bias, and are more aware of present bias in others. The participants in the online work experiment choose, on average, 3.50 rounds of work fewer when faced with decisions that have immediate consequences than when they make the decisions ahead of time. There is virtually no self-awareness regarding this time-inconsistency in preferences when participants are asked to predict their own future decisions. However, when asked to predict the decisions of others, participants expect others to choose an average of 1.49 rounds fewer when the choice is made for immediate work than when the choice is made for future work. The results are robust to posing the self- and other-prediction questions separately across participants and together to the same participants.

I exploit the rich and controlled setting of my experimental design to structurally estimate the extent of the participants' present bias, naïveté, and beliefs regarding others. I consider a standard β - δ model of quasi-hyperbolic discounting,⁵ coupled with a separable utility function consisting of a linear utility in money and a power cost of effort, and allow for a misunderstanding of the present bias parameter β when participants predict their own or others' future decisions. I use the participants' decisions and predictions for different dates at different wages to estimate the model's parameters.

Pooling across all of the participants' work decisions, I document a present bias parameter β of approximately 0.82, which is consistent with prior literature.⁶ The participants' self-predictions reveal no awareness of their own present bias: on average, they perceive their present bias parameter to be 1.03. By contrast, predictions regarding other participants indicate strong, albeit incomplete, awareness of others' present bias: participants perceive others' β to be around 0.87, higher than the true value of 0.82, but statistically significantly different from 1. These estimates are robust to excluding participants who do not complete

⁵See, for example, Laibson (1997), O'Donoghue and Rabin (1999a), DellaVigna and Paserman (2005), and Heidhues and Kőszegi (2010) and Kőszegi (2010).

⁶For example, Shui and Ausubel (2005) estimate β around 0.8, Laibson *et al.* (2008) estimate β at 0.71, and Augenblick and Rabin (2018) estimate β at 0.83.

the entire experiment, to allowing for different predicted baseline levels for self versus others, and to estimating the model separately for answers about self and others. The results are also corroborated by estimating the model individually for each participant. The median value of the present bias parameter β across individual participants is 0.92. The median self-prediction is at 1.00, and the median prediction of others is at 0.93.

To illustrate the practical relevance of the experimentally documented wedge in beliefs in a field setting, I run a secondary experiment in an undergraduate financial accounting class at the University of San Francisco. The students are assigned an Individual Project, which requires them to choose a publicly traded company and analyze its financial statements by May 2, 2016. The students must choose a company to analyze, download its financial statements ahead of time, and email their selection for instructor approval by April 2, 2016. On the first day of class, January 25, 2016, a survey is administered to the class, asking students to predict when they and / or their average classmates would submit an assignment. The voluntary and fully anonymous survey has a randomized structure with three arms: (i) students are asked to predict the date when they will email their selection to the instructor; or (ii) students are asked to predict the date when their classmates will email their selections to the instructor; or (iii) students are asked to make both predictions.

The results of the classroom experiments reveal the students' naïveté about their own procrastination coupled with more sophisticated beliefs about others. The students predict that they will send their chosen company to the instructor, on average, 22 days before the deadline. By contrast, the students expect that their peers will email the instructor an average of 9 days before the deadline. The actual dates when the students email the instructor occur, on average, 7 days before the deadline, indicating that the predictions for self are optimistic, while the predictions for others are well calibrated (the average prediction for others is not statistically different from the average actual date). The difference in predictions for self and others is highly statistically significant, and remains robust to posing the self- and other-predicting questions separately to different students or together to the same students.

The classroom experiment demonstrates the relevance of asymmetric naïveté in one

real-world setting, but the wedge in beliefs can also influence equilibrium outcomes across a wide spectrum of competitive, collaborative, and hierarchical environments. Relative performance metrics and tournament incentives are ubiquitous both in the workplace (e.g., bonuses for top performance) and in the classroom (e.g., grading on a curve). Across these situations, an individual's willingness to enter a tournament incentive scheme and her subsequent level of effort depend on her expectations regarding her peers' behavior. Similarly, in collaborative environments such as households or teams of coworkers, willingness to enter into commitment devices such as deadlines or savings contracts depends on each individual's perception of both her own and her partners' present bias. Lastly, in hierarchical settings, a teacher's ability to optimally structure class assignments hinges on his understanding of his students' present bias, while a manager's effectiveness at delegating, structuring tasks, and setting deadlines depends on her beliefs regarding her employees.

This paper contributes to the growing experimental literature on time preference and naïveté. Multiple prior studies experimentally assess the extent of individuals' present bias⁷ and participants' awareness of their own time-inconsistency.⁸ These studies document present bias in the domains of monetary rewards, food choice, and real effort, and find a fair amount of naïveté regarding one's own present bias. The present paper extends this line of work by jointly investigating beliefs about self and beliefs about others, and the extent to which the previously documented naïveté is a systematic underestimation of present bias in general or optimism specifically about one's own self-control. I offer experimental evidence in favor of the latter: individuals are generally aware of present bias in others, and are overoptimistic specifically about themselves.

The results on the wedge in beliefs also lay the foundations for theoretical studies of interactions between biased agents. Naïveté regarding one's own present bias has informed a number of theoretical works, including DellaVigna and Malmendier (2004) and Heidhues

⁷See Solnick *et al.* (1980), Read and Van Leeuwen (1998), McClure *et al.* (2004), Andersen *et al.* (2008), Tanaka *et al.* (2010), Bisin and Hyndman (2014), among others.

⁸See Ariely and Wertenbroch (2002), DellaVigna and Malmendier (2006), Skiba and Tobacman (2009), Acland and Levy (2015), and Augenblick and Rabin (2018).

and Kőszegi (2010).⁹ While beliefs regarding self suffice for these models of single biased agents and rational principals, recent theoretical studies have begun to model interactions between multiple present-biased agents. For example, Fahn and Hakenes (2014) consider fully sophisticated agents who are aware of their own and others' self-control problems. Fedyk (2015) assumes that agents are at least partially naïve about their own present bias, but hold more accurate beliefs regarding others. The investigation of individuals' awareness of others' present bias will serve to ground models of interactions between present-biased individuals with experimentally-tested assumptions.

The remainder of the paper proceeds as follows. Section 3.2 outlines the design of the online laboratory experiment, while Section 3.3 presents the reduced-form results. Section 3.4 presents the structural model and estimates the belief parameters for self and others. Section 3.5 presents field evidence of the wedge in beliefs regarding own and others' procrastination in the classroom. Section 3.6 discusses applications of the documented wedge in beliefs. Section 3.7 concludes.

3.2 Experimental Design

In this section, I detail the design of the online laboratory experiment used to evaluate participants' beliefs about their own and others' present bias. The experiment centers around a real-effort task, and the participants' predictions of their own and others' work decisions allow me to measure their beliefs about their own and others' present bias.

The experiment runs over the course of four weeks, recruiting participants from the Harvard Decision Sciences Lab. Each participant chooses a day of the week on which to participate, and needs to log in on that day of the week during each of the following four weeks. The instructions are presented on the first participation date, and the participants must pass a comprehension quiz in order to be eligible for the study. All instructions, questions, and assignments are catalogued in Appendix C.3; the informed consent language

⁹See also O'Donoghue and Rabin (1999a,b), Gul and Pesendorfer (2001), Gottlieb (2008), and Herweg and Müller (2011), among others.

can be seen in Appendix C.4.

I present the experimental design in five subsections. First, I describe the experimental task and the information that the participants receive about other participants in the experiment. Next, I present the experimental timeline and detail the payment scheme. The work decisions faced by the participants are detailed in the third subsection, while the predictions are discussed in the fourth subsection. The last subsection presents information on the experimental sample, including sessions, recruitment, and attrition.

3.2.1 Experimental task

The real-effort task consists of a random sequence of characters appearing (one by one) on an otherwise empty screen, where participants are asked to press a key every time an asterisk appears. The duration of each round is 60 seconds: 50 seconds of work (with a total of 25 characters appearing during that period), followed by a 10 second break. The participants must achieve an average accuracy of 80% across all rounds within a session to successfully complete the work and receive payment. Figure 3.1 displays a sample task screen.

This task is specifically designed with a two-fold objective. First, the task needs to be tedious, so that the participants are exposed to the dynamic tension between the cost of completing more rounds of the task now and the benefit of receiving a higher payment later. Second, the task must be relatively straightforward and simple to complete, so that there is no skill involved, ensuring that any differences between predictions of the participants' own and others' choices are indeed driven by a wedge in beliefs about present bias, rather than overconfidence regarding skill.

While the character-identification task satisfies the objectives of being tedious and not requiring any skill, it is somewhat artificial, which poses a concern that participants might be ill-equipped to make predictions regarding either their own or others' behavior. In order to alleviate this concern and ensure that the elicited beliefs reflect real-world belief formation, I do the following:

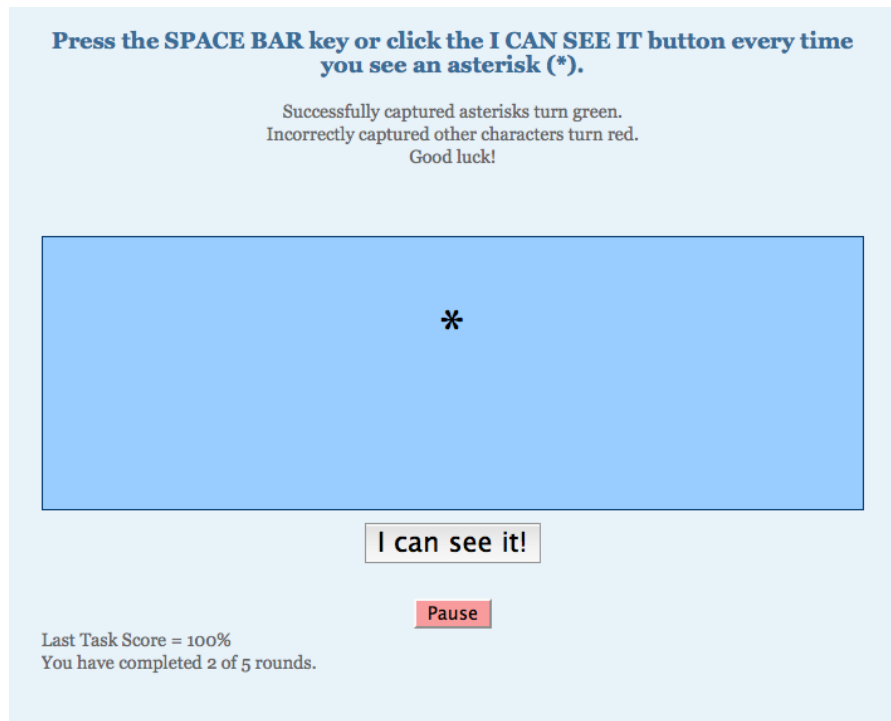


Figure 3.1: Screenshot of the experimental task.

1. All participants try the task for 5, 10, or 15 minutes before making any predictions, which ensures that they are familiar with what it is like to engage in the task.
2. A pilot study of the experimental design is run in October-November 2015, with participants recruited from the same Harvard Decision Sciences Lab pool as in the subsequent main experiment. Demographic data are gathered from all pilot study participants on the first participation date. Data on task enjoyment are gathered from the participants who complete the pilot study during a debrief questionnaire at the end of the last participation date.
3. Participants in the main experiment are presented with the data from the pilot study participants in an interactive display with break-downs by gender, race, marital status, age, education, and employment. A screenshot of this display is shown in Figure 3.2. The participants are encouraged to study these data as part of familiarizing themselves with the task.

Thus, when the participants are asked for predictions about others, they have some empirical familiarity with who the others are and how they feel about the task. The elicited beliefs then more closely correspond to beliefs in real-world scenarios, where individuals have familiarity with the general population of others and the assignment at hand.

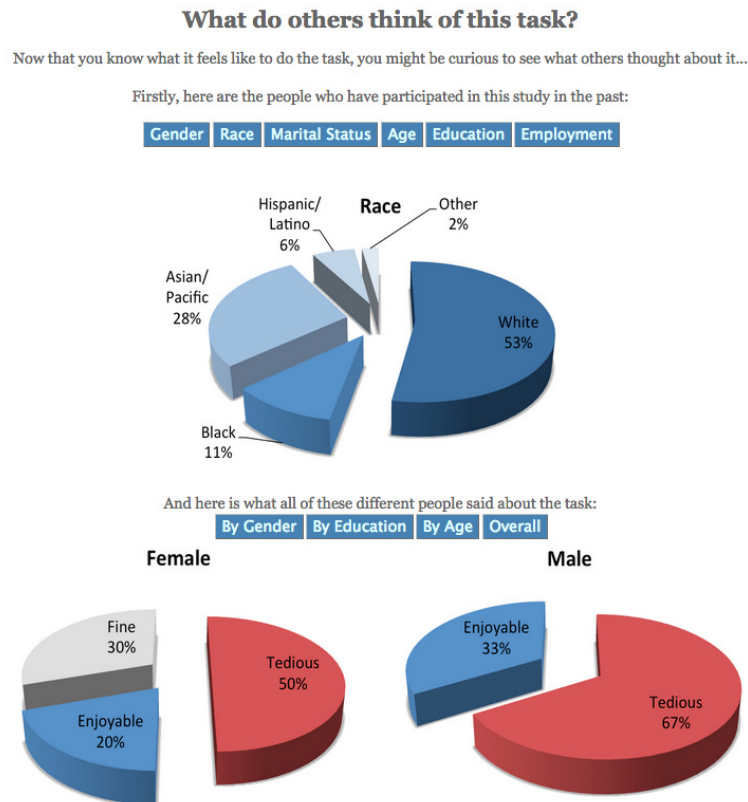


Figure 3.2: Breakdown of demographics and task enjoyment responses from the pilot study.

3.2.2 Experimental Timeline and Payments

Each participant logs into the experiment on her chosen day of the week during four consecutive weeks, denoted by Week 1, Week 2, Week 3, and Week 4. On each participation date, the participant must complete a mandatory warm-up of the task and answer all questions. At the end of the experiment, participants are paid based on the amount of work they do as well as completion of all mandatory items. The full experimental timeline is presented in Figure 3.3.

The first item on each participation date is a warm-up, which involves the participants having to do a mandatory number of rounds of the task. The warm-up amounts vary randomly across participants and consist of 5, 10, or 15 rounds. The differential warm-up amounts allow me to control for projection bias (see Loewenstein *et al.* (2003)), which might lead participants to underestimate the effort cost of doing the task when not significantly exposed to it.

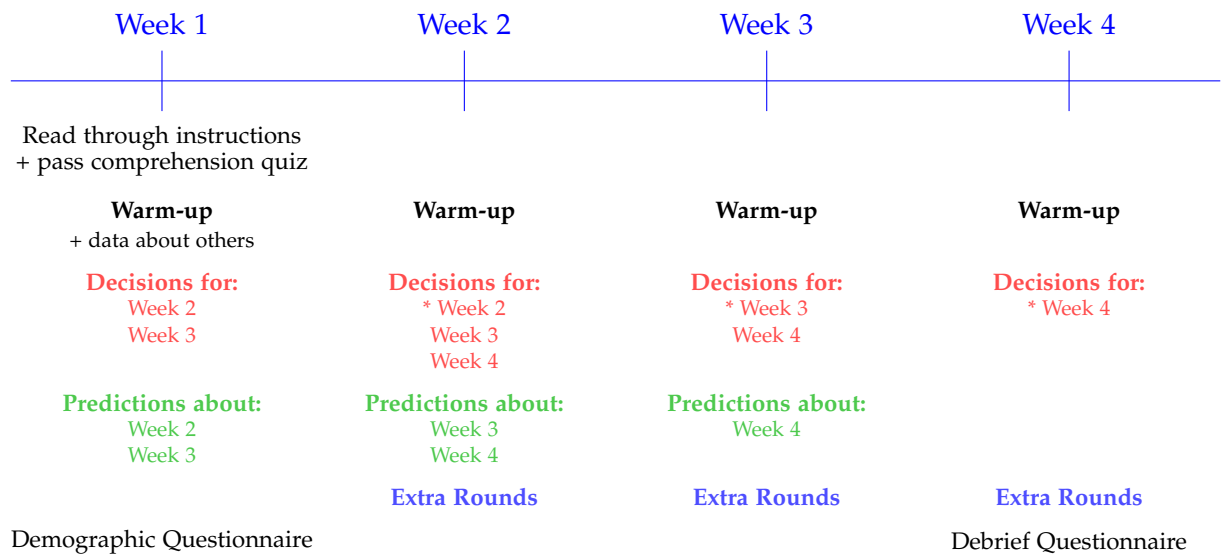


Figure 3.3: *Experimental timeline. Stars indicate questions regarding decisions for immediate work.*

After the warm-up, participants are asked how many extra rounds (between 0 and 70) of the task they would like to complete for additional pay at different wages, either on the same day or on future participation dates. The participants are also randomized into groups asked to predict either their own future decisions, or the average decisions of the other participants, or both.

Once all questions are answered, the next step is completing the chosen number of rounds of work. In particular, one of the decisions *for* the current date (made either *on* that date or earlier) is selected at random to be implemented, and the participant must immediately complete the number of extra rounds in that decision.

On the first participation date, participants also fill out a questionnaire consisting of

demographic questions and questions eliciting the participants' predictions regarding their own and their peers' psychological state, time constraints, and preferences over the next few weeks, presented in random order. Similarly, at the end of the last participation date, there is a short debrief questionnaire. Participants are asked for reasons behind their own and others' inconsistencies, as well as predictions on whether they would behave more consistently if offered another chance to participate. The debrief also elicits beliefs regarding one's own and one's peers' present bias in other domains: expected gym attendance, work procrastination, and healthy eating.

Each participant's payment consists of two components: the completion payment and supplemental wages. The participant receives the \$30 completion payment for logging in and completing all required work on each participation date. The supplemental wages are computed at the corresponding rates for any extra rounds that the participant completes, and any incentive bonuses earned for correct predictions. In order to be eligible for the \$30 completion payment, the participant must complete each warm-up, answer all decision and prediction questions, and then finish the additional rounds in her implemented work decisions. If the participant fails to complete any of these tasks on one of her participation dates, the participant is disqualified and foregoes the \$30 completion payment. Disqualified participants still receive payment for the additional rounds that they have completed before disqualification. The payments are dispensed in the form of Amazon.com gift cards on the Sunday one week after the end of Week 4.

3.2.3 Work Decisions

A critical component of the experimental design consists of the participants' decisions about how much of the real-effort task to do. The participants are asked to make these decisions for the current date and for future participation dates, and all decisions have an equal chance of being implemented. The differences in the participants' decisions for immediate versus future work are used to capture the participants' present bias.

The participants face work decisions on each of their four participation dates, indicated

'HOW MUCH TO WORK' DECISIONS: for today, 01/19/16

Today, you will do some number of extra rounds of the Task. You will have to complete the extra work **immediately** after the "Decision that Counts" is selected, with no more than 15 minutes of breaks.

Decisions made now for work to be done immediately

How many extra one-minute rounds would you like to do now at the following wages?

Wage	Rounds
\$0.10/round (\$6/hour)	<input type="text"/>
\$0.25/round (\$15/hour)	<input type="text"/>

EXIT **SUBMIT**

Figure 3.4: An example of a work decision screen.

in red in Figure 3.3. Each set of decisions consists of two questions for the same date but at different wages. The wages are drawn randomly from between \$0.10/round and \$0.30/round, in increments of \$0.05. This corresponds to \$6/hour-\$18/hour. All wages are equally likely to be drawn, with the restriction that the two wages on a single screen must be different. An example of a work decision screen is presented in Figure 3.4.

A few checks are in place to ensure that the participants' work decisions reflect their genuine preferences. First, if a participant enters the same number into both fields, she sees a warning enquiring whether she is certain that she would like to proceed with a decision to do the same amount of work regardless of the wage, or if she would like to reconsider. Second, if a participant enters a higher number into the field with the lower wage, she is asked whether she would really wish to do more work for lower pay, or whether she would like to reconsider her answers. These checks are in place to ensure that the participants are paying attention to the questions, rather than quickly entering random or repeating numbers into the fields. To ensure that the decisions correspond to permissible amounts, the participants must also enter an integer between 0 and 70 into each field to proceed.

Overall, each participant makes 16 work decisions – 6 immediate decisions for the same date and 10 ahead-of-time decisions for future dates. The full set of decisions is displayed

in Panel 1 of Table 3.1, with each cell corresponding to a two-question screen analogous to the one displayed in Figure 3.4. The blue row catalogues immediate decisions, while the green rows list ahead-of-time decisions.

During Weeks 2, 3, and 4, each participant’s actual amount of work is randomly selected from all of the decisions that the participant has made for that date. In particular, once her work decisions are complete on a given participation date, the participant is shown all decisions that she has made for that date – this includes decisions made only moments earlier as well as ahead-of-time decisions made on prior dates. Figure 3.5 displays a sample screen aggregating all work decisions for a particular date.

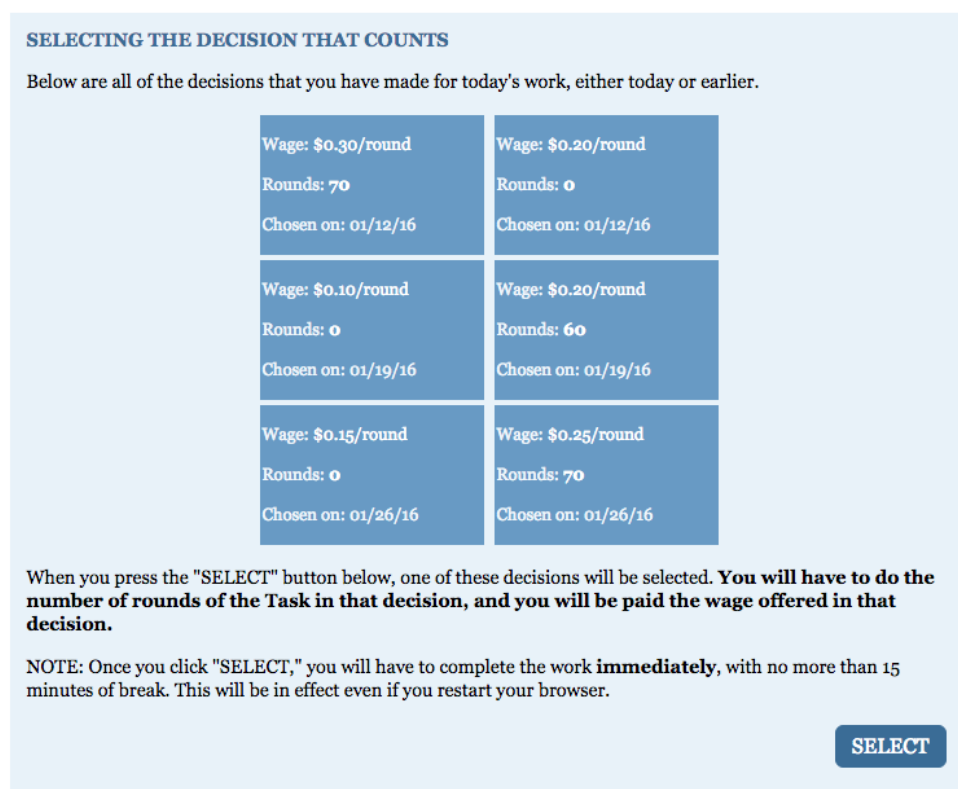


Figure 3.5: An example screen aggregating a participant’s work decisions for a given date.

The participant is reminded that, once one of these decisions is randomly selected to be implemented, she must complete the work in that decision immediately, with no more than a total of 15 minutes of break. This restriction serves to ensure that the decisions made for

immediate work are perceived as truly immediate, rather than, for example, decisions made in the morning for work to be completed in the evening.

Once the participant clicks on the "SELECT" button, a randomization is run and one of the decisions is selected as the "Decision that Counts." All decisions are ex ante equally likely to be selected. The selected decision is then marked in dark blue, and a counter appears on the webpage. In order to continue participating in the experiment and receive the completion payment, the participant must complete the amount of work in the selected decision immediately with no more than 15 minutes of breaks.

3.2.4 Predictions

In order to compare participants' beliefs about their own and others' present bias, the participants are asked to make a set of predictions. After making their work decisions, the participants are asked to predict either how much work they will choose for immediate completion on future dates, or how much work other participants will choose, or both.

For the prediction questions, I split the participants into the following three groups:

- **Group 1:** Throughout the experiment, participants in this group are asked how much work they anticipate choosing for immediate completion when various future dates actually arrive. Since the decision and prediction questions are quite similar, predictions appear side by side with the decision questions, in order to make the questions clearer and more straightforward. See Panel 1 of Figure 3.6 for an example screen presented to participants in this group.
- **Group 2:** Throughout the experiment, participants in this group are asked to predict how others make decisions. They are asked to predict the average of the other participants' current decisions for work on future dates, as well as the average of the other participants' choices for immediate completion when those future dates actually arrive. Participants in Group 2 are asked to make the predictions about others' current and future decisions side by side, as illustrated in Panel 2 of Figure 3.6.

Panel 1: Self-Predictions

'HOW MUCH TO WORK' DECISIONS: for 01/26/16

On 01/26/16, **after the warm-up**, you will do some number of extra rounds of the Task. You will have to complete the extra work **immediately** after the "Decision that Counts" is selected, with no more than 15 minutes of breaks.

Decisions made <u>now</u> for work to be done on 01/26/16		Decisions made <u>on 01/26/16</u> when the time to do the work comes	
How many extra one-minute rounds would you like to do on 01/26/16 at the following wages?		When the time comes to actually do the work on 01/26/16, how many extra one-minute rounds do you think you will want to do at the following wages?	
Wage	Rounds	Wage	Rounds
\$0.30/round (\$18/hour)	<input type="text"/>	\$0.25/round (\$15/hour)	<input type="text"/>
\$0.25/round (\$15/hour)	<input type="text"/>	\$0.20/round (\$12/hour)	<input type="text"/>

EXIT **SUBMIT**

Panel 2: Other-Predictions

OTHER SUBJECTS' DECISIONS: for 01/26/16

On 01/26/16, **after the warm-up**, each subject will do some number of extra rounds of the Task. Every subject will have to complete his or her extra work **immediately** after the "Decision that Counts" is selected, with no more than 15 minutes of breaks.

Decisions made <u>now</u> for work to be done on 01/26/16		Decisions made <u>on 01/26/16</u> when the time to do the work comes	
How many extra one-minute rounds do you think, on average, other subjects are choosing today to complete on 01/26/16 at the following wages?		When the time comes to actually do the work on 01/26/16, how many extra one-minute rounds do you think, on average, other subjects will want to do at the following wages?	
Wage	Rounds	Wage	Rounds
\$0.20/round (\$12/hour)	<input type="text"/>	\$0.30/round (\$18/hour)	<input type="text"/>
\$0.10/round (\$6/hour)	<input type="text"/>	\$0.25/round (\$15/hour)	<input type="text"/>

EXIT **SUBMIT**

Figure 3.6: Examples of screens eliciting participants' predictions of their own and others' work decisions.

- **Group 3:** Participants in this group are asked both sets of questions described above and illustrated in Figure 3.6. The order in which the participants see these questions is randomized across participants.

I test the robustness of the participants' predictions to posing the two sets of questions (about self and others) to two separate groups of participants (Groups 1 and 2) versus to the same participants (Group 3). On the one hand, asking participants to make predictions about both themselves and others may lead to anchoring effects analogous to those documented by Tversky and Kahneman (1974), where participants use their answers to the first set of predictions as an anchor for the second set of predictions. In this sense, the answers by participants in Groups 1 and 2 present cleaner, unanchored beliefs regarding self and others. On the other hand, the juxtaposed answers of participants in Group 3 more accurately reflect beliefs in situations where individuals explicitly evaluate themselves and others in the same context. Such scenarios arise in a variety of common environments, including relative performance compensation contracts in the workplace and curve-graded assignments in schools. As I show in the next section, the effects are consistent across posing the two sets of questions separately and together, suggesting that the above concerns do not play a significant role for elicited beliefs.

The structure of the decision and prediction questions for both groups of participants is illustrated in Table 3.1. Present bias can be estimated by comparing immediate decisions (blue row in Panel 1 of Table 3.1) to ahead-of-time decisions (green rows in Panel 1). Beliefs about one's own present bias are captured by comparing one's ahead-of-time decisions for future dates (green rows in Panel 1) against predictions of one's decisions when the future dates actually arrive (blue row in Panel 2 and the first set of blue rows in Panel 4). Beliefs about others' present bias are estimated by comparing predictions of others' ahead-of-time decisions for future dates (green rows in Panels 3 and 4) against predictions of others' decisions when the future dates actually arrive (blue rows in Panel 3 and the second set of blue rows in Panel 4).

I wish to elicit thoughtful, truthful answers to the prediction questions. For predictions

regarding others, this can be achieved by making the questions incentive-compatible with monetary rewards for correct predictions. Predictions about one's own behavior, however, are more subtle. In this case, there are feedback effects, since the correctness of these predictions is influenced by the participants' own subsequent behavior, which creates scope for strategic rather than truthful answers and behaviors. For example, participants may use their predictions as commitment devices to guide their future behavior.

To check that the monetary incentives do not prompt any commitment demand that would perversely affect participants' self-predictions, I randomly assign each participant into either the incentivized or the unincentivized treatment arm, with equal probability. Participants in the incentivized arm are given a monetary incentive for correct predictions about decisions that are eventually implemented. The monetary incentives are randomized across these participants, and vary from \$0.10 to \$0.40 – similar to the wages for one minute of work. Participants in the unincentivized arm are asked to state their predictions without any monetary incentive. In order to keep the design symmetric, this is implemented analogously for participants making predictions about self, those making predictions about others, and those making both sets of predictions. The incentive structure extends equally to all predictions made by a given participant.

For example, consider a participant from Group 1 or 3 who is randomly assigned to the incentivized group with a prediction bonus of \$0.20. Suppose that she is asked on her first participation date (Date 1) to predict how much work she will choose to do immediately at \$0.10/round on Date 2, and she answers 15 rounds. Then she receives a prediction bonus of \$0.20 if the following conditions are met: (a) on Date 2, she is asked how much work she would like to complete immediately at \$0.10/round, and she chooses 15 rounds; and (b) this decision is implemented as the "Decision that Counts." Similarly, consider a participant from Group 2 or 3, who is randomly assigned to the incentivized group with a prediction bonus of \$0.20. Suppose that on Date 1, she is asked to predict how much work, on average, other participants will prefer to do immediately on Date 2 at \$0.10/round, and she answers 15 rounds. Then she will receive a bonus of \$0.20 if: (a) on Date 2, at least one other participant

Table 3.1: Decision and prediction questions posed to participants in Groups 1, 2, and 3.

Panel 1: Decisions – All Participants			
Decisions on Date 1	Decisions on Date 2	Decisions on Date 3	Decisions on Date 4
	For Date 2	For Date 3	For Date 4
For Date 2	For Date 3	For Date 4	
For Date 3	For Date 4		

Panel 2: Predictions – Group 1 Participants			
Predictions on Date 1	Predictions on Date 2	Predictions on Date 3	Predictions on Date 4
Own decision on Date 2 for Date 2	Own decision on Date 3 for Date 3	Own decision on Date 4 for Date 4	
Own decision on Date 3 for Date 3	Own decision on Date 4 for Date 4		

Panel 3: Predictions – Group 2 Participants			
Predictions on Date 1	Predictions on Date 2	Predictions on Date 3	Predictions on Date 4
Others' dec. on Date 2 for Date 2	Others' dec. on Date 3 for Date 3	Others' dec. on Date 4 for Date 4	
Others' dec. on Date 3 for Date 3	Others' dec. on Date 4 for Date 4		
Others' dec. on Date 1 for Date 2	Others' dec. on Date 1 for Date 3	Others' dec. on Date 1 for Date 4	
Others' dec. on Date 1 for Date 3	Others' dec. on Date 1 for Date 4		

Panel 4: Predictions – Group 3 Participants			
Predictions on Date 1	Predictions on Date 2	Predictions on Date 3	Predictions on Date 4
Own decision on Date 2 for Date 2	Own decision on Date 3 for Date 3	Own decision on Date 4 for Date 4	
Own decision on Date 3 for Date 3	Own decision on Date 4 for Date 4		
Others' dec. on Date 2 for Date 2	Others' dec. on Date 3 for Date 3	Others' dec. on Date 4 for Date 4	
Others' dec. on Date 3 for Date 3	Others' dec. on Date 4 for Date 4		
Others' dec. on Date 1 for Date 2	Others' dec. on Date 1 for Date 3	Others' dec. on Date 1 for Date 4	
Others' dec. on Date 1 for Date 3	Others' dec. on Date 1 for Date 4		

is asked how much work he would like to complete immediately at \$0.10/round, and the average answer is 15 rounds; and (b) this decision is implemented as the "Decision that Counts" for at least one of the other participants.

3.2.5 Sample

The experiment runs over five non-overlapping sessions, with a total of 364 individuals taking part, recruited through the Harvard Decision Sciences Lab. In order to be eligible for the study, participants must pass a comprehension quiz after reading the instructions, testing the participants' understanding of the experiment. The comprehension quiz includes questions regarding payment, timeline, decisions, and predictions; the full quiz is catalogued in Appendix C.3.

The goal of running the experiment over multiple sessions is to minimize the effects of any unforeseen systematic shocks such as weather disruptions or university-wide events. Furthermore, since a large part of the Harvard Decision Sciences Lab subject pool consists of Harvard University undergraduates, the five sessions are explicitly timed to avoid the University's midterm exams (mid-March) and final exams (May). The five experimental sessions are run at the following times:

- Session 1: January 11 - February 7, 2016
- Session 2: February 8 - March 6, 2016
- Session 3: March 28 - April 24, 2016
- Session 4: June 6 - July 3, 2016
- Session 5: July 11 - August 7, 2016

In addition, a small-scale pilot study of the experimental design is run during October 12 - November 8, 2015. For the results of the pilot study, please refer to Appendix C.1.2.

A total of 198 participants complete the entirety of the experiment during the five experimental sessions, with an additional 166 participants consenting to participate but

Table 3.2: *Numbers of recruited participants and attrition rates across experimental sessions.*

	Consent	Week 1	Week 2	Week 3	Week 4
Pilot	27	23	21	19	19
Sessions 1-5	364	278	230	208	198
Session 1: Jan. 11 - Feb. 7, 2016	78	61	53	50	50
Session 2: Feb. 8 - Mar. 6, 2016	81	65	59	52	50
Session 3: Mar. 26 - Apr. 24, 2016	86	71	57	49	43
Session 4: Jun. 6 - Jul. 3, 2016	64	42	31	29	28
Session 5: Jul. 11 - Aug. 7, 2016	55	39	30	28	27

not finishing the entirety of the four-week-long experiment. A break-down of recruited participants and attrition rates by session is reported in Table 3.2. Since registering for the online study is virtually costless, a large number of participants drop out once they begin reading the instructions upon their first log in; of the 364 participants consenting to take part in the experiment, 86 (24%) do not complete the instructions, warm-up, and work decisions on the first participation date or fail the comprehension quiz. The attrition rates attenuate over the subsequent weeks, as exiting the experiment costs the participants their \$30 completion payments. Of the 278 participants who finish their first participation date, 230 (83%) complete the second participation date, 208 (75%) complete the third participation date, and 198 (71%) complete the entirety of the experiment. The results detailed below are robust to including attrited participants and to focusing solely on those participants who complete the entirety of the experiment.

3.3 Reduced-Form Results

The reduced-form results from the laboratory experiment suggest that participants have more awareness regarding present bias in others than in themselves. The results are robust to varying the incentive structure for the predictions and to posing the questions about self and others together to the same participants or to separate groups of participants.

3.3.1 Present Bias and Beliefs

Pooled results from all participants who finish the experiment indicate that (i) participants display present bias in their effort choices; (ii) participants do not anticipate their own present bias; and (iii) participants expect present bias in others. For robustness of these results to the inclusion of attrited participants who complete the preliminaries but do not finish the experiment, please refer to Appendix C.1.1.

The pooled sample consists of the 198 participants who complete the entirety of the experiment. Of these, 60 are in Group 1, 60 are in Group 2, and 78 are in Group 3. Each participant is asked to make a total of 6 decisions for immediate work and 10 decisions for future dates. In addition, each participant in Groups 1 and 3 answers 10 questions regarding her own decisions when the future dates actually arrive. Similarly, each participant in Groups 2 and 3 is asked a total of 10 questions regarding others' current work decisions for future dates and 10 questions about what others will choose when the future dates actually arrive.

Present bias is estimated by comparing participants' work decisions for future dates against their work decisions for immediate completion. The experimental participants choose to do, on average, 30.03 rounds per session when the choices are elicited ahead of time. When asked how much work they would like to complete immediately, the participants choose an average of 26.53 rounds per session. Figure 3.7 plots the ahead-of-time and immediate work decisions across the five possible wages from \$0.10/minute to \$0.30/minute, with standard error bars clustered by participant. As illustrated in the figure, participants choose to do more work when the decision is made in advance for all wages except for \$0.10/minute.

The difference between the two types of decisions is statistically significant and robust to controlling for wage fixed effects and participant fixed effects. The results are presented in Panel 1 of Table 3.3. Participants choose to do, on average, 3.50 rounds fewer when their decisions are for immediate work (3.33 rounds when controlling for wage fixed effects, 3.50 with participant fixed effects, and 3.36 including both fixed effects). The difference

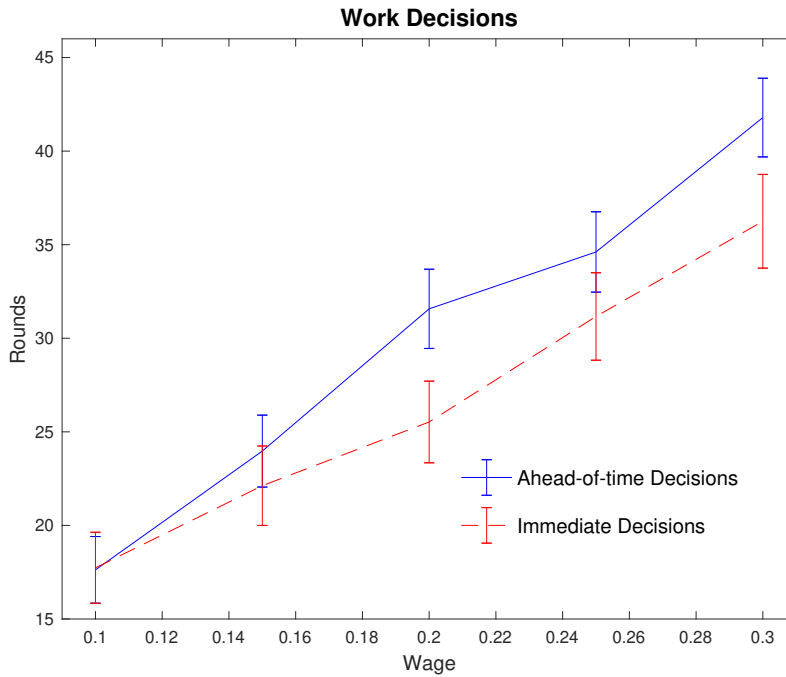


Figure 3.7: A comparison of work decisions for the current date (immediate decisions) and for future dates (ahead-of-time decisions).

is statistically significant at the 1% level across specifications, and consistent with prior evidence on present bias in real-effort tasks (see, e.g., Augenblick *et al.* (2015)).

Are the participants aware of this time inconsistency in their effort choices? The participants' naïveté regarding their own present bias is captured by comparing their work decisions for future dates against their *predictions* of the decisions they will make when those dates actually arrive. Forecasts of lower work decisions when the dates actually arrive would indicate experimental participants' sophistication regarding their present bias. On the other hand, if participants do not anticipate their decisions changing when the work becomes imminent, then they display naïveté regarding their present bias.

The results indicate that, on average, experimental participants display little anticipation of their own present bias. The average predicted differences in their effort choices, estimated from the 138 participants making self-predictions (i.e., participants from Groups 1 and 3), are displayed in Panel 2 of Table 3.3. The differences are small, and significant only in one

Table 3.3: *Reduced-form results from participants who complete the entirety of the experiment.*

Panel 1: Actual Difference in Ahead-of-time vs. Immediate Decisions

Actual Difference	3.50**	3.33**	3.50**	3.36**
Standard error	(0.59)	(0.55)	(0.61)	(0.57)
Controls:				
Wage FE		X		X
Participant FE			X	X

Panel 2: Predicted Difference in Own Ahead-of-time vs. Immediate Decisions

Self-Prediction	0.10	-0.43	1.15**	0.62†
Standard error	(1.13)	(1.12)	(0.40)	(0.34)
Controls:				
Wage FE		X		X
Participant FE			X	X

Panel 3: Predicted Difference in Others' Ahead-of-time vs. Immediate Decisions

Other-Prediction	1.27*	1.47**	1.27*	1.49**
Standard error	(0.52)	(0.42)	(0.53)	(0.44)
Controls:				
Wage FE		X		X
Participant FE			X	X

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

specification (participant fixed effects, no wage fixed effects). The differences also do not have consistent signs across the specifications. This is consistent with the findings of naïveté in Augenblick and Rabin (2018).

By contrast, predictions about others, made by the 138 participants in Groups 2 and 3, reveal awareness of others' present bias. The experimental participants expect their peers to choose less work when the decisions are for immediate completion than when the decisions are for future completion. Asked how many rounds others wish to do ahead of time, participants predict an average of 28.28 rounds. When asked about others' work decisions for immediate completion, the average prediction is 27.01 rounds. The difference, estimated in Panel 3 of Table 3.3, is significant at the 1% level if participant fixed effects are included,

and at the 5% level omitting participant fixed effects. The results are robust to inclusion of attrited participants, and are statistically stronger in this larger sample (see Appendix C.1.1).

Interestingly, while predictions regarding others reveal that participants expect present bias in others, participants do not correctly guess the magnitude of the effect. The average predicted difference in others' decisions is 1.49 rounds, whereas the actual average difference in participants' work decisions is 3.50 rounds. This contrasts with the classroom experiment detailed in Section 3.5, in which the students' predictions regarding their peers are remarkably well calibrated. The difference in accuracy across the two settings is most likely attributable to the participants' different levels of experience with the two settings: students have substantial experience observing their classmates procrastinate on assignments, but participants in the experiment have no experience observing others choose work decisions for the experimental task.

Overall, the reduced-form results from the online experiment paint the following picture. Experimental participants display almost full naïveté about their own present bias but some, although imperfect, awareness of others' present bias. In the remainder of this section, I explore the robustness of these findings to incentivizing predictions and to posing the questions jointly or separately.

3.3.2 Incentivizing Predictions

I test the robustness of the results to altering the incentivization mechanism for eliciting participants' predictions regarding their own and others' work decisions. The predictions are not significantly different when the questions are posed in an unincentivized manner versus when the participants are offered monetary bonuses for correct predictions.

The incentive structure is randomized across experimental participants. Each participant is randomly allocated, with equal probability, to either the incentivized or the unincentivized treatment arm. Within the incentivized arm, the size of the incentive is randomly selected from \$0.10, \$0.20, \$0.30, or \$0.40 per correct prediction, with equal likelihoods. Thus, of the 198 participants who finish the experiment, 95 are unincentivized, and 103 are incentivized,

with 23 participants receiving the \$0.10 bonus, 24 receiving the \$0.20 bonus, 30 receiving the \$0.30 bonus, and 26 receiving the \$0.40 bonus.

Table 3.4: *Predicted differences in one's own and others' decisions, sliced by incentive.*

Panel 1: Self-Predictions by Incentivized and Unincentivized Participants

	Incentivized	Unincentivized
Self-Prediction	0.59	0.65
Standard error	(0.48)	(0.46)
Controls:		
Wage FE	X	X
Participant FE	X	X

Panel 2: Self-Predictions by Incentivized Participants, Varying Size of Incentive

	\$0.10 incentive	\$0.20 incentive	\$0.30 incentive	\$0.40 incentive
Self-Prediction	0.03	0.66	0.67	0.57
Standard error	(0.98)	(1.11)	(0.58)	(0.98)
Controls:				
Wage FE	X	X	X	X
Participant FE	X	X	X	X

Panel 3: Other-Predictions by Incentivized and Unincentivized Participants

	Incentivized	Unincentivized
Other-Prediction	1.41*	1.70**
Standard error	(0.62)	(0.60)
Controls:		
Wage FE	X	X
Participant FE	X	X

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

The participants do not display significant awareness of their own present bias, regardless of the incentive structure. Panel 1 of Table 3.4 reports the average predicted differences in one's own decisions for the 72 participants asked to make incentivized self-predictions and the 66 participants making self-predictions without monetary incentives. The predicted differences are estimated with wage and participant fixed effects. The average predicted difference is 0.59 rounds with the incentive and 0.65 rounds without the incentive. In both

cases, the predicted difference is statistically indistinguishable from zero, indicating that the participants do not anticipate present bias in their own decisions, regardless of whether they are incentivized for correct predictions. For the incentivized group, the participants' naïveté is likewise robust across the size of the incentive. With the exception of the \$0.10 incentive, for which the predicted difference is 0.03 rounds, the point estimates of the predicted differences are approximately 0.60 rounds across the incentive amounts. For none of the incentives are these predicted differences statistically distinguishable from zero, although the sliced samples are too small to properly evaluate significance.

Incentivizing predictions also has no significant effect on the elicited beliefs about other participants. The predicted differences in others' decisions are, on average, 1.41 rounds when the predictions are incentivized and 1.70 rounds without the incentive. In both cases, the predicted differences are statistically different from zero. The former is significant at the 5% level and the latter at the 1% level.

Overall, the subsample analysis slicing by incentive indicates that the results are not driven by strategic responses to incentive structures. Instead, the participants' answers are robust to incentivized and unincentivized elicitation of beliefs. Across the board, participants display fairly precise awareness of others' present bias, and no significant awareness of their own present bias.

3.3.3 Juxtaposing Predictions about Self and Others

Next, I confirm the robustness of the results to posing the two sets of questions (predictions regarding self and others) to the same participants or separately to two groups of participants. Participants' answers do not systematically vary across the two methods of posing the questions, as evidenced by the results in Table 3.5.

Participants do not expect significant present bias in themselves, regardless of whether they are also asked to make predictions about others. The 60 participants who make only self-predictions anticipate that they will choose to do an average of 0.62 rounds fewer when the work decision has immediate consequences; this predicted difference is not statistically

Table 3.5: *Predicted differences in one's own and others' decisions, sliced by method of presenting the questions.*

Panel 1: Self-Predictions

	Self-Prediction Only	Both Sets of Questions
Self-Prediction	0.62	0.73
Standard error	(0.41)	(0.50)
Controls:		
Wage FE	X	X
Participant FE	X	X

Panel 2: Other-Predictions

	Other-Prediction Only	Both Sets of Questions
Other-Prediction	1.53**	1.46*
Standard error	(0.59)	(0.64)
Controls:		
Wage FE	X	X
Participant FE	X	X

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

different from zero. Similarly, the 78 participants who also face questions about others predict that they will choose an average of 0.73 rounds fewer for immediate completion, also statistically indistinguishable from zero.

Likewise, participants' expectations of others' present bias are similar across those who only make predictions regarding others and those who answer both sets of questions. The 60 participants who are only asked to make predictions about others expect that the average other participant will want to do 1.53 rounds fewer when the work decision is made for immediate completion. The 78 participants who are asked to make both sets of predictions anticipate that others will choose 1.46 rounds fewer when the work decision is made for immediate completion. The predicted differences by both groups are statistically significantly different from zero at the 5% level.

Overall, participants appear to be providing independent answers for the two sets of questions, and their answers are robust to seeing only one type of question or both. These results suggest that the participants' predictions regarding their own future decisions and

regarding the decisions of the other participants are not influenced by anchoring effects or strategic comparisons. Instead, the elicited predictions reflect the participants' underlying beliefs regarding self and others.

In Appendix C.1.3, I exploit differences in participants' warm-up amounts to explore whether the warm-up amount systematically affects the participants' displayed present bias and predictions, and find mixed results. Overall, the analysis sliced by warm-up indicates two patterns. First, my experimental design does not elicit significant projection bias in the participants. Second, the qualitative patterns of more critical beliefs regarding others tend to hold across the different warm-up amounts.

3.4 Structural Estimation

In this section, I present a model of decision-making and predictions within the framework of β - δ preferences and use my experimental data to estimate the model's parameters. I find estimates of present bias consistent with prior literature, virtually no awareness of one's own present bias, and robust albeit incomplete awareness of others' present bias.

3.4.1 Model

I begin by outlining the functional form assumptions for my structural estimation. I then present the resulting expressions for the participants' work decisions, their predictions regarding their own future decisions, and their predictions regarding the decisions of other participants.

Structural Assumptions. The model of individual decision-making and predictions is based on standard quasi-hyperbolic discounting. Each subsequent period is discounted relative to the preceding period by a time-consistent (daily) discount factor δ ; in addition, all periods outside of the current one are discounted by the present bias parameter β .

I denote each participant's beliefs regarding her own present bias by $\beta_{(s)}$, and beliefs regarding others by $\beta_{(o)}$. Aggregate parameter estimation constrains the discount factors, beliefs, and all functional form parameters to be the same across all experimental

participants.

I assume that each participant's utility function is separable in effort and money. Thus, when a participant performs e rounds of the task in addition to her x warm-up rounds, she incurs some effort cost $c(e + x)$ immediately and receives utility $U(w \times e)$ from the monetary payoff $w \times e$ on the predetermined future (post-experiment) payment date. For the purposes of estimation, I follow Augenblick and Rabin (2018) and assume a linear form for the utility in money with parameter ϕ (i.e., $U(w \times e) = \phi ew$) and a power form for the cost of effort with parameter γ (i.e., $c(e) = \frac{1}{\gamma}e^\gamma$).

When evaluating predictions regarding others, I assume that all participants perceive others' utility function to have the same functional form and parameters ϕ and γ as their own.¹⁰ I also assume that a participant with a warm-up amount of x understands all other experimental participants to have the same warm-up x .¹¹ Thus, the only feature that the participants perceive as different for others than for themselves is the feature of interest: present bias.

Decisions. The structural assumptions allow for the following expression of the participants' decisions. Let T denote the payment date, and consider an individual with a warm-up amount x making a decision about how much work to complete immediately at wage w on date τ . She discounts the future payment by $\beta\delta^{T-\tau}$, but does not discount the immediate cost of effort. As a result, her chosen number of extra rounds is given by:

$$e_{imm}^* = \arg \max_e \left\{ \beta\delta^{T-\tau} \phi w e - \frac{1}{\gamma} (e + x)^\gamma \right\} = (\beta\delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (3.1)$$

Now suppose that the participant is again considering how much to work on the same date τ , but the decision itself is now made ahead of time, on date $t < \tau$. In this case, both

¹⁰I relax this assumption in Appendix C.2.2, where I estimate self-predictions and other-predictions separately, allowing for different beliefs regarding others' parameters γ and ϕ . In Appendix C.2.1, I do not allow for differing γ and ϕ , but incorporate a difference in baseline levels of actual ahead-of-time work decisions and the predicted ahead-of-time decisions by others.

¹¹Since the participants are not told that the warm-up amounts can vary, and there is no communication between experimental participants, there is no reason for them to think that others have different warm-up amounts.

the monetary payment and the effort cost are incurred in the future, and both are discounted by the present bias parameter β . As a result, the participant chooses the following number of extra rounds:

$$e_{del}^* = \arg \max_e \{ \beta \delta^{T-t} \phi w e - \beta \delta^{\tau-t} \frac{1}{\gamma} (e+x)^\gamma \} = (\delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (3.2)$$

Self-Predictions. In order to model a participant's expectations regarding her own future decisions, consider a participant on date t asked what choice she would make for immediate work at a wage w when a future date τ actually arrives. Effectively, the participant is predicting, ahead of time on date t , the decision expressed in (3.1).

Under the structural assumptions outlined above, the participant is aware of her effort cost function and her utility from money, but may hold incorrect beliefs regarding her present bias parameter β . In particular, the participant thinks that her future self will be making the decision in (3.1) under a present bias parameter $\beta_{(s)}$. Hence, she expects to choose the following number of extra rounds:

$$e_{imm}^{(s)} = \arg \max_e \{ \beta_{(s)} \delta^{T-\tau} \phi w e - \frac{1}{\gamma} (e+x)^\gamma \} = (\beta_{(s)} \delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (3.3)$$

Other-Predictions. I now turn to expectations regarding the decisions made by other participants.

First, consider a participant in period t asked to predict how many rounds of the task others are choosing now for some future date τ , at a given wage w . The participant assumes that others have the same warm-up amount as her, x , as well as the same utility in money, effort cost function, and time-consistent discount factor δ . She perceives the others' present bias parameter to be $\beta_{(o)}$. Since the predicted choice is made for future work, the participant understands that others will discount both the effort cost and the monetary payoff by the present bias parameter. She hence perceives the others' decision for the delayed work as follows:

$$e_{del}^{(o)} = \arg \max_e \{ \beta_{(o)} \delta^{T-t} \phi w e - \beta_{(o)} \delta^{\tau-t} \frac{1}{\gamma} (e+x)^\gamma \} = (\delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (3.4)$$

Second, consider a participant in period t predicting how many rounds of the task others will want to do when the future date τ actually arrives. Since the predicted choice consists of a trade-off between immediate work and a delayed monetary payoff (in period T), the participant expects that only the monetary payoff will be discounted by the present bias parameter $\beta_{(o)}$. As a result, she expects others to make the following choice:

$$e_{imm}^{(o)} = \arg \max_e \{ \beta_{(o)} \delta^{T-\tau} \phi w e - \frac{1}{\gamma} (e+x)^\gamma \} = (\beta_{(o)} \delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (3.5)$$

The discussion above makes one key implicit assumption: that participants' predictions reflect their true expectations regarding their own and others' behavior. For predictions regarding others, this assumption should not pose any problems: participants have no strategic incentive to misstate their beliefs regarding others, and the prediction bonuses offer an unambiguous incentive to provide their best guesses. For predictions regarding one's own future behavior, the bonus might introduce a strategic incentive to use the predictions as a commitment device. However, the analysis in Section 3.4.2 indicates that the participants' answers do not vary significantly based on incentivizing predictions. Furthermore, Augenblick and Rabin (2018) find that allowing for the possibility that participants are optimally using their predictions as commitment devices does not alter the structural estimates of β and $\beta_{(s)}$. Hence, throughout this section, I assume that all predictions are stated truthfully.

3.4.2 Empirical Estimation

I now estimate the parameters in equations (3.1)-(3.5) using data from the experimental participants' responses to the decision and prediction questions.

Let $e(t, \tau, w, x, \mathbb{1}_s, \mathbb{1}_o)$ denote the stated number of extra rounds in response to a question posed to a participant with warm-up amount x on date t regarding work on date τ at wage w . The indicator variable $\mathbb{1}_s$ captures self-prediction responses, while the indicator variable $\mathbb{1}_o$ denotes responses about others. The remaining responses are the participant's actual decisions. The indicator $\mathbb{1}_{t=\tau}$ is equal to one if and only if the date when the response is elicited, t , is the same as the date for which the question is posed, τ .

Combining (3.1)-(3.5), I express the model's prediction of the participant's response, $\hat{e}(t, \tau, w, x, \mathbb{1}_s, \mathbb{1}_o)$, in terms of the model parameters:

$$\hat{e}(t, \tau, w, x, \mathbb{1}_s, \mathbb{1}_o) = (\beta^{\mathbb{1}_{t=\tau}} \beta_{(s)}^{\mathbb{1}_s} \beta_{(o)}^{\mathbb{1}_{t=\tau} \mathbb{1}_o} \delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (3.6)$$

I estimate the parameters in (3.6) using maximum likelihood. Since the participants cannot pick fewer than 0 rounds or more than 70 rounds in any of the questions, I use a two-limit Tobit regression, with censoring from below at 0 and from above at 70 rounds.

Intuitively, the model parameters are identified from the data as follows:

- Variation in the timing of the decisions identifies β . In equation (3.6), the parameter β is present for immediate decisions ($\mathbb{1}_{t=\tau} = 1$ when $t = \tau$) but not for future decisions. Thus, present bias is identified by the difference between the number of rounds chosen for future participation dates and the number of rounds chosen for immediate work. Higher β corresponds to a smaller difference between these two sets of decisions.
- Restricting attention to the decisions made for the future, variation in the proximity of these future dates identifies the time-consistent discount factor δ , which enters multiplicatively with each day separating the decision date t from the work date τ .
- The parameter $\beta_{(s)}$ is identified through a comparison of the participants' actual decisions for immediate work and their predictions regarding these decisions, since the term $\beta_{(s)}$ enters only for the latter set of responses. Higher $\beta_{(s)}$ thus corresponds to smaller predicted differences in one's own choices.
- Similarly, the parameter $\beta_{(o)}$ is identified through a comparison of choices for future dates against participants' predictions regarding others' choices when those dates actually arrive, since only the latter set of responses reflects the parameter $\beta_{(o)}$.¹² The higher is $\beta_{(o)}$, the less difference participants anticipate to see in others' choices.

¹²Note that in (3.6), the participants' current choices for future work and their predictions of others' current choices for future work are indistinguishable. I allow for the possibility of a difference in these baseline levels in Appendix C.2.1.

- Lastly, parameters γ and ϕ , which capture the shape of the participants' utility function in money and effort, are identified through variation in the wage w , which traces out the curvature and intercept of the utility function.

The parameter estimates are presented in Table 3.6. The first column reports the estimates of the two-limit Tobit regression of (3.6) restricting attention to the 198 participants who complete the entirety of the experiment. The second column includes all participants who provide at least one work decision or prediction response. Standard errors are bootstrapped and clustered by participant across the specifications. The estimated cost of effort is close to quadratic, with the estimate of $\hat{\gamma}$ at approximately 2.25, and the estimate of ϕ is around 380. The time-consistent discount factor δ is slightly above 1, indicating that when making decisions for the future, participants choose to do more rounds of work when the future is relatively nearer. This is possibly driven by participants being more certain regarding their schedules for the near future than for dates further in the future, and is consistent with the pattern documented by Augenblick and Rabin (2018).

The estimated model parameters reveal a significant extent of present bias among experimental participants. Estimates of the parameter β are around 0.82-0.86, with the difference from the null of $\beta = 1$ (no present bias) statistically significant at the 1% level. The estimates are consistent with prior evidence on present bias: for example, Laibson *et al.* (2008) estimate β around 0.71 using consumption choices, while Augenblick *et al.* (2015) document β around 0.89 for real-effort tasks. In the closest setting to the present paper, Augenblick and Rabin (2018) obtain estimates of β around 0.83.

Participants' beliefs regarding their own present bias, captured by the parameter $\beta_{(s)}$, display virtually complete naïveté, supporting the reduced form results. Estimates of $\beta_{(s)}$ are around 1.03-1.05, statistically indistinguishable from the complete-naïveté case of $\beta_{(s)} = 1$.

The participants display substantially more awareness of others' present bias than of their own. The parameter $\beta_{(o)}$, which captures the participants' beliefs regarding others' present bias, is estimated to be around 0.87, and strongly statistically significantly different from the null of $\beta_{(o)} = 1$. The estimated value of $\beta_{(o)}$ is significantly lower than that of β_s :

Table 3.6: *Parameter estimates from the structural model.*

Parameter	Without attrited participants	With attrited participants
Present bias β	0.8589 (0.0330)	0.8151 (0.0335)
Self-prediction $\beta_{(s)}$	1.0502 (0.0629)	1.0306 (0.0523)
Other-prediction $\beta_{(o)}$	0.8711 (0.0349)	0.8715 (0.0314)
δ	1.0147 (0.0024)	1.0154 (0.0028)
γ	2.2485 (0.1123)	2.2481 (0.0910)
ϕ	377.4181 (209.5252)	385.9745 (169.9324)

the bootstrapped t-statistic on the difference between these two parameters is 2.11 without attrited participants, and 2.14 including all participants.

The beliefs-about-others parameter $\beta_{(o)}$ is somewhat higher than the true present bias parameter β , but the difference is not statistically significant. Consistent with the reduced-form results, these parameter estimates indicate that participants are consistently aware of the fact that others will choose to do fewer rounds of the task when the work is imminent, even if they may underestimate the full extent of the difference.

The findings of significant present bias, naïveté about one’s own present bias, and awareness of present bias in others are robust to alternative specifications considered in Appendices C.2.1 and C.2.2: different baseline levels in ahead-of-time decisions for self versus predicted others and separate estimation of all model parameters for self and others. In these alternative specifications, I find consistent estimates of β around 0.78-0.84 (always significantly different from 1), $\beta_{(s)}$ between 0.99 and 1.01 (never significantly different from 1), and $\beta_{(o)}$ between 0.92 and 0.93 (always significantly different from 1).

Overall, my structural estimates confirm the intuition from the reduced form experimental results: although individuals tend to be naïve about their own present bias, they are more aware of present bias in others.

3.4.3 Individual Level Estimates

I estimate the structural model individually for each participant, and document two key findings. First, individual-level results confirm the results from pooled analysis: $\beta_{(s)}$ is centered around 1.00, while $\beta_{(o)}$ is centered around 0.93. Second, although the estimates of $\beta_{(s)}$ are perfectly naïve in absolute terms, individual-level estimates of $\beta_{(s)}$ and β are positively correlated, indicating some awareness of relative self-control.

I estimate $\beta^{(i)}$, $\beta_{(s)}^{(i)}$, and $\beta_{(o)}^{(i)}$ for each individual participant i as follows. For participants in Group 3, who face the full set of experimental questions, I estimate the full specification (3.6) individually for each participant, obtaining individual-level estimates $\beta^{(i)}$, $\beta_{(s)}^{(i)}$, and $\beta_{(o)}^{(i)}$.

For each participant i in Group 1, who makes predictions about herself but not others, I use her work decisions and predictions regarding her own future work to estimate $\beta^{(i)}$ and $\beta_{(s)}^{(i)}$. Denote the predicted values of individual i 's responses by $\hat{e}_{(s)}^{(i)}(t, \tau, w, x, \mathbb{1}_s)$. Then I estimate the following specification:

$$\hat{e}_{(s)}^{(i)}(t, \tau, w, x, \mathbb{1}_s) = \left((\beta^{(i)})^{\mathbb{1}_{t=\tau}} (\beta_{(s)}^{(i)})^{\mathbb{1}_s} \delta^{T-\tau} \phi w \right)^{\frac{1}{\gamma-1}} - x \quad (3.7)$$

Similarly, for each participant i in Group 2, I use her predictions regarding others' current and future choices, $\hat{e}_{(o)}^{(i)}(t, \tau, w, x)$, to estimate the following specification:

$$\hat{e}_{(o)}^{(i)}(t, \tau, w, x) = \left((\beta_{(o)}^{(i)})^{\mathbb{1}_{t=\tau}} \delta^{T-\tau} \phi w \right)^{\frac{1}{\gamma-1}} - x \quad (3.8)$$

which yields an individual estimate of the parameter $\beta_{(o)}^{(i)}$.

I then compile individual estimates $\beta^{(i)}$ and $\beta_{(s)}^{(i)}$ from participants making self-predictions (Groups 1 and 3), and individual estimates of $\beta_{(o)}^{(i)}$ from participants making other-predictions (Groups 2 and 3). All individual estimates are constrained to fall between 0.5 and 1.5. The median values and inter-quartile ranges of the three estimated parameters are displayed in Figure 3.8. These results reflect all participants, including attritors. Without attritors, the median values of the estimated parameters are 0.85 for $\beta^{(i)}$, 1.00 for $\beta_{(s)}^{(i)}$, and 0.94 for $\beta_{(o)}^{(i)}$.

Although individual-level estimates are imprecise due to small amounts of data per

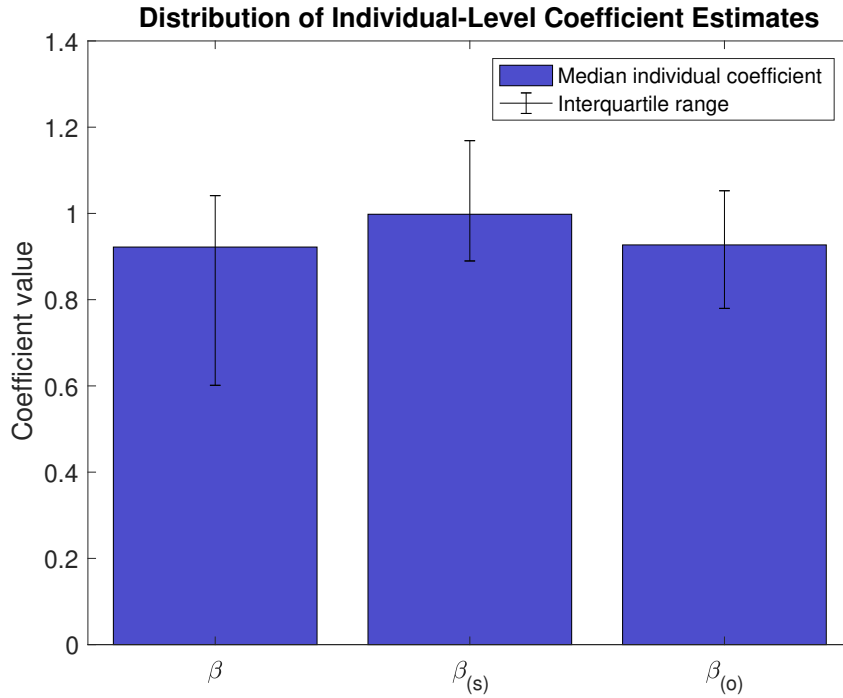


Figure 3.8: *Distribution of the structural parameter estimates across individual experimental participants.*

participant (e.g., each participant makes only 10 decisions for immediate work), the patterns are consistent with the pooled results. The median value of individual estimates of $\beta^{(i)}$ is 0.92, the median value of estimates of $\beta_{(s)}$ is 1.00, and the median value of individual-level estimates of $\beta_{(o)}$ is 0.93. The mean values of the three estimated parameters are 0.88, 1.02, and 0.94, respectively, consistent with the results in Table 3.6. The interquartile ranges, however, are quite wide. The 75th percentile of all three estimated parameters is above 1, and the 25th percentile is as low as 0.60 for estimates of $\beta^{(i)}$, 0.89 for $\beta_{(s)}^{(i)}$, and 0.78 for $\beta_{(o)}^{(i)}$.

Interestingly, despite the wedge in median estimates, it is not necessarily the case that predictions regarding randomly matched others are more accurate than self-predictions. I perform the comparison by matching each individual estimate $\beta^{(i)}$ to two predictions: (i) the same individual's self-prediction $\beta_{(s)}^{(i)}$; and (ii) a randomly drawn, with replacement, prediction $\beta_{(o)}^{(j)}$. I then calculate the mean squared error of each set of predictors. The self-predictions yield a mean squared error of 0.1237. I run 1,000 random matches to

estimate the mean squared error of the randomly matched other-predictions, and find a median value of 0.1445 and an inter-quartile range of [0.1368, 0.1525]. Thus, it is possible for the self-predictions to actually perform better in matching the present bias parameters $\beta^{(i)}$ than random matching with other-predictions, although this result is sensitive to the exact evaluation criterion used (mean squared error).

The reason for the relatively good performance of self-predictions is the following. While the median value of individual-level $\beta_{(s)}^{(i)}$ is perfectly naïve at 1.00, there is a *positive correlation* between the individual-level parameter estimates. In particular, the correlation between individual-level estimates of $\beta^{(i)}$ and $\beta_{(s)}^{(i)}$ is 0.28. Similarly, a regression of the extent of an individual's self-awareness, $1 - \beta_{(s)}^{(i)}$, on the extent of the individual's actual self-control problem, $1 - \beta^{(i)}$, yields a coefficient of 0.19, statistically significant at the 1% level. By construction, predictions regarding others are elicited regarding the *average* of others' behavior, and not regarding precise individuals. As a result, although other-predictions are substantially more accurate in general, the average wedge in beliefs may not necessarily compensate for precise knowledge of a given individual.

The last question I pose with the individual-level estimates is whether there is a positive relationship between individuals' beliefs regarding self and others. Conceptually, when forming beliefs regarding unknown others, individuals may project their expectations of their own behavior.¹³ This would induce a positive relationship between the estimates of $\beta_{(s)}^{(i)}$ and $\beta_{(o)}^{(i)}$ for those participants who make both sets of predictions. To test this conjecture, I look at individual-level estimates of $\beta_{(s)}^{(i)}$ and $\beta_{(o)}^{(i)}$ from participants in Group 3. Interestingly, I find no positive relationship between beliefs regarding self and others. If anything, correlation between individual-level estimates of $\beta_{(s)}$ and $\beta_{(o)}$ is negative at -0.21. Hence, it does not appear to be the case that my experimental participants project

¹³A number of studies document that people tend to self-project when forming beliefs regarding others in a variety of domains. For example, Van Boven and Loewenstein (2000) study self-projection of valuations, Van Boven and Loewenstein (2003) consider self-projection of transient drive states, Iriberry and Rey-Biel (2013) find evidence of self-projection of social preferences, and Ludwig and Nafziger (2011) observe self-projection of perceived ability. Madarasz (2012) proposes a model of self-projection of information, and Loewenstein *et al.* (2006) and Danz *et al.* (2015) offer experimental studies illustrating information projection.

their self-expectations when forming beliefs regarding others.

3.5 External Evidence: Classroom Experiment

This section presents an intuitive field survey conducted in a classroom, which serves to illustrate the wedge in beliefs regarding one's own and others' present bias in a real-world setting. I first outline the setting and design of the classroom experiment, and then present the results.

3.5.1 Design

The classroom experiment is administered to students in an undergraduate financial accounting course (BUS201) at the University of San Francisco. On the first day of class, January 25, 2016, the students are presented with the course syllabus and introduced to the Individual Project that they have to complete for the course, due on May 2, 2016. The project consists of analyzing accounting ratios of a publicly traded company. In order to proceed with the project, students must first choose a company to analyze and confirm that they can download the company's financial statements for the past three years from the Securities and Exchange Commission's website. The students need to email their chosen company and the downloaded financial statements for instructor approval by April 2, 2016. No two students can cover the same company, and approval is granted on a first-come-first-served basis.

Present bias is proxied by the time when the students email the instructor for approval (hereafter referred to as the students' "completion dates"). On the one hand, earlier submission is efficient in that it maximizes the chances of approval (i.e., that no other student has preempted the choice) and leaves more time to work on the project once approved. On the other hand, downloading financial statements carries an immediate effort cost, on which the students might wish to procrastinate. Thus, the more present-biased a given student is, the more likely she is to delay the completion date.

After the project is explained to the students, they are asked to fill out an anonymous, voluntary survey, featuring one or both of the following two questions:

- **Self-Prediction:** As you can see from the syllabus, the deadline for the Individual Project is on May 2, 2016. The last day to submit your chosen company for instructor approval is on April 2, 2016. When do you think you will email your chosen company to the instructor? (Enter a date)
- **Other-Prediction:** As you can see from the syllabus, the deadline for the Individual Project is on May 2, 2016. The last day to submit your chosen company for instructor approval is on April 2, 2016. On average, when do you think your classmates will email their chosen companies to the instructor? (Enter a date)

Each student receives one of four survey versions, distributed randomly among the students:

- Group 1: This version includes the self-prediction question only.
- Group 2: This version includes the other-prediction question only.
- Group 3: This version includes both predictions, with the self-prediction question posed first.
- Group 4: This version includes both predictions, with the other-prediction first.

A total of 57 students attended the class on January 25, 2016, all of whom filled out the voluntary survey. Of these students, 13 were in Group 1, 11 in Group 2, 15 in Group 3, and 18 in Group 4.

3.5.2 Results

The results of the classroom experiment confirm that the students are significantly more aware of their classmates' procrastination than of their own. While expectations for self are quite overconfident, expectations for others are, on average, correct.

I begin by assessing the differences in the students' predictions about themselves and their classmates graphically. Panel 1 of Figure 3.9 displays: (i) the distribution of answers among the students predicting for themselves (in dark blue); (ii) the distribution of the students' predictions about their classmates (in light blue); and (iii) the distribution of the actual completion dates (in grey). Only 37% of students making self-predictions expect their completion dates to fall within one week of the deadline (March 27 - April 2, 2016), but a substantially larger proportion (68%) of students expect that others' completion dates would fall, on average, in the last week.

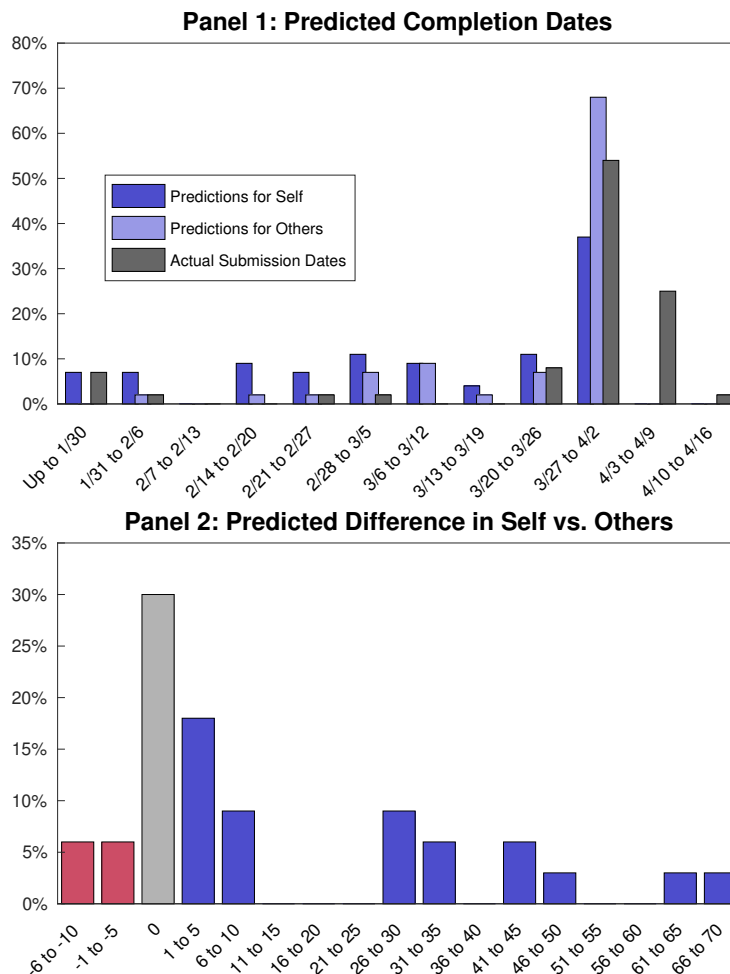


Figure 3.9: Results from the classroom experiment.

In order to compare the average predictions about self and others, I code predicted completion dates as the number of days before the April 2, 2016 deadline. Thus, for example, a student that predicts that she will email the instructor on March 25, 2016 is coded as making a self-prediction 8 days before the deadline. The average predictions for self and others across survey versions are presented in Panel 1 of Table 3.7. The average (median) predicted completion date for self is 22.28 days (15.5 days) before the deadline, while the average (median) prediction for others is only 9.07 days (1 day) before the deadline.

Table 3.7: Comparison of students' predictions of their own and their classmates' completion dates.

Panel 1: Summary Statistics

		Overall	Group 1	Group 2	Group 3	Group 4
Self-prediction	Mean	22.28	22.31	–	21.00	23.33
	SE	(3.27)	(6.58)	–	(5.90)	(6.01)
	Median	15.5	18	–	8	15.5
	# Obs	46	13	–	15	18
Other-prediction	Mean	9.07	–	7.91	9.13	9.72
	SE	(2.15)	–	(3.39)	(3.57)	(5.06)
	Median	1	–	1	3	1
	# Obs	44	–	11	15	18

Panel 2: Differences in Predictions for Self vs. Others

	Overall	Group 1 & Group 2	Group 3	Group 4
Difference	13.21	14.40	11.87	13.61
SE	(3.24)	(7.42)	(6.15)	(4.29)

Panel 3: Diff in Predictions from Actual Completion Dates

	Self-prediction	Other-prediction
Difference	14.82	1.61
SE	(3.94)	(3.36)

To more precisely evaluate the difference between the students' self- and other- predictions, I estimate the following specification:

$$\#DaysBeforeDeadline_i = \alpha + \gamma SelfDummy_i + \epsilon_i, \quad (3.9)$$

where the response variable $\#DaysBeforeDeadline_i$ denotes the number of days between the prediction i and the deadline (April 2, 2016), and $SelfDummy_i$ is a dummy variable equal to one if prediction i is made about self. In samples including students from Groups 3 and 4, standard errors are clustered by student. The estimates of the difference (coefficient γ) are reported in Panel 2 of Table 3.7.

The average difference between the students' predictions of their own and others' completion dates is 13.21 days. This result is significant at the 1% level when using larger samples (Group 4 or the combined Overall sample), and at the 10% level in smaller samples (Groups 1 & 2 combined, as well as Group 3).

Just as in the online laboratory experiment, the predictions are independent of either the set of questions asked or the order in which they are asked. In particular, posing the two questions side by side to the same students (responses from Groups 3 and 4) yields the same results as asking different groups of students to make the two sets of predictions (responses from Groups 1 and 2). Similarly, varying the order in which the students in Groups 3 and 4 see the two questions does not materially affect the results.

For the students who make both sets of predictions, I observe the distribution of the individual-level differences in predicted dates. This distribution, presented in Panel 2 of Figure 1, displays the incidence of individual students being more optimistic about themselves, being more optimistic about others, or holding identical beliefs about themselves and others. While a large portion of students (30%) make the exact same prediction for themselves as for their average classmates, the clear majority (58%) expect others to email the instructor later than they will. Only 12% of the respondents expect themselves to email the instructor later than their classmates. Thus, the individual-level results confirm the patterns of the pooled analysis: the students expect themselves to display, on average, less present bias (i.e., have earlier completion dates) than their peers.

Which set of predictions is more correct? A cursory examination of the distributions in Panel 1 of Figure 3.9 indicates that the more critical expectations about others (in light blue) more closely match the real distribution of completion date (in grey). In fact, a sizable

proportion of the students (26%) email the instructor for approval a few days *after* the April 2, 2016 deadline.

The notion that beliefs about others more closely match real completion dates is confirmed by a statistical comparison of the average predictions from students making self- and other- predictions against the average actual completion dates, presented in Panel 3 of Table 3.7. The students' predictions of their own completion dates are, on average, a full two weeks off from actual completion dates. The difference between self-predictions and actual completion dates is statistically significant at the 1% level. Predictions about others, however, are remarkably spot-on. The average difference between the students' predictions of when their classmates would email the instructor and actual completion dates is economically negligible (1.61 days) and statistically indistinguishable from zero.

Altogether, the results of the classroom experiment illustrate that the wedge in beliefs documented in my laboratory experiment is operative in a real-world setting: the classroom. Among college students, there is a strong tendency to anticipate last-minute assignment completion from their classmates, but not from themselves. A more detailed discussion of applications of the documented wedge in beliefs is presented in the next section.

3.6 Applications

In this section, I offer examples of settings featuring interactions among present-biased individuals, where beliefs regarding others' present bias could affect equilibrium outcomes. I split these illustrative scenarios into three broad categories: competitive environments, collaborative group decision making, and principal-agent problems.

3.6.1 Competitive Environments

Beliefs regarding peers are instrumental for behavior in competitive settings. Under competitive incentive structures such as tournaments,¹⁴ optimal level of expended effort depends

¹⁴See Lazear and Rosen (1981) and Green and Stokey (1983), among others.

on one's expectations regarding the behavior of others.

Ex ante participation. Relatively more optimistic beliefs regarding one's own versus others' present bias makes entering competitive incentive schemes appear more attractive. For example, consider an employee faced with a tournament incentive contract. The wedge in beliefs makes the employee overestimate his future effort relative to his peers, inflating perceived chances of receiving the prize. This relaxes his participation constraint, potentially resulting in acceptance of contracts with negative net present value, and can lead to exploitative situations. O'Donoghue and Rabin (1999b) allude to the potential for firms to exploit naïve workers with individual performance-based pay. The wedge in beliefs regarding self versus others creates scope for further exploitation through relative performance contracts.

Ex post efficacy. However, asymmetric naïveté can also reduce the ex-post efficacy of tournaments as incentive devices. For example, consider an employee receiving compensation based on his relative performance. Due to asymmetric naïveté, he expects his peers to procrastinate, but holds overconfident beliefs regarding his future self. As a result, the employee can underestimate the cost of delaying the work, and expend less effort initially. Similar logic applies in contexts such as queueing, trading in financial markets in response to new information, or otherwise expending effort to secure a limited resource in a competitive marketplace. Overall, the wedge in beliefs regarding own vs. others' present bias can exacerbate procrastination in competitive environments.

3.6.2 Group Decision Making

Beliefs regarding others play a role not only in competitive situations, but also in collaborative settings. Below, I outline a few examples of joint decision making where each party's expectations regarding the other's present bias can influence equilibrium behavior.

Household consumption. A number of studies document that individuals display present bias in their day-to-day household decisions including credit card usage (Meier and Sprenger (2010)), saving for retirement (see Laibson *et al.* (1998); Choi *et al.* (2011)),

and exercise (DellaVigna and Malmendier (2006)). Asymmetrically naïve individuals are likely to make more sophisticated decisions on behalf of their partners or children than on their own behalf. This predicts higher willingness to enroll in commitment devices such as savings accounts when the decision is made jointly with a spouse versus by oneself, as well as potential gains from joint time management. The wedge in beliefs can also increase efficiency gains from group commitment devices such as weight loss programs and group-lending in microcredit markets.

Teamwork. Understanding of others' present bias can also serve as a valuable disciplining and commitment device for teams in the workplace (Gans and Landry (2016)). Fedyk (2015) highlights one channel through which team assignments can improve performance in the face of asymmetric naïveté. When a naïve present biased individual is assigned a task to be completed within a predetermined amount of time, her overoptimistic beliefs regarding her future present bias cause her to overestimate the option value of postponing the task until later. By contrast, awareness of each other's present bias makes each member of a team more willing to commit to completing the work early on, so as not to leave scope for her teammate's future procrastination. More generally, awareness of other's present bias makes commitment devices such as preplanned meetings, milestones, and deadlines more attractive to teams than to standalone workers.

Collaboration and delegation. However, asymmetric naïveté can also have detrimental effects on performance. In arenas where teamwork is not mandated, the wedge in beliefs regarding own versus others' present bias makes teamwork appear (erroneously) less appealing. This can lead to inefficiently low levels of collaboration among peers. Similarly, in existing teams, asymmetric naïveté can create the temptation for an overconfident individual to take on too much, displaying suboptimally low levels of delegation of tasks across teammates.

3.6.3 Principal-Agent Settings

Most principal-agent models featuring present bias assume a rational and omniscient principal interacting with present-biased agent(s).¹⁵ However, the extent to which firms, employers, and governments can correctly assess consumers', employees', and individuals' present bias is an empirical question. The wedge in beliefs documented in the present paper gives some credence to the standard theoretical assumption of omniscient principals. Below I summarize the implications in the classroom and in the workplace.

Classroom. A common feature of classroom instruction is deadlines, including homework assignments, in-class presentations, and intermediate exams. Without these paternalistic mechanisms, a student's naïveté regarding his own present bias would prompt him to set overly flexible deadlines (Ariely and Wertenbroch (2002)). Yet a teacher's ability to assign work in a way that maximizes her students' effort and performance hinges on her understanding of the students' present bias. The present paper indicates that naïveté regarding present bias is asymmetric. In the teacher-student setting, this implies that teachers hold more critical beliefs regarding their students' present bias, and hence impose more effective deadlines.

Workplace management. Analogous to the paternalistic commitment devices offered by teachers in the classroom, managers serve a similar function in organizations. In a field experiment conducted at a Colombian bank, Cadena *et al.* (2011) show that greater paternalistic incentives, such as goal reminders and managerial monitoring, lead to not only superior on-the-job performance and earnings, but also higher ex post employee satisfaction and lower stress levels. As Laibson (2018) highlights, individuals do not explicitly seek out more paternalistic and restrictive workplaces (consistent with naïveté regarding own present bias), but work environments invariably provide restrictions such as intermediate deadlines and monitored attendance (consistent with general awareness of others' present bias). The

¹⁵See, for example, DellaVigna and Malmendier (2004) and O'Donoghue and Rabin (2006). Similarly, Heidhues and Köszegi (2010) assume that firms can directly observe either consumers' β or $\hat{\beta}$ parameters, as well as the structural relationship between the two. O'Donoghue and Rabin (1999b) assume that the principal knows the agent's present-bias parameter β and model the principal's uncertainty regarding the agent's idiosyncratic completion costs.

wedge in beliefs creates the scope for such private paternalism: through their more critical awareness of others' present bias, managers serve as more effective organizational devices than employees' own planning.

Altogether, beliefs regarding others' present bias can inform our understanding of equilibrium outcomes across competitive, collaborative, and hierarchical environments. The examples above discuss several potential applications, but the list is by no means exhaustive, since present bias has been shown to be operative in a variety of domains, and strategic interactions feature in multiple types of economic decision-making.

3.7 Conclusion

This paper investigates whether individuals are aware of present bias in others. Both the online laboratory experiment and the field survey in the classroom reveal a wedge in beliefs: individuals are fairly naïve about their own present bias, but anticipate present bias in others. This finding is robust to incentivizing the predictions with monetary payments, and to asking the two sets of predictions – about self and about others – to the same experimental participants versus separately to different groups of participants. The wedge in beliefs explored in this paper opens two avenues for future work.

First, further investigation of the documented wedge in beliefs can shed light on the mechanisms of belief formation. My findings indicate that present bias is subject to relative overconfidence akin to that documented in several other domains.¹⁶ In addition, my results support the notion of bias blind spots documented in the social psychology literature: that individuals are, in general, more perceptive of others' biases than of their own.¹⁷ Probing further into how beliefs regarding one's own and others' present bias evolve depending on the setting, task, or experience can help shed light on the extent to which the documented

¹⁶For example, Svenson (1981) documents overconfidence regarding driving skills, while Weinstein (1980) finds overconfidence and overoptimism about a host of potential life events. Alicke (1985) documents that participants deem positive (negative) adjectives to be more (less) characteristic of themselves than of their average peer.

¹⁷See, for example: Pronin *et al.* (2002), Ehrlinger *et al.* (2005), and West *et al.* (2012).

wedge in beliefs reflects motivated thinking, blindspots, or other frictions.

Second, the documented wedge in beliefs lays the foundations for understanding interactions between present-biased individuals in the workplace, in the classroom, in households, and in markets. Differential awareness of one's own and others' present bias is likely to impact how groups of present-biased individuals schedule their joint work, seek external commitment devices, or evaluate their own and their peers' performance. Investigating these effects, both theoretically and empirically, could constitute a fruitful avenue for future research.

References

- ACLAND, D. and LEVY, M. R. (2015). Naïveté, projection bias, and habit formation in gym attendance. *Management Science*, **61** (1), 146–160.
- AHERN, K. R. and SOSYURA, D. (2014). Who writes the news? Corporate press releases during merger negotiations. *Journal of Finance*, **69** (1), 241–291.
- AJINKYA, B. B., ATIASE, R. K., and GIFT, M. J. (1991). Volume of trading and the dispersion in financial analysts' earnings forecasts. *Accounting Review*, **66** (2), 389–401.
- ALICKE, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, **49** (6), 1621–1630.
- ANDERSEN, S., HARRISON, G. W., LAU, M. I. and RUTSTROM, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, **76** (3), 583–618.
- ANDERSON, E. W., GHYSELS, E. and JUERGENS, J. L. (2005). Do heterogeneous beliefs matter for asset pricing? *Review of Financial Studies*, **18** (3), 875–924.
- ANDREI, D. and CUJEAN, J. (2017). Information percolation, momentum and reversal. *Journal of Financial Economics*, **123** (3), 617–645.
- and HASLER, M. (2014). Investor attention and stock market volatility. *Review of Financial Studies*, **28** (1), 33–72.
- ANTWEILER, W. and FRANK, M. Z. (2006). Do us stock markets typically overreact to corporate news stories?, Working paper.
- ARIELY, D. and LOEWENSTEIN, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, **19** (2), 87–98.
- and WERTENBROCH, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, **13** (3), 219–224.
- ASNESS, C. S., MOSKOWITZ, T. J. and PEDERSEN, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, **68** (3), 929–985.
- ATIASE, R. K. and BAMBER, L. S. (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.

- ATKINS, A. B. and DYL, E. A. (1990). Price reversals, bid-ask spreads, and market efficiency. *Journal of Financial and Quantitative Analysis*, **28** (4), 535–547.
- AUGENBLICK, N., NIEDERLE, M. and SPRENGER, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, **130** (3), 1067–1115.
- and RABIN, M. (2018). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, **forthcoming**.
- BACKUS, D., CHERNOV, M. and ZIN, S. (2014). Sources of entropy in representative agent models. *Journal of Finance*, **69** (1), 51–99.
- BALI, T., BODNARUK, A., SCHERBINA, A. and TANG, Y. (2017). Unusual news flow and the cross-section of stock returns. *Management Science*, **forthcoming**.
- BALL, R. and BROWN, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, **6** (2), 159–178.
- BAMBER, L. S., BARRON, O. E. and STEVENS, D. E. (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.
- BANERJEE, S. and KREMER, I. (2010). Disagreement and learning: Dynamic patterns of trade. *Journal of Finance*, **65** (4), 1269–1302.
- BARBER, B. M. and LOEFFLER, D. (1993). The “Dartboard” column: Second-hand information and price pressure. *Journal of Financial and Quantitative Analysis*, **28** (2), 273–284.
- and ODEAN, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, **21** (2), 785–818.
- BARBERIS, N., SHLEIFER, A. and VISHNY, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, **49** (3), 307–343.
- BAUGUESS, S. W., COONEY, J. and HANLEY, K. W. (2013). Investor demand for information in newly issued securities, Working paper.
- BEN-REPHAEL, A., DA, Z. and ISRAELSEN, R. D. (2017). It depends on where you search: A comparison of institutional and retail attention. *Review of Financial Studies*, **30** (9), 3009–3047.
- BERNARD, V. L. and THOMAS, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research*, **27** (Supplement), 1–36.
- BISIN, A. and HYNDMAN, K. (2014). Present-bias, procrastination and deadlines in a field experiment, NBER Working paper No. 19874.
- BLANKESPOOR, E., DEHAAN, E. and ZHU, C. (2017). Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. *Review of Accounting Studies*, **forthcoming**.

- BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, **3** (1), 993–1022.
- BOLLEN, J., MAO, H. and ZENG, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, **2** (1), 1–8.
- BOULLAND, R., DEGEORGE, F. and GINGLINGER, E. (2017). News dissemination and investor attention. *Review of Finance*, **21** (2), 1–31.
- BREIMAN, L. (2001). Random forests. *Machine Learning*, **45** (1), 5–32.
- BUSSE, J. A. and GREEN, T. C. (2002). Market efficiency in real time. *Journal of Financial Economics*, **65** (3), 415–437.
- CADENA, X., SCHOAR, A., CRISTEA, A. and DELGADO-MEDRANO, H. (2011). Fighting procrastination in the workplace: An experiment, NBER Working paper No. 16944.
- CAO, H. H., COVAL, J. D. and HIRSHLEIFER, D. (2002). Sidelined investors, trading generated news, and security returns. *Review of Financial Studies*, **15** (2), 615–648.
- and OU-YANG, H. (2009). Differences of opinion of public information and speculative trading in stocks and options. *Review of Financial Studies*, **22** (1), 299–335.
- CARLIN, B. I., LONGSTAFF, F. A. and MATOBA, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, **114** (2), 226–238.
- CARVALHO, C., KLAGGE, N. and MOENCH, E. (2011). The persistent effects of a false news shock. *Journal of Empirical Finance*, **18** (4), 597–615.
- CHAN, W. (2003). Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics*, **70** (2), 223–260.
- CHANG, Y. C., HONG, H., TIEDENS, L., WANG, N. and ZHAO, B. (2015). Does diversity lead to diverse opinions? Evidence from languages and stock markets, Working paper.
- CHOI, J., LAIBSON, D. and MADRIAN, B. (2011). \$100 bills on the sidewalk: Suboptimal investment in 401(k) plans. *Review of Economics and Statistics*, **93** (3), 748–763.
- CHORDIA, T., ROLL, R. and SUBRAHMANYAM, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, **65** (1), 111–130.
- CONLIN, M., O'DONOGHUE, T. and VOGELSANG, T. J. (2007). Projection bias in catalog orders. *American Economic Review*, **97** (4), 1217–1249.
- COOKSON, J. A. and NIESSNER, M. (2016). Why don't we agree? Evidence from a social network of investors, Working paper.
- CORTES, C. and VAPNIK, V. (1995). Support-vector networks. *Machine Learning*, **20** (3), 273–297.
- CORWIN, S. A. and COUGHENOUR, J. F. (2008). Limited attention and the allocation of effort in securities trading. *Journal of Finance*, **63** (6), 3031–3067.

- CRONQVIST, H., SIEGEL, S. and YU, F. (2015). Value versus growth investing: Why do different investors have different styles? *Journal of Financial Economics*, **117** (2), 333–349.
- CURTIS, A., RICHARDSON, V. and SCHMARDEBECK, R. (2014). Investor attention and the pricing of earnings news, Working paper.
- DA, Z., ENGELBERG, J. and GAO, P. (2011). In search of attention. *Journal of Finance*, **66** (5), 1461–1499.
- , — and — (2015). The sum of all fears investor sentiment and asset prices. *Review of Financial Studies*, **28** (1), 1–32.
- DANIEL, K., GRINBLATT, M., TITMAN, S. and WERMERS, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance*, **52** (3), 1035–1058.
- DANZ, D., MADARÁSZ, K., and WANG, S. W. (2015). Anticipating information projection: An experimental investigation, Working paper.
- DAS, S. R. and CHEN, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, **53** (9), 1375–1388.
- DEHAAN, E., SHEVLIN, T. and THORNOCK, J. (2015). Market (in)attention and the strategic scheduling and timing of earnings announcements. *Journal of Accounting and Economics*, **60** (1), 36–55.
- DELLAVIGNA, S. and MALMENDIER, U. (2004). Contract design and self-control: Theory and evidence. *Quarterly Journal of Economics*, **119** (2), 353–402.
- and — (2006). Paying not to go to the gym. *American Economic Review*, **96** (3), 694–719.
- and PASERMAN, M. D. (2005). Job search and impatience. *Journal of Labor Economics*, **23** (3), 527–588.
- and POLLET, J. M. (2009). Investor inattention and friday earnings announcements. *Journal of Finance*, **64** (2), 709–749.
- DIETHER, K. B., MALLOY, C. J. and SCHERBINA, A. (2002). Differences of opinion and the cross section of stock returns. *Journal of Finance*, **57** (2), 2113–2141.
- DOUGAL, C., ENGELBERG, J., GARCÍA, D. and PARSONS, C. A. (2012). Journalists and the stock market. *Review of Financial Studies*, **25** (3), 639–679.
- DRAKE, M. S., ROULSTONE, D. T. and THORNOCK, J. R. (2012). Investor information demand: Evidence from google searches around earnings announcements investor information demand. *Journal of Accounting Research*, **50** (4), 1001–1040.
- , — and — (2015). The determinants and consequences of information acquisition via edgar. *Contemporary Accounting Research*, **32** (3), 1128–1161.
- , — and — (2016). The usefulness of historical accounting reports. *Journal of Accounting Research*, **61** (2), 448–464.

- DUGAST, J. and FOUCAULT, T. (2017). Data abundance and asset price informativeness, Working paper.
- EDERINGTON, L. H. and LEE, J. H. (1995). The short-run dynamics of the price adjustment to new information. *Journal of Financial and Quantitative Analysis*, **30** (1), 117–134.
- EHRLINGER, J., GILOVICH, T. and ROSS, L. (2005). Peering into the bias blind spot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin*, **31** (5), 680–692.
- ENGELBERG, J., SASSEVILLE, C. and WILLIAMS, J. (2012). Market madness? The case of mad money. *Management Science*, **58** (2), 351–364.
- ENGELBERG, J. E. (2008). Costly information processing: Evidence from earnings announcements, Working paper.
- and PARSONS, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance*, **66** (1), 67–97.
- FAHN, M. and HAKENES, H. (2014). Teamwork as a self-disciplining device, Working paper.
- FAMA, E. F. and FRENCH, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, **105** (3), 457–472.
- FANG, L. and PERESS, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance*, **64** (5), 2023–2052.
- FARBOODI, M., MATRAY, A. and VELDKAMP, L. (2017). Where has all the big data gone?, Working paper.
- FEDYK, A. (2015). Overcoming overconfidence: Teamwork and self-control, Working paper.
- and HODSON, J. (2015). When can the market identify stale news?, Working paper.
- FEENBERG, D., GANGULI, I., GAULE, P. and GRUBER, J. (2017). It's good to be first: Order bias in reading and citing nber working papers. *Review of Economics and Statistics*, **99** (1), 32–39.
- FEHLE, F., TSYPLAKOV, S. and ZDOROVTSOV, V. (2005). Can companies influence investor behaviour through advertising? Super bowl commercials and stock returns. *European Financial Management*, **11** (5), 625–647.
- FOSTER, G., OLSEN, C. and SHEVLIN, T. (1984). Earnings releases, anomalies, and the behavior of security returns. *Accounting Review*, **59** (4), 574–603.
- FOUCAULT, T., HOMBERT, J. and ROŞU, I. (2016). News trading and speed. *Journal of Finance*, **71** (1), 335–382.
- FRAZZINI, A. (2006). The disposition effect and underreaction to news. *Journal of Finance*, **61** (4), 2017–2046.
- FREY, B. J. and DUECK, D. (2007). Clustering by passing messages between data points. *Science*, **315** (5814), 972–976.

- FUNG, A. K. W., MOK, D. M. Y. and LAM, K. (2000). Intraday price reversals for index futures in the us and hong kong. *Journal of Banking & Finance*, **24** (7), 1179–1201.
- GANS, J. S. and LANDRY, P. (2016). Procrastination in teams, NEBR Working paper No. 21891.
- GARCÍA, D. (2013). Sentiment during recessions. *Journal of Finance*, **68** (3), 1267–1300.
- GARFINKEL, J. A. (2009). Measuring investors' opinion divergence. *Journal of Accounting Research*, **47** (5), 1317–1348.
- and SOKOBIN, J. (2006). Volume, opinion divergence, and returns: A study of post-earnings announcement drift. *Journal of Accounting Research*, **44** (1), 85–112.
- GIANNINI, R. C., IRVINE, P. J. and SHU, T. (2015). The convergence and divergence of investors' opinions around earnings news: Evidence from a social network, Working paper.
- GILBERT, T., KOGAN, S., LOCHSTOER, L. and OZYILDIRIM, A. (2012). Investor inattention and the market impact of summary statistics. *Management Science*, **58** (2), 336–350.
- GOLUB, B. and JACKSON, M. P. (2012). How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*, **127** (3), 1287–1338.
- GOTTLIEB, D. (2008). Competition over time-inconsistent consumers. *Journal of Public Economic Theory*, **10** (4), 673–684.
- GREEN, J. and STOKEY, N. (1983). A comparison of tournaments and contracts. *Journal of Political Economy*, **91** (3), 349–364.
- GRIFFITHS, T. L. and STEYVERS, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, **101** (Supplement), 5228–5235.
- GRINBLATT, M. and MOSKOWITZ, T. J. (2004). Predicting stock price movements from past returns: The role of consistency and tax-loss selling. *Journal of Financial Economics*, **71** (3), 541–579.
- GUL, F. and PESENDORFER, W. (2001). Temptation and self-control. *Econometrica*, **69** (6), 1403–1435.
- HARRIS, M. and RAVIV, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies*, **6** (3), 473–506.
- HEIDHUES, P. and KÓSZEGI, B. (2010). Exploiting naïveté about self-control in the credit market. *American Economic Review*, **100** (5), 2279–2303.
- HERWEG, F. and MÜLLER, D. (2011). Performance of procrastinators: On the value of deadlines. *Theory and Decision*, **70** (3), 329–366.
- HESTON, S. L., KORAJCZYK, R. A. and SADKA, R. (2010). Intraday patterns in the cross-section of stock returns. *Journal of Finance*, **65** (4), 1369–1407.
- HILLERT, A., JACOBS, H. and MÜLLER, S. (2014). Media makes momentum. *Review of Financial Studies*, **27** (12), 3467–3501.

- HIRSHLEIFER, D., LIM, S. S. and TEOH, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance*, **64** (5), 2289–2325.
- and TEOH, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, **36** (1), 337–386.
- HO, D. E. and IMAI, K. (2008). Estimating causal effects of ballot order from a randomized natural experiment. *Public Opinion Quarterly*, **72**, 216–240.
- HOFMANN, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 50–57.
- HOLDEN, C. W. and SUBRAHMANYAM, A. (1992). Long-lived private information and imperfect competition. *Journal of Finance*, **47** (1), 247–270.
- HONG, H., LIM, T. and STEIN, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, **55**, 265–295.
- and STEIN, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, **54** (6), 2143–2184.
- and — (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, **21** (2), 109–128.
- HUANG, X., NEKRASOV, A. and TEOH, S. H. (2017). Headline salience and over- and underreactions to earnings, Working paper.
- HUBERMAN, G. and REGEV, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance*, **56** (1), 387–396.
- IRIBERRI, N. and REY-BIEL, P. (2013). Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do? *Quantitative Economics*, **4** (3), 515–547.
- JACOBS, H. and HILLERT, A. (2015). Alphabetic bias, investor recognition, and trading behavior. *Review of Finance*, **20** (2), 693–723.
- KANDEL, E. and PEARSON, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, **103** (4), 831–872.
- KANIEL, R., LIU, S., SAAR, G. and TITMAN, S. (2012). Individual investor trading and return patterns around earnings announcements. *Journal of Political Finance*, **67** (2), 639–680.
- and PARHAM, R. (2017). WSJ Category Kings—The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, **123** (2), 337–356.
- KARPOFF, J. M. (1986). A theory of trading volume. *Journal of Finance*, **41** (5), 1069–1087.
- KAUFMANN, M. (2017). Projection bias in effort choices, Working paper.

- KEMPE, E., MANCONI, A. and SPALT, O. (2016). Distracted shareholders and corporate actions. *Review of Financial Studies*, **30** (5), 1660–1695.
- KLIBANOFF, P., LAMONT, O. and WIZMAN, T. A. (1998). Investor reaction to salient news in closed-end country funds. *Journal of Finance*, **53** (2), 674–699.
- KONDOR, P. (2012). The more we know about the fundamental, the less we agree on the price. *Review of Financial Studies*, **79** (3), 1175–1207.
- KOUDIJS, P. (2016). The boats that did not sail: Asset price volatility in a natural experiment. *Journal of Finance*, **71** (3), 1185–11226.
- KYLE, P. (1985). Continuous auctions and insider trading. *Econometrica*, **53** (6), 1315–1335.
- LAIbson, D. (1997). Hyperbolic discounting and golden eggs. *Quarterly Journal of Economics*, **112** (2), 443–477.
- (2018). Private paternalism, the commitment puzzle, and model-free equilibrium. *American Economic Review: Papers & Proceedings*.
- , REPETTO, A. and TOBACMAN, J. (1998). Self-control and saving for retirement. *Brookings Papers on Economic Activity*, **1**, 91–196.
- , — and — (2008). Estimating discount functions with consumption choices over the lifecycle, NBER Working paper No. 13314.
- LAWRENCE, A., RYANS, J., SUN, E., and LAPTEV, N. (2017). Earnings announcement promotions: A Yahoo Finance field experiment, Working paper.
- LAZEAR, E. and ROSEN, S. (1981). Rank order tournaments as optimum labor contracts. *Journal of Political Economy*, **89** (5), 841–864.
- LEHAVY, R., LI, F. and MERKLEY, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review*, **86** (3), 1087–1115.
- LEWIS, D. D., YANG, Y., ROSE, T. G. and LI, F. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, **5** (4), 361–397.
- LI, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, **45** (2), 221–247.
- LIU, L., SHERMAN, A. and ZHANG, Y. (2014). The long-run role of the media: Evidence from initial public offerings. *Management Science*, **60** (8), 1945–1964.
- LOEWENSTEIN, G. and ADLER, D. (1995). A bias in the prediction of tastes. *Economic Journal*, **105** (431), 929–937.
- , MOORE, D. and WEBER, R. (2006). Misperceiving the value of information in predicting the performance of others. *Experimental Economics*, **9** (3), 281–295.

- , O'DONOGHUE, T. and RABIN, M. (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics*, **118** (4), 1209–1248.
- LOH, R. (2010). Investor inattention and the underreaction to stock recommendations. *Financial Management*, **39** (3), 1223–1252.
- LOUGHRAN, T. and McDONALD, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance*, **66** (1), 35–65.
- and — (2014). Measuring readability in financial disclosures. *Journal of Finance*, **69** (4), 1643–1671.
- LUDWIG, S. and NAFZIGER, J. (2011). Beliefs about overconfidence. *Theory and Decision*, **70** (4), 475–500.
- LUMSDAINE, R. L. (2010). What the market watched: Bloomberg news stories and bank returns as the financial crisis unfolded, Working paper.
- MADARASZ, K. (2012). Information projection: Model and applications. *Review of Economic Studies*, **79** (3), 961–985.
- MADSEN, J. M. and NIESSNER, M. (2016). Is investor attention for sale? The role of advertising in financial markets, Working paper.
- MARSHALL, B. R., VISALTANACHOTI, N. and COOPER, G. (2014). Sell the rumour, buy the fact? *Accounting & Finance*, **54** (1), 237–249.
- MCCLURE, S., LAIBSON, D., LOEWENSTEIN, G. and COHEN, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, **306** (5695), 503–507.
- MCPHERSON, M., SMITH-LOVIN, L. and COOK, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, **27** (1), 415–444.
- MEIER, S. and SPRENGER, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, **2** (1), 193–210.
- MILLER, B. P. (2010). The effects of reporting complexity on small and large investor trading. *Accounting Review*, **85** (6), 2107–2143.
- MOSKOWITZ, T. J., OOI, Y. H. and PEDERSEN, L. H. (2012). Time series momentum. *Journal of Financial Economics*, **104** (2), 228–250.
- NIESSNER, M. (2015). Strategic disclosure timing and insider trading, Working paper.
- O'DONOGHUE, T. and RABIN, M. (1999a). Doing it now or later. *American Economic Reviews*, **89** (1), 103–124.
- and — (1999b). Incentives for procrastinators. *Quarterly Journal of Economics*, **114** (3), 769–816.
- and — (2006). Optimal sin taxes. *Journal of Public Economics*, **90** (10-11), 1825–1849.

- PENG, L. and XIONG, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, **80** (3), 563–602.
- PERESS, J. (2008). Media coverage and investors' attention to earnings announcements, Working paper.
- (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *Journal of Finance*, **69** (5), 2007–2043.
- PHILIPPATOS, G. and GRESSIS, N. (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.
- PORTER, M. F. (1980). An algorithm for suffix stripping. *Program*, **14** (3), 130–137.
- PRITCHARD, K. J., STEPHENS, M. and DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155** (2), 945–959.
- PRONIN, E., LIN, D. Y. and ROSS, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, **28** (3), 369–381.
- RAO, J. and SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, **76** (374), 221–230.
- READ, D. and VAN LEEUWEN, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational Behavior and Human Decision Processes*, **76** (2), 189–205.
- ROGERS, J. L., SKINNER, D. J. and ZECHMAN, S. L. C. (2013). The role of the media in disseminating insider-trading news. *Review of Accounting Studies*, **21** (3), 711–739.
- ROUWENHORST, K. G. (1998). International momentum strategies. *Journal of Finance*, **53** (1), 267–284.
- SCHMIDT, D. (2013). Investors' attention and stock covariation, Working paper.
- SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 656–715.
- SHUI, H. and AUSUBEL, L. (2005). Time inconsistency in the credit card market, Working paper.
- SKIBA, P. and TOBACMAN, J. (2009). Payday loans, uncertainty and discounting: Explaining patterns of borrowing, repayment, and default, NBER Working paper No. 14659.
- SOLNICK, J., KANNENBERG, C. H., ECKERMAN, D. A., and WALLER, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation*, **11** (1), 61–77.
- SOLOMON, D. H. (2012). Selective publicity and stock prices. *Journal of Finance*, **67** (2), 599–638.

- SVENSON, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, **47** (2), 143–148.
- TANAKA, T., CAMERER, C. F. and NGUYEN, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*, **101** (1), 557–571.
- TETLOCK, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, **62** (3), 1139–1168.
- (2011). All the news that’s fit to reprint: Do investors react to stale information? *Review of Financial Studies*, **24** (5), 1481–1512.
- , SAAR-TSECHANSKY, M. and MACSKASSY, S. (2008). More than words: Quantifying language to measure firms’ fundamentals. *Journal of Finance*, **63** (3), 1437–1467.
- TVERSKY, A. and KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185** (4157), 1124–1131.
- UHL, M. W. (2014). Reuters sentiment and stock returns. *Journal of Behavioral Finance*, **15** (4), 287–298.
- UMAR, T. (2017). Attention grabbers when seeking alpha, Working paper.
- VAN BOVEN, L. and LOEWENSTEIN, G. (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology*, **79** (1), 66–76.
- and — (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, **29** (9), 1159–1168.
- VAN DER MAATEN, L. and HINTON, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, **9**, 2579–2605.
- VON BESCHWITZ, B., KEIM, D. and MASSA, K. (2015). First to “read” the news: News analytics and high frequency trading, Working paper.
- WANG., B. (2017). Ranking and salience, Working paper.
- WANG, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy*, **102** (1), 127–168.
- WEINSTEIN, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, **39** (5), 806–820.
- WEST, R. F., MESERVE, R. J. and STANOVICH, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, **103** (3), 506–519.
- YOU, H. and ZHANG, X. (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting Studies*, **14** (4), 559–586.

YUAN, Y. (2015). Market-wide attention, trading, and stock returns. *Journal of Financial Economics*, **116** (3), 548–564.

ZAWADOWSKI, A. G., ANDOR, G. and KERTÉSZ, J. (2006). Short-term market reaction after extreme price changes of liquid stocks. *Quantitative Finance*, **6** (4), 283–295.

Appendix A

Appendix to Chapter 1

A.1 Technical Details

A.1.1 Latent Dirichlet Allocation

I briefly present the Latent Dirichlet Allocation methodology for identifying representative topics covered by the financial news in the training corpus. For additional details on the methodology, please refer to Blei et al (2003).

Let D denote the set of financial news documents in the training corpus, with $d \in D$ representing an individual document. Each document d is a sequence of N words: $d = (w_1, \dots, w_N)$, where w_n is the n th term to appear in the document d . All terms come from the vocabulary W , which is constructed as described in Appendix A.2.

The latent set of topics is denoted by T , where each element $t \in T$ is a unit vector in k -dimensional space. The parameter k is the desired number of topics, specified by the researcher.

The Latent Dirichlet Allocation algorithm conceptualizes each document d as a sequence of words drawn from a latent distribution \mathcal{D}_d over topics. The distribution \mathcal{D}_d is itself randomly determined for each document: in particular, for each document d , \mathcal{D}_d is a multinomial distribution whose parameters are a random variable drawn from a pre-specified Dirichlet prior.

Specifically, the generative process assumed by the Latent Dirichlet Allocation algorithm is as follows.

- Pre-specify model parameters: ζ, α, β .
- To construct each new document d :
 1. Choose the document length $N_d \sim \text{Poisson}(\zeta)$.
 2. Choose a distribution over topics $\theta_d \in \text{Dir}(\alpha)$.
 3. Fill the N words in the document d by sequentially choosing each word w_n as follows:
 - (a) Choose a topic $t_n \sim \text{Multinomial}(\theta_d)$.
 - (b) Choose a word w_n from $\mathbb{P}\{w_n|t_n, \beta\}$, the conditional probability distribution over words in the vocabulary given the chosen topic t_n .

The model relies on three parameters: ζ, α , and β . The parameter ζ is chosen to best match the set of document lengths in the corpus, assuming that the lengths are drawn from a Poisson distribution. This parameter is independent of the rest of the process, and therefore I forego it in the remainder of the discussion.

The key model parameters of interest are α and β : α is a k -dimensional vector that governs the relative frequencies of the k topics, and β is a k -by-2,000 matrix that specifies the likelihood of each word in the vocabulary conditional on each of the k topics. Thus, the element in the i th row and j th column of β is $\beta_{i,j} = \mathbb{P}\{w_j = 1|t_i = 1\}$.

In theory, the parameters α and β are estimated to maximize the likelihood of observing the actual corpus of documents D . The conditional probability of observing a document d given the model parameters α and β is given by:

$$\mathbb{P}\{d|\alpha, \beta\} = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \int \left(\prod_{i=1}^k \theta_i^{\alpha_i - 1} \right) \times \left(\prod_{n=1}^N \sum_{i=1}^k \prod_{j=1}^V (\theta_i \beta_{i,j})^{w_n^j} \right) d\theta, \quad (\text{A.1})$$

where w_n^j denotes the j th component of the n th word vector w_n , and $\Gamma(\cdot)$ is the Gamma function.

Note that unlike other methods such as the probabilistic Latent Semantic Indexing approach (see Hofmann (1999)), the Latent Dirichlet Allocation method does not require the parameters to be estimated individually for each document; there is a single set of parameters ξ, α, β for the entire model. This offers two advantages highlighted by Blei et al (2003). First, by reducing the number of estimated parameters, the Latent Dirichlet Allocation approach reduces the computational complexity of the estimation problem. Second, and more importantly, the Latent Dirichlet Allocation method allows for the generation of any arbitrary document and facilitates the evaluation of the likelihood of out-of-sample documents. This ability to represent out-of-sample documents in terms of the identified topics is essential to the application in this paper.

A.1.2 Text Preprocessing

The Latent Dirichlet Allocation algorithm takes as its input a set of documents, each represented by a sequence of terms from a pre-specified vocabulary. Before applying the topic modeling methodology, I need to identify a relevant vocabulary to represent the financial news documents. I proceed in three steps.

First, in order to focus on the set of relevant terms, I begin by stripping out all “stop words.” To identify “stop words,” I use the list provided by the University of Glasgow Information Retrieval Group.¹

Second, I construct the vocabulary using not only single words appearing in the TRC2 news corpus, but also common pairs of words. The Latent Dirichlet Allocation method is a bag-of-words method, meaning that the algorithm ignores the ordering of terms within a document and treats each term as an independently drawn random variable. In theory, this may be a problematic assumption, particularly for financial news, where some concepts are captured by phrases, for example “traditional enterprise” or “stock exchange.” In order to account for this feature of the data, I augment the vocabulary of unigrams (single words) appearing in the corpus with bigrams (pairs of words).

¹The full list of stop words can be accessed at http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words.

Lastly, I limit my attention to the most common and representative terms. In particular, I focus on the terms that appear in at least two distinct documents and that appear in no more than 70% of the documents in the training corpus. Furthermore, the terms are ranked according to frequency in order to capture relative importance; the final vocabulary is comprised of the top 2,000 terms.

A.1.3 Topic Model Estimation

I estimate the model varying the number of iterations and the number of identified topics. The model's fit flattens out at around fifteen topics.

In practice, the expression in (A.1) is intractable, and hence parameter estimation relies on approximate inference methods. Following Griffiths and Steyvers (2004), I estimate the parameters from the training corpus of documents using a collapsed Gibbs sampling algorithm.

I vary the number of iterations of the sampling algorithm from 30 to 1,000, and find that the marginal improvement in the model's fit is largest up to approximately 250 iterations, and mostly flattens out after 500 iterations (see Panel 1 of Figure A.1, which plots the log likelihood as a function of the number of iterations for a model with $k = 15$ topics). The results in the paper come from the estimation algorithm with 500 iterations for all considered specifications.

The results from estimating the parameters of the Latent Dirichlet Allocation topic model for $k \in \{10, \dots, 40\}$ indicate that the model's fit is best at around $k = 15$ topics. Panel 2 of Figure A.1 plots the model fit for $k \in \{10, \dots, 40\}$. For each number of topics, the model is estimated using the collapsed Gibbs sampler with 500 iterations. The figure shows the final log likelihood for each specification. The model's fit improves somewhat as the number of topics increases from ten to fifteen, with an increase in log likelihood from -6.01×10^5 to -6.00×10^5 . Increasing the number of topics to 20, 25, or 30 does not offer marginal improvements over the $k = 15$ specification. Increasing the number of topics further to 35 or 40 markedly decreases the estimated log likelihood. Overall, the $k = 15$ specification

achieves the best fit after 500 iterations.

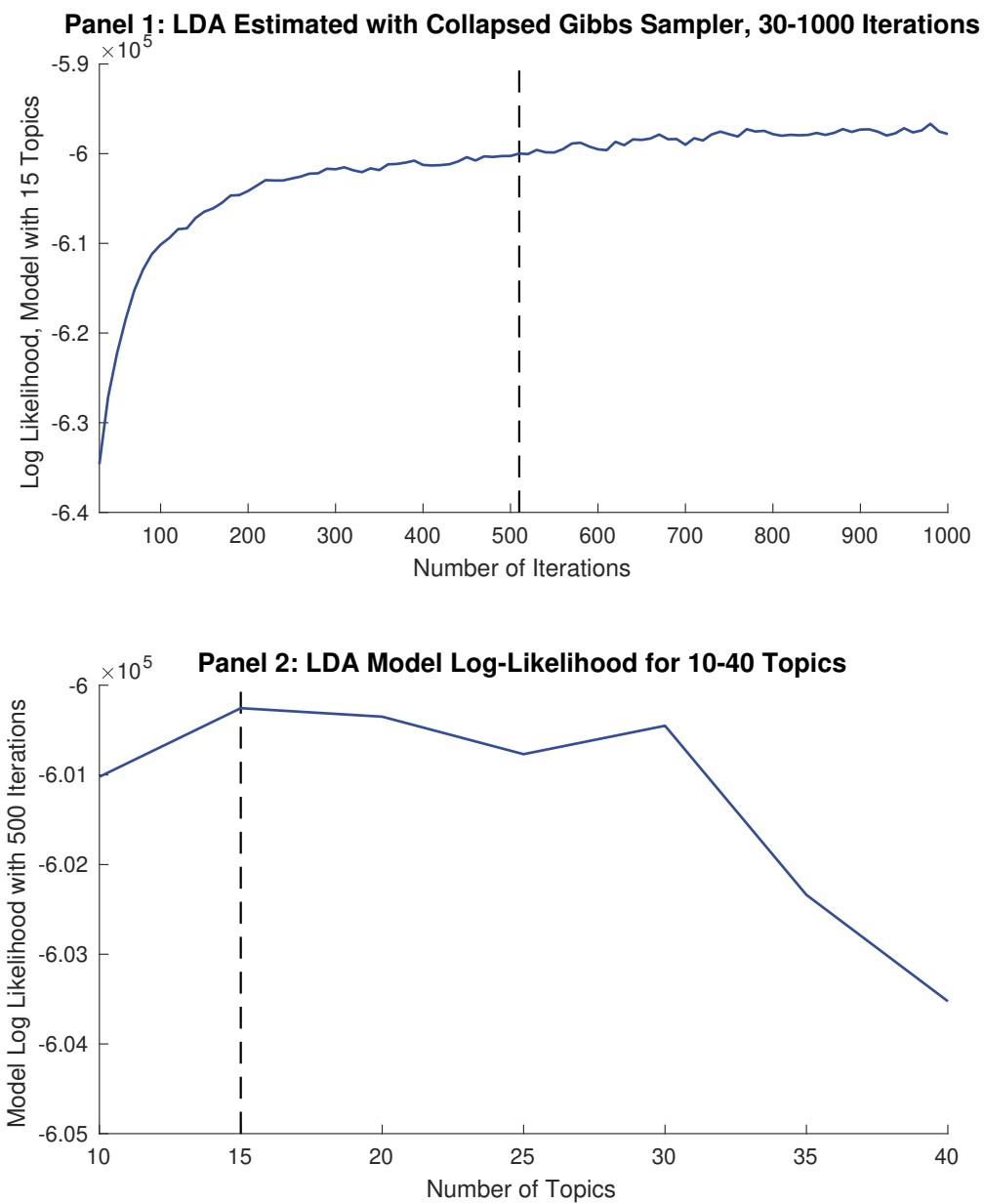


Figure A.1: Log likelihood for different estimations of the Latent Dirichlet Allocation model.

Appendix B

Appendix to Chapter 2

B.1 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.

B.1.1 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 :

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

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 B.1. 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.

Table B.1: Relationship between immediacy of attention and immediacy of price variance.

		Immediate price variance window (τ)		
		30 seconds	1 minute	2 minutes
Click window (τ)	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.

B.1.2 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 2.2, 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 *et al.* (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 *et al.* (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 $CAR_{i,[t+2,t+2]}$ over the earnings announcements (i, t) within the portfolio. The rows of Table B.2 display the relationship between SUE and the abnormal returns within a particular attention quintile.

Following Foster *et al.* (1984), statistical significance is determined by comparing the observed average CAR ($ACAR$) for each portfolio against an empirical distribution of

Table B.2: Cumulative abnormal returns two to twenty days following earnings announcements, by earnings surprise and attention speed.

		SUE quintile					
		1 (bottom)	2	3	4	5(top)	Diff (5-1)
ImmClicks quin.	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.

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 B.2, 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.

B.1.3 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 *et al.* (2012) and Asness *et al.* (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 *et al.* (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 *et al.* (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 $AbnRet_{i,[t_1,t_2]}$ denote firm i 's cumulative return over months t_1 to t_2 , 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.

Table B.3: Cross-sectional regression of firm-level momentum against proxies of attention.

	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.

Since momentum varies with firm size (see Hong *et al.* (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 B.3 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.

B.2 Technical Details

B.2.1 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.2.2 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.2.3 Reader Classification into Clusters: Random Forest

In this subsection, I briefly describe the random forest classification algorithm for classifying the remaining readers.¹ The reader-type categories are constructed using affinity propagation clustering on a subset of the sample. This reduces computational complexity 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 2.4, the first node splits the data according

¹For more details on random forests and their convergence properties, see Breiman (2001).

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.

Appendix C

Appendix to Chapter 3

C.1 Additional Reduced-Form Results

C.1.1 Results Including Attrited Participants

The experimental results are robust to the inclusion of attrited participants. The reduced-form findings presented in the main body of the paper (Section 4) restrict attention to participants who complete the entirety of the experiment. I repeat the analysis over the full set of participants who complete the consent process, including those who attrited between Week 1 and Week 4, and report the results in Table C.1. These results are very consistent with those reported in Table 3.4, and slightly stronger due to the larger sample sizes.

C.1.2 Pilot Study

I briefly present the results from the pilot run of the online experiment, which are broadly consistent with the results from the subsequent main experimental sessions.

I estimate the difference between immediate and ahead-of-time decisions for two samples: using the full set of decisions, and using only the decisions made by participants who complete the entire four-week experiment. The results are reported in Panel 1 of Table C.2. The estimation is done with wage and participant fixed effects. In all specifications, standard errors are clustered by participant. On average, participants choose to do 2.39-2.59

Table C.1: Pooled results from all experimental participants, including participants who attrited between Week 1 and Week 4.

Panel 1: Actual Difference in Ahead-of-time vs. Immediate Decisions

Actual Difference	3.46**	3.22**	3.56**	3.34**
Standard error	(0.60)	(0.57)	(0.53)	(0.50)
Controls:				
Wage FE		X		X
Participant FE			X	X

Panel 2: Predicted Difference in Own Ahead-of-time vs. Immediate Decisions

Self-Prediction	0.98	0.41	0.89**	0.35
Standard error	(1.01)	(0.99)	(0.33)	(0.28)
Controls:				
Wage FE		X		X
Participant FE			X	X

Panel 3: Predicted Difference in Others' Ahead-of-time vs. Immediate Decisions

Other-Prediction	1.19**	1.44**	1.19**	1.45**
Standard error	(0.43)	(0.35)	(0.44)	(0.36)
Controls:				
Wage FE		X		X
Participant FE			X	X

** denotes significance at the 1% level.

rounds fewer when the decision is immediate than when the decision is made ahead of time. The difference is statistically significant at the 10% level in the sample including attrited participants, despite the small sample size (only 23 participants make at least one work decision).

Participants' beliefs regarding their own present bias, estimated as the difference between the participants' ahead-of-time decisions for a given date and their predictions of the choices they would make when that date actually arrives, are estimated in Panel 2 of Table C.2. Participants' predictions of the changes in their choices are statistically indistinguishable from zero, indicating that the participants are naïve about their own present bias.

Beliefs about others appear to be quite sophisticated, as indicated by the estimates in

Table C.2: Results from the pilot run of the online experiment.

Panel 1: Actual Difference in Ahead-of-time vs. Immediate Decisions

	With attrited participants	Without attrited participants
Actual Difference	2.59†	2.39
Standard error	(1.40)	(1.49)
Controls:		
Wage FE	X	X
Participant FE	X	X

Panel 2: Predicted Diff in Own Ahead-of-time vs. Immediate Decisions

	With attrited participants	Without attrited participants
Self-Prediction	0.25	0.45
Standard error	(0.70)	(0.74)
Controls:		
Wage FE	X	X
Participant FE	X	X

Panel 3: Predicted Diff in Others' Ahead-of-time vs. Immediate Decisions

	With attrited participants	Without attrited participants
Other-Prediction	2.15*	2.22*
Standard error	(0.99)	(0.99)
Controls:		
Wage FE	X	X
Participant FE	X	X

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

Panel 3 of Table C.2. Participants expect others to do an average of 2.15-2.22 rounds of work fewer when the work decision concerns immediate completion than when the work decision is made ahead of time. This result is statistically significant at the 5% level. The predicted differences for others in Panel 3 are also very close to the actual differences in Panel 1, suggesting that the participants in the pilot sample are almost perfectly aware of present bias in others, even as they remain naïve about their own present bias.

C.1.3 Warm-up Amounts and Projection Bias

I exploit the differences in the participants' warm-up amounts to explore the possibility that participants' work decisions and predictions reflect projection bias over the cost of effort in doing the task. I do not find significant projection bias and discuss potential explanations for why this bias is not operative in my setting.

Individuals subject to projection bias project utility in their current state onto decisions made in other states.¹ For example, while hungry, an individual may overestimate the utility she would experience from consuming dessert after a filling meal. In my setting, when an individual has only just begun working on the experimental task, she may perceive it as relatively easy, and erroneously expect to feel the same way after doing the task for an hour.

The variation in the participants' warm-up amounts – 5 rounds, 10 rounds, or 15 rounds at the start of each participation date – provides a framework for capturing the effect of projection bias on the participants' decisions and predictions. In particular, since all decisions and predictions are made after the warm-up and assuming a convex cost of effort for the experimental task, projection bias predicts that participants with higher warm-up amounts would project their higher current marginal cost of effort and choose (and predict) fewer rounds of work.

I do not find the predicted effect of projection bias in my experimental data. Participants with higher warm-up amounts actually choose to do *more* work, as can be seen in Panel 1 of Table C.3, which pools immediate and ahead-of-time work decisions but slices the sample by the warm-up amount. After a 5-minute warm-up, participants choose, on average, 24.27 rounds of work. With a 10-minute warm-up, the average work decision is 5.40 rounds higher, albeit not statistically significantly different. The difference in decisions following a 15-minute warm-up versus a 5-minute warm-up is 7.64 rounds, significant at the 10% level. These patterns are robust to controlling for wage fixed effects.²

¹For studies exploring projection bias, see Loewenstein and Adler (1995), Read and Van Leeuwen (1998), Loewenstein *et al.* (2003), Ariely and Loewenstein (2006), Conlin *et al.* (2007), Acland and Levy (2015), and Kaufmann (2017).

²Note that I do not include participant fixed effects in this analysis, since the explanatory variable of interest,

Table C.3: Pooled decisions, actual differences, and predicted differences, sliced by warm-up amount.

Panel 1: Decisions (Immediate & Ahead-of-Time)

	Warm-up = 5 rounds	Warm-up = 10 rounds	Warm-up = 15 rounds
Baseline	24.27	+5.40	+7.64†
Standard error	(2.92)	(4.01)	(4.09)
With Wage FE	–	+5.31	+7.43†
Standard error	–	(3.96)	(4.02)

Panel 2: Actual differences

	Warm-up = 5 rounds	Warm-up = 10 rounds	Warm-up = 15 rounds
Difference	4.04**	3.78**	2.33*
Standard error	(1.27)	(0.84)	(1.13)
Controls:			
Wage FE	X	X	X
Participant FE	X	X	X

Panel 3: Self-Predictions

	Warm-up = 5 rounds	Warm-up = 10 rounds	Warm-up = 15 rounds
Self-Prediction	0.94	1.22*	-0.54
Standard error	(0.69)	(0.58)	(0.49)
Controls:			
Wage FE	X	X	X
Participant FE	X	X	X

Panel 4: Other-Predictions

	Warm-up = 5 rounds	Warm-up = 10 rounds	Warm-up = 15 rounds
Other-Prediction	1.08†	2.48*	0.84
Standard error	(0.57)	(0.98)	(0.61)
Controls:			
Wage FE	X	X	X
Participant FE	X	X	X

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

There are several potential explanations for the lack of predicted effects of projection bias in my setting. First, the predicted effects rely on the implicit assumption of convex effort costs, which could be violated if the task becomes relatively easier for participants after engaging in it for longer. This is unlikely to be the case, since participants generally choose interior options such as 30 rounds or 60 rounds, rather than the boundary solutions of 0 and 70 rounds. The second possibility is that the warm-up amounts introduce anchoring effects, so that after doing 15 rounds of the task, a choice of 20 additional rounds seems too insubstantial, whereas the same 20 rounds appear to be a reasonable amount of work if the warm-up consists of 5 rounds. These anchoring effects are not at play in the design of Augenblick and Rabin (2017), who do not vary the warm-up amounts across participants but instead vary whether the decisions are elicited before or after the warm-up, and who do find evidence of projection bias in their setting. Since projection bias is not a focus of the present study, I leave more in-depth exploration of these effects for future work.

I check whether the warm-up amount systematically affects the participants' displayed present bias and predictions, and find mixed results. There is slight suggestive evidence that present bias is reduced with an increased warm-up amount, as the difference between ahead-of-time and immediate decisions is, on average, 4.04 rounds for participants with warm-ups of 5 rounds and 2.33 rounds for participants with warm-ups of 15 rounds (see Table C.3, Panel 2). The differences in self- and other-predictions across warm-up amounts do not reveal any consistent patterns. Participants with warm-up amounts of 10 rounds make, on average, more accurate predictions regarding others, while predictions regarding self are less accurate for warm-up amount of 15 rounds than for either 5 rounds or 10 rounds (see Panels 3 and 4 of Table C.3). Overall, the warm-up amounts do not appear to affect beliefs regarding present bias in a consistent systematic way, and the qualitative pattern of more accurate expectations regarding others holds across warm-up amounts.

the warm-up amount, is constant across each participant.

C.2 Additional Structural Estimates

C.2.1 Different Baseline Levels of Effort for Self and Others

I begin the structural analysis by allowing for one additional difference in beliefs regarding self and others, relative to the results in the main body of the paper: that participants expect, on average, others to choose a different amount of effort even when the choice is for future work (and hence present bias is not operative). To allow for this difference in baseline levels of effort, I modify the model specification as follows:

$$\hat{e}(t, \tau, w, x, \mathbb{1}_s, \mathbb{1}_o) = (\beta^{\mathbb{1}_{t=\tau}} \beta_{(s)}^{\mathbb{1}_s} \beta_{(o)}^{\mathbb{1}_{t=\tau} \mathbb{1}_o} \delta^{T-\tau} \phi w \zeta^{\mathbb{1}_o})^{\frac{1}{\gamma-1}} - x \quad (\text{C.1})$$

The parameter ζ governs the extent to which each participant expects others to generally choose more or less work than herself. It is identified by comparing the participants' ahead-of-time decisions (for which $\mathbb{1}_o = 0$) against their predictions of others' ahead-of-time decisions (for which $\mathbb{1}_o = 1$). Note that the parameter ζ is distinct from the beliefs regarding others' present bias $\beta_{(o)}$, as the latter is tied to the timing of the decisions (through the indicator $\mathbb{1}_{t=\tau}$).

The results from estimating (C.1), reported in Table C.4, confirm that the main findings are robust to allowing differences in beliefs regarding baseline effort. The estimates of β continue to indicate robust presence of present bias, and estimates of $\beta_{(s)}$ display no awareness of one's own present bias. The estimates of $\beta_{(o)}$ indicate participants' recognition of present bias in others, although in this specification the expectations of others' present bias are further from the true value of β than in the main specification. Bootstrapped comparisons indicate that the estimate of $\beta_{(o)}$, at 0.93, is statistically different from 1 (t-statistic on the difference: 3.53 excluding attriters and 4.02 including all participants), but also different from the corresponding estimates of β (t-statistic: 3.27 excluding attrited participants and 4.50 including all participants). The fact that the additional parameter ζ falls below 1 indicates that participants expect others to choose, on average, to do less work than themselves. Combined with the beliefs regarding present bias, this points towards

Table C.4: Parameter estimates from the structural model, allowing for different perceived baseline levels for self versus others when the decision is made for future dates.

Parameter	Without attrited participants	With attrited participants
Present bias β	0.8234 (0.0285)	0.7833 (0.0303)
Self-prediction $\beta_{(s)}$	1.0067 (0.0545)	0.9904 (0.0410)
Other-prediction $\beta_{(o)}$	0.9318 (0.0193)	0.9316 (0.0160)
δ	1.0149 (0.0026)	1.0155 (0.0028)
γ	2.2452 (0.1121)	2.2458 (0.0909)
ϕ	387.3583 (224.3250)	396.4874 (177.2949)
ζ	0.8964 (0.0516)	0.8992 (0.0439)

relative overconfidence as a driver for the wedge in beliefs: participants (incorrectly) expect others to do less work overall, (incorrectly) believe themselves to have no present bias, but (correctly) anticipate others to display present bias.

C.2.2 Different Utility Function Parameters for Self and Others

I allow for further differences in perceptions of self versus others by estimating all model parameters separately for responses regarding self and responses regarding others.

I pool all responses about self, including all participants' work decisions and all self-predictions by participants who make predictions regarding their own future work (i.e., participants in Groups 1 and 2). I use these responses to estimate the parameters β and $\beta_{(s)}$ with the following specification:

$$\hat{e}_{(s)}(t, \tau, w, x, \mathbb{1}_s) = (\beta^{1-t-\tau} \beta_{(s)}^{1-s} \delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (\text{C.2})$$

The results are displayed in columns marked with (1) in Table C.5. The leftmost column presents estimates without attrited participants; the column marked with (1) on the right include attrited participants. The results indicate that participants display substantial present bias (estimates of β fall between 0.80 and 0.84) and no awareness of their own present bias

Table C.5: *Parameter estimates from the structural model, allowing for different model parameters for responses about self versus other.*

Parameter	Without attrited participants		With attrited participants	
	(1)	(2)	(1)	(2)
Present bias β	0.8354 (0.0306)		0.7952 (0.0496)	
Self-prediction $\beta_{(s)}$	1.0051 (0.0466)		0.9904 (0.0406)	
Other-prediction $\beta_{(o)}$		0.9244 (0.0216)		0.9244 (0.0189)
δ	1.0142 (0.0225)	1.0160 (0.0046)	1.0153 (0.0543)	1.0149 (0.0045)
γ	2.0821 (0.1330)	2.5405 (0.1609)	2.0953 (0.1698)	2.5158 (0.1356)
ϕ	212.9095 (135.9622)	1011.4998 (1223.6143)	225.5672 (114.2770)	984.6902 (750.3936)

($\beta_{(s)}$ is indistinguishable from 1). These estimates are consistent with those from the main specification.

In columns marked with (2), I focus on the responses regarding others from those participants who make other-predictions (i.e., participants in Groups 2 and 3). I use the following specification for predicted responses regarding others, $\hat{e}_{(o)}(t, \tau, w, x)$:

$$\hat{e}_{(o)}(t, \tau, w, x) = (\beta_{(o)}^{1-t-\tau} \delta^{T-\tau} \phi w)^{\frac{1}{\gamma-1}} - x \quad (\text{C.3})$$

The results show that participants anticipate present bias in others. The estimated $\beta_{(o)}$ is around 0.92-0.93, significantly different from 1 at the 1% level. The estimates of the other model parameters, δ , ϕ , and γ indicate that participants expect a steeper cost of effort in others than in themselves (γ of 2.5 compared to γ of 2.1), and correspondingly higher linear utility in money (ϕ of 985-1,011 compared to ϕ of 212-226), although the estimates for others, especially of ϕ , are very noisy.

Qualitatively, the results in the separate specifications confirm naïveté regarding one's own present bias coupled with some but incomplete awareness regarding others. Quantitatively, the estimates are very similar to the estimates in Table C.4 from the joint specification allowing for different baseline levels.

C.3 Experimental Instructions

C.3.1 Study Sign-Up

Sign-Up

Welcome to the study "Doing Work over Time"!

If you sign up for this study:

- You will be asked to participate for at least 20 minutes on four different dates, all on the same day of the week.
- We will send you reminders to log in and do the minimum required work on each of the dates.
- You will be paid \$30 for completing all of the work over these four dates.
- In addition, you will have the option to perform tasks for us at various wages to earn extra money.

If you are interested in participating, please pick the set of dates for your participation. On the first date, we will explain the study in more detail. You will have the option of declining to participate at any time.

Note: you will not be able to change your participation dates, so please choose wisely.

- MONDAYS: [display next four Monday dates]
- TUESDAYS: [display next four Tuesday dates]
- WEDNESDAYS: [display next four Wednesday dates]
- THURSDAYS: [display next four Thursday dates]
- FRIDAYS: [display next four Friday dates]
- SATURDAYS: [display next four Saturday dates]

- SUNDAYS: [display next four Sunday dates]

Please enter your email address if you would like to sign up: [Text entry box]

["NO, THANKS" and "SIGN UP!" buttons]

C.3.2 Instructions and Questions: First Day

Welcome to our experiment!

ELIGIBILITY FOR THIS STUDY: To be in this study, you need to meet the following criteria:

- You must be at least 18 years old.
- **You will need to participate TODAY, and on the NEXT THREE [Day of Week], [Insert the dates: DATE 2, DATE 3, and DATE 4].**
- After reading the instructions today, you will need to complete a comprehension quiz. You must answer at least 8 out of 10 questions correctly to be eligible for the study.
- Participation will require you to log in between 12:01AM EST and 10:30PM EST on each participation date, and complete a **warm-up consisting of [5, 10, or 15] rounds of work**. Additional work can be assigned and completed at your discretion.
- You must be willing to receive all earnings from this experiment as **one single payment at the end of the study**. The payment will be made through an Amazon.com gift voucher on [END DATE].

If you do not meet these criteria, please click EXIT below. Otherwise, please click I AGREE to proceed.

["EXIT" and "I AGREE" buttons]

Consent

Please read through the information below and certify your agreement to participate in this study by clicking on I AGREE at the bottom of the text. If you do not wish to participate in the study, please click on EXIT at the bottom of the text.

[See Appendix D for the text of informed consent, which appears on the participants' screens.]

["EXIT" and "I AGREE" buttons]

What to Expect Today

You will read through the instructions for the experiment. At the end, you will need to pass a **comprehension quiz**. You must answer **no more than TWO questions incorrectly** in order to be eligible for this study. If you answer more than two questions incorrectly, you will be directed to exit the study.

Upon passing the comprehension quiz, you will practice the experimental task, and then you will answer a series of questions about how much work you would like to do at different wages during the future participation dates. At the end of today's session, you will be asked some survey questions.

["EXIT" and "NEXT" buttons]

[Loading bar to show the participant's progress through the instructions – "Instructions: page 1/8"]

The Task

Each round of the Task lasts 60 seconds: 50 seconds of work and 10 seconds of rest. During the work phase, you will be presented with characters appearing one by one on the screen, at the pace of one character every two seconds. You need to press the SPACE key or click the 'I can see it!' button every time an ASTERISK (*) appears on the screen. Do NOT press the key when any other character appears – only the asterisk.

Your score will be calculated as the percentage of characters that you identify correctly. For example, if during one round, which consists of 25 characters, you miss one asterisk and incorrectly capture one other character, your accuracy will be 23/25 or 92%. To successfully progress, you must achieve an average of at least 80% accuracy on the Task within each session, so please pay attention!

You will do a [5, 10, or 15] minute warm-up of the Task today, after you finish reading

the instructions.

["EXIT" and "NEXT" buttons]

[Loading bar – "Instructions: page 2/8"]

Your Earnings

On every participation date, you have to do the warm-up consisting of [5/10/15] rounds and answer questions. For this work, you will receive a **single \$30 completion payment**. In addition, you will have the option of doing extra work for additional payment [*for incentivized participants*: "and bonuses for correct predictions of [your future / other subjects' / your own and other subjects'] work"].

Overall, the payment structure can be summarized as follows:

Payment Type	Amount	What you need to do to receive it
Completion payment	\$30	Complete all mandatory tasks and all chosen extra rounds with at least 80% accuracy
Wages	\$0.10-0.30/round (0-210 rounds in total)	Complete extra rounds at the relevant wage with at least 80% accuracy
Prediction bonuses	\$0.10-0.40/prediction (10 predictions)	Correctly predict other subjects' choices if those choices are then randomly selected to be implemented

Figure C.1: *Payment schedule.*

[Note: Last row of Figure C.1 appears only for incentivized participants. The example in the figure includes bonus payments for participants making predictions regarding others. Bonus payments for participants who make self-predictions are analogous.]

You will receive all of your earnings (the \$30 completion payment + earnings from the extra rounds) from this experiment in a single lump-sum payment on [PAYMENT DATE].

Once again, it is very important to note that in order to receive the \$30 completion payment, **you must log in and do the assigned work including your chosen extra rounds**

on all participation dates, [DATE 1], [DATE 2], [DATE 3], and [DATE 4]. If you miss a participation date, you will still receive the wages for the additional rounds of work that you have already completed up until that point; **however, you will be removed from the experiment, and forego your completion payment.**

["EXIT" and "I UNDERSTAND" buttons]

[Loading bar – "Instructions: page 3/8"]

"How Much to Work" Decisions

As we have told you, the work schedule will be up to you. Each day we will ask you how many extra rounds you would like to do at randomly generated wages. Some of the questions will be for the same day. Some will be for days in the future.

We will vary the wages between \$0.10/round and \$0.30/round. Since each round lasts one minute, this corresponds to hourly wages between \$6/hour and \$18/hour.

On each day we will collect all the decisions you made for that day – some of these will be immediate decisions you made on that date, some will be decisions you made ahead of time on prior participation dates. We will **randomly** select one of your choices for the day; this will be the wage and the amount of work that you will have to do.

You will have to complete **exactly** the number of extra rounds in the selected decision. This is **after** the [5 / 10 / 15] rounds of warm-up. If you do not complete the selected extra work, you will be **disqualified** and forego the \$30 completion payment. Note that **all of your choices have some chance of being selected, so it is in your best interests to always answer truthfully.** Today, you will not be doing any work, so we will only ask for your preferences about future dates.

You will specify your preferences by entering the number of rounds you would like to do next to the potential wage. **The minimum amount of work is 0 rounds and the maximum is 70 rounds.** Here is an example of what the decisions will look like:

Decisions made now for work to be done on [DATE 3]

How many extra rounds would you like to do on [DATE 3] at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

["EXIT" and "NEXT" button]

[Loading bar – "Instructions: page 4/8"]

"How Much to Work" Decisions: Practice

Let's practice making the work decisions! As you know, you will not be doing the work today. So these are merely hypothetical decisions for you to try out.

[Decisions made now for work to be done immediately](#)

How many extra rounds would you like to do now at the following wages?

Wage	Rounds
\$0.10/round (\$6/hour)	[Entry box]
\$0.20/round (\$12/hour)	[Entry box]

["EXIT" and "GOT IT" button]

[Loading bar – "Instructions: page 5/8"]

Selecting the Decision that Counts

On each participation date, when the time comes to do the work, we will gather all of the wages for which you have made decisions. **This includes ALL decisions made for that day – either on that day or earlier.**

For example, here are the practice choices you made today:

Wage: [\$0.10/round]	Wage: [\$0.20/round]
Rounds: [CHOICE I]	Rounds: [CHOICE II]

Then, we will randomly select ONE of your decisions for this day. This is the **"Decision that Counts."** All decisions are equally likely to be selected. You will then have to do the number of rounds you chose in that decision. **You will have to complete the work immediately after the selection, with no more than 15 minutes of breaks.** Note that the

total break allowance is fixed at 15 minutes regardless of the amount of work you choose to do.

For example, suppose that today, we ran the random selection of the "Decision that Counts," with the following result (highlighted):

Wage: [\$0.10/round]	Wage: [\$0.20/round]
Rounds: [CHOICE I]	Rounds: [CHOICE II]

Then you would have to complete [CHOICE II] rounds of the Task within [CHOICE II +15] minutes ([CHOICE II] minutes for the [CHOICE II] rounds + 15 minutes of break) of the time when the "Decision that Counts" is selected.

Note: any decision you make has a chance of being the "Decision that Counts." So it is in your best interests to make every decision carefully and truthfully.

["EXIT" and "NEXT" buttons]

[Loading bar – "Instructions: page 6/8"]

Predictions

[For participants in Group 1, who make predictions regarding themselves:]

We are also interested in your predictions about your future decisions.

So, we will ask you to predict how many extra rounds of work you will choose to do **after the warm-up** on a future date at various wages. The predictions will look like this:

Decisions made on [DATE 3] when the time to do the work comes

When the time comes to actually do the work on [DATE 3], how many extra rounds do you **think** you will want to do at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

Your prediction will be considered correct if your decision on [DATE 3] is the same as your prediction. For example, suppose that you answered 40 rounds to the prediction above. Then, suppose that on [DATE 3] you are asked how many rounds of the Task you

would like to complete immediately on that day at \$0.20/round, and you answer 40. In this case, your prediction will prove to be correct. [*For those in the incentivized treatment: You will receive a bonus of [X; randomly distributed across participants between \$0.10, \$0.20, \$0.30, and \$0.40] for every prediction that proves to be correct, and is then randomly selected as the "Decision that Counts."*]

[For participants in Group 2, who make predictions regarding others:]

We are also interested in your predictions about other subjects' future decisions.

So, we will ask you to predict the average number of extra rounds of work that the other subjects will choose to do **after the warm-up** at various wages. The predictions will look like this:

Decisions made now for work to be done on [DATE 3]

When the time comes to actually do the work on [DATE 3], how many extra rounds do you think on average, other subjects will want to do at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

Your prediction will be considered correct if the average of the other subjects' answers on [DATE 3] is the same as your prediction. For example, suppose that you answered 40 rounds to the prediction above. Then, suppose that on [DATE 3] two subjects are asked how many rounds of the Task they would like to complete immediately on that day at \$0.20/round, and one of them answers 20 and the other says 60. In this case, the average of the other subjects' answers will be 40, and your prediction will prove to be correct. [*For those in the incentivized treatment: You will receive a bonus [X] for every prediction that proves to be correct, and is then randomly selected as the "Decision that Counts."*]

[Participants in Group 3, who make both sets of predictions, see the following text. Note that the order of questions is randomized across these participants. The sample displayed here shows self-predictions first.]

We are also interested in your predictions about your future decisions, and the decisions of other subjects.

First, we will ask you to predict how many extra rounds of work you will choose to do **after the warm-up** on a future date at various wages. These predictions will look like this:

Decisions made on [DATE 3] when the time to do the work comes

When the time comes to actually do the work on [DATE 3], how many extra rounds do you **think** you will want to do at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

Then, we will ask you to predict the choices of other subjects. For example, we might ask you this:

Decisions made now for work to be done on [DATE 3]

When the time comes to actually do the work on [DATE 3], how many extra rounds do you think, on average, other subjects will want to do at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

Each of your predictions will be considered correct if your choice or the average choice of others is the same as that prediction. For example, suppose that you are asked to make the above prediction about other subjects on [DATE 3], and you answer 40 rounds. Then, suppose that on [DATE 3] two subjects are asked how many rounds of the Task they would like to complete immediately on that day at \$0.20/round, and one of them answers 20 and the other says 60. In this case, the average of the other subjects' answers will be 40, and your prediction will prove to be correct. [*For those in the incentivized treatment: You will receive a bonus of [X] for every prediction that proves to be correct, and is then randomly selected as the "Decision that Counts."*]

["EXIT" and "NEXT" buttons]

[Loading bar – "Instructions: page 7/8"]

Timeline

TODAY:

- **Take the comprehension quiz.** If you forget any part of the instructions, you will be able to view the relevant instruction page. Remember: if you make more than 2 mistakes on the quiz, you will be disqualified from the study.
- **Practice the Task.** To see what the Task is like, you will do [5 / 10 / 15] rounds of the Task.
- **Learn about the other subjects.** After you complete the Task, we will show you the breakdown by age, gender, etc. of the other subjects participating in the experiment, and what previous participants thought of the Task.
- **"How Much to Work" Decisions For Future Participation Dates.** You will make four different decisions about extra rounds, for two different future participation dates.
- **Predictions.** You will also make [four / eight / twelve] predictions about how much work you think [you / on average, the other subjects / you and the other subjects] will want to do at various wages on two different dates.
- **Questionnaire.** We will finish today's session with a quick questionnaire.

FUTURE PARTICIPATION DATES [DATE 2, DATE 3, DATE 4]:

- **Log in.** You must log in between 12:01AM EST and 10:30PM EST on each participation date. All work must be finished by 11:59PM EST.
- **Warm-Up Rounds.** You will start each session with [5 / 10 / 15] warm-up rounds of the Task.
- **"How Much to Work" Decisions For Current Participation Date.** On each participation date, you will be asked how much work you want to do RIGHT ON THAT DAY, at different wages.

- **"How Much to Work" Decisions For Future Participation Dates.** On each participation date other than the last one, you will also make decisions for FUTURE participation dates.
- **Predictions.** You will also make predictions about how much work you think [you / on average, the other subjects / you and the other subjects] will want to do at various wages on different dates, which can be the current participation date or future ones.
- **Completion of Extra Work.** One of your decisions for each participation date will be selected as the "Decision that Counts." You will then need to complete the amount of work you had chosen in that decision with no more than 15 minutes of breaks.
- **Questionnaire.** On the last participation date, [DATE 4], the study will end with a brief questionnaire.

["REREAD ALL INSTRUCTIONS" and "GO TO QUIZ" buttons]

[Loading bar – "Instructions: page 8/8"]

Quiz

[Below, the correct answers are marked in bold for reference. On the actual quiz screen, next to each question, participants see a button that opens a pop-up of the relevant instructions page.]

Please feel free to click on "Check Instructions" to see the relevant sections of the instructions.

1. Will you have to complete a warm-up on every participation date?
 - (a) **Yes**
 - (b) No
2. Including today, on how many participation dates must you log into the study?
 - (a) One
 - (b) **Two**

- (c) **Four**
 - (d) Seven
3. When will you receive your payment?
- (a) At the end of today
 - (b) [DATE 3]
 - (c) [DATE 4]
 - (d) [**PAYMENT DATE**]
4. If you fail to participate on one of your participation dates, what payment will you receive?
- (a) Nothing
 - (b) \$30
 - (c) **Payment for the extra rounds already completed up to that point.**
 - (d) \$30 + payment for the extra rounds already completed up to that point
5. How many additional rounds of the Task can you choose to do on each participation date?
- (a) 0-10 rounds
 - (b) **0-70 rounds**
 - (c) 0-100 rounds
 - (d) 50-70 rounds
6. Suppose that you are asked today how many extra rounds of the Task you would like to do on [DATE 2], and you answer 32. If this decision is selected as the "Decision that Counts" on [DATE 2], how many extra rounds of the Task will you be doing on [DATE 2]?
- (a) At least 32 rounds

- (b) **Exactly 32 rounds**
- (c) Between 0 and 32 rounds
- (d) Either 0 or 32 rounds
7. You will make many decisions about how much extra work to do for any given participation date. Only one of those decisions will be selected as the "Decision that Counts." How will this be selected?
- (a) The first decision will be selected.
- (b) The latest decision will be selected.
- (c) **The decision will be selected randomly, and all decisions will have the same likelihood of being selected.**
- (d) The decision will be selected randomly, with more weight given to more recent decisions.
8. Suppose that on [DATE 4], the selected "Decision that Counts" involves you doing 18 rounds of the Task. Once the "Decision that Counts" is selected, how much time do you have to complete this work?
- (a) 15 minutes of breaks
- (b) 18 minutes for the Task
- (c) **33 minutes: 18 for the Task and 15 minutes of breaks**
- (d) 1 hour: 18 minutes for the Task and 42 minutes of breaks
9. Suppose that you log in on [DATE 3]. It is time for the "Decision that Counts" to be selected. Suppose that you have made the following two decisions for [DATE 3] (in reality, you will be making many more decisions for each day, so this is a simplified example):
- On [DATE 2], you chose to do 50 rounds on [DATE 3] at \$0.20/round.
 - On [DATE 3], you chose to do 40 rounds on [DATE 3] at \$0.10/round.

Which of the following are possible?

- (a) You will have to do 0 rounds on [DATE 3]
- (b) You will have to do 40 rounds on [DATE 3]
- (c) You will have to do 50 rounds on [DATE 3]
- (d) Both (a) and (b) are possible
- (e) Both (a) and (c) are possible
- (f) **Both (b) and (c) are possible**
- (g) (a), (b), and (c) are all possible

10. *[For participants in Group 1:]*

Suppose that on [DATE 2], you are asked to make the following prediction:

Decisions made on [DATE 3] when the time to do the work comes

When the time comes to actually do the work on [DATE 3], how many extra rounds do you **think** you will want to do at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

And you answer 60 rounds.

In which of the following cases will your prediction prove to be CORRECT?

- **On [DATE 3], you are asked how many extra rounds of the Task you would like to do at \$0.20/round on that date ([DATE 3]), and you answer 60.**
- On [DATE 3], you are asked how many extra rounds of the Task you would like to do at \$0.20/round on that date ([DATE 3]), and you answer 50.
- On [DATE 3], you are asked how many extra rounds of the Task you would like to do at \$0.20/round on that date ([DATE 3]), and you answer 70.
- On [DATE 3], you are not asked to make any decisions at \$0.20/round.

[For participants in Groups 2 and 3:]

Decisions made on [DATE 3] when the time to do the work comes

When the time comes to actually do the work on [DATE 3], how many rounds to you think, on average, other subjects will want to do at the following wages?

Wage	Rounds
\$0.20/round (\$12/hour)	[Entry box]

And you answer 60 rounds.

In which of the following cases will your prediction prove to be CORRECT?

- On [DATE 3], two other subjects are asked how many extra rounds of the Task they would like to do at \$0.20/round on that date ([DATE 3]). One of them answers 50 and the other says 70.
- On [DATE 3], two other subjects are asked how many extra rounds of the Task they would like to do at \$0.20/round on that date ([DATE 3]). One of them answers 30 and the other says 70.
- On [DATE 3], one other subjects is asked how many extra rounds of the Task he would like to do at \$0.20/round on that date ([DATE 3]), and he chooses 70.
- On [DATE 3], none of the other subjects are asked to make any decisions at \$0.20/round.

Failed Quiz screen

[The following message is displayed to participants who do not pass the quiz with at least 8/10 correct answers.]

You answered [# INCORRECT] questions incorrectly, which indicates that you did not fully understand the experimental procedure. Unfortunately, this means that you do not qualify to participate in this study.

["EXIT" button]

Quiz Results

You have answered [# CORRECT] of the 10 questions correctly. This means that you passed the quiz!

[If the number of questions answered correctly is 8 or 9 rather than 10, the participant is shown the question(s) that (s)he missed, with the correct answer(s) highlighted in blue.]

["EXIT" and "CONTINUE" buttons]

Warm-Up

And now, you can practice doing the Task for yourself.

During this warm-up, you will spend [5 / 10 / 15] minutes familiarizing yourself with the experimental Task. You will be presented with [5 / 10 / 15] rounds of the Task. Each round lasts 60 seconds including 10 seconds of break. Remember: you need to achieve at least 80% accuracy to pass, so please pay attention!

Press the SPACEBAR key or click on the "I can see it!" button every time you see an asterisk.

["EXIT" and "PRACTICE" buttons]

Warm-Up

[The participant has to complete [5 / 10 / 15] rounds of the Task as warm-up – see Figure 2 for the Task screen.]

Great Job

Congratulations on successfully completing the [5 / 10 / 15] warm-up rounds of the Task!

["EXIT" and "CONTINUE" buttons]

What do Others Think of the Task?

Now that you know what it feels like to do the Task, you might be curious to see what others thought about it...

[The page then displays a pie chart break-down of pilot participants by gender, race, marital status, age, education, employment, followed by a pie chart break-down of pilot participants' opinion of the Task (tedious / enjoyable / fine) by gender, race, marital status, age, education, employment. See Figure 3 for a view of this page.]

["EXIT" and "CONTINUE" buttons]

Work Decisions and Predictions

[The participants are next faced with a series of screens eliciting their work decisions and predictions.

- For participants in Group 1, who make predictions regarding themselves:
 - Participants in Group 1 see the screen displayed in Panel 1 of Figure 7 for two dates: [DATE 2] and [DATE 3].
- For participants in Group 2, who make predictions regarding other participants:
 - Participants in Group 2 see the screen displayed in Panel 2 of Figure 7 for two dates: [DATE 2] and [DATE 3], and the screen displayed in Figure 5 for two future dates: [DATE 2] and [DATE 3].
- For participants in Group 3, who make both sets of predictions:
 - Participants in Group 3 see the screen displayed in Panels 1 and 2 of Figure 7, in random order, for two dates: [DATE 2] and [DATE 3].

All wages are randomly drawn from \$0.10 to \$0.30 in \$0.05 increments, and the two wages are always different within each box.

Participants can only input integers between 0 and 70 into the fields, and must answer all questions before proceeding. If a participant inputs the same answer to both questions, she is asked whether she is sure that she wishes (or predicts for herself or others) to do the same amount of work regardless of the wage. If the participant chooses fewer rounds at the higher wage, she is also alerted to this inconsistency, and asked to confirm whether she would like to proceed with this answer.

Incentivized participants in Groups 1 and 3 see the following message at the bottom of their self-prediction screens:]

Bonus: on [FUTURE DATE], if your answer is the same as the prediction you make today **and** this decision is selected as the “Decision that Counts,” you will receive a bonus of \$[BONUS]. If you do not answer such that the previous prediction is correct or this decision is not selected as the “Decision that Counts,” you will not receive any bonus.

[Incentivized participants in Groups 2 and 3 see the following message after their other-prediction screens:]

Bonus: if the average of the other subjects’ answers is the same as your prediction **and** this decision is selected as the “Decision that Counts” for at least one subject, you will receive a bonus of \$[BONUS]. If your prediction is not correct or this decision is not selected as the “Decision that Counts” for any subject, you will not receive any bonus.

Demographic Questionnaire

Congratulations on finishing the tasks for the first participation date! Now, all that remains is a quick questionnaire.

Please answer the following questions to the best of your ability. Note that all wages and choices of “Decisions that Count” are random, and will in no way depend on the answers you give today. **Please answer all questions truthfully.**

[All questions appear separately on the screen, one by one. The order of the questions is randomized for each participant.]

[Demographic questions:]

Question 1 How old are you? [Text entry box]

Question 2 Gender: — Female — Male — Decline to Answer

Question 3 Please specify your ethnicity

— White — Hispanic or Latino — Black or African American

- Native American or American Indian — Asian / Pacific Islander
- Other — Decline to Answer

Question 4 What is the highest degree or level of schooling you have completed? If currently enrolled, highest degree received

- Nursery school to 8th grade
- Some high school, no diploma — High school graduate
- Some college credit, no degree — Trade/technical/vocational training
- Associate degree — Bachelors degree — Master's degree
- Professional degree — Doctorate degree — Decline to Answer

Question 5 What is your marital status?

- Single, never married — Married or domestic partnership
- Divorced — Separated — Widowed — Decline to Answer

Question 6 What is your current employment status?

- Employed — Self-employed — Student
- Military — Retired — Out of work and looking for work
- Out of work but not currently looking for work — Decline to Answer

[Time budgeting and Task-enjoyment questions:]

Question 7 How busy do you expect to be in the next few weeks?

[Can choose one of 5 loci] Not busy at all – (1) – (2) – (3) – (4) – (5) – Very busy

Question 8 How busy do you expect your colleagues or classmates to be in the next few weeks?

[Can choose one of 5 loci] Not busy at all – (1) – (2) – (3) – (4) – (5) – Very busy

Question 9 How much did you enjoy doing the Task during today's warm-up?

[Can choose one of 5 loci] Not enjoy it at all – (1) – (2) – (3) – (4) – (5) – Enjoy it very much

Question 10 On average, how much do you think the other subjects enjoyed doing the Task during today's warm-up?

[Can choose one of 5 loci] Not enjoy it at all – (1) – (2) – (3) – (4) – (5) – Enjoy it very much

Question 11 How easy do you think it will be for you to find the time to work on additional rounds of the Task in the coming weeks?

[Can choose one of 5 loci] Very Easy – (1) – (2) – (3) – (4) – (5) – Very Difficult

Question 12 On average, how easy do you think it will be for the other subjects to find the time to work on additional rounds of the Task in the coming weeks?<

[Can choose one of 5 loci] Very Easy – (1) – (2) – (3) – (4) – (5) – Very Difficult

Question 13 How productive do you think you will be at your work in the next few weeks?

[Can choose one of 5 loci] Not productive at all – (1) – (2) – (3) – (4) – (5) – Very productive

Question 14 How productive do you think your classmates or colleagues will be at their work in the next few weeks?

[Can choose one of 5 loci] Not productive at all – (1) – (2) – (3) – (4) – (5) – Very productive

Question 15 How good do you think your overall mood will be in the next few weeks?

[Can choose one of 5 loci] Very bad, depressed – (1) – (2) – (3) – (4) – (5) – Excellent, euphoric

Question 16 On average, how good do you think your colleagues' or classmates' mood will be in the next few weeks?

[Can choose one of 5 loci] Very bad, depressed – (1) – (2) – (3) – (4) – (5) – Excellent, euphoric

[Overconfidence questions:]

Question 17 What percentage of your classmates or colleagues do you think has attention span that is at least as good as yours? [Text entry box]

Question 18 What percentage of your classmates or colleagues do you think has reflexes that are at least as fast as yours? [Text entry box]

Question 19 What percentage of your classmates or colleagues do you think are at least as hardworking as you? [Text entry box]

You're Done for the Day

Congratulations, you have successfully completed the assignment for the first participation date!

We look forward to your return on [DATE 2], [DATE 3], and [DATE 4]. Remember, in order to receive your \$30 completion payment, you must participate on ALL participation dates. You will receive email reminders on each of your participation dates. Have a lovely rest of the week!

C.3.3 Instructions and Questions: Dates 2, 3, and 4

Welcome Back

Welcome back to the experiment!

Today, you will start out with a warm-up of [5 / 10 / 15] rounds of the Task. We will then ask you to make some work decisions and predictions. We will put together all of the decisions you have made for today at various wages - this will include both today's decisions and past decisions. Then we will select **one** of these hypothetical decisions as the "Decisions that Counts." This will be the decision we implement, and you will have to immediately do the number of extra rounds in this decision.

["EXIT" and "PROCEED TO WARMUP " buttons]

Warm-up

[Participants are faced with a warm-up consisting of [5 / 10 / 15] rounds of the Task – see Figure 2 for the Task screen.]

Warm-up complete

Great job on the warm-up! You finished the [5 / 10 / 15] warm-up rounds of the Task.

Now, let's proceed to today's work decisions and predictions.

["EXIT" and "CONTINUE " buttons]

Work Decisions and Predictions

[First, each participant sees the questions about immediate work on that date, displayed in Figure 5. Then, on [DATE 2] and [DATE 3], the participants are presented with the following questions depending on their treatment group.

- *For participants in Group 1, who make self-predictions:*
 - *Participants in this group see the screen displayed in Panel 1 of Figure 7. On [DATE 2], they see these questions for [DATE 3] and [DATE 4]. On [DATE 3], they see these questions for [DATE 4].*
- *For participants in Group 2, who make predictions regarding others:*
 - *Participants in this group see the screen displayed in Panel 2 of Figure 7 and the screen in Figure 5 for the future date. On [DATE 2], they see these questions for [DATE 3] and [DATE 4]. On [DATE 3], they see these questions for [DATE 4].*
- *For participants in Group 3, who make both sets of predictions:*
 - *Participants in this group see the screens displayed in Panels 1 and 2 of Figure 7. On [DATE 2], they see these questions for [DATE 3] and [DATE 4]. On [DATE 3], they see these questions for [DATE 4].*

All wages are randomly drawn from \$0.10 to \$0.30 in \$0.05 increments, and the two wages are always different within each box.

Participants can only input integers between 0 and 70 into the fields, and must answer all questions before proceeding. If a participant inputs the same answer to both questions, she is asked whether she is sure that she wishes (or predicts for herself or others) to do the same amount of work regardless of the wage. If the participant chooses fewer rounds at the higher wage, she is also alerted to this inconsistency, and asked to confirm whether she would like to proceed with this answer.

Incentivized participants in Groups 1 and 3 see the following message at the bottom of their self-prediction screens:]

Bonus: on [FUTURE DATE], if your answer is the same as the prediction you make today **and** this decision is selected as the “Decision that Counts,” you will receive a bonus of \$[BONUS]. If you do not answer such that the previous prediction is correct or this decision is not selected as the “Decision that Counts,” you will not receive any bonus.

[Incentivized participants in Groups 2 and 3 see the following message at the bottom of their other-prediction screens:]

Bonus: if the average of the other subjects’ answers is the same as your prediction **and** this decision is selected as the “Decision that Counts” for at least one subject, you will receive a bonus of \$[BONUS]. If your prediction is not correct or this decision is not selected as the “Decision that Counts” for any subject, you will not receive any bonus.

Selecting the “Decision that Counts”

[The selection screen is presented in Figure 6. The displayed screen is for [DATE 3] or [DATE 4]. The screen for [DATE 2] is analogous, but with four decisions instead of six.]

“Decision that Counts” Selected

[The selection screen displayed in Figure 6 now has a randomly selected decision highlighted.]

Work

[Participant sees a screen analogous to Figure 2, only this time it's for the selected number of rounds of the Task and features a timer.]

You're Done for the Day

[On [DATE 2] and [DATE 3], participants see the following:]

You have finished all of the work for today – great job! For the supplementary rounds of the Task you did today, you will receive an extra \$[Y].

We look forward to your return on [REMAINING PARTICIPATION DATES]. Remember, in order to receive your \$30 completion payment, you must participate on ALL participation dates. You will receive email reminders on each of the remaining participation dates. Have a lovely rest of the week!

["EXIT" button]

[On [DATE 4], participants see the following:]

Congratulations! You have completed all of the work on the Task for this study. You are very, very close to being done with the study.

All that remains is for you to answer a couple of questions. Please press NEXT to proceed.

["NEXT" button]

Debrief Questions at the end of Date 4

1. There are [X] other subjects participating in this experiment. **Guess:** how many of them were at least as consistent in their decisions (same day compared to ahead-of-time) as you?

Note: if you guess correctly, you will receive a bonus of \$5!

2. How much did you enjoy doing the Task?

Not at all – (1) – (2) – (3) – (4) – (5) – Very much

3. How difficult was the Task?

Very easy – (1) – (2) – (3) – (4) – (5) – Very difficult

4. How difficult was it for you to find the time to work on the Task?

Very easy – (1) – (2) – (3) – (4) – (5) – Very difficult

5. Do you wish you had time to do more rounds of the Task on your participation dates?

– Yes – No

6. Were your choices consistent? For example, if you made a decision on [DATE 1] for [DATE 2], and then the same wage came up on [DATE 2] – did you make the same decision?

– Yes, always – Almost always – Usually – Half the time – Sometimes – Rarely – Never

7. When your decisions for the same day differed from your decisions ahead-of-time, why was it?

- I did not feel like doing as much as I had originally planned
- Something came up unexpectedly
- I felt like doing more than I had originally planned
- I realized that the Task was more unpleasant than I had originally thought
- Other: [text input box]

8. When the other subjects' decisions for the same day differed from their decisions ahead-of-time, why do you think it was?

- They did not feel like doing as much as they had originally planned
- Something came up unexpectedly
- They felt like doing more than they had originally planned

- They realized that the Task was more unpleasant than they had originally thought
 - Other: [text input box]
9. If you participated in this experiment again, do you think your decisions for the same day would match your decisions ahead-of-time more or less than the first time around?
- *Less – Same – More*
- Why? [TEXT ENTRY BOX]
10. How much do you think you will procrastinate on your school or work assignments over the coming month?
- *Very often – Often – Sometimes – Rarely – Not at all*
11. How much do you think your classmates or colleagues will procrastinate on their school or work assignments over the coming month?
- *Very often – Often – Sometimes – Rarely – Not at all*
12. How many times do you think you will go to the gym next week?
- *Once – Twice – Three times – Four times or more*
13. How many times do you think your friends will go to the gym next week?
- *Once – Twice – Three times – Four times or more*
14. How healthy do you think you will eat over the next few months?
- *Very health – Fairly healthy – So-so – Quite unhealthy*
15. How healthy do you think your peers will eat over the next few months?
- *Very health – Fairly healthy – So-so – Quite unhealthy*

You are finished with the study!

You have completed the entirety of the study “Doing Work over Time"! Thank you very much for your help with this experiment.

You will receive all of your earnings from the experiment via an Amazon.com gift voucher on [PAYMENT DATE].

If you would like to request a copy of the study you have just participated in, please check the box below. We will be sure to email you a copy when the study is ready.

[CHECK BOX] Please send me a copy of the study.

["CONTINUE" button]

Thank you

Your payment (and a copy of the study, if you requested one) will be emailed to: [PARTICIPANT'S EMAIL ADDRESS]

["EXIT" button]

C.4 Informed Consent

Study Title: Doing Work over Time

Researcher: Anastassia Fedyk

Participation is voluntary

It is your choice whether or not to participate in this research. If you choose to participate, you may change your mind and leave the study at any time. Refusal to participate or stopping your participation will involve no penalty or loss of benefits to which you are otherwise entitled.

What is the purpose of this research?

The purpose of this study is to see how much people want to work at different wages. We are interested in these work decisions at different points in time.

How long will I take part in this research?

You will need to participate on four different dates: today, [DATE 2], [DATE 3], and [DATE 4].

Today, reading the instructions and doing the mandatory work will take approximately 30 minutes of your time. Mandatory tasks on each participation date after today will require approximately 20 minutes of your time. In addition, you will have the option of doing extra work for additional payment. The choices of how much extra work to do will be up to you.

What can I expect if I take part in this research?

Today, you will first be introduced to the experiment. In order to be eligible for this study, you will need to read the instructions carefully. You will be quizzed to ensure your understanding of the experimental procedures. To be eligible to participate in the study, you must answer at least 8 out of the 10 quiz questions correctly after reading the experimental instructions.

You will then be asked to do a warm-up consisting of [5 / 10 / 15] rounds of work on the following Task. Each round of the Task consists of 60 seconds: 50 seconds of work followed by 10 seconds of rest. During the 50 seconds of work, characters appear on the screen one by one, at the rate of one character every two seconds. You need to press the SPACE bar or click on the "I can see it!" button every time an ASTERISK appears.

On future participation dates, you will also need to complete a warm-up consisting of [5, 10, or 15] rounds of the Task each time. After that, you will complete some number of extra rounds of the Task on each date. You can choose between 0 and 70 extra rounds for each participation date. Basically, on each day, we will present you with different wages – ranging from \$0.10/round to \$0.30/round – and ask you how much work you would like to do either immediately or on a future participation date at these wages. Then, on each day, we will gather all of your decisions for that day and choose one of them to implement. You will then do the number of rounds you had chosen in that decision.

For continued eligibility, you will need to achieve an average of at least 80% accuracy on the Task in each session. You will also need to be sure to log in by 10:30PM EST on each of your participation dates.

What are the risks and possible discomforts?

There are no foreseeable risks besides those that normally may be experienced while

working on a computer [5 / 10 / 15] minutes on each date, plus any additional rounds you choose to do.

Are there any benefits from being in this research study?

This study aims to obtain a better understanding of how people complete their work. We hope that the insights from this study will enable institutions such as universities and workplaces to offer more productive work environments and incentives.

Will I be compensated for participating in this research?

Yes, you will receive a \$30 completion payment. This is for logging in each week, and completing all warm-up rounds and all additional rounds in your implemented choices.

Furthermore, you will receive additional payment for doing the extra rounds in your implemented choices, at the corresponding wages (ranging from \$0.10/round to \$0.30/round).

[*For incentivized participants:*] Lastly, we will ask you to predict some of [your / other subjects? / your and other subjects?] future decisions, and pay you a \$[BONUS] bonus for every correct prediction.

Overall, the payment structure is as follows: [*The participants are shown Figure C.1.*]

[*Note: The last row appears only for incentivized participants, and differs across treatment groups – shown above is the view for incentivized participants in Group 2, who make predictions regarding others. The screens for participants in Groups 1 and 3 and analogous.*]

Once again, it is very important to note that in order to receive the \$30 completion payment, you must log in and do the assigned work with at least 80% accuracy on all participation dates, today, [DATE 2], [DATE 3], and [DATE 4]. If you miss or do not complete the work with at least 80% accuracy on one of your participation dates, you will still receive the wages for the extra rounds that you have successfully completed up until that point. However, you will be removed from the experiment, and forego your completion payment.

All of your earnings will be paid out in a single payment on [PAYMENT DATE], in the form of an Amazon.com gift certificate. Note that your email will be provided to Amazon.com to send the payment.

If I take part in this research, how will my privacy be protected? What happens to the information you collect?

Your name and email will be collected in order to contact you with reminders on your participation dates and to distribute payment. This information will be stored in a password-protected database and then transferred to a password-protected computer. All personal information will be destroyed as soon as the study is completed and the personal information is no longer needed for payment. Only de-identified data will be kept; these de-identified data will be stored indefinitely, in case of requests for further analysis by journal referees or other academics in the field. There are currently no plans for data transmission, but should other academics request the experimental data for replication purposes, the de-identified data will be made available.

In addition, since the study is administered online, you will be able to control the level of privacy you experience by choosing when and where you participate.

If I have any questions, concerns or complaints about this research study, whom can I talk to?

The researcher for this study is Anastassia Fedyk who can be reached at 609-755-4859; at Baker Library 244C, 25 Harvard Way, Boston, MA 02163; or at afedyk@hbs.edu for any of the following:

- If you have questions, concerns, or complaints,
- If you would like to talk to the research team,
- If you think the research has harmed you, or
- If you wish to withdraw from the study.

The researcher will be available at any time during the experiment, so please do not hesitate to contact her with any questions.

This research has been reviewed by the Committee on the Use of Human Subjects in Research at Harvard University. They can be reached at 617-496-2847, 1414 Massachusetts

Avenue, Second Floor, Cambridge, MA 02138, or cuhs@fas.harvard.edu for any of the following:

- If your questions, concerns, or complaints are not being answered by the research team,
- If you cannot reach the research team,
- If you want to talk to someone besides the research team, or
- If you have questions about your rights as a research participant.

Statement of Consent

I have read the information in this consent form. All my questions about the research have been answered to my satisfaction.

Copy for your records

Please print this page or retain a screen shot for your records.

SIGNATURE

By clicking "I AGREE" below, you will provide an electronic signature indicating your permission to take part in this research. If you would prefer not to participate, please click on "EXIT".

["EXIT" and "I AGREE" buttons]