COMPUTATIONAL METHODS IN THE STUDY OF INDIVIDUALS' ATTENTION ONLINE

A Dissertation

Presented to Faculty on the Student's Special Committee

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nir Grinberg

May 2017

COMPUTATIONAL METHODS IN THE STUDY OF INDIVIDUALS' ATTENTION ONLINE

Nir Grinberg, Ph.D.

Cornell University 2017

This dissertation uses computational methods to study individuals' attention online with the explicit goal of enabling information systems to support better use of people's attention. As consumption of information shifts to digital means, systems are playing a increasing role in shaping both the information we pay attention to and the practices for paying attention. Computer scientists are uniquely positioned to explore this unprecedented opportunity to design systems that impact millions of people, and support more efficient and effective use of human attention. However, incomplete measures of online attention and little research on the determinants of attention in online settings hamper the ability to design better information systems. To this end, this dissertation develops measures and methods to investigate individuals' attention online as it manifest in two of the most important domains of online activity: online news and social media. We devise new Web scale measures for capturing individuals' attention using non-invasive digital traces of online activity. In addition, we design novel computational methodology for studying the social, cognitive, and technological factors that affect attention online. Overall, this dissertation lays the foundation for assessing the impact information systems have on human attention, and provides guidelines for the design of better information systems in the future.

BIOGRAPHICAL SKETCH

Nir Grinberg completed his Ph.D. in the Computer Science Department at Cornell University in 2017. Prior to Cornell, Nir received a Masters of Science in Computer Science from Rutgers University in 2013, and a double major Bachelor of Science (B.Sc.) in Physics and Computer Science from Tel-Aviv University, Israel in 2010. Following the completion of his Ph.D., Nir continued to purse academic research as a Postdoctoral Research Fellow at the Harvard Institute for Quantitative Social Science (IQSS) jointly with the Lazer Lab at the Network Science Institute of Northeastern University.

At Cornell, Nir worked in the Social Technologies lab under the supervision of Prof. Mor Naaman, and was one of the first students to join the Jacobs Institute at Cornell Tech. During his Ph.D., he collaborated with researchers in both academia and industry, and held researcher positions in Facebook's Core Data Science team, Yahoo Research, SocialFlow Inc. and Bloomberg L.P. He frequently serves as program committee member in machine learning, social media, and HCI conferences including KDD, WSDM, IJCAI, ICWSM, CHI, CSCW, SocInfo, WebSci.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Mor Naaman, who taught me how to do research and has been a source of inspiration for me in shaping my identity as a researcher. Mor has been a phenomenal mentor in helping me navigate the hardships of pursing a Ph.D., always available and supportive, and radiant in his enthusiasm for research and can-do attitude. Mor had taught me how to think big, set ambitious goals for myself, and find practical paths for achieving these goals. I learned from Mor how to distill my ideas and communicate them more clearly. Without Mor's support and mentorship during the course of this Ph.D., I would not have achieved nearly as much as I did. Mor helped me grow as a researcher, a person, and a parent. I am deeply grateful for the tremendously positive influence Mor had been in my life and looking forward to many more years of thoughtful advice from him.

I would also like to thank the incredible members of my Ph.D. special committee: Daniel Huttenlocher, Dan Cosley, and Haym Hirsh. I am honored to have such an accomplished set of researchers on my committee, and humbled by the fact that they were able to find time to attend to my work. The advice and guidance I received from them helped me see my work in a new light, crystallize the contributions to the field of computer science, and redefine my research interests.

Many thanks to my other mentors, collaborators, colleague, and fellow researchers who helped directly or indirectly shape my thinking around the topics of this dissertation. In particular, I would like to thank Lada Adamic, Alex Dow, Shankar Kalyanaraman, Jeiran Jahani, Dan Valente, Josh Schwartz, Jonah Berger, Blake Shaw, and Gilad Lotan. I am thankful for all current and past members of the Social Technologies lab at Cornell Tech (and in its past incarnation as SMIL at Rutgers University) who contributed to a fun and conducive environment for conducting research.

Last but not least, I am extremely grateful to my family who supported me throughout this process. In particular, to my parents, who instilled in me a sense of perseverance from an early age and taught me to always strive for excellence. To my wife, Sarit Szpiro, who inspired me to pursue a Ph.D. in the first place, and whose love and encouragement all along the way enabled me to successfully reach the finish line. To my lovely daughters, Tamar and Rotem, who helped me find new purpose in life and see beyond this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The widespread adoption of platforms for consuming and sharing information offers a unique opportunity to broaden our understanding of individuals' attention and to design better systems for processing information. With more and more of human activity mediated by computers, and with finer measures of user interaction, we can begin to draw inferences about individual's attention in real-world settings, identify factors that affect allocation of attention and its sustainment, and devise guidelines for designing systems that enable more effective use of people's attention.

Our working definition for attention is based on the one given by Davenport and Beck, which captures the common elements of most functional models of attention: "Attention is focused mental engagement on a particular item of information. Items come into our awareness, we attend to a particular item, and then we decide whether to act" [65]. Therefore, attention involves the selection to further process some information but not other, and the decision of individuals to direct this selection process over time. The definition enables us to focus on two fundamental aspects of attention that exist in all user-facing information systems, allocation and sustainment of attention, while abstracting the targets of attention that may change from one domain to another.

It is instrumental to return to the famous words of Nobel Prize laureate, Herbert Simon, in order to fully understand the importance of studying attention in the realm of online information systems: "a wealth of information creates a poverty of attention and a need to allocate that attention efficiently" [202]. We will return to Simon's first assertion about the poverty of attention later, but let us first unpack his second argument about the need to allocate attention more efficiently. What is an efficient allocation of attention? Efficiency requires a target, an objective or a goal. By carefully studying how individuals pay attention online we can learn about their objectives and goals, and design better systems that adapt to people's dynamic needs. In a constantly changing world, the effectiveness of online information systems hinges on the ability of these systems to sense changes in people's attention and respond accordingly.

No other medium provides a greater "wealth of information" and a potential to improve the experience to so many people as our digital information systems currently do. With close to 50 percent of the world's population connected to the Internet[1], search engines offering instantaneous access to exceeding amounts of information, and computers mediating much of human communications – the role of technology in shaping the human experience cannot be underestimated. In 2015 alone the average adult in the US spent close to five and half hours each day consuming digital media, about a third of people's awake time during the day and more than an additional hour spent on any other form of media[2]. Last July Pew research estimated that 38% of Americans often get their news online, which is almost double the amount of people who read it in print [162]. If current trends persist, people are likely to spend more time consuming media in the future, shifting from traditional mediums like television and print to digital platforms. Therefore, improvements in the efficiency of information systems can greatly affect the lives of millions of people in the near future and many more people in the long haul.

Information systems, through people's attention, have far reaching implications both for the individuals paying attention and for society as whole. At the individual level, attention to online information affects people's perceptions and actions. Exposure to online information (like other forms of information) was shown to

[1] `http://www.internetworldstats.com/stats.htm`

[2] `https://goo.gl/GSqgJ7`

contribute to people's mental representation of the world, anywhere from affecting people's emotions, opinions, and attitudes to impacting the perception of self and others [11, 37, 62, 97, 109, 179]. Information from weak ties is considered one of major ways through which social media provides new opportunities to individuals and boosts social capital [77, 98, 143]. Once information is processed it can change the course of action: guiding subsequent behavior and affecting interactions with others [5, 78, 167, 215]. At the societal level, there is growing awareness (and concern) about the aggregate effect that information systems have on society-level outcomes such as political polarization and civic engagement [9, 25]. Just recently Facebook was the subject of public scrutiny over claims that the company's feed ranking algorithm contributed to the spread of fake news during the US presidential campaign, which tipped the election results in favor of elected president Donald Trump[3]. Without people's attention none of the aforementioned outcomes could have materialized. Thus, as computer scientists who design information systems we have the moral and social obligation to study the implications of these systems on people's attention, which will ultimately lead to improvements in existing systems and future ones.

A lack of gold-standard measures and methodologies for studying online attention prevents a more rigorous examination of the relationship between information systems and attention at Web scale. As we describe next, better measures and more robust methodologies can enable better assessment of bias in existing information systems that will facilitate curbing inefficiencies in future systems.

Systems that use incomplete measures of attention create a biased market for ideas. If indeed we live in the era of the Attention Economy [65], information is competing for the scarce commodity of human attention. However, if information

[3]`http://fortune.com/2016/11/10/facebook-blame-trump/`

systems consistently underestimates (or overestimates) the value of certain types of content they introduce a bias into the market. Crude measures of attention online do exactly that, equate content that received substantial allocation of attention with those that only got a quick glimpse. For example, consider the case of *Clickbaits* – stories with sensationalist headlines that are not informative of the actual story content [75]. Clickbaits lure individuals to visit the story page more often than other stories, produce higher click-through rates, and lead recommendation systems to further increase exposure of Clickbaits to other individuals. The absence of more substantive measures of attention in systems that optimize for higher click-through rate gives unfair advantage to *clickbaits* over content that merits deeper forms of engagement. Similar argument can be made for ranking based on other incomplete measures such as *likes*, *comments*, *shares*, and the combination of such measures. The more substantive measure of time spent on a page, known as dwell time, was found to be effective against *clickbaits* and helpful for recommendations [75, 229]. However, a measure of time spent on a page does not take into account the characteristics of the content, the individual paying attention, how attention is distributed on the page, or the context of individuals' actions. The development of more complete measures for online attention is crucial for evaluating bias in existing systems and paving the way for correcting these biases in the future.

Moreover, lack of methodologies for robustly identifying the factors affecting attention online impedes systems' ability to adapt to the changing needs and desires of people. Knowledge about the determinants of attention was established so far mostly through careful lab studies, which serves as the basis of our investigation. However, this foundation of knowledge does not immediately translate or generalize to online attention that takes place in more diverse, varying, and complex sociocultural environment. For example, lab studies showed that cognitive load plays a role in low-level attention processes [138], but theoretical constructs such as cognitive load are not readily available to system designers outside lab settings. Even if such constructs were easy to quantify, controlling for the large exogenous variation present in natural online settings requires careful methodology. Due to these methodological challenges, the factors affecting attention in even the most basic online tasks performed by millions of people every day, such as reading news articles or communicating with friends, are not yet specified or well understood. Without the ability to capture the precursors of attention change, information systems miss an opportunity to direct people's attention more efficiently and *design for* the dynamic nature of human attention.

Several aspects of attention make it an extremely difficult research subject. First and foremost, measuring individuals' attention is a difficult task even in well-controlled lab environments. Attention is a complex latent cognitive process that is not directly observable and is intertwined with many other cognitive processes: from motor control and language processing to perception and memory [54,139,181, 199]. As a result, one can only study attention through the traces of attention being paid such as eye movements, activity in certain brain regions or people's ability to retrieve infromation from memory. Different individuals differ significantly in their ability to use attention effectively, which also depend greatly on task difficulty and cognitive load at the time of tackling the task [48, 138, 207, 214]. As previously mention, attention is a highly dynamic process, which makes capturing it over time difficult since attention can shift quickly [80, 168].

Measuring attention through people's use of online systems introduces additional challenges. Recorded online activity is often ambiguous with respect to the underlying user intent and several steps removed from the psychological process behind it [102, 184]. A *like* on a news article shared on Facebook may express an agreement with the author, support of the subject of the article or simply an acknowledgment of noticing the post. A long amount of time spent on a page with no interactions could represent careful reading of the content or attending to another stimuli or a distraction. It is not exactly clear what signals like Facebook *likes*, time spent on page, and other online measures carry for attention, especially compared to more direct measurement of attention in lab settings. Further complicating the matter is the fact that online services are subject to rapid change both in their functionality and user base. The quick pace of development of new features and their differential adoption by people makes it hard to assume that the demand or supply for attention are static over time. Even if system affordances and population remain the same, the ways people utilize existing capabilities change over time [133]. Therefore, measuring attention through people's use of online systems requires careful consideration of the context of users' actions, the changing affordances of online platforms, the people who actively use these platforms and their objectives.

In addition to the challenges in measuring online attention directly, there are methodological issues in studying the determinants of attention online. Many of the factors identified as affecting attention in lab studies are equally difficult to measure online as attention itself. For example, quantifying the cognitive load associated with online actions by different people at different times is as complex as measuring individuals' attention. Small-scale experiments of online attention suffer from the same limitations as other lab studies: limited ability to generalize beyond the particular setup, inherently small study population, and questionable ability to capture the true relation between different factors and attention as it manifests in natural settings. Large-scale online experiments are costly and face growing public scrutiny over ethical concerns [103, 130, 208]. It is also difficult to argue that informed consent and participation in online experiments do not affect individuals' attention in any way. Moreover, randomized experiments are not suitable for all type of research questions, in particular those that involve counterfactual conditions [119]. For instance, an experiment cannot answer how attention would have changed if an individual had not taken certain actions (e.g. posting a message to friends). Methods for causal inference using observational data are making strides in addressing counterfactuals and recovering causal relations [6, 182]. However, causal inference on large datasets of observed behavior is still very much an open computational problem, let alone inference of causality on a latent processes such as attention that are not directly observable.

Furthermore, the large scale of information systems we wish to impact by this dissertation work requires us to constrain the computation of the measures and methods we propose. Information systems such as the Facebook News Feed tackle an incredibly difficult computational problem – deliver personalized ranking to many millions of people every day, and do so in a split of a second [8]. Further complicating the problem is the fact that most content is new[4]and unlabeled. The sheer size of new content being generated on social or news media platforms requires models to be updated frequently in order to capture emerging "trends" as they happen and produce up-to-date and relevant information. Thus, machine learning models that support systems at this large scale need to be strictly linear in the size of input, and highly efficient in both the model training and generation of predictions. Generalized Linear Models (GLM) such as linear or logistic

[4]On Facebook alone there was an average of 4.75 billion new posts *every day* in May 2013, an increase of 93% from August 2012 `https://goo.gl/vx5CMV`.

regression are among the most widely used models for large datasets since they are well understood and interpretable, they perform comparably to other more complex models, and have scalable inference algorithms with good convergence properties [26, 84, 88]. Therefore, we focus in this dissertation on measures that can be computed in real time or near real time, and use GLM over more complex models in order to make our findings easily portable to a wide range of existing systems.

Before describing our contributions in the study of online attention, it is imperative to describe prior work and identify areas where the literature is still lacking. There are four lines of related work, with the first two focused on the key issue of measurement – how can the attention of individuals be quantified. First, we describe the "gold standard" methodologies used in psychology to measure individuals' attention and then describe their extension into the measurement of attention through people's use of digital media. A third line of work covers the literature on using implicit signals of attention to improve information retrieval, personalization, and recommendation systems. Last, we describe research on diffusion processes that both motivated our investigation and can benefit from a more complete description of individuals' attention.

The first and most comprehensive description of individuals' attention can be found in the psychology literature, which provides the foundation of knowledge about human attention. Psychologists established the measures for quantifying attention in lab settings and mapped out various factors that affect attention. Over the years, the methodologies for measuring attention increased in their level of sophistication. Early studies used comprehension of information as a measure of attention. For example, participants were asked to direct their attention to only one of two audio messages played to them, and were then tested on their ability to recall the messages [28,29,50]. Recent studies in psychology use more advanced measuring techniques such as eye-tracking or functional magnetic resonance imaging (fMRI) to identify determinants of attention processes [81, 124, 157, 177, 224]. For example, arousal and emotional valence are among the factors linked to increased focus in visual attention [148, 185]. Despite the fact that this large body of research does not directly translate to or fully describe online attetion, it did inspire some of the hypotheses we test in Chapter 3 and the methodologies we use for measuring attention in Chapter 2.

A second line of work, which extends the first in psychology, is focused on the measurement of attention through individuals' use of online systems. Several studies developed explicit measures and computational models in order to quantify people's attention. For example, studies examined how people divide their attention across different page elements on the web. Buscher et al. conducted eye-tracking experiments in order to infer salient regions of web pages [42]. Other works set to infer attention using mouse cursor activity, first by linear models relating eye-mouse positions [110], then through non-linear transformations [171], and more recently using more complex mixture-models [131]. Studies that sought to connect user interactions to high-level cognitive tasks such as reading or skimming of web pages relied on modified text layout, expensive eye-tracking, and manual labeling by experts [21, 44, 173]. Chapter 2 in particular takes somewhat similar approach, and proposes a computational model that could joint learn from small-scale labelled data and much larger unlabeled data observed outside lab settings.

Other research used implicit signals of attention to improve the performance of information retrieval, personalization and recommendation systems. One of the most common measures of implicit attention is a click on a page hyperlink. Clicks were shown to be highly effective in improving the performance of search engines [115], and recommendation systems using Collaborative Filtering [146]. However, as mentioned before, clicks are also very crude proxy for attention, which give rise to undesired phenomena such as Clickbaits. A growing body of work investigates the potential of post-click measures of attention. Dwell time, the amount of time a user spends on a page, has been shown to correlate with the explicit ratings people give to content items and as a good indicator for satisfaction with a given search result [56, 125, 145]. Guo and Agichtein showed that speed of scrolling can improve estimates of document relevance in information retrieval tasks and Yi et al. demonstrated that dwell time can further improve click-optimized recommendation systems [102, 229]. This line of work demonstrates that better understanding of online attention has practical implications for recommendation systems, personalization and search. This dissertation will introduce measures and methods that could contribute to this line of work and similarly improve a wide range of information systems.

Finally, models of diffusion processes and collective attention can both inform and benefit from better models of individuals' attention online. Numerous works modeled the spread of information in networks (see [172] for a comprehensive overview). Several studies focused on examining the factors that affect the diffusion of content in social networks [69, 95, 111]. For example, Sharad et al. found that the propagation of viral content depends on the type of content (e.g. petition versus news) and the content's medium (e.g. image versus text). Other works examined the properties of collective attention to items. For example, Wu and Huberman showed that the attention a story on the site digg.com receives (measured in total number of "diggs") can be described by a single novelty factor [225].

The same factors affecting diffusion and popularity of content may also affect individuals' attention since sharing a story requires at least a minimal amount of attention. However, sharing and paying attention do not necessarily coincide – widely shared articles are not always backed by significant amount of attention at the individual level and vice versa. Clickbaits are a prime example of content that spreads widely but does not warrant substantial amount of attention. Other types of content such as long-form articles may require significant amount attention to fully comprehend but result in little circulation. Therefore, closer investigation of individuals' attention can both benefit from existing literature about diffusion processes and collective attention, and contribute to this body of work through more precise description of processes at the individual level.

This dissertation aims to fill the gap in prior literature with respect to measurement of individuals' attention online, and identification of the key factors affecting it. Addressing these issues enables information systems to learn more accurate representations of people's interests and adapt more quickly to changes over time, which brings us back to the definition of attention by Davenport and Beck. As described earlier, attention involves the selection of further processing some information over other information, and the direction of this cognitive process over time. Accordingly, the studies at the core of this dissertation provide a new persepctive on the allocation and sustainment of attention by examining people's interactions with online systems. While attention is an important component for any system that delivers information to people, the two parts of this dissertation focus on two prominent domains of information online, news and social media, that can benefit the most from careful modeling of individuals' attention. By studying these domains we examine factors that affect how people get to information in current day and age, and the extent to which they consume that information, as we describe next.

The first part of this dissertation focuses on attention to online news media. As mentioned earlier, a large percentage of the US population gets their news online, a shift that disrupted the traditional model of journalism in many ways. One of the most common claims about the digitiazation of news consumption is that technology hampers people's ability to sustain attention. This concern was eloquently articulated in an essay by playwright Richard Foreman[5] and described at length in Nicholas Carr's book "The Shallows" [46]. The central idea is captured by Foreman's critical comparison of modern-age information consumers to "pancake people": spread out in their coverage of information but shallow in depth. Herbert Simon's "poverty of attention" can be viewed as pointing to the same effect – abundance of information leads to poor quality of attention, which directly ties to people's ability to sustain attention in reading. Anecdotal evidence that supports these ideas is the decline in book sales[6] and reports showing that the attention span of people is shortening [61]. It is important to note that the alleged diminishing ability to maintain attention is not a force of nature: it is the result of man-made systems and human culture that developed around information consumption. While a tremendous amount of work on recommendation and personalization systems for news focused on getting the right article to the right person, relatively little computational work examined what is actually read and what are the ingredients that affect reading after clicking on an article. The first two studies in this dissertation aim to establish the foundation of understanding needed to build better measures of online reading and better systems that foster good habits for consuming online news.

[5] https://www.edge.org/3rd_culture/foreman05/foreman05_index.html

[6] http://newsroom.publishers.org/publisher-book-sales-were-537-billion-in-the-first-half-of-2016/

In addition, online news is a good point in case for studying how people maintain attention in reading online. Reading is one of the most fundamental cognitive tasks that people do online, and reading of news offers a good balance of different tradeoffs in studying it. First, reading news is a common task that involves different levels of focused attention (e.g. skimming, reading in-depth) that may help understand the consumption of other forms of textual content (e.g. books, manuals, instructions, e-mail) and perhaps even other forms of consumption online (e.g. watching videos, audiobooks, e-learning). Reading news online requires people to sustain attention, since processing a news article cannot be done in a glimpse, but usually does not span more than a single reading session. Focusing on just a single session simplifies the research subject, makes it much more self-contained, and eliminates many of the externalities that may occur in between sessions (though, acknowledgedly, not all externalities). News articles also vary in content considerably and are often read by multiple people, which enables us to study jointly how different people sustain attention to different types of content. The two studies in this part investigate the key factors associated with sustained attention to online news and develop new means for measuring different modes of attention in reading.

The first study about attention in online news introduces a simple new measure for quantifying attention to news articles, demonstrates how this quantity varies across people and articles, and assesses the ability of this new measure to be used in practice. The work complements existing and widely used measures of implicit attention to content online such as clicks, likes, and dwell time [102,125,145,146,229]. The study focuses on an often overlooked aspect of user engagement with a news article: the user's scrolling within an article. We use the scroll depth in a subset of interactions with a news article[7] as a proxy for the amount of attention sustained in

[7] A subset of interactions that are more likely to reflect reading of the article content.

an article. We examine the relation of our measure to existing measures and show how scrolling depth depends on different factors such as the reader's past behavior, the article topic, and more. We find that our new measure captures important variations in attention that are not captured by existing measures. Last, we evaluate the predictability of scroll depth on held-out articles and out-of-sample individuals, prior to publication time and shortly afterwards in order to demonstrate the potential of this new measure for recommendation systems.

The second study about online news extends the first one by taking a closer look at detection of reading from user interactions with online news article pages, and propses a principled computational model to distinguish different reading modes outside lab settings. The study contributes to previous work on explicit measures of attention [21, 42, 44, 110, 131, 171, 173] by developing a new explicit measure for individuals' attention to textual news content. The work also extends our first study and other implicit post-click measures [102, 125, 132, 145, 229] by taking a more direct look at reading as a process, thus modeling individuals' attention in reading with fewer assumptions about the way individuals spend their time on the page. Our goal is to develop a new measure for detecting reading modes of textual news content that has tighter guarantees for its validity outside lab settings. To that end, we devise a semi-supervised model that discerns in-depth reading of paragraphs from skimming and other forms of interaction online. The model draws on patterns of engagement with news articles outside lab settings (unlabeled) as well as paragraph-level labels obtained in lab settings. The goal of this model is to become the gold-standard for estimating the extent to which people pay attention to the content of news articles online.

The second part of this dissertation investigates attention in online social settings. We focus on attention to social media and on social media because of two main reasons: the major role social media platforms have in the dissemination of information, and the increasing amounts of content calling for people's attention on these platforms. Just last November, Pew research reported that 79% of online Americans are Facebook users, a 7-percentage-point increase from the previous year [100]. As Facebook and other social media platforms continue to grow, more content is being generated on these platforms[8]. Whether the rapid growth in the production of social content is due to increase in reach, density, or prominence of social networks as a medium for information diffusion, it is evident that major social network sites are turning to algorithmic ranking as a way to help people direct their attention more efficiently[9,10]. From the perspective of individuals who post content on social media, the abundance of information on social media poses a challenge to their the ability to be heard by the people they care about. The two studies in this part investigate the factors that affect the allocation of attention in social network sites as well as the expectations for attention from others. Addressing these questions will contribute to the understanding of the changing goals, needs, and preferences of people in online social setting. In addition, more accurate knowledge about the allocation of attention paves the way for the development of more adaptive ranking algorithms that incorporate this knowledge in system design. Both outcomes will help fulfill the need, identified by Herbert Simon, for information systems to direct attention more efficiently in an environment that is increasingly flooded with information.

[8]For example, Twitter in 2013 had to redesign their systems in order to support a record high of 143,199 new tweets *per second* and more than 500 million tweets per day `https://blog.twitter.com/2013/new-tweets-per-second-record-and-how`.

[9]`http://newsroom.fb.com/news/2016/06/building-a-better-news-feed-for-you/`

[10]`https://blog.twitter.com/2016/never-miss-important-tweets-from-people-you-follow`

The first study about attention in online social settings examines how the distribution of attention changes in different circumstances. We examine the allocation of attention to Facebook and on Facebook from the same individual at two different times – one when she composed an original post and the other when she liked, commented or shared someone else's content. We devise a quasi-experimental methodology to robustly study the changes in the distribution of attention from large-scale observational data. By drawing this comparison we show that the allocation of attention varies considerably based on the context of users' actions, suggesting that an underlying social and cognitive state is different. Moreover, we show that the allocated attention of others is valued and appreciated at large scale. The study provides a new perspective for understanding individuals' attention in context, and complement previous research that relied on self-reported measures (e.g. [43, 109, 154]). The study also highlights how the understanding of attention allocation can potentially better system design. In particular, our findings can be used to design nudges for contribution, introduce new measures in recommendation of social content, and differentially value feedback from contributors.

Changes in the allocation of attention on social media affect not only the individual who pays attention but also the people receiving that attention. The last study in this dissertation focuses on the expectations people have for getting attention from their online social ties. Prior research mostly approached the abundance of information in social media as a challenge for ranking algorithms, much like a query to a search engine [4, 20]. Only few works studied people's own preferences when sharing content in online social settings and people's ability to accurately target their friends interests [18, 198]. This work examines the often overlooked end of the attention "transaction": the content producer's expectation to be heard. This study offers both a conceptual framework for thinking about attention expectations and a computational model that can be used in practice. Our conceptual framework describes the factors affecting expectations for getting attention from others and the implications for the individual of fulfilling those expectations. By combining surveys and observational data analysis we provide evidence for the relation between observed behaviors and expectations. We also show that the fulfillment of expectations contributes to people's sense of connectedness to their friends, an important outcome for individuals' well being [41,128]. In addition, our simple and easily portable predictive models demonstrate how the expectations of content producers can be incorporated in recommendation systems in practice, a necessary step towards building less ego-centric recommendation systems for the consumption of social content.

In summary, this dissertation offers a computational perspective to the study of individuals' attention online with the explicit goal of improving the design of information systems. The abundance of information and wide adoption of information technologies create an unprecedented opportunity for systems to support more efficient and effective allocation of people's attention. As computer scientists who design information systems we have the opportunity, and moral and social obligation to study the relationship between human attention and information systems, which have consequential outcomes for individuals, communities, and society as a whole. Several aspects of the problem space make this into an extremely difficult research area: from the inherently dynamic nature of human attention, through methodological issues in measurement and inference of latent cognitive processes, to scalability issues of computation in systems that support millions of people at the same time. Our work builds on the foundation of knowledge about attention in psychology, the literature on explicit and implicit measures of attention online, and more broadly on works about diffusion process and collective attention. Equipped with this knowledge, we examine the allocation and sustainment of attention of individuals in prominent domains of online activity: online news and social networks. The studies in this dissertation advance prior work by devising new measures to quantify attention, and develop novel computational methods to robustly study the factors that affect attention in online settings. Our findings highlight practical ways for information systems to better assess individuals' attention and effectively contribute to the emergence of more attention-friendly information systems in the future.

Part I:

ATTENTION TO ONLINE NEWS

The two chapters to follow revolve around individuals' attention to online news. Reading, as a form of sustained attention, is an important skill that impacts people's knowledge, world view, and ultimately their ability to learn. Reading of news is particularly important for an informed civic society and a healthy democracy. As information systems take a bigger role in affecting what people read and how they read it, it is critical to develop measures and methods to assess reading as it happens in this new medium. The studies in this part focus on reading of online news as a point in case for starting to unpack the complex relation between technology and people's ability to sustain attention online.

The first chapter in this part examines a simple, yet often overlooked, measure of scroll depth as a proxy for reading in online news and the factors that affect it. The second chapter eliminates some of the assumptions of the first chapter about when people read, and directly models reading using non-invasive traces of user interaction with online news articles. The part as whole offers methodology for modeling reading of online news and highlights how information systems can utilize our findings to better support reading and readers.

CHAPTER 1

THE INGREDIENTS OF LONGER READS

1.1 Introduction

Our reading habits have turned from physical media (books, magazines and newspapers) to their digital counterparts (e-readers, websites, and apps). As a result, there is increasing opportunity to understand and model reading behavior as it occurs outside lab settings. Where previously publishers had to rely on gross measures of success such as book sales and page views, we now have finer and finer measures of individual readers' engagement with the content itself. For example, e-readers can tell us the exact pages that people read, and which parts of the text they highlighted. In our case, online news sites can track how much of an article's content was visible to the reader, and how deep the reader had gone into the article page. While these metrics are increasingly available, they are not yet well understood.

The Web, of course, has a long history of tracking user behavior on Web pages for purposes of improving usability of services or ranking of content (e.g., [102,125]). Eye tracking studies such as [47, 178] provided the foundation for understanding how users allocate their attention to Web pages. Other lab experiments instrumented browser behavior [219] in order to study user engagement with pages and search results. Using such instrumentation, Nielsen [173] found that people only read 28% of the content of an average Web page. Large-scale studies of online user engagement often apply machine learning and data-mining techniques on server logs (e.g. [3]) or client-side plugin logs (e.g. [57]). Server logs ignore the subtlety and richness of client-side interactions with a page such as scrolling or highlighting.

Client-side plugins often have limited user base, which may yield a biased estimates of user behavior. Previous approaches looked at user interactions with web pages in general (not just news media), and focused mostly on tracking visits and clicks, not reading depth of content.

The opportunity to understand reading depth can help writers, publishers and system designers (e.g. of search engines and recommendation systems) understand the factors that impact engagement with content. For example, using dwell time, the time users spend on a page, was shown to improve personalization [229] and help estimating user satisfaction from search results and improve relevance [102, 125]. Content recommendation systems, such as Facebook's newsfeed, are using dwell time to reduce the salience of *Clickbaits* [75]. Better modeling of reading behaviors can provide another dimension for models of virality and content propagation in a network [15, 104, 149, 204]. For publishers and writers, better understanding of user engagement can inform editorial decisions and help writing content that is better received. More broadly, finer measures of interactions can shed light on sustained attention in reading an article, guide efforts to increase reader satisfaction, and perhaps contribute to long-term success of media services.

In this chapter we present the first examination of a large-scale dataset of individuals' scrolling depth within news articles. We use a select set of 2.3 million user interactions with online news articles from eight different popular news publishers. We provide a first look into reading depth distributions and investigate how reading depth is affected by the article length, the device used, and how the reader got to that page (e.g. via search or social).

Moreover, we formulate a prediction problem for reading depth of an article, and investigate the features and models that help prediction. Our features include, in addition to article length, article metadata such as the site it was published on and author history; content features such as topic, readability measures, and sentiment; and features of readers such as device used and referral information. In particular, we are interested in predicting reading depth based on data available before and shortly after the article was published. We study the factors associated with both the article averaged reading depth and the reading depth for individual reads (i.e. predicting how deep a specific person will read a specific article). We compare the predictions of a standard Linear Model and Beta-Regression model, and highlight the factors that consistently and significantly affect reading depth.

Our contributions are therefore:

- A first large-scale examination of the factors associated longer reads.
- A model to predict average article reading depth.
- A model to predict reading depth for individual reading event.

We begin by surveying the relevant work, both offline and online, about reading of news.

1.2 Related Work

Naturally, the interest in how people read and respond to individual pieces of content has a long history, in both offline and online settings. Recently, we have seen renewed focus on reading and engagement measures, partially due to availability of data, e.g. from e-Readers [86]. At the same time, new Web and mobile technologies exposed some of shortcomings of simple measures like clicks and page views, and brought renewed interest in understanding reading behavior as a proxy for attention (and thus, importance and relevance).

The study of how well readers receive content dates back to the birth of mass printing [2, 55], when reading became the commonwealth of the ordinary people and with it, the study of reading practices. Most research on reading focused on improving processes of acquiring reading skills, assessing reading "level" and comprehension [59, 60, 140]. Reading decisions, such as when to stop reading, received relatively little attention, mostly in studies comparing reading on paper versus screen [175].

The development of eye-tracking devices brought a new level of rigor into the study of attention and reading practices, first examining reading in traditional print media and later in digital media. Even before digital media, eye tracking emerged as a powerful tool for studying visual attention [227]. When the digital platforms became more prevalent, eye tracking studies began to address attention on those platforms as well, initially focusing on studies geared to "optimize" outcomes such as users paying attention to the right information on the page [178]. Other small-scale studies focused on reading entry points, reading paths [107], and reading behaviors including reading in-depth and scanning [106]. In addition to eye tracking studies, other qualitative studies compared reading practices online and offline [140, 153]. Only a few studies examined text-reading practices online at Web scale. The study described in this chapter takes a first look, at scale, into the consumption of written media in "natural" settings.

In digital media, page views and clicks have been the dominating metrics used to quantify engagement with articles, but those incomplete measures fail to capture subtlety of the attention paid to content once clicked. Page views have been used for anything from studying the distributions of user activities on the Web [96] to recommendation of content (e.g. [64]) to ranking (e.g. [116]). Yet page views or clicks being the sole metric leaves out post-click behaviors therefore does not capture user engagement fully. For example, emphasis on clicks had led to the rise of Clickbaits as mentioned before. In news media, journalists have began to use web analytics to adjust their writing and experiment with headlines [197, 205]. Clicks as the only metric has the danger of resulting in more sensational and forward-referring headlines (and perhaps text) that offer inferior user experience [23, 164]. Clearly, the media is ready to explore more meaningful measures of engagement and attention, and here we explore one such measure.

Recently, several studies have focused on more advanced measures such as dwell time and cursor movements, and their usefulness was shown in a number of contexts. Longer dwell time has been used as an indicator for search satisfaction to improve search quality [125,145]. A study of cursor movements of 30 participants further showed speed of scrolling can distinguish reading and scanning behavior and predict document relevance [102]. Several studies that tracked user behavior over time on the Web [53, 57, 219], but used small samples of users and did not focus on news media. Based on the data from [219], Nielsen [173] estimated that people read 20-28% of the content on average web page (again not media) but the analysis was limited to pages between 30 and 1,250 words – different use case than media, where people often intend to read an article when they land on the page. Our work extends previous work by studying depth as a post-click behavior measure that compliments and refines dwell time. To the best of our knowledge, this is the first large-scale study to examine the depth variable with real world behavior data.

Richer page metrics are not just useful for assessing relevance but could also significantly improve the quality of search and recommendation systems through collaborative filtering. Collaborative filtering uses implicit behavior data of similar users and clicks proved to be highly efficient [146]. For example, according to [229], incorporating dwell time as a proxy for user satisfaction into Yahoo's recommendation system yielded better performance than click-optimized system.

Another relevant line of work studied properties of content that affect people's response to it, for example by sending a link to their friends. Diakopoulos and Zubiaga showed that socially deviant content is more likely to be shared by gatekeepers on Twitter [67]. Another work showed language features that are more likely to result in a Twitter retweet [204]. Emotional (both positive and negative), and physiological arousal was shown to increase sharing of New York Times articles [15, 16]. While sharing provides good signal for studying network propagation of content, it does not necessarily provide accurate picture of users' engagement with the content itself. In other words, sharing data is valuable but cannot help assess the attention or quality of an article in a way that reading depth perhaps can.

1.3 Dataset

Our dataset consists of more than 2.3 million likely *page-read events* (or *read events* for short) for articles published on eight popular online news sites, collected by Chartbeat[1]. The sites cover a wide range of topics (daily news, finance, sports, technology and science); target audience (women[2], local or subscribers only), and form (short and long form). A likely page-read event $e(r, p)$ occurs when a reader r

[1] http://www.chartbeat.com

[2] as defined in the publication's tagline

loads a publisher article page p and has some minimal amount of interaction with the page as we define in the next section. For each read event, we have information about the page (URL, contents, metadata such as author and published time, etc), and the user (a user ID that uniquely identifies a user within site). Most importantly, we have a proxy for the user's sustained attention on the article page, the reading depth, which we define next.

1.3.1 Measuring Reading Depth

The key measure we focus on in this work is the *reading depth*: how far a user scrolled down the visited article page. Since scrolling through a page does not guarantee reading we only work with a subset of page interactions that we call *read events*, and that are more likely to reflect actual reads. In order for a user interaction to be considered a read event it has to fulfill certain criteria, described in detail below. Notice that even in well-controlled eye-tracking experiments, determining what an individual actually read it an extremely difficult task (as we investigate more closely in the next chapter). While we cannot verify that our likely read events represent reading in every single case, we believe that the selection criteria makes it likely that, on average, deeper reads involve more sustained attention.

We obtain the reading depth by taking the maximal pixel position a user reached on the page as measured by Chartbeat's client-side logging system[3], and converting it into relative portion of the article's content. The *pixel depth* at any point is the top pixel position within the page that is visible on the user's screen.

[3]The data thus excludes certain browsers, and users that are not allowing javascript code to run.

For example, a user looking at a page that is 1600 pixels long would start at pixel depth 0. As they scroll down, the pixel depth will increase, potentially up to 1600. We developed a custom CSS selector, unique to each site, to identify the page content and compute the length of content. For each of the articles in the dataset, we visually rendered the page using the PhantomJS javascript library, extracted the text and the top pixel position of the last text paragraph $l(p)$ (of article p). The pixel depth of reading event $e(r, p)$ was turned into relative measure of reading depth by taking the maximal pixel depth of $e(r, p)$ and dividing it by $l(p)$. We manually verified that the article length $l(p)$ is accurately retrieved for several articles from each site in our data.

In order to focus on reading events and provide a valid comparison of reading across devices and platforms, we follow a four-part process that we describe next.

The first filtering step of the raw interaction data from Chartbeat was to require that the article page was visible on the user's screen for at least six seconds. At an average rate of reading of 250 words per minute a person reads about 25 words of the article, which is less than five percent that articles in our dataset.

Second, we only include page views where there was at least one user interaction with the page. In other words, our dataset of reading events consists of page visits where the user interacted with the page at least once via a mouse click, key-stroke or any other form of input. This condition eliminates cases where the user left the page open for more than six seconds, but did not engage with the content – an action that involves some page interaction [47, 173].

Third, we exclude likely reads where the user scrolled below the article content (i.e., explored page elements beneath the end of the text like ads, recommendations, comments or other page elements). We exclude these interactions because we believe that a considerable number of people read some portion of the text before skipping to the comments, recommendations, or other useful page elements that appear after the content. Since we only have the maximal pixel depth of people who view pages we cannot prove or disprove this claim, and thus choose to exclude scrolls that went beyond the content. While this decision may eliminate some legitimate reads of the content, we are more certain that the interactions included are likely to reflect reads. Therefore, our definition of reading events consists of cases where users reached a point before the end of the article's content or right at its end.

Lastly, we only include article pages with a known and fixed page layout that has no advertisement in between paragraphs. We wrote a custom set of rules for each site for extracting the content of articles from their HTML structure. Using these rules, we eliminated non-article page visits (e.g visits to the Sports section) and non-textual articles (e.g. video or image only). In addition, all sites in our dataset were "non-responsive", meaning that the pixel depth remains constant across devices, screen resolutions and window sizes. Non-responsive pages do not scale content or re-flow text based on the user's screen resolution or browser window size. Last, we also exclude sites that have inline ads breaking up the content as these are likely to disrupt the flow of reading.

Overall, our measure provides an upper limit for the proportion of content read. In order to increase the likelihood of studying actual reads, we focus on a subset of pages and page views that passed our four-step filtering process, which we call reading events. Our dataset consists of user interactions with articles pages from eight major news publishers, where the article was visible for at least six seconds, had at least one user interaction with it, and had not been scrolled past the content. Pages in our dataset render consistently across devices, screen resolutions and window sizes.

1.3.2 Page and User Metadata

In addition to reading depth, for each page in our dataset we collected and computed various features. The features came either from Chartbeat directly or from analyzing the article text as downloaded from the Web. For example, for each read event we have a unique (per-site) user identifier from Chartbeat; other information about the user-agent such as browser and device type; and referral information on where the reader came from to the article. We also downloaded and analyzed the article text for each article in our dataset and provide more details on these features in Section 1.5.

1.3.3 Data Summary

Our data was collected over the last two weeks of October 2014 from eight popular news sites. Four of the sites cover national and world news on a daily basis and the rest focus on a single topic or audience. The four news sites differ slightly on their emphasis: financial news, local news, long "magazine" form news. Hereafter, we refer to news sites by the characteristics that differentiates them from one another. Table 1.1 summarizes key statistics of the different sites.

We filtered out read events where the user did not meet a minimal amount of interaction with a page or went on to explore other page elements. After filtering,

Site	# Read Events	# Articles	# Readers	Avg. Length (words)
Financial	226,588	5,820	123,571	623
Science	436,190	8,137	287,330	732
Sports	295,223	5,664	240,400	501
Local	20,470	3,103	13,463	610
Tech 1	200,389	6,725	193,935	539
Tech 2	166,450	8,919	163,972	587
Women	396,293	6,855	239,710	650
Magazine	298,867	3,402	279,168	1,111

Table 1.1: Summary statistics of reading events analyzed in this work from eight popular news sites.

our dataset consists of 2.3 million reading events $e(r, p)$, where users spent at least six seconds on a news article page, interacted with the page at least once and did not go beyond the content. Again, these conditions do not "guarantee" reading, but increase the likelihood of reading taking place.

1.4 Patterns of Reading Depth

Over all our data, article pages are 66% read (or, to be more careful, potentially viewed), with a median page read event clocking at 71% of the page. These figures remain relatively robust across sites, with the lowest site having the average page 63% read (Financial news) and the highest being 69% (Sports). Of course, these sites have different article length distribution, visual layout, audiences and serve different user needs. In this section we examine the relation between reading depth article length, where the reader came from, and the reading device they used.

First, though, we examine the relation of reading depth with dwell time. We find that dwell time only has a weak linear relationship with reading depth. The Pearson correlation between article averaged reading depth and dwell time is 0.37, and the correlation for an individual read is as low as 0.11. In other words, there is reasonable (but far from perfect) correlation between the average amount of time spent on a page and the reading depth. The low correlation on individual reads relative to article averages suggests important individual differences in page interactions across readers, which we investigate more fully in the next chapter. We leave it for future work to further investigate the cases where dwell time and reading depth align well, and where differences are found. Overall, we conclude that our measure of reading depth provides information that is not captured by dwell time, which varies considerably from one individual to another.

1.4.1 Article length

The next analysis looks at the reading depth as a function of both article length and the site featuring the content. For this analysis we use a sample of read events, one for each article, in order to give different articles (e.g. popular and less popular) the same weight. In other words, for each article p we choose at random one read event $e(r, p)$ and use the reading depth for that event[4].

Figure 1.1 shows the distribution of reading depths for different article lengths and sites. Each curve is a probability density function for a particular set of pages, summing up to 1. The X-axis is the portion of a page that was read, and the Y-axis represents the density of read events in the sample that reached that point. For example, the red curve on the top row shows that the short pages on the Financial

[4] Averaging across reads and reweighing would have "smoothed" popular articles more than less popular ones, which would have biased our results.

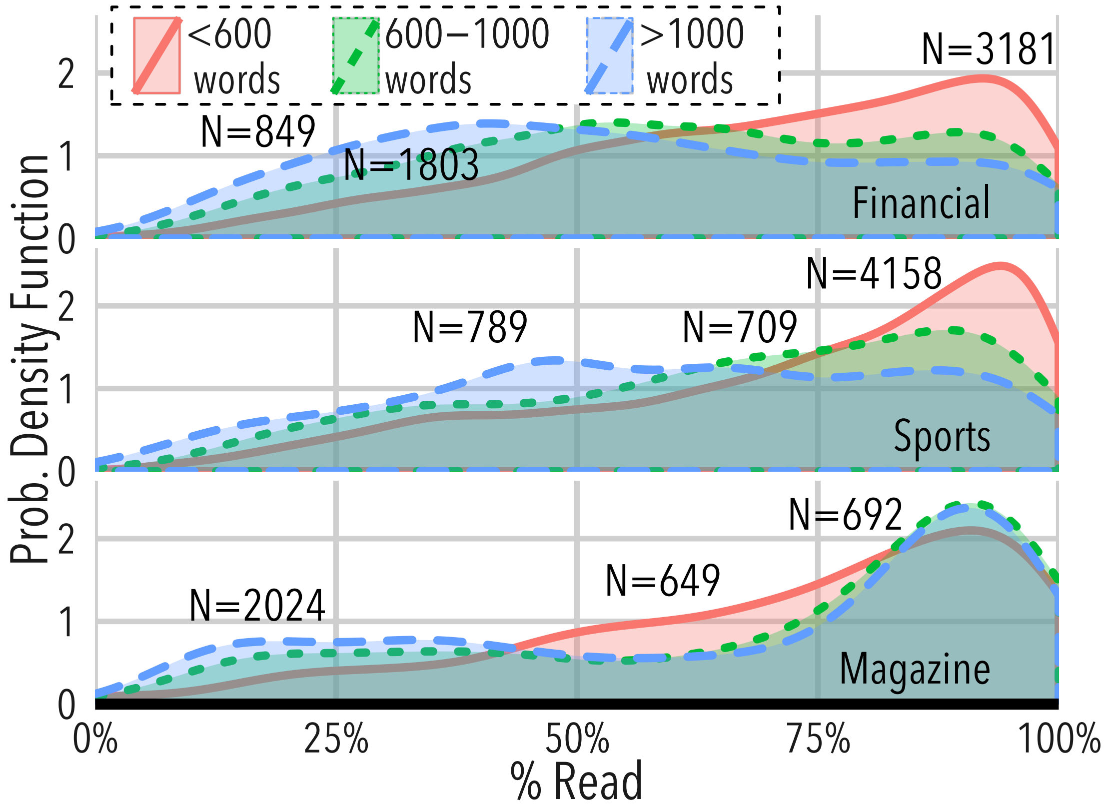


Figure 1.1: Distribution of read proportion for short (red), medium (green) and long (blue) articles on three exemplary sites of Financial, Sports and Magazine News.

site are often read to the end (read events concentrate on the right side towards the 100% mark), based on $N = 3181$ pages p and the same number of reading events $e(r, p)$ for these pages. Two-sample Kolmogorov-Smirnov tests found all three distributions in each site significantly differ from each other ($p < 0.01$).

Figure 1.1 demonstrates that not all sites and content are attended to equally. The Financial site demonstrates the general pattern that appeared in all sites in our dataset: short articles are read more to completion than medium or long articles. Long articles require more effort and time on the reader's part and therefore more likely to be abandoned. An interesting exception to the above rule can be seen in the long and medium densities of the Magazine site. Long form articles are not only more common in the Magazine site, there are more likely to be read to completion with a mode at around 90%. We believe that these findings highlight an important point about readers and expectations – long form content can succeed if paired with the right audience and meeting their expectations. While we do not have information about people's expectations going into articles, Chapter 4 of this dissertation investigates the aspect of attention expectations in online social settings.

1.4.2 Device differences

Figure 1.2 explores the effect of different device types on reading events in our dataset. In this case, we include for each page p in our dataset exactly three read events $e(r, p)$, one event (chosen at random) for each device type. Similar to the analysis of reading depth for articles of different length, the sampling of three reading events per article gives equal weight to reads on different devices. The figure shows the distribution of reading depths for three different sites (different panel rows) as observed on different device types (designated by curves with different line types and fill color). For example, the curves on the Magazine row show the probability of reaching X percent of the article on mobile (red), tablet (green) or desktop (blue), for *the same* set of articles. As before, the X-axis is the portion of page p that was read by reader r, and the Y-axis represents the density of people that read to that point. Two-sample Kolmogorov-Smirnov tests found all three distributions in each site significantly different from each other ($p < 0.001$), except for the Tablet and Desktop distributions in the Financial site, that show similar trends but at a significance level that is only slightly above 0.05.

The pattern emerging from examining reading depth reached on different devices is that mobile users (red) drop earlier and in almost-uniform manner com-

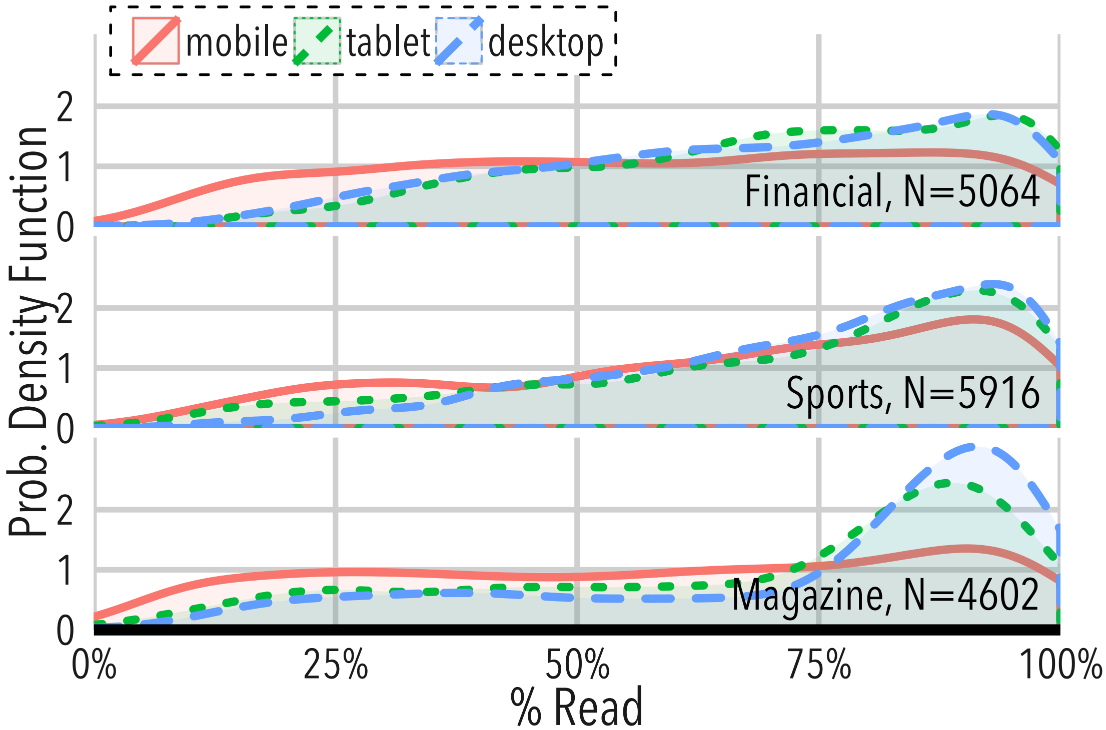


Figure 1.2: Distribution of read proportion for the same articles on mobile (red), tablet (green) and desktop (blue) devices for 3 exemplary sites.

pared to tablet and desktop readers for the same articles. For the larger-form devices, tablet and desktop, the reading depths track very similarly, with higher likelihood of reading to completion than in mobile devices. This pattern is evident in Figure 1.2 and was also present for the five sites not shown in the figure. There are several potential driving forces behind this finding of reading less on mobile on different sites, as we discuss later in this chapter.

1.4.3 Referrer differences

Does reading depth change depending on where the readers come from? Figure 1.3 provides some insight into reading depth based on the user's source of referral to the article page. Similar to previous analyses, for each page p in our dataset we include exactly four read events $e(r, p)$, one event (chosen at random) for each referral type. The four different referral source types we consider are *search*, *social*, *news*, or *internal*. These sources capture whether the reader came to the article page p from a search engine (e.g., Google), from a social media site (e.g. clicking on a link on Twitter or Facebook), or from other news site (e.g. BBC) or news aggregation (e.g. Google News). The internal category is for read events when the reader navigated to the page from another page on the same site. Note that, technically, we identify internal referrals as read events with no referral data, which in most cases indicates intra-site traffic.

Figure 1.3 shows that only the Social referral source significantly differ from the rest of sources, and only in the 25%-50% region ($p < 0.05$ according to two-sample Kolmogorov-Smirnov test). For example, magazine readers that come from Social sources (dashed green line) are more likely to drop earlier than readers from other sources. The pattern of early drop off for readers from Social appeared in for most sites we examined (including those not in the figure). This finding is consistent with findings by Pew Research, showing that people who arrive at news from Facebook spend considerable less time on articles than those who get to the article directly [163]. However, our findings also show that once a user spent a minimal amount of time and interacted with the article, the referring source does not matter much anymore.

In this section we showed the impact of several key factors on reading depth: site differences, article length, user device and referring source. The next section investigates the ability of aforementioned factors and additional features to predict reading depth.

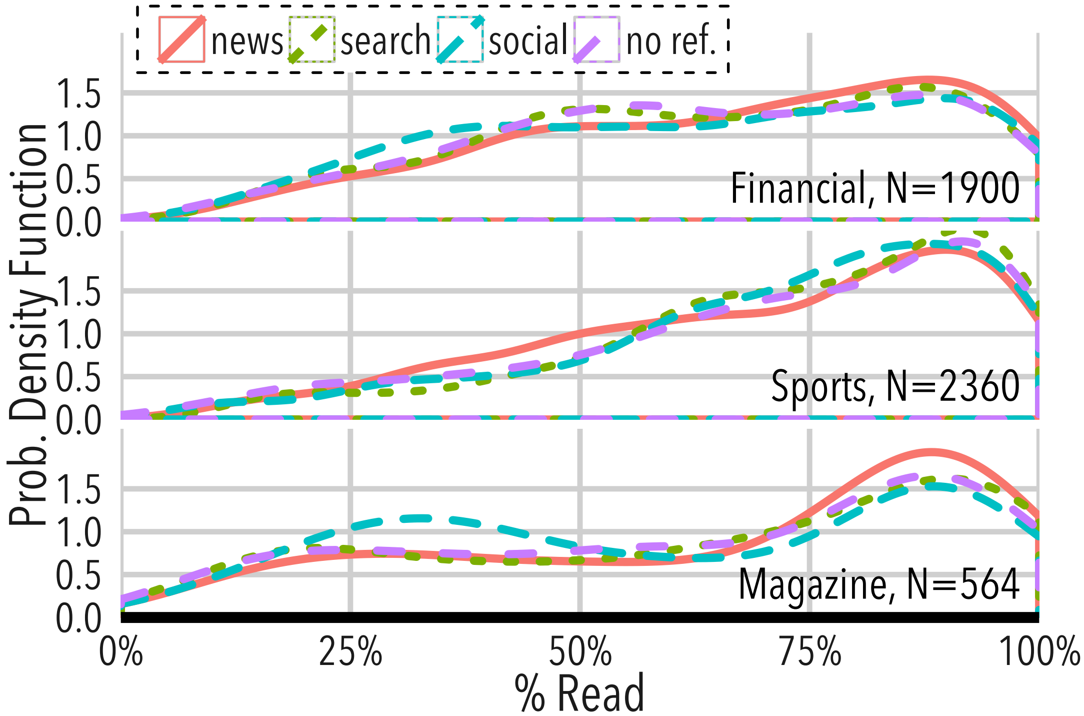


Figure 1.3: Distribution of read proportion for the same articles of readers coming from News (red), search (green), social (tile) and no referrer (purple) sources.

1.5 Predicting Reading Depth

In this section we formulate two prediction tasks, develop models and features, and evaluate the accuracy of our models. In the first task, we predict the average reading depth for article p. The average is computed over all read events $e(r, p)$ for that page, and the features we use for the prediction include the article metadata and text, as well as aggregate properties of readers. The second, arguably harder prediction task is for the depth of an individual read event $e(r, p)$ by individual reader r of article p. Here, we can also use information about the individual reader and reading event, such as the exact time when it took place.

1.5.1 Predicting Article Average Reading Depth

We begin by approaching the simpler prediction problem of article average reading depth. Our dependent variable in this task is, for each article p, the average of reading depths over all reading events e of the article. In order to learn more robust associations between features and average reading depth, we use articles that had sufficient traffic to them and remove outliers. We set 30 reads as our threshold for articles with "sufficient traffic" and use the more robust 95% trimmed mean[5] instead of a regular average that is more susceptible to outliers.

Models and Features

We test a number of prediction models with different feature sets that represent the information available about the article p at different points of the article's life-cycle. We create models that use features available before the publication (*Pre-publication model*), shortly after publication (*Post-publication model*) or, perhaps only theoretically interesting, after all the data was collected (*Final model*). Our *Baseline model* is using a single, but key feature: the length of an article. Following the natural progression of time, the feature sets are nested: we start with the simplest model (our Baseline), add Pre-publication features, then Post-publication features, and finally the Final model features, which was evaluated with all available features. Table 1.2 lists the features added by each model. We refrained from including interaction features in order to keep the dimensionality of the features space small relative to the number of data points (number of articles in this prediction task).

The features we include in each model can be grouped into four families of

[5] Excluding 2.5% of the lowest and highest reading depths.

Pre-publication Model
Length (baseline): pixel depth of the last paragraph on the page. Log-transformed in prediction models.
Site: indicator variable for 1 out of 8 sites in our dataset.
Author: average author reading depth on other articles. The maximum value in case of multiple authors.
Topic: most prominent topic for each article from a 50-topics latent-dirichlet allocation (LDA).
Readability: Flesch-Kincaid grade level score; mean and std of sentence lengths.
Quotations: num. of quotations (log-transformed); num. quotations over num. sentences; num. of words inside quotations (log-transformed); num. of words in quotations over num. words in the full text.
Sentiment: mean and std of the cumulative sum of individual sentences' sentiment.
Lede: Sentence length and Sentiment features computed on the article head (first 3 sentences).
Post-publication Model (first hour)
Readers: num. reading events in the first hour (log-transformed).
Avg. Depth: average reading depth of reading event in the first hour.
Final Model
Device: proportion of readers using mobile, tablet or desktop devices.
Referrer: proportion of readers coming from search, social, other news source, etc.
Local Time: proportion of reads taking place in user local time at night (midnight-6am), morning (6am-noon), afternoon (noon-6pm) or evening (6pm-midnight).
Buzz: proportion of reads in the first hour.
Overall Buzz: average time since the first article reading event.

Table 1.2: Features used in predicting reading depth by each model. Notice that models are additive, including all features listed above them.

features as we describe next: *article metadata*, *content features*, *audience features* and *article-audience features*.

Article metadata features consist of site and author information. We use an indicator variable per site, designating the publisher of the article. For authors we use the average reading depth on their articles, excluding the current one. In case of multiple authors we take the author with highest average reading depth.

Content features were developed in inspiration of journalism practices of engaging readers [193], and are listed next; they include the article topic, ease of reading, quotations, sentiment and lede (first few sentences).

Topic. Different topics convey different types of information, using different writing styles and are targeted at different populations. Therefore, we expect topics to effect reading depth. We use Online LDA [105] run on the entire dataset of articles to extract the most prominent topic for each article in a 50-topic model.

We use a relatively small number of topics in order to keep the topics fairly general and interpretable. We manually label each of the 50 topics for ease of reference.

Readability. One can imagine that the readability of the text will have an impact on how deep people read into an article. To measure text readability, we use the Flesch-Kincaid (F-K) Grade Level measure [126] for each article p. The F-K is a measure for readability presented as U.S. grade level, based on the following formula: $0.39(\frac{|w(p)|}{|s(p)|} + 11.8(\frac{|syl(p)|}{|w(p)|} - 15.9$ where w(p), s(p) and syl(p) are, respectively, the number of words, sentences and syllables in the article. Other features that capture readability include average sentence length (and standard deviation), expecting that shorter sentences would be easier to read.

Quotations. Quotations are widely used in news reporting to give authenticity and flavor to the story by humanizing the reporting. Research has shown that quotation is a powerful persuasion tool [93] and individuals tend to pay more attention and to vivid examples [230].In our case, if the text between two quotation marks contains six words or more, we consider it a quote. We then compute the features listed under "quotations" in Table 1.2.

Sentiment. Reading has an emotional aspect to it - highly emotional piece may influence the reader and even regular content may evoke certain feelings. In our work we used an empirically validated rule-based sentiment analysis tool VADER [114] to compute the sentiment scores of for each sentence in the article. Then we calculated the mean and standard deviation of the cumulative sum of sentences' sentiment.

Lede. The last content feature is trying to capture text structure. "Don't bury the lede", is perhaps the most well known "rule" for journalistic writing [193].

Ledes are usually the first few sentences of an article. We compute the sentence length and our sentiment features on the text first three sentences.

Audience features consist of information about the size of the audience, where people are coming from, how are they reading the content and when. Audience features are not available pre-publication and thus excluded from that model. We compute the proportion of users coming from different sources of referral (such as search, social, other news source, etc.) and devices being used (mobile, tablet or desktop). To capture the temporal aspect of when reading is taking place we compute the percentage of reads in users' local time, binned into four equal parts of the day. In other words, the percentage of reads taking place between midnight to 6am, 6am to noon, noon to 6pm, and 6pm to midnight according to the reader's time zone.

Article-audience features are aggregate features of article performance, or reception, by its audience. We include the average reading depth of readers in the first hour in our Post-publication model. An hour is a small percentage of the 24-hour cycle of news and thus allow us to peek into the early adoption of an article [201]. In addition, we try to capture article temporal popularity or buzz by calculating the proportion of reads in the first hour and the average reading time relative to the first read.

Prediction Models

Our focus on understanding reading behaviors aligns well with using interpretable prediction models. The benefit of fitting interpretable models is that in addition to the model predictive power we can assess the significance of features and examine the contribution of single factors. We evaluate two regression models in this work:

Linear Regression and Beta-Regression. Both models were fitted to data in R using standard and *betareg* [63] software packages.

Our dependent variable in this section, the article averaged reading depth, is bounded in the $[0, 1]$ range. However, the linear model is not bounded to the unit range and may predict values outside $[0, 1]$. To amend this, we simply truncate the predictions of the linear model to 1 if above one and 0 if below zero.

Beta-regression is a form of Generalized Linear Model that support dependent variables in the range $(0, 1)$ and assumes that observation are drawn from a beta distribution parametrized by mean (μ) and precision (ϕ) parameters:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \tag{1.1}$$

Where y is the response variable (reading depth in our case) and Γ is the gamma function. In regression setting, μ and ϕ are linear functions of the covariates x_i where i is the i-th data point. We follow the practice of using the logit function for μ and log function for ϕ as it was found to improve convergence of the learning algorithm. Compared to the linear model, beta-regression offers the ability to model non-symmetric distributions and fit a wide range of distribution "shapes".

We first experimented with including per site fixed effect versus running our prediction models separately for each site. A fixed effect per site is a single site feature that offsets the prediction for each site separately. In contrast, running separate predictions per site optimizes all model parameters per site. We did not find any improvement in prediction accuracy of the per site prediction over the fixed effect model, therefore resort to using only the simpler fixed effect model.

Results

Figure 1.4 summarizes the performance of the different models. We report the results in term of *prediction error*: the error in predicting article average reading depth. On the Y-axis is the prediction error RMSE obtained using 5-fold cross-validation with standard errors. The X-axis details the different feature sets included in each prediction, with shape and color of points designate the prediction model (green triangles for linear model and blue squares for beta-regression). The figure shows that the baseline prediction for both the linear and beta-regression is close to 0.1, which is 10% away from the observed article average. As we include more and more features the linear model does not consistently reduce the prediction error while the beta-regression model reduces the error to 0.082, or 20% error reduction.

About half of the error reduction is obtained by using pre-publication features by the beta-regression model. In other words, using article metadata and content information, we were able to reduce prediction error by 10%, before any reader has seen the actual page. While site differences play a major role, both author and the properties of the text contribute to the prediction accuracy (a closer analysis of individual features is below). A second improvement in accuracy is attained using the average reading depth in the first hour, which lowers the RMSE by another 8%. The difference between the baseline RMSE and our most accurate model is not large in absolute terms (about 0.02 error reduction). For prediction purposes and for estimating whether the article received more or less attention than expected, using just the article length may be sufficiently accurate. However, in order to better understand the factors behind article averaged reading depth our models do provide a more nuanced view, as we describe next.

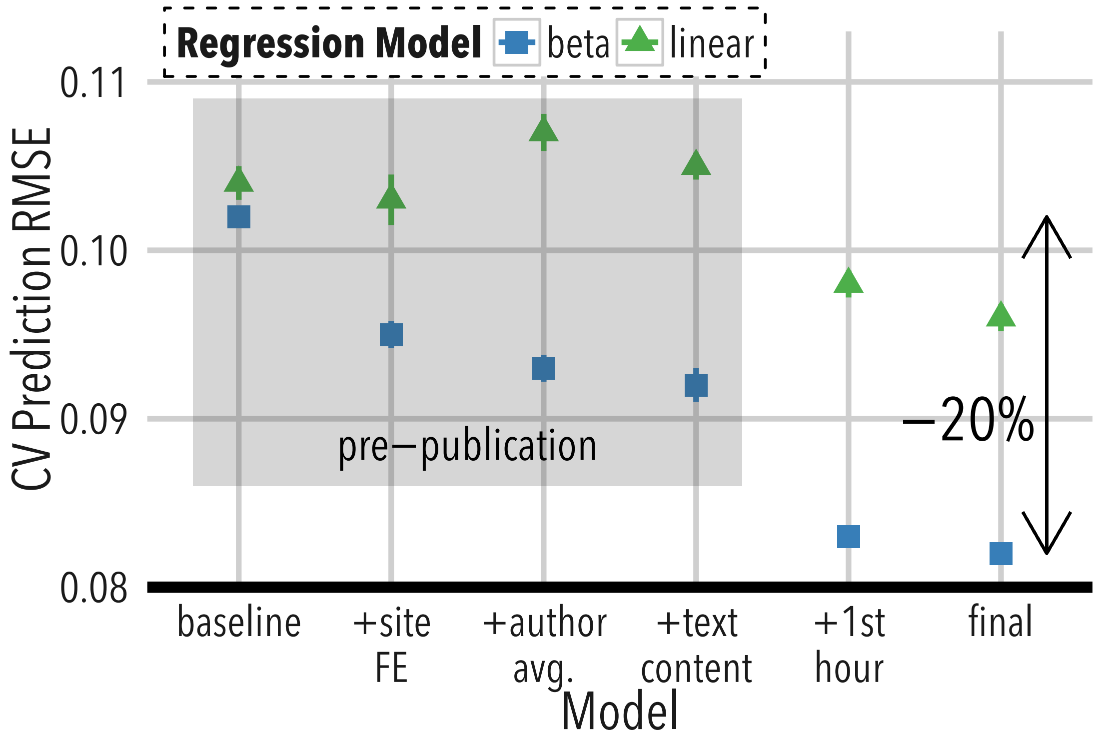


Figure 1.4: Prediction Error for article average reading depth. Relative error reduction of 20% is half due to pre-publication features and the rest due to first hour features.

Significant Features: the linear and beta-regression models identified roughly the same set of features as significant. For brevity, we only report significance based on the beta-regression model, which predicted the results more accurately. In addition, we only report on features significance in the most complete Final model (includes all features) since "earlier" models contained only a subset of features.

The coefficients of the logit-transformed beta-regression are in the space of log-odds, which complicates reading their magnitude. Therefore, Table 1.3 only show to the sign of the effect (positive or negative) of individual factors that were found significant with $p < 0.05$.

The same features that improved prediction accuracy were found to be

Pre-publication Model
↓length** ↑author avg.**
topics: ↑fashion** ↑food** ↑url parts** ↑work-related** ↑foreign languages** ↑gaming** ↑cameras* ↑film* ↑michael brown* ↑history ↑home improvement ↓hong kong*
↑num. quote words** ↑% quotes**, ↓num. quotes** ↑avg. sentiment ↓sentiment std.** ↓sentence len.** ↓head sentence len.**
Post-publication Model (first hour)
↑num. readers first hour ↑first hour reading depth**
Final Model
↑% desktop readers** ↓% mobile readers** (relative to tablet)
↑% search readers** ↑% social readers** ↑% news readers** (relative to no referral)
↑num. readers** ↑buzz* ↓overall buzz

Table 1.3: Significant features found in predicting article average depth. Positive or negative association with reading depth marked by (↑, blue) and (↓, red). All features are significant with $p < 0.05$. '*' designate $p < 0.01$ and '**' $p < 0.001$

highly statistically significant. These features include site, author and the "autoregressive" reading depth in the first hour. Among the content features, we found small but interesting effects. Certain LDA topics were associated with reading more than average, for example, topics that we labeled as "fashion", "film" or "food" were read more. Longers quotes and positive sentiment contributed to reading depth, while sentence-by-sentence variance in sentiment reduced it.

In the Post-publication model, the dominant feature was the one based on early reads of the article in the first hour after publication. The average article had 8% of its reads taking place in the first hour. Thus, a small fraction of the early reads carry significant predictive power for of the majority of reads that follow.

The final model adds additional information about readers interaction with content. In line with the descriptive results presented in Section 1.4, reading devices add significant information, with readers on desktop and tablet devices read more than on mobile. In contrast to the descriptive results *per read* (in Section 1.4), in our model, a greater proportion of readers from search, social or news source are associated with *increased* reading depth relative to no referral (internal) traffic. The discrepancy of the social referral results merits further investigation but may be due to the fact that the social referral distribution is bi-modal, or due to significant differences between sites. The article popularity, in number of readers and buzz contributed positively and significantly to reading depth. Local time proportions (at which times of day the article was read) were not found to have a significant effect.

1.5.2 Predicting Individual Reading Depth

We now turn to our second prediction task: predicting reading depth of an individual reader and article in one reading event e. Our dependent variable in this case is simply the reading depth in a single reading event e of a specific article by a single user. We expect more variability in our dependent variable, making this prediction task harder. For this part of the analysis, because of the large number of events, we sample 100,000 reading events per site.

Features

We use the same sets of features as described in Section 1.5.1 and detailed in Table 1.2, with few distinctions as below. All "audience" features collapse into reader features, namely, categorical variables about the reader device, referrer and local time. In addition, we include infomration about the reader, including their average reading depth in past reading events. All other features were identical to those in the average reading depth models.

As before, we start with a baseline of the single article length feature and gradually add features to our prediction models following the article life-cycle. Since the focus in this prediction task is an individual, we add reader average reading depth before introducing the Post-publication features.

Similarly to Section 1.5.1 we use linear and beta-regression models to predict reading depth, this time where the dependent variable is reading depth of an individual reading event.

Results

The prediction errors for individual read event depth are shown in Figure 1.5. On the Y-axis is the prediction error RMSE obtain using 5-fold cross-validation with surrounding standard errors. The X-axis details the different feature sets included in each prediction, with shape and color of points designate the prediction model (green triangles for linear model and blue squares for beta-regression). The baseline prediction errors for both the linear and beta-regression are close to 0.29, which is 29% away from the observed reading depth. Except for the final prediction, the beta-regression outperforms the linear prediction model. As we include more and more features the beta-regression reduces the error to 0.239 – a 15% error reduction.

Note that in the prediction of individual reads the Pre-publication model obtains most of the error reduction, with site fixed-effects and textual features significantly improving prediction accuracy. Then, reader information, post-publication and final features further improve prediction accuracy.

Significant Features: similarly to the previous section Table 1.4 summarizes the significant features that contributed to the mean part of the beta-regression model.

Among the content features, most of the earlier findings reappear from the

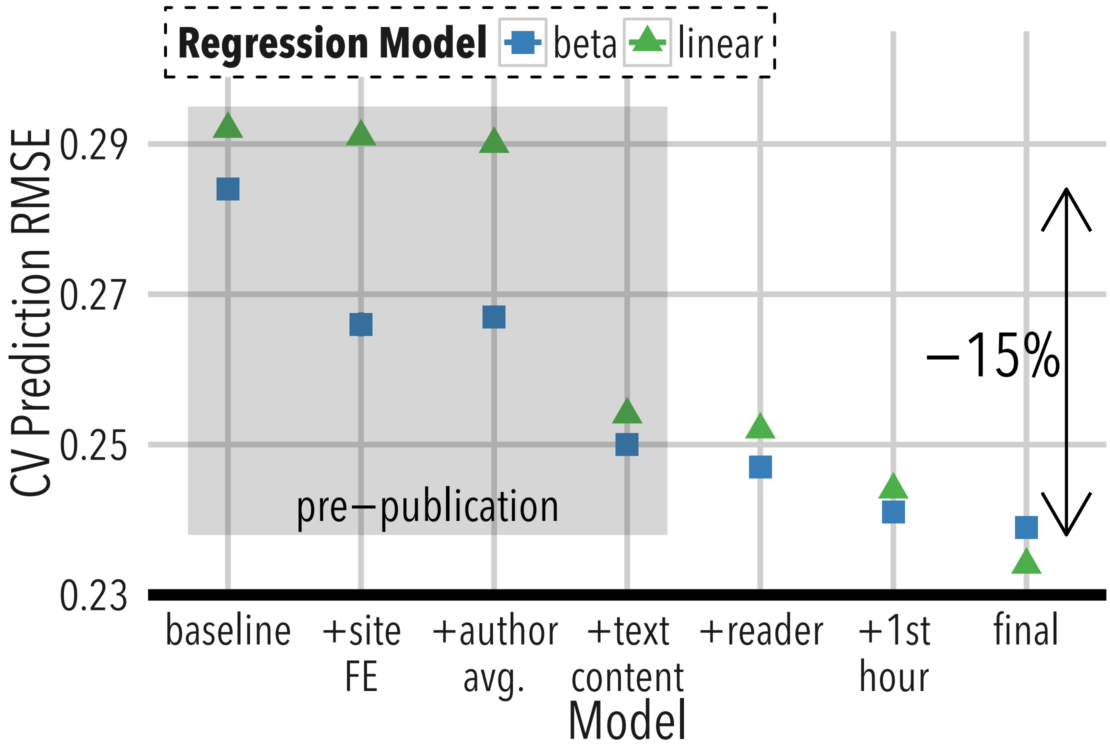


Figure 1.5: Prediction Error for individual reads. Relative error reduction of 15% over the baseline was obtained mostly by incorporating site, textual and after publication features.

Pre-publication Model
↓length** ↑author avg.**
topics: ↑sports ↓tech** ↓nightlife ↓film** ↓science** ↓past-related** ↓people-verbs
↓flesch-grade* ↑num. quote words ↑% quotes**, ↓head sentiment** ↑head sentiment std.* ↑head sentence len.* ↓sentiment std.**
Post-publication Model
↓reader avg. ↓hours since first read** ↓num. readers first hour** ↑first hour reading depth**
Final Model
↑desktop reader** ↓mobile reader** (relative to tablet)
↑search reader** ↑social reader** ↑news reader** (relative to no referral)
↑num. readers** ↑avg. reading depth** ↓buzz**

Table 1.4: Significant features found in predicting individual reading event depth. Positive or negative association with reading depth marked by (↑, blue) and (↓, red). All features are significant with $p < 0.05$. '*' designate $p < 0.01$ and '**' $p < 0.001$

article average prediction task, but with larger predictive power in the individual prediction task. As before, consistent sentiment and long quotes are associated with deeper reads. However, this time higher Flesch-Kincaid grade level was found to be negatively associated with reading depth ($p < 0.01$). On the other hand, longer head sentences contribute to reading depth. The larger predictive power of textual features in the prediction of individual reads suggests interesting variation across readers that is diminished when averaging across read events.

The reader average, as well as features from the Post-publication and the Final models were found significant ($p < 0.05$), though their magnitude was small. As before, reading on desktop devices is associated with higher percentages of reading relative to mobile and tablet. Readers coming from search and social read more than internal readers, in line with the findings of the article-level prediction. Local time was again not found to be predictive. As for the time of reading, the more time passed between the article being published and the read event, reading decreases. Lastly, the article average reading depth both in the first hour and overall (excluding the current read) were found highly significant for predicting individual reading depth.

1.5.3 Discussion and Conclusions

In this chapter we used the simple measure scrolling depth in a large dataset of likely reads on news articles. Both descriptive analysis and predictive modeling show that the length of articles is a strong and significant predictor of reading depth, but other factors also affect the reading depth. In particular, content features remain significant for reading depth, even after including information about site, author's past readership and article's audience composition. The content features that were significant in several of our models include shorter sentences, positive (and consistent) sentiment, and longer quotes.

In our models, reading on larger screens such as desktop or tablet is positively associated with reading depth. It is important to note that our analysis focused on reading of content in "non responsive" layouts, which may be harder to read on a mobile device than pages that re-flow text. Without a more precise model of reading at the level of paragraphs or other sub-document elements, we cannot further investigate the reasons behind shorter reads on mobile (more on this in Chapter 2).

Our findings highlight a complex relationship between the referral source and reading depth. The predictive models indicate that for users coming from external sources, including social, has a positive average effect on reading depth, which is consistent across all nested models in both prediction tasks. On the other hand, the descriptive results show that in certain cases readers are more likely to drop, e.g. after 25-50% of the article when coming from social. The ultimate test for the effect of referrers on reading depth should be evaluated in an experiment or causal analysis.

Finally, we note that the number of reads per article contributes to explaining average and individual reading depths. This finding is encouraging – as it may suggest a link between popularity of content and level of engagement with it.

For publishers, our measure of reading depth can provide additional information about readers' engagement with their content and potentially inform editorial decisions. Publishers could use reading depth as additional metric for evaluating the success of a story, and decide to promote it to their homepage, for example.

Our predictive models can a help guide editors about an article reception even before it is publish and shortly afterwards.

As a data-driven study, our study has several limitations. As we point out above, we call our dependent variable “reading depth” and focus on a subset of page views that are more likely to reflect reading activity, but we only know that the users had scrolled down the page. A better understanding of the reader’s activity on the page, a task we take on more fully in the next chapter, will enable a much more precise analysis of the factors affecting sustained attention in reading. Moreover, while our dataset consists of a diverse set of sites and probably more heterogeneous population than any small-scale lab study, our results may still be biased due to the unrepresentative nature of readers and content on these sites/ For example, the fact that fashion articles tend to be read in more depth may just mean that fashion has a different set of readers than sports, with different interests, and not that fashion itself is inherently more interesting than sports.

Future work could address some of these limitations and expand to other forms of content. First, it would very useful for publishers and system designer to know whether the trend we found in this work (e.g. shorter reads on mobile devices) are due to technological factors (e.g. screen resolution) or human factors. For example, publishers and recommendation systems should probably take different action if shorter reads on mobile devices are due to additional effort involved in scrolling through a small screen versus different time constraints of people reading on mobile. Second, many news articles online use multimedia (e.g. images, videos or ads) that may have an effect on reading depth, in different contexts and for different audiences. Last, a longer snapshot of people’s reading habits over time could help researchers study trends in people’s ability to sustain attention in reading and understand the role of systems in effecting reading.

As mentioned above, the next chapter in this dissertation extends the work presented thus far by tackling the core issue of measurement of reading using noisy page interaction data, and by proposing a computational model for learning about reading jointly from lab and large-scale observational data.

CHAPTER 2

MODELING READING AND SKIMMING

The previous chapter focused on scrolling depth in a subset of interactions with news article pages as a proxy for reading, but without any guarantees that reading actually took place. In this chapter we take a close and more direct look at measuring reading in online news from signals of user interaction with article pages. Similar to other works that draw inferences about people's attention from online interactions [21, 42, 44, 110, 131, 171, 173], we develop a new measure for sensing reading of individuals at the level of paragraphs. By tackling the core issue of measurement, this chapter complements the previous chapter as well as other research on implicit post-click measures [102, 125, 132, 145, 229] by modeling individuals' attention in reading with fewer assumptions.

2.1 Introduction

Reading is perhaps the most fundamental way to engage with news articles and yet existing online measures provide little information about what is being read and to what extent. A model for quantifying reading online can improve recommendation and personalization systems and potentially alleviate deficiencies in existing systems. For example, a granular view of the attention spent within an article can provide a more accurate description of a person's interests and help recommend relevant content to them, similar to improvements gained by using other post-click measures in recommendation systems [102, 125]. A richer description of what is actually read, and not just shared or clicked, can help social systems combat the spread of low quality content like Clickbaits [75]. In addition, tying online user interactions to interpretable constructs of reading can help journalists write more compelling stories and advance a scientific discussion about people's ability to sustain attention in the digital age. However, without knowing the target of people's attention on screen (if at all) and over time it is extremely difficult to make accurate inferences about individuals' reading of online content.

In this work, we take the first step towards building a semi-supervised model that estimates the extent of reading by an individual at the level of paragraphs. The model draws inferences about reading from a sequence of user interactions with a page (e.g. in scrolling, moving the mouse, etc.), which can be easily collected online. Based on these signals alone and without more intrusive measures such as eye-tracking as done in other works [21], the model classifies individual's reading of paragraphs as in-depth reading, skimming, or no reading at all. We conduct a lab experiment to obtain the ground-truth labels for reading of individual paragraphs and evaluate the predictive accuracy of different features set in a supervised manner. We conclude by describing a semi-supervised model that could potentially integrate the labels obtained in our experiment with the patterns of user interaction observed in natural online settings.

Our contributions are therefore:

- A novel semi-supervised model for inferring reading modes from mostly unlabeled user interaction logs.
- Robust evaluation of the ability of existing models to identify reading and reading modes at the level of paragraphs.
- Identification of the important features associated with different reading modes of paragraphs.

2.2 Related Work

Researchers have long been interested in strategies that people apply in order to deal with the abundance of information available to them, in particular, in reading for comprehension. Three of the most prominent reading strategies described in the literature are skimming, scanning, and reading in-depth. According to Grellet, both reading and skimming are necessary techniques for efficient reading [101]. The definition given by Grellet, which we use in this chapter, describes skimming as processing text quickly in order to get the gist of it, to understand its organization, or to get an idea of the tone or intention of the writer. Scan reading is a strategy that focuses more on finding specific details as defined by Nuttall: "glancing rapidly through a text either to search for a specific piece of information (e.g. a name or date) or to get an initial impression of whether a text is suitable for a specific purpose [...]" [176]. In contrast to skimming or scanning, the goal of reading in-depth is to extract the maximum amount of information possible from the passage and integrate it with prior knowledge [70]. While the ability to identify scanning of news articles is a worthy research topic, our focus in current work is on skimming and reading in-depth, which involve more extensive interaction with the article content.

Evaluating what people read and how they read it is a difficult task. In fact, people only master the metacognitive skill of assessing their own reading when reaching an intermediate level of reading [90]. Prior research utilized mostly three approaches to study reading processes, sometimes in conjunction with each other: eye-tracking [72, 190, 191], brain-imaging (e.g. PET, fMRI, MEG) [58, 89, 188], and comprehension tests [117, 121, 123]. Accurate eye-tracking recordings usually require fixed viewing angle and distance, headset mounted devices, or modified text interface showing limited number of words at a time in large fonts. Brain-imaging techniques such as fMRI are even more restrictive, requiring the participants to be placed inside a magnetic chamber that is expensive to acquire and operate. Fundamentally, eye-tracking and brain-imaging techniques trade off the ability to capture reading as it is in natural settings for accuracy and precision. Comprehension tests offer an alternative that is less intrusive since people are only presented with a series of questions about the text they read. However, the accuracy of comprehension tests depends on a variety of factors including the cognitive load, proficiency of working memory, fatigue, prior knowledge about the reading subject, and level of detail required by test questions [117]. The lab experiment described in this chapter attempts to capture reading as naturally as possible and thus uses comprehension questions to assess reading and log data to trace it.

Numerous studies have shown that comprehension varies when reading on screen and on paper. In one of the early studies about reading behaviors in the digital environment [147], Ziming notes "screen-based reading behavior is characterized by more time spent on browsing and scanning, keyword spotting, one-time reading, non-linear reading, and reading more selectively, while less time is spent on in-depth reading, and concentrated reading". Birkerts further states that younger generations who are mostly exposed to content in digital form lack the ability to read content in depth and sustain reading for long periods of time [22]. Ackerman and Lauterman conclude that self-regulatory factors, much more than technology-related factors of the medium, lead to inferior comprehension on screen [1]. Therefore, a measure that accurately captures attention to news outside lab settings could facilitate further investigation into the affect of technology on reading.

Prior research also examined how users divide their attention among page elements and in reading online content, but mostly in lab settings and for fixed layouts. For example, Buscher et al. conducted eye-tracking experiments in order to infer salient regions of web pages [42]. Other works set to infer attention using mouse cursor activity, first by linear models relating eye-mouse positions [110], then through non-linear transformations [171], and more recently using more complex mixture-models [131, 132]. Nielsen estimated that people read 20-28% of the content on average web page [173], but the analysis was limited to pages between 30 and 1,250 words – different use case than news media, where people often have stronger intent to actually read the article they click on. Closest in nature to the current study is the work of Biedert et al. who built a classifier for detecting in-depth and skim reading using expert-generated labels for sequences of gaze movements [21]. The current work focuses on estimating people's attention in reading organic content (i.e. unmodified, mostly textual news content) using digital traces that are readily available outside lab settings.

Another line of work investigated the potential of post-click measures, similar to the one developed in this chapter, to improve information systems. Post-click measures were shown to be beneficial for information retrieval, personalization and recommendation systems. Previous work found that scrolling and the amount of time a user spends on a page are associated with subjective assessment of page relevance and interest by the user [56, 102]. Other works showed that dwell time is a good indicator for user satisfaction with search results, which can improve ranking in a search engines [125, 145]. Similarly, post-click measures of engagement were used to improve the quality of personalization and recommendation systems through collaborative filtering. For example, according to Yi et al., incorporating dwell time as a proxy for user satisfaction into Yahoo's recommendation system provided better performance than already click-optimized system [229]. This work focuses on developing more accurate post-click measure for reading and leaves the assessment of likely improvements in information systems to future work.

Next, we describe our methodology for obtaining labels for reading and the models used to learn from these labels.

2.3 Methods

In this section we describe our two-step methodology for learning the patterns of user interaction that represent the different reading modes at the level of paragraphs. The first step consists of a lab experiment that focused on getting labels about reading at a paragraph level. The second step uses the data from the experiment to learn classifiers that can distinguishes in-depth reading from skimming and other non-reading activities. We begin with a description of the experiment.

2.3.1 Experimental procedure

We devised an experimental procedure[1] that asks study participants to read in-depth or skim a pre-selected set of news articles, and answer a set of comprehension questions about specific paragraphs in the article they just read. Similar to other experiments on skimming [71, 156], we elicit different reading modes by instructing participants to read in-depth or skim articles and vary the time allocated for reading based on the individual's reading speed, the length of the article, and the reading condition (skimming or in-depth) as calibrated in a pilot

[1] Approved by Cornell IRB protocol number 1603006226.

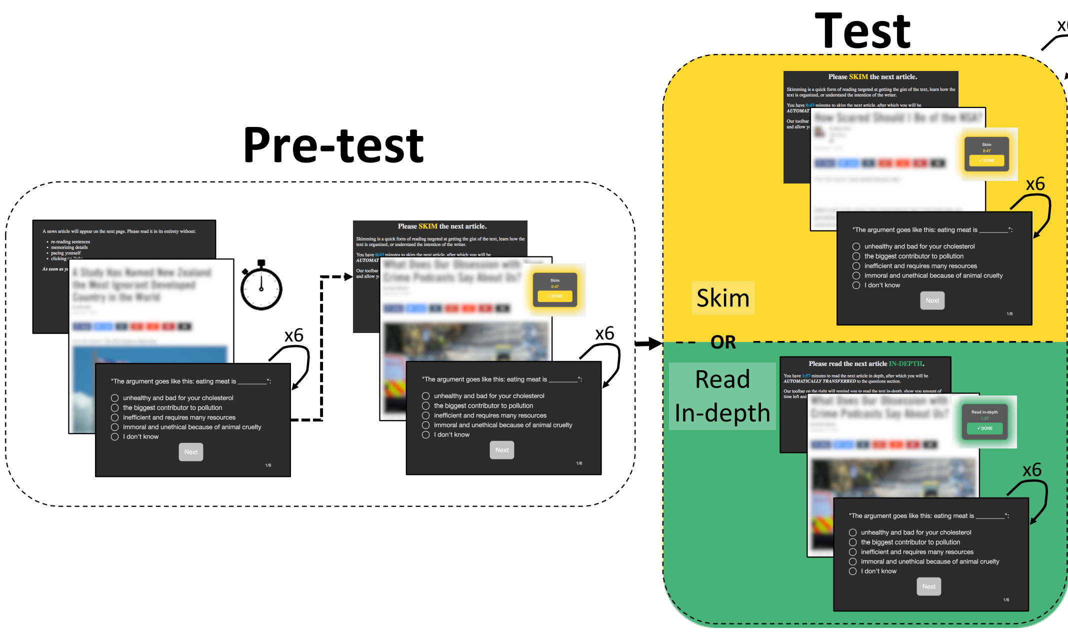


Figure 2.1: The experimental procedure devised and carried out in this study. Two pre-test trials measured participants' reading speed and familiarized people with the experimental flow, followed by six test trials. Each trial consists of one article, preceded by instructions and followed by comprehension questions. Participants interactions with the article page were recorded and as they were instructed to read articles in-depth or skim. Informed consent as well as basic demographic questions appeared at the onset of the experiment (not shown in the figure)

study. The comprehension questions allow us to verify whether participants read in-depth, skimmed, or did not read at all specific paragraphs. All articles came from Vice.com, a popular US-based news site that features both long and short articles. We used a single site in order to minimize the effect of adjustment to new layouts. The articles we pick spanned a range of topics that required little or no prior domain knowledge.

The experiment procedure is described in Figure 2.1. The experiment started with standard inform consent form and basic demographic questions (e.g. age, gender), and was followed by two pre-test trials and six test trials. Each trial consisted of reading one article, preceded by instructions and followed by six comprehension questions. The first pre-test trial asked participants to read the article in-depth, avoid rereading sentences, memorize details, pacing oneself or clicking links, and indicate when completed reading. Based on the duration of this trial we estimated the participant's reading speed in words per minute that was used for allocating time for all other trials. The second pre-test trial asked participants to skim the next article, included a definition of skimming[2], and indicated the amount of time allocated for reading. A small overlay (shown to the right of articles in Figure 2.1) gave participants information about the amount of time left for reading and allowed them to proceed with the experiment if they finished reading earlier. This second pre-test trial allowed participant to practice skimming and familiarize themselves with the operation of the experiment.

Six test trials commenced after the first two pre-test trials. Test trials included articles that ranged in length from short (~600 words) to medium (~900 words) or long (~1400 words). Participants were asked to read those article in one of two conditions: in-depth or skimming. Instructions for reading in-depth were identical to the ones given for skimming, except that there was no definition of in-depth reading. Participants were asked to read in-depth and were allocated sufficient time to do so. Skimming instructions were identical to the ones given in the second pre-test trial and described previously. Reading conditions were counter-balanced across the different length categories such that each participant read in-depth and skimmed exactly one article in each length category. The order of articles was randomized in order to allow controlling for fatigue during the experiment.

Following each article, participants answered six multi-choice questions about the passage they just read. Questions immediately followed the article in order

[2] "Skimming is a quick form of reading targeted at getting the gist of the text, learn how the text is organized, or understand the intention of the writer" [101].

to minimize the effect of diminishing ability to recall information over time. The questions focused on three paragraphs of the article that were long enough (more than a couple of sentences) and contained simple but memorable details. We assessed the memorability of questions' details using a pilot study described next. Each paragraph had exactly two questions associated with it: one covering information mentioned in the opening sentence of the paragraph and another question regarding information mentioned in the body of the paragraph. When participants answered correctly the question about the body of the paragraph we considered the paragraph as read in-depth (regardless of the correctness of the skim question), and when only the skim question was answered correctly we considered the paragraph skimmed. Otherwise, the paragraph was labeled as not read. Our labeling is based on findings by Duggan and Payne, which showed that when people skim they spend more time on the beginning of paragraphs than the second half of paragraphs [72]. While certain reading strategies (e.g. scanning for numbers) could clearly glean details from text without reading the text in-depth, the pilot study, described next, confirmed that the labels correspond closely to way people described what was read and the extent to which they were reading it.

Before launching the experiment protocol, we conducted a pilot study in two phases to identify problematic questions and instructions, test the memorability of details in our questions, and adjust variables in the experiment. We recruited volunteers from the Cornell Tech campus to participate in the pilot without any compensation. After reading an article and before being presented with comprehension questions participants were asked to describe memorable details from the article. This first phase of the pilot helped us identify questions and instructions that were not clear enough, and memorable details of articles. In the second phase of the pilot we corrected these issues and composed additional questions that covered details mentioned by multiple participants in the first phase. We also introduced an “I do not know” answer option and added instructions to avoid guessing since participants reported that they guessed answers when they were not sure of the correct answer. Based on participants’ assessments of time pressure we set the skimming rate to be 2.5 faster than reading in-depth and increased by 10% the time allocation for two articles that participants found slightly more difficult than others. At the end of the pilot, the experimenter asked participants to describe their reading of the specific paragraphs and reviewed their labels. The vast majority of descriptions of paragraphs reviewed corresponded to the labels obtained from participants’ answers.

We recruited participants for the main experiment on Cornell Tech’s campus through flyers and email announcements. Participants came to lab with their personal computers and were compensated $10 for their participation. The study protocol was carried through a Google Chrome browser extension that we developed and participants installed on their personal computers. The extension guided participants through the entire flow of the experiment, and rendered articles in the user’s browser exactly as they would have rendered had the participant navigated to the article page directly[3]. While participants read articles in their browser, the extension recorded their interactions with the page every 100 milliseconds, without straining the browser resources too heavily. This sampling rate is higher than the average eye-fixation duration of 200-300 milliseconds, which is necessary for visual processing of text [121]. Interactions were collected client-side using javascript and buffered before periodically sending back to our servers. Every tracking sample included the participant’s viewpoint position on the page, cursor position, and any mouse of key strokes that occurred since the last sample. Participants’ answers

[3]With the only caveat of a small hovering toolbar was present to keep track of time.

Number of participants	30
Gender	18 Females, 12 Males.
Age	Averaged 29.2 (min: 21, max: 56)
Highest level of education completed	11 college, 19 masters.
Measured reading speed (words per minute)	Averaged 252.5 (min: 165, max: 375)

Table 2.1: Demographics and reading statistics of study participants.

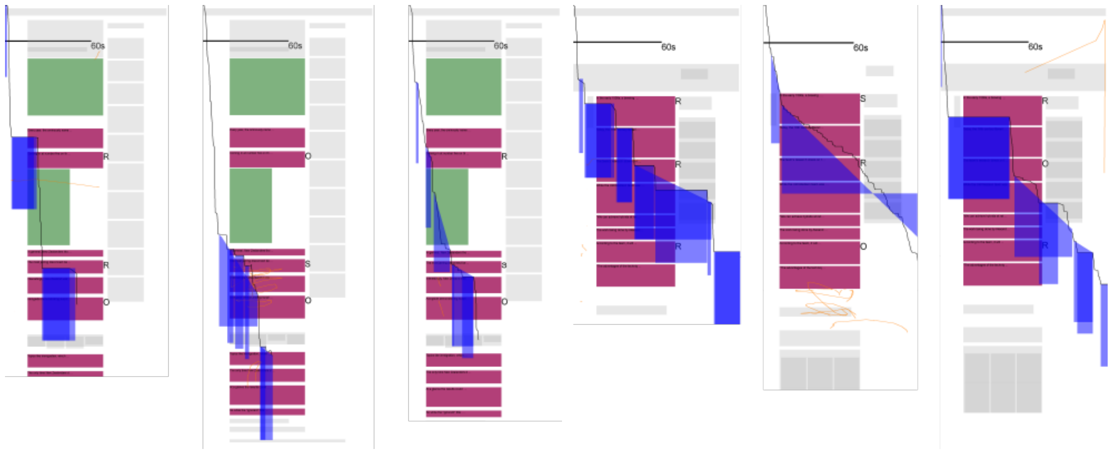
and page interaction data were stored anonymously on our servers with no identifiable information. Upon completion of the study protocol, participants followed instructions to remove the extension from their browsers.

Table 2.1 provides summary statistics about participants in the experiment. After excluding one participant who did not have recorded page interactions, we had 30 participants with valid recordings, who took between 23 to 45 minutes to complete the experiment. None of the participants reported visual impairments (beyond short sightedness) or dyslexia. The sample had more females than males (18 vs. 12) and included people mostly in the 20's and 30's. The average participant was 29.2 years old. All the people in the study held at least a college degree, with 19 of them completed higher education program at the level of a masters program. Based on the assessment of reading speed in the experiment, we found that the average participant read at a rate of 252.5 words per minute. Most people in the study read at a rate of 200-300 words per minute with only a few exceptions below and above this range. This reading speed is very much in line with the average rate of approximately 250 words per minute for average readers described in the literature [123,156]. Overall, we conclude that people in the study are slightly more likely to be female than male, about 20-30 years old, college educated, and reading at a normal rate for their level of education. Yet, there are large differences in reading speed between individuals where some people read a at rate that is 2.5 times faster than others.

Exp. condition \ Reading Label	**In-depth**	**Skim**	**Other**	**Total**
In-depth	162	50	58	270
Skim	68	57	145	270

Table 2.2: The effectiveness of the experimental manipulation shown as number of labels obtained from comprehension questions (columns) in each experimental condition (rows).

Table 2.2 demonstrates that the experimental manipulation was effective in changing participant's ability to answer questions about the text, presumably due to different reading strategy being employed. Recall that we assigned "in-depth" labels if and only if participants correctly answered the question about the body of paragraphs, otherwise assigned "skim" if they correctly answered the question about the head sentence or "other" for all other cases. The table shows that when participants were asked to read articles in-depth (first row) most of the resulting labels are "in-depth" (162), which corresponds to participants being able to answer correctly questions about the body of paragraphs. In the skimming condition (second row), participants attained considerably less "in-depth" labels (68), much more "other" labels (145), and slight more skimming labels (57). On the face of it, it may seem like the skimming condition most converted in-depth labels to non-reads without a significant effect on skimming. However, a closer examination of the individual answers revealed that participants in the skimming condition were correct fewer times, more uncertain (choosing the "I do not know" option more frequently), and wrong about the same number of times. This indicates a shift in comprehension that is in line with previous research, which found that comprehension generally deteriorates as reading speed increases [120, 122]. Despite the lower comprehension in the skimming condition the number of skim labels actually slightly increased (from 50 to 57), suggesting that participants did indeed shift their attention to the beginning of paragraphs in the skimming condition.



(a) Skimming (b) Reading in-depth

Figure 2.2: Page content along with page interactions of three individuals skimming the same article (a) or reading in-depth (b). Magenta rectangles represent textual paragraphs, green rectangles represent images, and gray areas represent other page elements (e.g. ads, recommendations, headers). Overlaid on top of the content is the user interaction, scaled horizontally by time (legend appears on top as black bar equal to 60 seconds). The black line crossing the content represents the top of the user's screen over time and orange line represents the mouse position. Purple rectangles represent times when the user was not scrolling for at least a second, height shows viewport height and width shows duration (scaled). Semi-transparent purple trapezoids, overlaid over the purple rectangles, represent an attempt to flex the definition of static viewport view to a dynamic counterpart that is anchored by the content of the page.

Figure 2.2 shows how three different individuals interacted with two articles in (a) the skimming condition, and (b) in-depth reading condition. The difference between in-depth and skim reading is evident in the figure as the speed that users moved through the content is clearly different between skimming (2.2a) and in-depth reading (2.2b). In addition, there are noticeable differences in the interaction patterns of different individuals with the same article in the same reading condition. In the skimming condition (2.2a), the first participant on the left seems to have spent only a few seconds on the first paragraph and read in-depth the second one. The middle participant in (2.2a) seems to have completely skipped the first two paragraphs and move in short strokes through later paragraphs. The third participant (right of 2.2a), spent more time on the first two paragraphs and moved in longer strokes. Initially we expected skimming to occur in short bursts of movement. However, the general pattern that appeared is more similar to the middle and right participants in 2.2a. People were skimming while moving through the page and rarely stopping for more than a second. In terms of reading in-depth (2.2b) we also observe large differences across people. Some appear to read the content in the middle of their screen while being static while others appear to read at the top of the screen, moving line by line. Some move their mouse over the text while others mostly scroll with it. The central question we address in the next section is whether a classifier can correctly identify the reading mode at the level of *a single paragraph* for participants in the experiment.

2.3.2 Reading modes classification

Based on the data obtained in the experiment, we evaluated using 5-fold cross-validation the ability of different feature sets and four different classifiers to predict reading modes. Class labels can take one of three labels that are in-depth reading, skimming or non-reading (other). We test the following predictive models as implemented in R: Ordinal Regression (from MASS package [210]), Regularized Ordinal Regression (from glmnetcr package [7]), Support Vector Machine (from e1071 package [161]), and Random Forest (from randomForest package [142]). Regression model have better numerical stability when variables are standardized so we center and scale by two standard deviations[4] all continuous variables. For Regularized Ordinal Regression we use l_1 regularization to encourage sparsity and we

[4]Scaling by two standard deviation puts continuous variable on roughly the same scale as binary variables [91].

tune the parameters of SVM prior to prediction. We train Random Forest with 1,000 decision trees. Before we describe the different feature sets we experimented with a note about overfitting is in order.

The relatively small scale of the dataset (30 participants × 18 paragraphs per participant = 540 data points) requires greater care with respect to features used for classification. Even with regularization on the number of features used and cross-validation, certain features may lead to overfitting. For example, a classifier may pick up that a paragraph of a particular length (e.g. 68 words) is more likely to be read in-depth than similar length paragraphs. Cross-validation may not be able to correct such bias because the skew may be spread across different strata of the data. Our approach to minimizing overfitting is to limit the number of features used and categorize variables of high cardinality. We also limit the minimum size of leaves in our Random Forest model to five such that no label assigned by the model is based on less than five examples.

We experimented with the following feature sets:

Article properties: sustaining attention in reading is a challenging cognitive task and therefore we must take into account both order at which articles appeared in the experiment and their length. Therefore, we include information about the article ordinal position in the experiment (ranging from 0 to 5) and length categorized into short (~600 words), medium (~900 words), or long (~1400 words).

Paragraph properties: the characteristics of paragraphs we consider include ordinal position (measured in screens[5] to avoid overfitting), word length quantile

[5]The article was divided into equal size screens based on the average participant window height and each paragraph was associated with a screen based on the top pixel.

(short (48-62 words), medium (63-91) words, or long (92-127 words)), visual height[6] quantile (short (100-150 pixels), medium (151-186) pixels, or long (187-321 pixels)). Other visual aspects of paragraphs such as width or left position did not have sufficient variability in our data in order to be included. Because we thought people might have a preference to read the last paragraph as the article summary we included a binary variable indicating whether a paragraph is the last one in the article.

Page interaction: a paragraph that appears on the user's screen can appear in full or be only partially visible. Therefore, we compute the same set of user interaction features (below) separately on the time series resulting from the paragraph being partially visible and fully visible. Based on a close examination of the interaction patterns that emerged in the experiment (see Figure 2.2), we hypothesize that sometimes when paragraphs are read in-depth or skimmed they are the first ones to appear in full on the user's screen. Therefore, we include a third set of interaction features based on the time series of the paragraph appearing first on the user's screen.

The page interaction features computed for all three time series (full view, partial view, or first on screen) include the amount of time visible (also known as dwell time), time the cursor was inside the paragraph's bounding box, time the cursor was vertically within the paragraph's bounds, number of seconds when the mouse was moving, and number of mouse or other keys pressed. We also compute the paragraph's inverse "reading" speed (as dwell time divided by number of words) and the ratio of dwell time to expected dwell time based on the participant's

[6] Our site of choice was non-responsive, therefore content did not rescale or reflow based on screen resolutions. However, people with lower screen resolutions would require more scrolling to cover the same total amount of pixels/content.

reading speed and paragraph length.

We note that all of the above features are easily accessible outside lab settings with the only exception of the dwell time ratio, which requires some estimate of the user's reading speed. A user's reading speed can potentially be estimated her previous interaction with articles, if such visits exist.

In order to evaluate the contribution of different features to the overall predictive task we use a measure of feature importance. Strobl et al. showed that the standard measure of importance in Random Forest, based on decrease in node impurities averaged over all trees, is biased when variables are correlated [203]. In our case most page interaction variables are correlated because user activity tends to cluster over time. For example, a user highlighting a line of text would result an increase in mouse clicks and mouse movement in addition to the added dwell time on the paragraph. We use the conditional importance introduced by Strobl et al. in the R package *party* to address this issue. We assess the variable importance for each class label separately and for the different feature set. In order to obtain the best performing model we retain only features whose average importance across the three class labels is positive.

2.4 Results

Based on the features described in the previous section we trained four different classifiers: Ordinal Regression, Regularized Ordinal Regression, SVM, and Random Forests as implemented in R. Despite the counter-balancing of reading conditions in the experimental design, the way people actually read specific paragraphs naturally deviated from these instructions. In other words, our dataset consists of some class imbalance where more paragraphs were read in-depth (230 instances) or not read at all (203 instances) than skimmed (107 instances). One way to address this imbalance in class labels is to introduce class weights into the classification model such that errors on the skimming label, for example, are more costly than errors on the in-depth instances. However, different classifiers handle weights differently and therefore we chose to balance our dataset instead. Hence, our dataset consisted of 321 instances altogether, exactly 107 instances of each class label, which were stratified during the 5-fold cross-validation.

Table 2.3 summarizes the predictive results of three different classifiers on different feature sets. We found the results of the Regularized Ordinal Regression very similar to the ones obtained for Ordinal Regression and therefore exclude the regularized version for brevity. Each cell in the table represents the Area Under the Curve (AUC) in a three classes ROC plot characterizing the precision and recall of a classifier. Notice that for three class labels AUC is no longer bounded by 1/2 from below as in the case for binary classification, but rather bounded by 1/3 from below. For example, a random forest classifier trained on random labels (the first row of Table 2.3) obtained an average AUC of 42.71, which is based on the AUC of the three individual class labels (38.21, 37.67, and 52.26).

We can make several observations from Table 2.3. First, we see that the different classifiers performed comparably to each other, with SVM and Random Forest usually outperforming Ordinal Regression. The AUC for identifying other interactions is mostly higher than reading in-depth or skimming, but generally improvements in the average AUC achieve improvements in the identification of all reading modes. All feature sets shown in different rows yield significant improvements over the random baseline of 42.71 AUC. The *Best* feature set, consists of a

	Reg	SVM	Random Forest			
Reading mode AUC(%)	*Avg.*	*Avg.*	*Avg.*	*In-depth*	*Skim*	*Other*
Random	45.54 46.31 47.08	46.26 46.97 47.69	40.98 42.71 44.45	34.55 38.21 41.88	33.75 37.67 41.60	51.06 52.26 53.45
Article	49.64 **50.33** 51.01	49.30 50.00 50.71	48.87 49.67 50.47	48.26 49.37 50.48	49.92 50.91 51.90	47.53 48.73 49.92
Paragraph	55.11 55.95 56.80	54.77 55.53 56.30	56.89 **57.78** 58.68	53.54 54.89 56.23	48.90 50.22 51.54	67.17 68.25 69.32
Interaction partial view	56.71 57.47 58.23	57.25 **58.01** 58.77	53.87 54.68 55.49	52.67 53.67 54.67	48.20 49.17 50.14	59.98 61.20 62.42
Interaction full view	52.31 53.15 53.99	52.33 53.14 53.96	53.98 **54.90** 55.81	52.99 54.12 55.25	50.27 51.45 52.63	57.90 59.11 60.33
Interaction 1st in view	54.84 55.70 56.55	55.25 **56.01** 56.77	54.38 55.26 56.14	53.96 55.34 56.71	53.84 55.08 56.33	54.33 55.35 56.38
Best	57.00 57.81 58.63	56.79 57.70 58.61	62.08 **62.91** 63.75	59.47 60.53 61.58	56.85 57.99 59.13	69.17 70.23 71.28

Table 2.3: The predictive power of different features sets and different classifiers in predicting reading mode. Classifiers include Ordinal Regression (Reg), SVM, and Random Forest. The best performing model for each feature set, averaged over the three class labels, appears in bold. For Random Forest, we also show the AUC for individual class labels in order to demonstrate that the accuracy is not dominated by a single class label. Numbers above/below in each cell represent 95% confidence intervals.

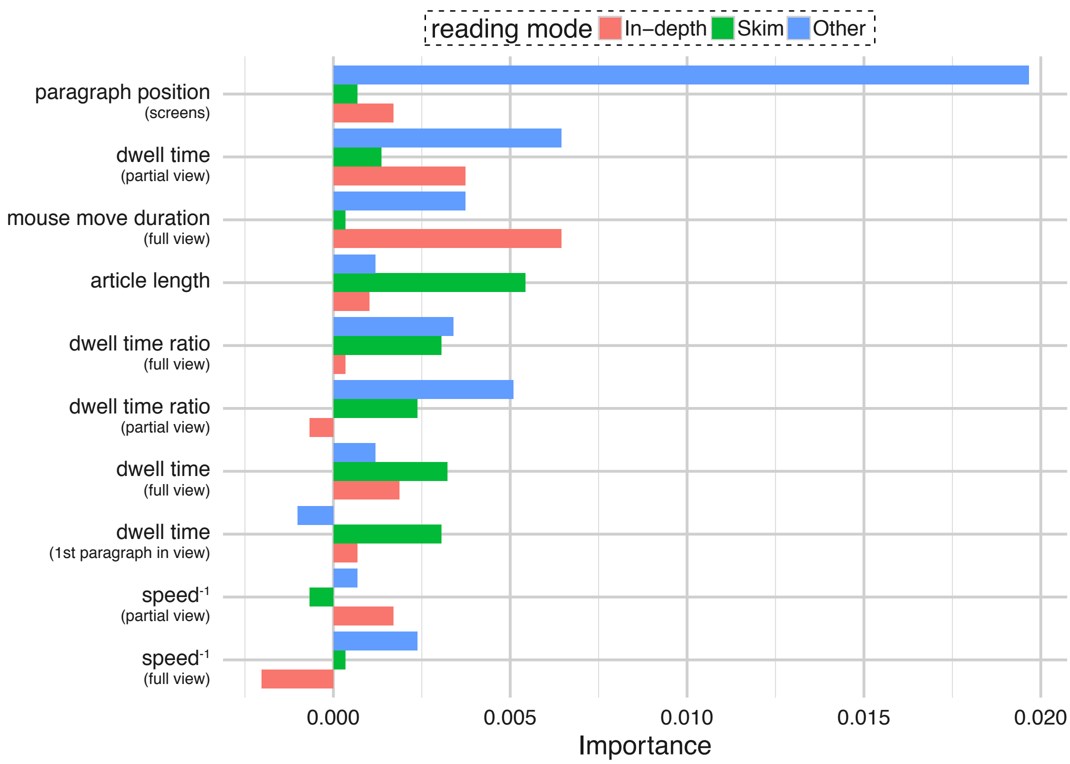


Figure 2.3: Variables' importance (based on average decrease in impurity in the forest) in the prediction of reading mode using Random Forest.

subset of features described next, achieved the highest performance overall using Random Forest with 62.91 AUC (+42% improvement over the random baseline).

Figure 2.3 shows the conditional importance (X-axis) of different variables (Y-axis) in our *Best* performing model using Random Forest. The features included in this model consist of only the features that contributed to the predictive accuracy in other feature sets (e.g. article or paragraph features) as analyzed through conditional importance. As mentioned before, we used conditional importance due to the correlated nature of variables in our dataset. The conditional importance scale still measures mean decrease in accuracy for individual features (in more stable and less biased than the unconditional measure), but the scale is more interpretable in terms of relative importance of features as we discuss next.

Based on Figure 2.3 we can interpret the important features for classifying each reading mode. Starting with the non-reading label (other) shown as blue bars, we see that the most important features are the position of the paragraph on the screen, the dwell time in partial view, and the ratio of time in partial view relative to fully reading it by the participant. Further examination revealed that later paragraphs are less likely to be read in-depth or skimmed, as well as paragraphs that had little time on the screen, even in partial view. In other words, when a paragraph is not read people usually spend less time viewing it and it tends to be positioned later in the article. It is plausible that greater time pressure and/or fatigue contributed to less reading of later paragraphs, but our data does not provide direct ways to measure such factors.

For paragraphs that were skimmed (green bars) we see that the article length is the most predictive feature, followed by dwell time variables and their ratio to expected reading time. We find that when people skimmed, all dwell time related measures were 20%-37% higher on average than the same measures when not reading and 11%-16% lower than reading in-depth. This places skimming close to in-depth reading and further away from non-reading interactions. We also note that paragraph dwell time while it is the first fully visible on screen is more important for skimming than other modes of reading, suggesting that skimming is more strongly associated with scrolling. The article lengths that were associated with more skimming than other labels are long articles, and surprisingly short ones. While more skimming in long articles is reasonable as a strategy to cover more of the text, its prominence in short articles (but not medium ones) is surprising. It is possible that the time allotment in the experiment was effectively shorter for short articles (in relative terms) and thus people had to skim more, which is something we can potentially corroborate in the future with observational data.

In-depth reading designated by red bars in 2.3 is associated with mouse movement while the paragraph is in full view, and dwell time in partial view. As before, average dwell time measures in all forms (partial view or not) are significantly higher than when skimming or not reading at all. More significantly, we find that mouse movements while the paragraph is in full view correlate more strongly with reading in-depth than other reading modes, indicating that people are not only viewing the paragraph longer, but are also more active (i.e. moving the mouse) when doing so. This aligns with prior work that found a strong correlation between eye and mouse position, suggesting that people as indeed reading [47].

Overall, we found that content properties and user interaction jointly contribute to the ability to distinguish different reading modes. We developed features that significantly improved the accuracy of our predictions and identified the features associated with different reading modes. Next, we highlight avenues for further improving our results and detail a model that will combine labels obtained in lab settings with observational data available at a much larger scale.

2.5 Future Work

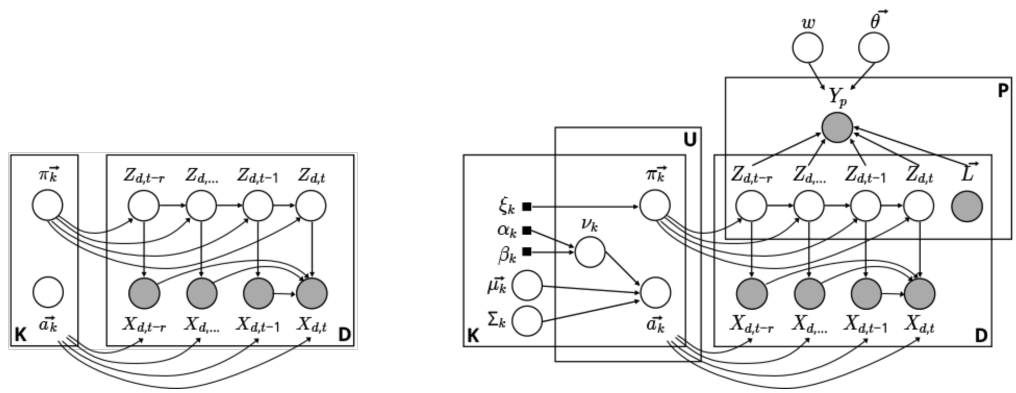
The work presented in this chapter laid the foundation for modeling the different reading modes of individuals from their interactions with news article pages. In this section we outline the ways for future work to extend these models to include “organic” users interactions with news articles as they occur in natural settings, at a much larger scale, and with only a small set of labeled examples.

The approach presented in this chapter thus far has several noticeable shortcomings. First, an experiment is limited in its ability to simulate interruptions in reading. Of course, we could have designed interruptions in the experimental procedure, but these would not necessarily be representative of the type of interruptions people experience online. In addition, while our models used individuals' reading speed, they were not fully personal. In other words, our models did not fully capture the "style" of interaction a person is utilizing as demonstrated in Figure 2.2. We obtained significant improvements in the predictive accuracy of reading mode, but at about 63% AUC our results are still far from perfect.

In order to address these issues we propose a semi-supervised probabilistic Bayesian model to identify the different reading modes. The key idea in using semi-supervised learning is to learn patterns of user interaction that have strong support in large scale observational data (i.e. without intervention) that can better separate and identify reading modes. A Bayesian model would not only provide predictions, but also provide a way to express the model confidence in those predictions. Ideally, the model will provide good fit for the data and a useful abstraction of user interactions it that is interpretable and meaningful with respect to the attention paid to content.

We call our proposed model *pBAR-HMM* – a personalized, Bayesian, Autoregressive Hidden Markov Model that will jointly infer reading modes using labeled and unlabeled data. The goal of pBAR-HMM, just like other predictive models used in this chapter, is to determine for every paragraph p of a news article d the probability that individual u has been reading it in-depth, skimming, or not reading at all, given the engagement time-series $x_{d,u}$ and the rendered page layout $L_{d,u}$.

We base pBAR-HMM on Switching Autoregressive HMM (SAR-HMM), introduced by Ephraim and Roberts [79], which is a variant of Autoregressive HMM



(a) SAR-HMM (b) pBAR-HMM

Figure 2.4: Plate diagrams for SAR-HMM (a), and pBAR-HMM (b). Empty circles designate latent variables, shaded circles designate observed variables, and full square points represent hyperparameters. Rectangular plates indicate repetition per document D, paragraph P, state K or user U.

(AR-HMM) by Poritz [186]. SAR-HMM is shown in Figure 2.4(a). The SAR-HMM model is similar to standard HMM in having K latent states that generate observations and a transition matrix $\Pi_{k,k}$ between states. However, SAR-HMM goes beyond standard HMM in describing the dependency between consecutive observations using a linear autoregressive function of order R:

$$x_t = \sum_{r=1}^{R} a_r(z_t) x_{t-r} + e_t \quad \text{with} \quad e_t \sim N(0, \sigma^2) \tag{2.1}$$

Where $a_r(z_t)$ is the r$^{\text{th}}$ autoregressor when in state $z_t \in \{1 \dots K\}$ at time t. The SAR-HMM switches between K hidden states that parameterize the dependency between observation at time t and r previous observations. Mesot and Barber introduced a Bayesian extension of the SAR-HMM [160] by putting a prior on the autoregressive parameters of each state k:

$$\begin{aligned} \vec{a_k}|\nu &\sim N(\vec{\mu}_k, \nu^{-1}\Sigma_k) \quad \text{and} \quad \nu|k \sim \gamma(\alpha_k, \beta_k) \\ \vec{\pi_k} &\sim Dir(\vec{\xi_s}) \end{aligned} \tag{2.2}$$

Where N is the multivariate normal, γ is the gamma distribution, and Dir is the Dirichlet distribution for the transition probability from state k. The hyperparameters $\left\{\vec{\mu}_k, \Sigma_k, \alpha_k, \beta_k, \vec{\xi}_k\right\}$ are for each hidden state s.

pBAR-HMM, shown in Figure 2.4(b), extends the Bayesian SAR-HMM in two fundamental ways: it is hierarchical (and thus allowing dependency across different time-series) and introduces supervision at the time-series level. First, since reading varies considerably between people (see Figure 2.2 for example), we would like pBAR-HMM to be able to "adapt" to the different reading style of individuals. Therefore, instead of the global $\vec{a_k}$ and $\vec{\pi_k}$ we now have personalized variants of these parameters sampled per person u: $\vec{a}_k^{(u)}$ and $\vec{\pi}_k^{(u)}$. As a Bayesian model the inferred parameters per person could deviate more from the prior when additional evidence supports it and the model will become more confident in those parameters.

In addition to personalization, pBAR-HMM aims to infer latent states of reading in the same manner defined in the lab experiment. Supervision in standard HMM, however, is given in the form of a known state sequence over time (z_t in our notation). However, in our case we do not have time stamped labels for whether a person was reading in-depth, skimming, or not reading at all[7]. Therefore, pBAR-HMM must utilize the label associated with the entire time series of the paragraph. A paragraph p viewed by individual u can get one out of three labels $y_{p,u} = l \in \{\text{not read, skimmed, read in-depth}\}$. We can further assume an order among the label categories (similar to the Ordinal Regression used in the Results Section), such that skimming is greater comprehension level than not reading at all, and that in-depth reading is greater than skimming. Then, supervision can be introduced through the latent state sequence $z^{vis}(p,u) = \{z_t | p \text{ visible at time } t \text{ by person } u\}$ and the rendered page layout $L_{d,u}$

[7]Even expensive fMRI studies only provide approximate estimates for whether a person is reading.

in an ordinal regression:

$$\begin{aligned} P(y_{p,u} = l | z^{vis}, L_{d,u}) =& \Phi(\theta_l - w \cdot f(z^{vis}, L_{d,u})) \\ & - \Phi(\theta_{l-1} - w \cdot f(z^{vis}, rpl_{d,u})) \end{aligned} \tag{2.3}$$

Where Φ is the cumulative distribution function of the standard normal distribution, w is the regression coefficients, f is the feature extraction function based on user interaction and page layout and θ_l are thresholds between class labels. We emphasize that during inference one should jointly infer the regression parameters along with the emission and transition probabilities of pBAR-HMM, such that the labels "percolate" to influence the choice of clusters and transitions between states. The benefit of such joint inference is that the learned hidden states can better separate the reading labels, ideally, without hurting the goodness of fit for the unsupervised user interaction data.

Initial experimentations with fitting a Bayesian SAR-HMM model to user interaction data revealed that there are only two dominant elements in the Autoregressive model: the last data point and its empirical derivative from the previous one. In other words, the best predictors for a user interaction at time t (position, mouse movement, mouse click, etc.) are the user's interaction at time $t-1$ and how it changed from interaction at time $t-2$. The SAR-HMM is picking a very short time scale for the model, which of course is reasonable because it is not geared towards reading of paragraphs in any way. However, when further developing pBAR-HMM one must be aware of this tendency of Autoregressive models and should make attempt to address it. One possibility for addressing the time scale issue is to experiment with Hidden Semi-Markov Models (HSMM's) that model the duration of staying in a state using a probability distribution (e.g. Poisson) [169]. Another more extreme alternative is to use standard HMM instead of the Autoregressive one and do a Line Search over all reasonable time scale to test the right scale for capturing the different reading labels.

2.6 Discussion and Conclusion

In this chapter, we described a first step towards building a semi-supervised measure of reading of news articles at a paragraph level using non-intrusive means (i.e. user interactions with the article page). We developed experimental methodology to manipulate the "depth" at which people read and used the traces of user interaction with news article pages to train classifiers for detecting reading at a paragraph level. We developed a set of visualizations to examine the user interactions data and found distinct "styles" of reading when reading in-depth and when skimming that varied from one person to another. Based on these findings we devised a set of features for classifying reading of paragraphs as reading in-depth, skimming, or non-reading. We evaluated the ability of Ordinal Regression, Regularized Ordinal Regression, SVM, and Random Forest models to predict reading of paragraphs. The classifiers performed comparably in most cases, with Random Forest slightly outperforming other models with the highest AUC of 62.91% (+42% improvement over the random baseline). Investigating the important features in the prediction task revealed that later paragraphs and paragraphs that were briefly visible are less likely to be read. We found that the time spent on paragraphs while skimming was longer than non-reads but shorter than in-depth reading, and that skimming was more tightly coupled with scrolling. Reading in-depth, in addition to require more time and attention, was associated with more mouse movements while the paragraph is in full view. Finally, we outlined pBAR-HMM, a personalized Bayesian Autoregressive Hidden Markov Model, that aims to learn the patterns of user interaction that corresponding to different reading modes from both lab and observational data.

The experimental procedure devised in this chapter reflects an attempt to capture reading as naturally as possible, but has some clear limitations that future work could try to tackle. First of all, reading in lab settings, under time pressure, and for monetary compensation may affect people's reading practices. In addition, the website and articles we used may not accurately represent reading more broadly. Different layouts and text introduce different demands of human attention, and we cannot argue that the ones included in the experiment represent news articles at large. The unit of analysis, reading at the level of a paragraph, may not generalize to particularly long or short paragraphs, other genres of text than news (e.g. prose), or texts of different level of difficulty (e.g. scientific articles). The study population, drawn from Cornell Tech's campus, is perhaps representative of a college-educated sector of the population, but not likely to be representative of the wider population of internet readers. The experimental manipulation was geared towards eliciting reading in-depth and skimming, but is not complete with respect to the other types of non-reading interactions. For example, distractions and long interruptions that require people to shift their attention from the text for long periods of time (e.g. receiving a phone call) are underrepresented in the experiment. An evaluation of pBAR-HMM using both lab and observational data could address some of these issues. Another noteworthy extension of this work could explore interruptions more directly by experimenting with frequency, duration, and saliency of interruptions that mimic real-world distractions.

In terms of prediction, the different predictive models provided significant improvements over random baseline, but still leave large room for improvements. Most, if not all, of our features summarize the entire time series of interaction with a paragraph, whether it is partially or fully visible, or the first in view. Future work could investigate the classification errors more closely and derive features that better separate the different classes. The different reading "styles" evident in Figure 2.2 lead us to believe that better quantification of interaction motifs could also improve predictive accuracy. Moreover, additional heuristics like the "first fully visible paragraph" could be useful for prediction. Knowing which paragraph is attended to inside the viewport view could help quantify the attention paid in reading and avoid conflating it with attention to neighboring paragraphs.

There are several important directions for future work to further investigate. First, as more internet traffic is taking place on mobile devices it is important to develop measures that quantify reading on these devices. The approach we took for estimating reading, using comprehension questions and non-invasive page interactions, was partially chosen with mobile devices in mind. Since mobile screens take smaller portion of the visual field they pose a greater challenge for accurate eye-tracking measurement, which is further exacerbated by the wider range of possible viewing angles and distances on mobile. Carefully designed comprehension tests offer a good alternative to eye-tracking on mobile devices. Another important direction for future work to pursue is modeling the prior knowledge of people when reading. Many news stories follow a stream of developments of a certain topic (e.g. Brexit), which is likely to affect how people read about new developments. For recommendation systems a more granular sensing of sub-document reading can help better assess the value of a story for an individual. Last, there is merit in exploring designs and interventions that assist people to achieve their goals in reading more effectively. Such interventions could potentially reverse some of the negative impacts of information systems, such as shortening attention span and greater susceptibility to distractions.

Part II:

ATTENTION IN SOCIAL SETTINGS

In this part we move on to consider attention in online social settings. We focus on the domain of social media because more and more people turn to social media for information these days and because exceeding amounts of information is calling for people's attention on these platforms. Moreover, the demands for attention are likely to increase in the future as social networks continue to reach new markets, connect more people than ever before, and rise in prominence as the leading channel for information dissemination. The two chapters in this part investigate complimentary aspects of attention in online social settings: how people's attention changes in different circumstances, and the expectations people have for getting attention from their friends. Both chapters provide evidence for the dynamic nature of attention and offer concrete ways for social systems to help people direct their attention more efficiently.

CHAPTER 3

THE DYNAMICS OF PAYING ATTENTION

In this chapter we devise a quasi-experimental methodology to study how the attention changes at different times. We use this methodology to examine how the attention of millions of individuals on social media shifts as a result of people's own actions, without any intervention. In particular, we anchor our analysis around the action of posting to Facebook, an action that many millions of people take every day, and investigate the changes before and after posting in individuals' engagement with Facebook as a whole and with content on it. The chapter demonstrates how careful quantitative methodology for analyzing large-scale observational data can provide new perspective on people's goals, values, and desires online, which were mostly studied through qualitative methods. The chapter also describes how systems can utilize the changing pattern of attention and behavior around posting to account more for the context of people's actions and dynamically adapt to people's needs.

3.1 Introduction

The affordances of information sharing on Social Network Sites (SNS) [27] determine the experience for contributors, their community and the dynamics of the network as a whole. As a result, much research has focused on people's motivations to post on social media [66, 118, 170, 174, 180, 189]. However, to date, little research has examined posters' behavior and attention *directly* after (or before) the act of posting.

We draw on existing theories from communication and social psychology to formulate hypotheses about contributors' behavior on SNS. We address three different questions in this work. First, we test whether contributors (those who post their own content at a given point in time) are intrinsically feedback-seeking and visit the site more often after contribution even when no knowledge of feedback exists. Second, we examine whether contributors allocation of attention to content change both in quantity and selectivity. Lastly, we investigate changes in interaction rates with others' content, and quantify the effect of reciprocity in interactions with friends.

Better understanding of the mechanisms behind contribution is important for both theoretical and practical reasons. The underlying processes that accompany contribution to SNS are not yet well understood [40, 49, 136]. Studying the relation between contribution and user attention in large-scale observational datasets can provide a new perspective for understanding individuals' behavior in context, and complement previous research that relied on self-reported measures (e.g. [43,109,154]). Examining user engagement around posting can identify changing needs and preferences of contributors, as well as indicate expectations for attention from others. Practically, better understanding of contributors' behavior can help encourage posting, better support users at times of contribution, and may even be used to improve personalized recommendations.

We devise a within-subject, observational data analysis of de-identified log data of Facebook activity from a sample of 2.4 million people over a period of nine days. In our design, we observe individuals' actions on Facebook around times of contribution (without any intervention) and another comparable activity, like liking or commenting on another's post. Specifically, we consider when an individual posts a piece of content, e.g. writes a post or posts a photo, and compare her activity around that time to a different time when she gives feedback on someone else's content. We use measures of activity such as site visits, number of stories read and number of stories interacted with in the 48 hours surrounding contribution, in order to learn about the relation between posting and contributors' behavior.

Our contributions are therefore:

- First large-scale evidence for within-subject differences in engagement around times of contribution, e.g. when posting content to Facebook rather than commenting on others' posts.
- Empirical evidence for an increase in site visits, reading more stories from friends and interacting more with friends in the 24 hours after posting.
- Potential design implications for better supporting contributors on social network sites.

To further motivate this study, we describe the theoretical framework used to draw hypotheses about changes in contributors' behavior.

3.2 Background

We build on theories from various fields to examine behavioral changes of contributors in SNS. These theories help us reason about the ways in which posting content can affect how individuals use Facebook, consume content, and interact with others on it. But first, we need to describe the motivating factors for contribution on SNS.

Previous research identified key motivating factors for participation in online communities, and gratifications contributors draw from it. For example, Dholakia et al. [66] identified five motivating factors for contribution online: purposive value (exchange of information), self-discovery (acquiring knowledge), entertainment, enhancing social status and maintaining relationships. Other studies [174,189] examined the motivations for active participation on Wikipedia, finding similar motivations and gratifications. Preece and Shneiderman [187] describe contributors' recognition and ability to build reputation as a major motivating factor for social contribution. Several studies examined contribution to SNS, and Facebook in particular. Both Joinson et al. [118], and Papacharissi and Mendelson [180] provided evidence that Facebook contribution helps support expressive information sharing and maintaining relationships.

While previous research mostly relied on self-reported measures for studying *why* people contribute online, we focus in this work on the ways in which contribution may affect user behavior and attention, using a large-scale dataset of contributors' actions that are free of any intervention.

3.2.1 Feedback Expectations and Site Activity

Feedback is a key component of any social exchange: it is important both for motivating contributions in the first place [40, 51, 135] and for evaluating social relationships over time [98, 143, 206]. Most, if not all, of the motivating factors for contribution identified by Dholakia et al. [66] depend on feedback from the online community, which suggests that contributors will expect some feedback. For example, *purposive value* is the value people derive from achieving a pre-determined purpose with the help of the community such as planning a trip or selling items. Similarly, if people post on Facebook to maintain relationships as suggested by previous research [118, 180] then it is reasonable that contributors expect responses. We investigate contributors' expectations more fully in Chapter 4. Here, our hypothesis is that in anticipation of new interactions, contributors will visit Facebook more frequently after posting. We refer to site visits that are not initiated by a notification (e.g. email sent by Facebook) as *self-motivated site visits* and hypothesize that:

H1 *Following a post, self-motivated site visits will increase.*

3.2.2 Shifts in Content Consumption Patterns

Contrasting theoretical explanations can be argued for changes in consumption of content from others after posting. On the one hand, contributors already spent time crafting their message, which may directly compete with the limited amount of time or attention they have to spend online after posting. On the other hand, contribution may take place at times when people are more free in the first place, and posting may be associated with a further increase in their consumption of content. The later argument is consistent with an account of participation taking place in a more active state [187] or aroused state in psychological terms, which was shown to be associated with increased levels of activity [83, 137, 199, 228].

At the same time, alertness or arousal may also mean more selective distribution of attention. Easterbrook hypothesized, based on studies of cue utilization, that arousal would lead to narrowing of attention [73], a finding that was later verified in an eye movement experiment [148]. If the act of posting makes one more selective, it is feasible that contributors would focus more on content from friends, as opposed to pages or other broadcast sources that are less specific to them.

The fact that habitual time-passing behavior is a major motivation for social media use (see [118, 180] for more details) leads us to believe that contribution would not come at the expense of content consumption, but rather enhance and make it more selective. Therefore, our hypotheses for content consumption are:

H2.a *Following a post, contributors will consume more content.*

H2.b *Following a post, contributors will consume more content from friends.*

3.2.3 Interaction Rates and Reciprocity

Contributing content is likely to have an effect on subsequent interactions with others, but different factors may positively or negatively affect the overall rate of interactions over time. On the one hand, higher interaction rates after posting may occur due to greater time availability, more active state or reciprocity. On the other hand, fatigue or a fixed-quota for interactions may result in a lower interaction rate after posting. We describe each of these arguments next and consider how these factors may affect interaction rates jointly.

Two of the arguments presented before, regarding contribution happening at more flexible times and more active state, can also explain an increase in the rate of interactions. For example, if people post when they have more free time then they may continue to interact more with content after posting. If contributors are more active and selective, as suggested before, they may choose to interact more in general, and with friends in particular.

In addition, ***reciprocity*** as the social norm of returning a favor, can also lead to higher interaction rate with friends after posting. In the realm of computer-mediated communication, even simple one-way communications such as a like or short "composed communication" (as defined by Burke et al. [38]) bare value. Therefore, receiving feedback from friends on a post, perhaps similarly to receiving a gift, creates indebtedness and calls for reciprocation. Reciprocity in social exchanges can take one of two forms: direct or indirect (also known as generalized reciprocity) [141, 195]. Direct reciprocity in our settings implies that contributors would interact more with the friends who responded to their post, while indirect reciprocity suggests more interactions with friends in general. In both cases, reciprocity results in more interactions with friends after posting.

In contrast to the above theories, fatigue or a fixed-quota policy may explain a decrease in interaction rates after contribution. If contributors consume more content, as postulated in the previous section, they may experience fatigue over time and engage in fewer interactions. Similarly, if people have a fixed amount of interaction they can engage in, and more content is consumed, the rate of interaction would decrease. We believe that the additional amount of content consumed would be relatively small and thus neither fatigue nor interaction limits would be dominant in our case. Therefore, our hypotheses are:

H3.a *Following a post, contributors are more likely to give feedback to friends.*

H3.b *Contributors are more likely to give feedback to those who responded to their content than to other friends.*

3.3 Methods

To test the hypotheses listed above, we devised a quantitative, within-subject, observational data analysis of Facebook activity logs. We wanted to isolate the effect of contribution as much as possible while controlling for other variables. To that end, we devised a comparative analysis of activity before and after posting on Facebook with a baseline of activity from the same individual at another time. We used feedback actions such as liking or commenting on someone else's content as our baseline because those are similar times where people are on Facebook and actively engage with others. As we will show in the results section, there are no material differences in the context in which feedback and contribution actions take place. But first, we describe the dataset and the measures used in our analysis.

3.3.1 Dataset

Our dataset consists of the activity a sample of Facebook users engaged in, without any intervention, around two types of actions: contributing content (C), and providing feedback to others (F). The data were de-identified and content of posts was not analyzed. ***Contribution*** is defined as the act of posting content to Facebook, for example, an individual posting a status update, sharing of a link or uploading a photo. ***Feedback*** is defined as reacting to someone else's content on Facebook: a like, a comment or re-share of others' content. Identifying such pairs of actions from the same individual allows us to compare behavior around contribution with a baseline of activity around feedback.

Figure 3.1 illustrates the setup of our dataset. Each individual had one contri-

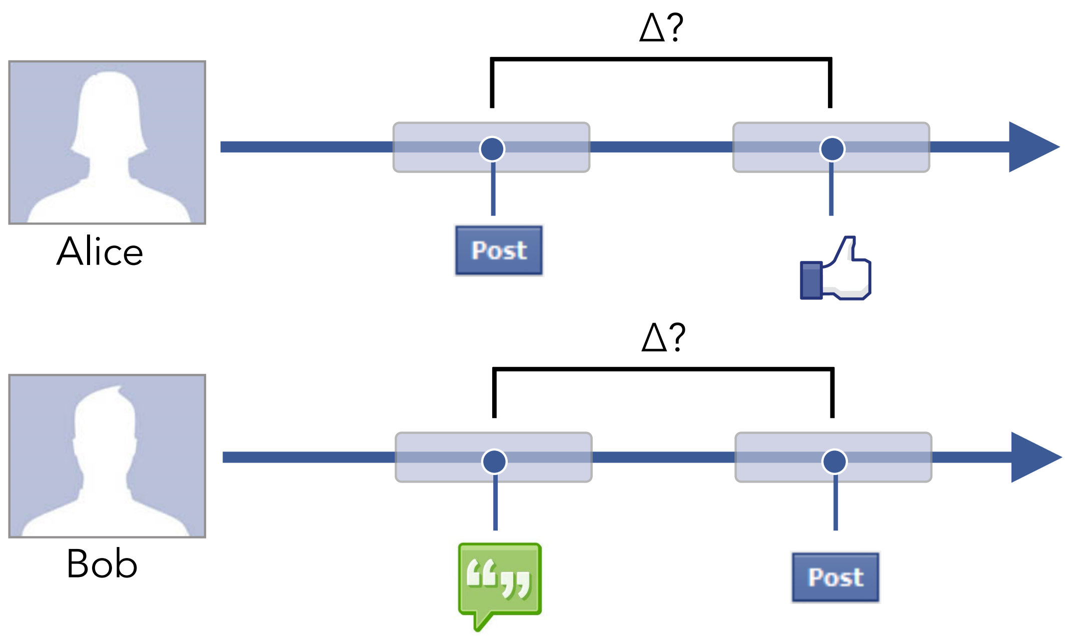


Figure 3.1: Research design: observational analysis comparing individuals' activity in the 48 hours centered around either a contribution action C (e.g. posting a status update) or feedback action F (e.g. a like or a comment). We chose pairs of anchoring actions C and F that took place a week apart, with equal number of pairs having contribution followed by a feedback action (as in Alice's case) and vice versa (as in Bob's case).

bution action C and one feedback action F that happened on the same day-of-week, one week apart from each other, in any order, using Facebook's web interface on a desktop device. In Figure 3.1, Alice posted a status update first and liked a friend's photo a week later, while Bob commented on a friend's post first and posted his own photo a week later. Both such sequences were included in this study.

We wanted to control, as much as possible, for external factors driving changes in individuals' engagement other than contribution. In cases where individuals had multiple pairs of actions we randomly selected one pair in order to equally represent people in our dataset. We further balanced the dataset such that there is an equal number of pairs with contribution happening first (like Alice) and feedback first (like Bob). We required both actions to have been performed on a mid-week workday (some time during the 24 hour span of Wednesday Pacific Standard time) to reduce bias from day-to-day variation. Our comparison of activity around actions included Facebook use through any device (mobile or not), but we required posting and feedback actions to have happened on Facebook's web interface using a desktop device. We focused on contributions happening on the web interface in order to reduce bias stemming from differences in device capabilities, screen resolutions, and versioning, all of which vary more on mobile.

Given the selected actions C and F for each individual, we compared their behavior 24 hours before and after each action. We chose a window of 48 hours around actions in order to respect the natural and regular periodicity of human behavior. The matching of actions did not exclude the other type of action from occurring around that same time. For example, it is possible that a given individual posted content some time before or after the feedback action F selected for the analysis, and vice versa. Stricter filtering, requiring no contribution by the user around the time of the F action selected for analysis, would have resulted in a much smaller dataset, which would have been less representative of the general population of contributors. Our non-strict selection criteria are noisier, but provide a less biased lower bound on the actual effect size of contribution versus feedback.

Our selection criteria of two actions per contributor yield a sample of individuals who are slightly more active than a reference population (RP) who used Facebook's web interface to post that week. The median person in our dataset is 37 years old (RP median=35), has 400 friends (RP median=344), has been using Facebook for 4.2 years (RP median=4.0), and has logged into Facebook 26.8 days out of the last 28 (RP mean = 24.3). Our sample is 55.7% female (RP: 51.9%).

In summary, our dataset includes C and F actions for 2.4 million individuals who posted content to Facebook or gave feedback to others using the web interface on two specific dates, February 11th and February 18th of 2015. The dataset is balanced in terms of the order in which contribution and feedback actions appear in it. Each individual included in the analysis has exactly one contribution action and one feedback action, where actions took place on the same day-of-week, interface and device. This set of individuals and actions is a sample of all users with actions that aligned with the selection criteria for those dates. Except for the analysis of self-motivated site visits that uses a subset of contributors, the rest of analysis uses the complete dataset.

3.3.2 Measures

We now turn to define the key measures used in our analysis.

Self-Motivated Site Visits

The measure of self-motivated site visits refers to the number of site visits that are not initiated by a notification, before any knowledge of feedback is available to contributors. We count site visits in terms of sessions, where each session is a sequence of actions of a logged-in user where actions are less than 30 minutes apart; if the individual was not active for 30 minutes, we count a subsequent action as a new session and a "site visit" [1].

When measuring site visits and sessions we want to ignore those visits that are

[1]We chose relatively long (30 minutes) sessions in order to enhance resilience for short-term attention shifts. We experimented with shorter spans and found similar results.

due to offline notifications – users getting e-mail, SMS or mobile push notifications about Facebook activity that invites them to come back to the site. Therefore, we examined a subset of contributors for whom Facebook did not generate any offline notifications in the two days preceding an action and the day following it. This subset of contributors did not receive notifications because they disabled offline notifications explicitly in their profile preferences or there was no activity that led to a notification being generated for them. While this sub-population many not represent the entire population of people who post on Facebook, it allows us to focus on a large sample of more than 150,000 people and rule out notifications as the factor affecting behavior.

Stories Read

We measure content consumption by examining the number of News Feed stories read by contributors in the 24 hours preceding or following an action. Facebook's News Feed is the landing page for people browsing to facebook.com or opening the mobile app, where content from friends and followed accounts is algorithmically ranked. A story is considered read if it was visible in the central portion of the user's screen for a least two seconds. Note that we explicitly exclude stories that originated from the contributor herself as this may appear in her News Feed. In addition, our measure of stories read is not directly impacted by notifications because stories read as a result of clicking on a notification (on any platform) are logged separately and thus not counted towards our measure of stories read[2].

[2]Notifications may affect the number of stories read *indirectly* by encouraging people to visit their Facebook profile more often even if they do not directly follow the link on the notification. However, these changes in engagement are moderated by the individual and therefore an integral part of the behaviors we wish to study.

Interaction Rate

We define interaction rate as the proportion of likes or comments given per News Feed story read by the contributor in 24 hours before or after activity. Interaction rate is the portion of stories read from others (as defined above) that contributors liked or commented on directly from the News Feed. In other words, our measure of interaction rate excludes likes and comments that occur in other parts of Facebook such as Timeline or groups. Here again, any interactions with the contributor's own content (reads, likes, comments) were excluded.

3.3.3 Statistical Analysis

For most of the analyses described below, we use Difference in Differences (DID) analysis in order to estimate the effect size of contribution while accounting for exogenous variation external to contribution. DID is a common statistical analysis technique used in observational data analysis to mimic a random assignment experimental design. DID estimates the effect of "receiving treatment" (in our case choosing to post) by controlling for a trend evident in the control group (feedback action in our case). In particular, DID analysis for our measure of stories read would be calculated as follows:

$$DID_{reads} = \left(R_C^{after} - R_C^{before}\right) - \left(R_F^{after} - R_F^{before}\right) \tag{3.1}$$

Where R is our measure of stories read in this case, and indices of after/before designate period relative to contribution C and feedback F actions for which the measure was computed. The underlying assumption in DID is that the treatment and control groups are comparable in every respect other than the assignment to treatment or control. Recall that we compare activities from the same individuals, day-of-week, interface, device, and comparable context as we will show in the next section. Therefore, we believe DID approach is particularly adequate for our settings since it highlights differences in engagement after contribution and contrasts them with the trend in engagement around comparable feedback action from the same person.

Two elements in the way we apply the DID help reduce selection bias and bias due to ordering effects. First, DID is often suspected for a selection bias in the assignment of individuals into treatment and control groups. In our analysis, however, both control and treatment groups include the *same* people, which eliminates individual differences between groups by design. Second, we reduce bias due to ordering effects by choosing a long gap in between the actions we examine (C and F) and balance the occurrence of actions in any particular order (contribution or feedback first). While we cannot rule out that one action may effect another action over a long period of time, our preliminary analysis suggest a diminishing difference in activity after 24 hours from posting or giving feedback. We use a much longer gap, of one week in between actions C and F, to further eliminate such interactions. In addition, the balanced order at which contribution and feedback actions appear in our dataset reduces the bias that observed effects are due to a one-time external event that affects only one of the conditions, or other time-based trends like increase in use over time.

All of our statistical tests were done using the standard technique of bootstrapping, with 10,000 replicas. We estimated means and 95% confidence intervals around them using the bootstrapped samples. Bootstrapping is more stable, asymptotically more accurate than estimates of confidence intervals based on a single empirical sample, and do not require normality assumptions [68]. We also favor bootstrapping over traditional paired t-tests since the latter tends to yield highly-significant p-values in all cases due to the sheer size of the sample (hundreds of thousands people in our smallest sample).

3.4 Results

In this section we present the results of our comparative analysis of individuals' behavior around contribution and feedback actions. Before we address the hypotheses described in the Background section, we first establish the validity of comparing activity around feedback and posting actions to each other.

3.4.1 Preliminary Analysis

We performed a series of descriptive and comparative analyses to better understand the data, and verify that there are no material differences between the contexts in which people performed the different actions (contribution and feedback).

A central question to our analysis is how active people on Facebook are before and after different actions. Figure 3.2 addresses exactly this question by presenting on its top panel the percentage of people in our sample who were active on Facebook as a function of time, for 24 hours before and after each of the two actions that they took. The figure shows activity around contribution action (solid red line) and feedback action (dashed black line). Data points in the figure correspond to the percentage of the 2.4M people in our dataset that had used Facebook during each 20 minute time bin on the x-axis. For example, at the exact time of an action (time

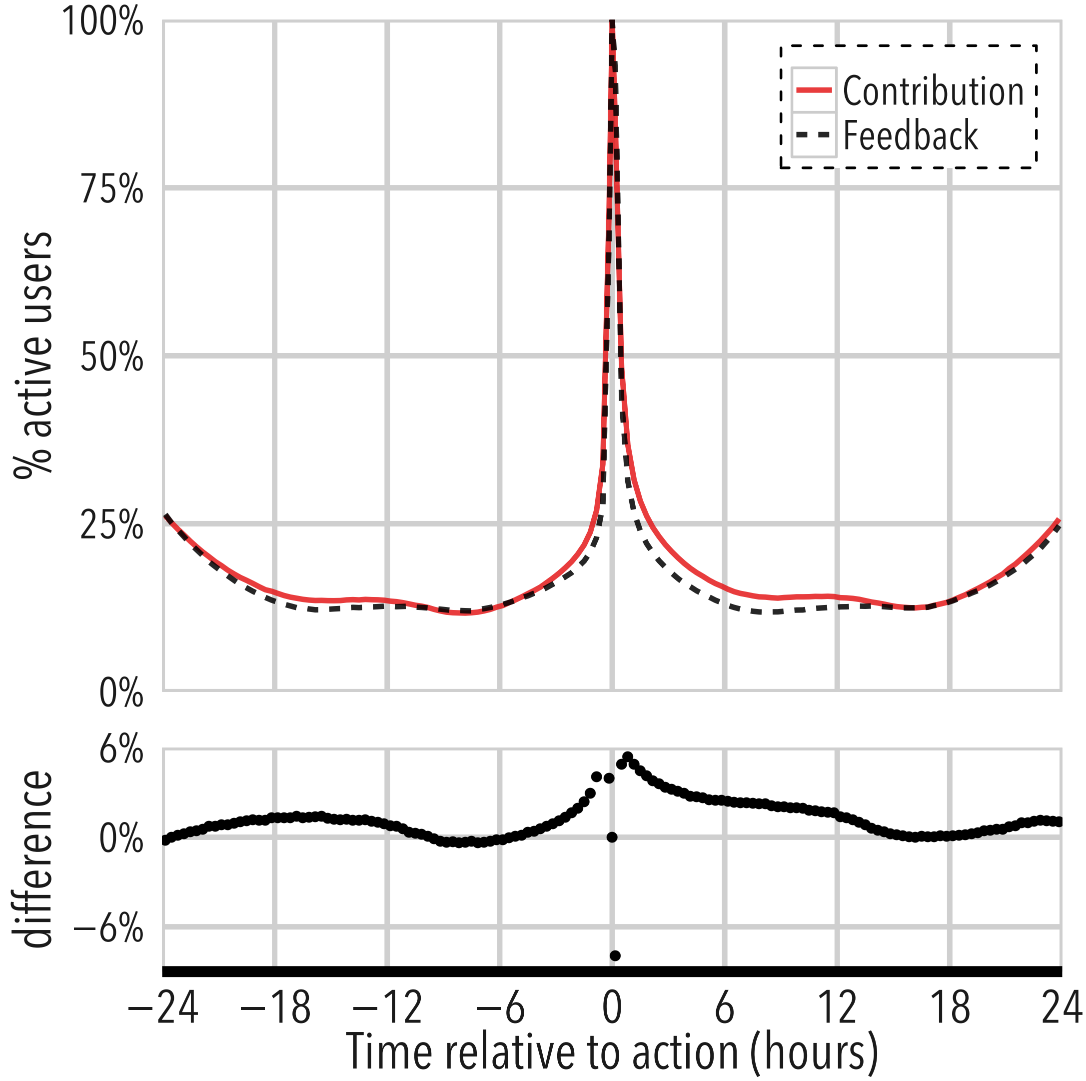


Figure 3.2: Percent of active users (top panel) in the 48 hours around contribution and feedback actions, with the differences ($A_C - A_F$) visible on the bottom panel. The 95% confidence intervals were too small to be visible.

0) all of the individuals in our dataset were active on Facebook since they either posted content or gave feedback. As a result, the plot spikes for both conditions at exactly 100%. The bottom panel shows the differences between the percent of active sessions around contribution and feedback (in other words, the difference between the solid red and dashed black lines on top). Figure 3.2 clearly shows that except for the 20 minutes immediately following an action, contributors are more

active for several hours both before and after posting content compared to their activity around feedback at the same time frame. The only exception is the 20 minutes shortly after feedback where people are more likely to continue to engage with News Feed content rather than leave Facebook, as 7% of contributors do immediately after posting content.

The discontinuity observed at around zero in Figure 3.2 informed our decision to exclude the hour immediately following or preceding an action from our analysis. The fluctuation visible in the differences panel about an hour before the action and about an hour afterwards indicate short-term differences, probably stemming from the different sequence of user interactions at which feedback and contribution occur in. Therefore, for the rest of our analysis *we use a window of 48 hours around an action, but exclude the 2 hours centered around an action.*

Three interesting findings emerge from Figure 3.2 regarding the higher activity levels around contribution, and its return to baseline levels at the ±24 hour period. First, we see that higher activity levels start as early as six hours before contribution and last more than 12 hours afterwards. The fact that contributors are more active even six hours before contribution is interesting and cannot be simply explained by the additional time necessary to conceive and articulate a post. The higher levels of activity after contribution are likely to be driven, at least in part, by notifications that contributors get due to feedback on their content. Below, when we address hypothesis H1, we show that notifications are not the only factor that explains higher level of user engagement after contribution. Second, Figure 3.2 shows uptick in activity in the 24 mark before and after each action. The increased activity indicates regular patterns in user activity and justifies the choice of 24 hours for analysis. Lastly, the diminishing differences at the ±24 hours

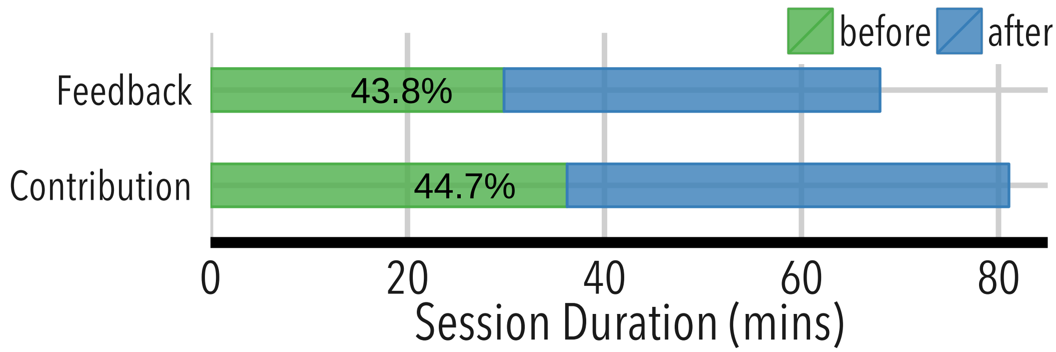


Figure 3.3: Average duration of sessions around feedback/contribution actions, with the percent of time spent before the action.

relative to actions demonstrate that the effect is largely dissolved in a day.

We further examined the data to make sure posting sessions are not fundamentally different than feedback sessions. We looked at the length of sessions and the position within a session where actions were recorded. As before, our definition for a session is a sequence of actions that are less than 30 minutes apart from each other. Figure 3.3 shows the average duration of contribution and feedback sessions. While a contribution session lasts more than 80 minutes on average, feedback sessions are significantly shorter, lasting only close to 68 minutes (95% confidence intervals were 20 seconds long, too short to be visible on the relevant scale). The long duration of sessions is likely to be a result of the long sessionization window used, but the relative position of actions within sessions are more robust. The figure shows that contributions are positioned similarly within a session, with 43.8% of the session time passing by before feedback occurs and 44.7% for contribution. A one percent increase in the relative position of contribution within session is equivalent to $\sim$ 50 seconds, which is relatively small and could potentially be explained by the extra time required to compose a post.

We also verified that contribution and feedback actions occur at comparable time of day. For example, we wanted to make sure our dataset is not biased such that contribution takes place in the morning and feedback at night. By computing the difference in time of day for each pair of user actions, we find no statistically significant difference. The average difference in time of day is bound by a 95% confidence interval of $(-3.2, 3.0)$ minutes. No difference (difference of zero) is well within the 95% confidence interval. Therefore, we conclude that contribution and feedback actions occur at roughly the same time of day.

In summary, the preliminary analysis provided evidence for the adequacy of our comparative analysis of individuals' engagement in the 24 hours before and after contribution and feedback actions. This initial analysis informed our decision to exclude the hour right before and after an action for the rest of the analysis and established that contribution and feedback actions are positioned comparably within sessions and within the day.

3.4.2 Site Visits

We test hypothesis H1 about an increase in self-motivated site visits by conducting DID analysis on our measure of site visits. Recall that for this analysis, we wish to neutralize the effect of notification. To this end, we focus on a sub-sample of 150,000 people for whom Facebook did not send any offline notifications in the 48 hours preceding an action and 24 hours after. These people either chose not to receive offline notifications or there was no activity that led Facebook to generate a notification for them.

Figure 3.4 shows the average number of site visits for the same set of people before and after contribution and feedback actions. For instance, we see that in

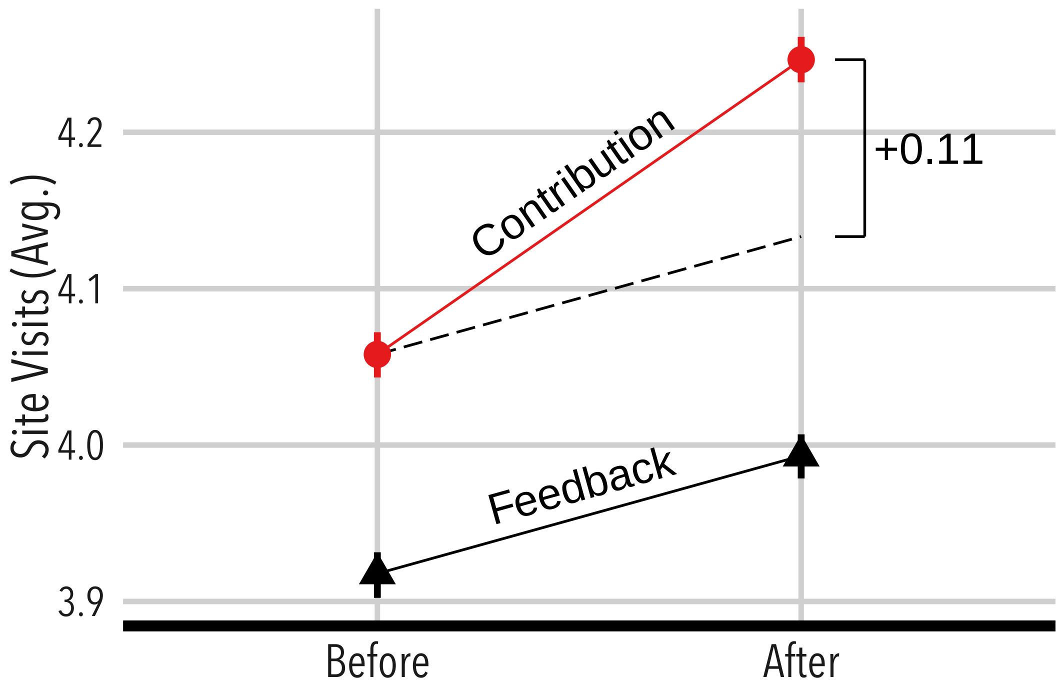


Figure 3.4: Difference in Differences analysis shows a significant increase in self-motivated site visits in 24 hours before/after activity (95% CIs). The dashed line designates the DID prediction for levels after contribution based on the trend evident in the Feedback condition. Square brackets highlight the significant increase in site visits of +0.11 on average.

the 24 hours before feedback individuals had an average of 3.9 self-motivated site visits, while closer to 4 site visits after taking a feedback action (excluding the feedback/posting session itself). The dashed line designates the projected number of site visits, if the general trend apparent in the feedback condition occurred at times of contribution.

The evidence from Figure 3.4 is that the number of self-motivated site visits after contribution exceeds the projection by 0.11 site visits on average, the difference is statistically significant, and is in line with hypothesis H1. The figure shows that in the 24 hours after both feedback and contribution actions, people are visiting Facebook more often even without getting any offline notifications. On top of the projected increase in site visits from giving feedback, individuals visit Facebook 0.11 (+2.6%) more often on average when posting. These findings show that there is a small increase in site visits not stemming from notifications or from merely taking an action on the site. We further discuss the implications of this increase for system design in the discussion section.

3.4.3 Content Consumption

We now examine how contribution affects an individual's attention to content. Hypotheses H2.a and H2.b postulate that contributors will consume more content overall and particularly more content from friends, respectively. We test these hypotheses using a DID analysis on the measure of stories read, counting stories viewed for at least two seconds in the central portion of the user screen[3]. The measure of stories read will increase if people reading more pieces of content or decrease if they are skipping content.

Figure 3.5 presents the DID analysis of, separately, stories read from friends and stories read from other sources like Facebook Pages. While fewer stories are read on average after giving feedback (evident in the decreasing trend in black), the trend for contribution is positive for content from friends and neutral for pages. Similar patterns emerge when we do not distinguish between friend and page content – people read (on average) slightly fewer stories after engaging in feedback actions and about three more stories (+2.1%) after contribution. The number of stories read from pages also increases on average by 1.2 (+1.8%) compared to the DID projection, and are statistically significant as explained above. These find-

[3]We note that there is potential for more neunced measure of reading of stories presented in social feeds, much like the work presented in Chapter 2.

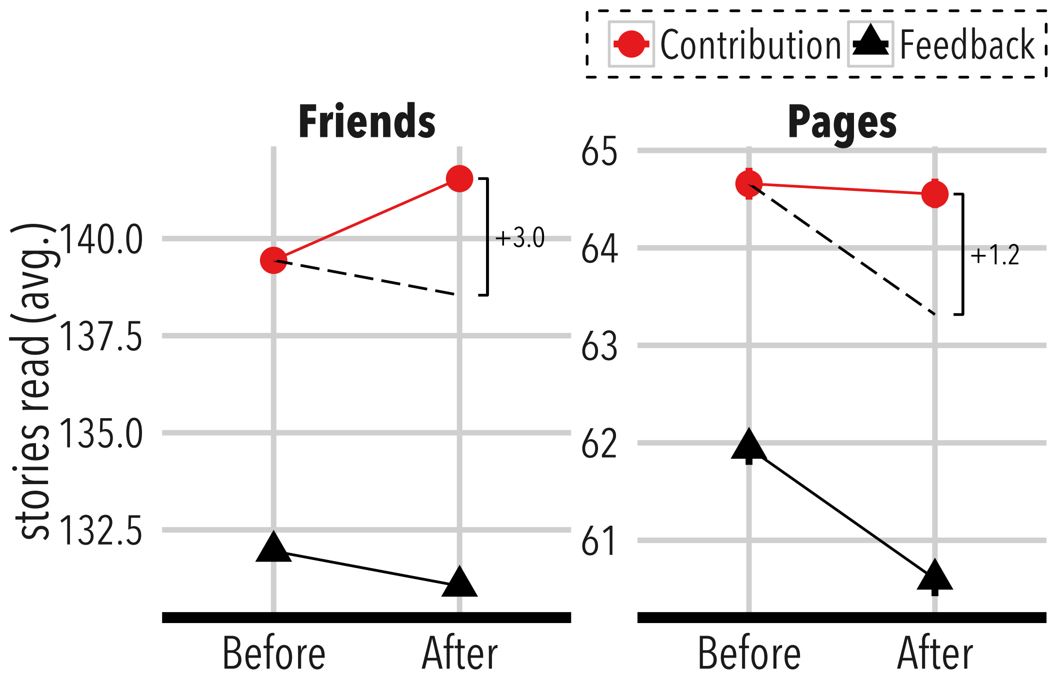


Figure 3.5: Difference in Differences in number of Newsfeed stories read in 24 hours before/after activity (95% CIs). Dashed lines designate the DID prediction for levels after contribution based on the trend evident in the Feedback condition. Square brackets highlight the significant increase in the number of stories read.

ings support an increase in overall content consumption and consumption of friend content.

Interestingly, only the number of stories read from friends increases compared to pre-contribution levels, suggesting a shift in attention towards friends but not others. These results are consistent with the previous observation that contributors remain active for longer periods of time after contribution, but also indicate that the additional time is spent more selectively on friends' content.

We performed further analysis to verify that the above changes in individual consumption habits do not simply stem from differences in the content available to people in the News Feed around feedback actions and contribution actions. As a crude measure of content availability, we test whether the distribution of content available from weak and strong ties changes before and after the C and F actions in our dataset. We use a measure of tie strength that is based on the frequency of past communication between two individuals and we simply associate the tie strength of the friend authoring the post with the content viewed by the contributor. We conducted DID analysis on the tie strength associated with content and found no significant difference. Therefore, we conclude that the content available for consumption around contribution is not significantly different than the content around feedback actions.

In conclusion, we find that contributors consume more content before posting, but increase consumption even further after posting, particularly of friends' content. We rule out an explanation that those changes in consumption habits simply arise from News Feed ranking or other differences in the availability of content at different times.

3.4.4 Interaction Rates and Reciprocity

Previous sections established that contributors are more engaged around contribution and consume more content, even though the content itself remains the same. We now examine whether posting content affects individuals' decisions to interact with others as postulated by hypotheses H3.a and H3.b.

Figure 3.6 shows our DID analysis of interaction rates with content from friends and pages. The bottom left panel, for example, shows that before posting, users comment on 0.74% of the stories they read from friends and that this rate significantly increases to 0.77% after posting. DID that were statistically significantly are

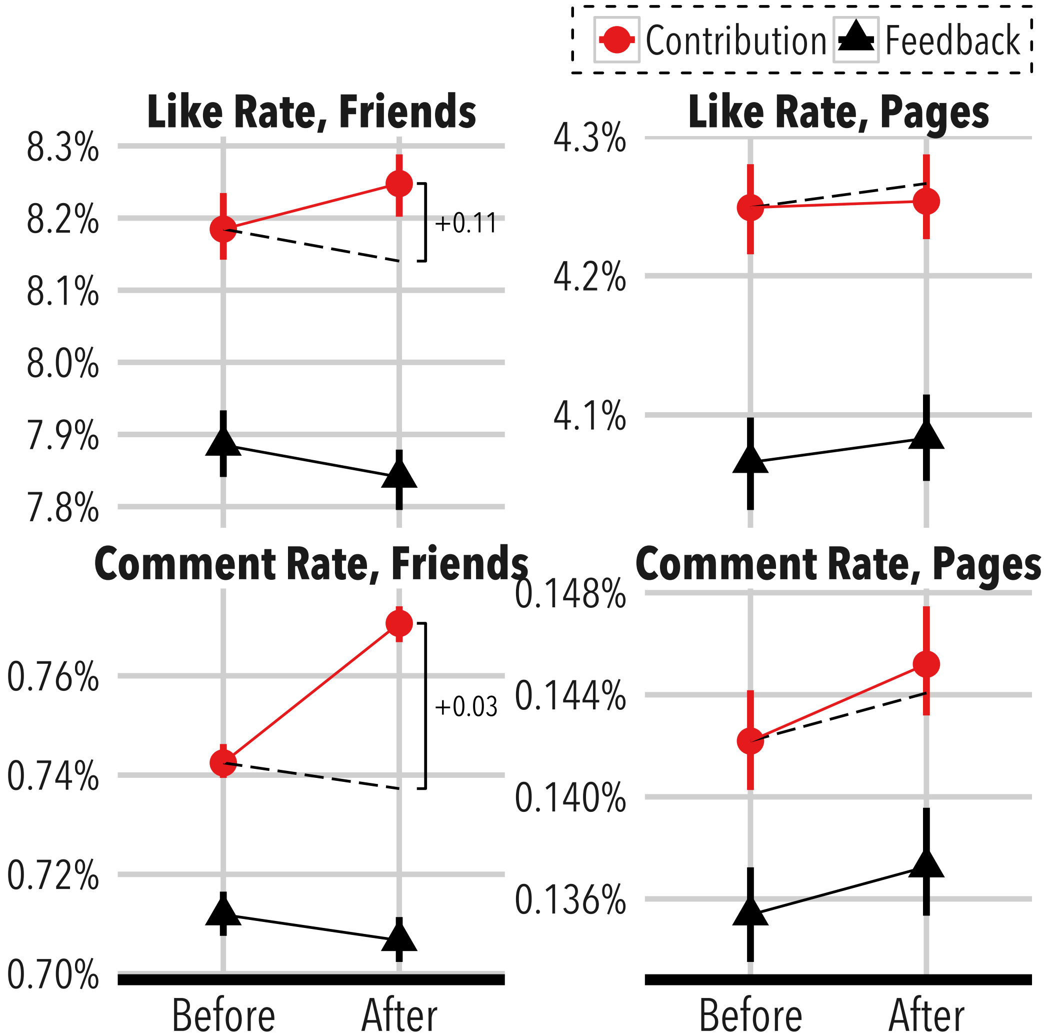


Figure 3.6: Difference in Differences in Liking and Commenting rates with friend/page content in 24 hours before/after activity (95% CIs). Dashed lines designate the DID prediction for levels after contribution based on the trend evident in the Feedback condition. Square brackets highlight the effect size when significant.

highlighted in the figure by square brackets as can be seen in the left two panels.

Several interesting findings should be noted about Figure 3.6. Across the board, the interaction rate before and after posting is significantly higher than the rate when simply giving feedback to others. After giving feedback, there is no significant change in the interaction rate or even a slight decrease compared to pre-feedback levels. In contrast, the interaction rate with friends after posting increases significantly for both likes (an absolute gain of +0.11%, which is a 1% gain relative to the "before" level) and comments (+0.03%, 4% gain). The changes in interaction rates are statistically significant, substantial, and even more interesting given that there are no significant changes in interaction rate for pages (right side of Figure 3.6). These findings provide supporting evidence for hypothesis H3.a about increase in interaction rate with friends, and provide counter evidence to the idea of a "fixed-quota" or decision fatigue over time.

Next, we provide a deeper examination of the interaction rate with friends to understand the role of reciprocity in these interactions. For example, consider an individual, Anna, who posted a status update on Facebook and later saw stories from two of her friends, Brian and Colin. Hypothesis H3.b suggests that if Colin gave feedback to Anna's original post, she would be now more likely to comment on Colin's post than on Brian's. Of course, it is possible that Anna and Colin are simply more likely to interact with each other in general, for example, because they are closer friends. To control for this difference in relationship, we use Propensity Score Matching (PSM) with a score based on tie strength (as described in the previous section). For every person posting and friend who commented/liked their post (designated as indebted), we match an equally close friend who did not comment/like the person's post (control). We verified that the average tie strength in the indebted and control groups is not significantly different. We can then compare the interaction rates of contributors with content viewed from the two groups, where the only difference between group is whether the friend previously responded to the contributor's post or not.

Figure 3.7 shows liking and commenting rates for indebted and control groups

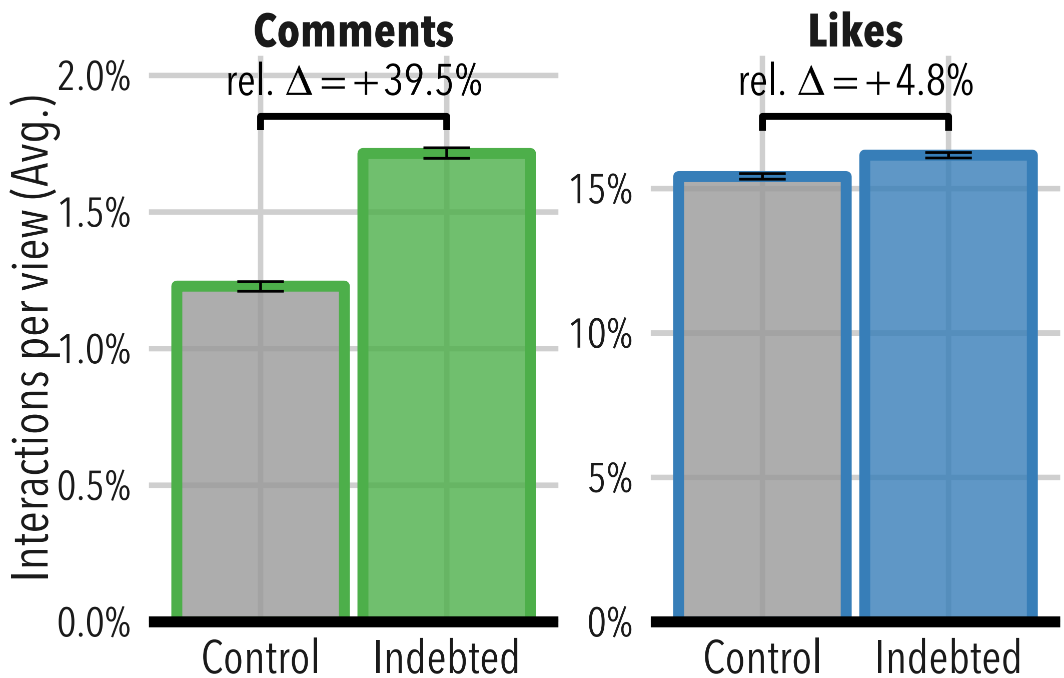


Figure 3.7: Commenting and Liking rates (95% CIs) on friends' content who responded to the contributor's post (indebted) or not (control), controlling for tie-strength.

of friends. On the left part of the figure, we see the rate in which contributors commented on content they saw, split to friends who responded to the contributor's post (indebted condition) and equally close friends who *did not* respond to contribution (control group). As the figure shows, commenting on content from friends who responded to contribution (indebted condition) occurs at an average rate of 1.71 comments per 100 stories read (1.71%, solid green bar, second from the left in Figure 3.7). In contrast, contributors only commented on content from friends who did not respond to their contribution (control condition) at a rate of 1.22 comments per 100 stories read (1.22%, first bar from the left).

The relative change in commenting rate in the indebted condition is a large 39.5% increase over the control group and a more modest 4.8% increase for likes.

These findings are highly significant, align well with the theory regarding direct reciprocity, and supportive of hypothesis H3.b. As a side note, notice that rates of interaction in Figure 3.7 are much higher than on the left side of Figure 3.6. This observation is reasonable since the interactions in Figure 3.7 are with friends who responded to contribution, which are more likely to be friends one frequently interacts with, and thus results in higher rates.

We note that the interaction rates increase also for friends who did not respond to contribution, but at smaller rates than those in Figure 3.6. This result is consistent with indirect (generalized) reciprocity as we described in the background section. However, more complex analysis is needed to substantiate indirect reciprocity in this case because it requires careful control for the activity of others in addition to the actions taken by the contributor herself.

In summary, we see an overall higher interaction rate around times of contribution, with further increase after posting, especially for friends and not others. We find that whether equally close friends respond to contribution or not affects the likelihood of future interactions with their content, resulting in substantially more likes and comments for friends who responded to contribution.

3.5 Discussion

In this study we examined individuals' behavior when posting to Facebook and found significant changes in engagement both before and after contribution. We discuss here why we think these shifts in individuals' engagement occur and what SNS can do to better support contributors.

3.5.1 Contribution and Changes in Engagement

Higher Engagement Before Posting

A salient theme across all of our findings is that contribution is associated with more active engagement even before contribution takes place. These findings can be explained by external factors that influence both posting and engagement, or by higher engagement leading to contribution. External factors may include contribution taking place when people have more free time to spend on Facebook, being in a more active and alert state, or simply when people attend an event together with their friends. All of the above can increase engagement and be associated with posting. Alternatively, higher engagement can also lead, through various means, to contribution. For example, being exposed to interesting content from others can inspire or simply remind people to post.

The fact that posting is positioned similarly within session to feedback actions suggests that people often spend considerable amount of time on others' content before posting their own content. These findings are consistent with the notion of influence from Social Learning Theory [10], which posits that people learn by observing others and gradually behave more similarly to them, even without any external incentives. Whether increased engagement leads to posting remains an open question, with implications for newcomers [40] as well as contributors in general.

Higher Engagement After Posting

Our findings show that contributors are not only more engaged before posting, they also increase their engagement after posting at a higher rate than they do after feedback activities.

The result showing an increase in site visits after posting (without notifications) supports the idea of self-motivated changes in individuals' engagement. We believe that some of the additional site visits are motivated by anticipation of feedback and that similar changes occur for people who do get notifications. On platforms at the scale of Facebook, an effect of 2.6% increase in site visits translates into hundreds of thousands of additional site visits each day that are presumably motivated by anticipation of feedback.

Most consistent with all of our findings, both before and after contribution, is that posting is associated with a more active and alert state. These results interact with ideas from attention theories looking at how we allocate attention [73]. Key recent theories of attention deal with selection processes (what do people pay attention to) and vigilance (how do we sustain attention over time) [54].

Other alternative explanations for the increased engagement after posting are consistent with some of our findings, but not all of them. Some of this higher level of activity can simply be tied to contributors interacting with the responses on their post. However, it was not established until now that other activities on Facebook, unrelated to contribution, also rise. In addition, if people post when they have more time to spend on Facebook it is feasible that they will continue to engage even after posting. However, looser time constraints around contribution do not immediately explain the changes in selectivity of consumption and interactions with content. Similarly, attending an event with friends and posting about it on Facebook could explain some of the increases in interaction rates with friends, but not the persistently high levels of engagement with non-friend content. Reciprocity, as we will discuss in greater detail next, does not explain the high engagement levels before posting or with page content afterwards.

Contribution and Reciprocity

Once contribution is made and responses come in, it is reasonable that contributors will reciprocate, but the magnitude and speed at which it occurs is somewhat surprising. In the 24 hours after contribution, commenting rate on content from friends who responded to contribution increased close to 40% more than the control, compared to a more modest (but still substantial) 4.8% increase for likes over the control. While reciprocity is a well-documented and replicated phenomenon, this is the first time the immediacy of the effect is shown in social media settings and at large scale.

An important question is whether the reciprocity effect is deliberate. In other words, do contributors seek out opportunities to comment or like the content from those who gave feedback on their content? Or are they implicitly and unconsciously inclined to reciprocate because they have positive feelings towards those who gave them feedback? In offline settings, a well-established result shows that we are more likely to like people who evaluate us positively [14, 200], or in other words, "we like those who like us." [87][4]. These previous findings may suggest that individuals develop more positive feelings towards those who give them positive feedback, and as a result may be more inclined to like or comment on their content. A

[4] As [87] shows, we even like those who positively evaluate *others* – "everybody likes a liker".

dual processing mechanism can perhaps support both a deliberate and conscious reciprocating behavior as well as a more implicit and unconscious response [82]. Our working assumption is that both deliberate and more implicit mechanisms are in effect here, and we leave it for future work to distinguish between these two potentially competing mechanisms.

Under the assumption that at least some of the feedback is due to a deliberate attempt at reciprocity, these findings are in line with the claims of the hyper-personal model in Social Information Processing [215]. In particular, the model captures how interactions in CMC get amplified over face-to-face communication, which can then turn into greater indebtedness and reciprocity. This theory aligns well with asymmetric increase in comments versus likes; the different time investment and meaning for comments over likes has been well documented, and the fact that contributors choose to comment more than like content may indicate a greater sense of indebtedness on their part. These findings are in line with the changes in tie strength highlighted in [38] and the preference for "composed communications".

3.5.2 Limitations

While we attempted to carefully design our analysis and control for key factors, the study still has several limitations.

First, as a purely observational study our findings are only suggestive of the causal relations between posting and user engagement. We believe that posting does lead to an increase in overall activity and changes the composition of actions contributors take on the site. Similarly, we think that seeing more engaging content can encourage, inspire, or simply remind people to post their own content.

Nevertheless, by merely observing the actions people take on Facebook we cannot definitely discern these causal explanations from other alternative explanations that were mentioned before.

Second, by focusing on aggregate measures of activity over a period of 24 hours we average out some of the behaviors that only occur at shorter time spans and lose the ordinal aspect of activity. For example, our measures are likely to smooth effects that happen on the next session immediately following a post, especially since we are excluding the one hour before and after posting.

Lastly, including in our analysis contributors who were active on Facebook at two different times a week apart introduces some selection bias. While we did work with a sample of millions of people, our methodology is not suitable for drawing inferences about less active contributors. In addition, the dynamic changes we observed on Facebook may not generalize to contributions on other SNS. There is certainly room for arguing that different forces will prevail in other social networks and we look forward to see future work using the methodology of this chapter to investigate these effects in other social media platforms.

3.5.3 Future Work

Future studies could examine the role of feedback in modulating contributors' behavior and the time-range for these effects. While most feedback received on SNS is positive, even in sites with a weaker sense of identity and friendship than Facebook (see Cheng et al. [49] for details), the question remains as to how feedback affects behavior. A closer investigation can examine the temporal aspect of the behavioral changes we identified and try to link the short-term changes in engagement with long-term effects on relationships. Even more challenging is the fact that the effect of feedback is likely to depend heavily on contributors' expectations, which are subjective and not directly observable. Chapter 4 pursues this direction to an extent; it uses a mixed-methods approach to examine subjective expectations of contributors, identifies factors affecting those expectations, and evaluates how well can expectations be predicted in order for systems to utilize them in practice.

Other extensions of this work can investigate how engagement changes as a result of individual differences as well as differences in form and substance of the posted content. Different populations (e.g. women and men, young and old) engage differently with SNS [113,127] and analyzing the effect for different sub-populations can reveal additional differences. Posted content may very well differ in form, style, content, effort and intent embedded in it, which all call for further exploration of their effect on contributors' behavior.

3.5.4 Design Implications

Our findings suggest a potential for designing adaptive systems that encourage social participation, help contributors focus on the content that is important to them, and recommend content based on the context of actions. First, the observation that contributors are more active six hours before posting opens possibilities for researchers to design nudges for contribution at times of high engagement and evaluate whether these are perceived as beneficial. Second, the importance of feedback from friends may call for refinement of user experiences around feedback interactions and rethinking how to surface these to contributors. Lastly, we demonstrated that individuals' engagement with content depends on the context in which it occurs (e.g. posting on Facebook), a finding that recommendation systems can use to differentially value explicit feedback from people. Further research is needed to better serve the naturally-changing needs, expectations and preferences of contributors.

3.5.5 Conclusions

In this chapter, we examined the short-term engagement of individuals when posting to Facebook and contrast it with their activity at another time when they give feedback to others. Our within-subject comparative analysis resulted in two major findings. First, we found the people are more active around posting actions than feedback actions for about six hours before posting and more than 12 hours afterwards. The deeper engagement happens both before and after the time of posting and across all the measures we examined: self-motivated site visits, stories read and interaction rates with content. Second, contributors not only start more engaged, but also further increase their engagement after posting at a higher rate than any other feedback action. Self-motivated site visits increase after posting as well as the consumption and interaction with friends' content, but not others.

We highlighted a few areas in which interface design can better support contributors, encourage social participation and possibly improve content ranking in recommendation systems. Taken together, the findings in this chapter identify an important pattern of engagement that is consistent with key behavioral and social theories. It is possible that underlying all of these is a distinct cognitive state that is associated with contribution, greater desire for social connection and more willingness to engage with friends. However, we believe that additional evidence needs to accumulate before a more holistic theory could emerge, explaining individuals' attention in the complex social context in which it is embedded.

The next chapter in dissertation continues this line of work on attention in social setting, and was very much inspired by the findings in this chapter. The increase found in self-motivated site visits after posting helped us make two important observations: (1) the literature is missing an account of people's expectations when communicating on social media, and (2) information about contributors' expectations for attention are not available for recommendation systems of social content. The next chapter addresses both of these aspects by taking a mixed-methods approach to study expectations of attention from others on social media, and by providing accurate predictive models for incorporating expectations in social systems.

CHAPTER 4

EXPECTATIONS FROM THE EGO NETWORK

The previous chapter as well as other previous research examined the attention people pay to social media as a whole and to content items on social media more specifically [4,20]. This chapter investigates a complimentary aspect of attention allocation that is unique to social interactions: the expectations people have for getting attention from their communication partners. We conduct two large-scale surveys that cover different aspects of people's expectations for feedback on Facebook both at the level of a single post and from specific people in the ego's social network. We combine survey responses with log data of people's activity on Facebook to study the factors associated with feedback expectations. We demonstrate the potential for improving individuals' well being by fulfilling feedback expectations, and describe accurate predictive models that social networks could potentailly use to improve recommendations of social content.

4.1 Introduction

Posting and receiving feedback shapes the experience of people on social media. Feedback, whether it is expressed via lightweight one-click communication or more, carries social value that motivates people to post, provides social and emotional support, and shapes relationships over time [37, 38, 77]. While a considerable body of work has studied the role of feedback in social network sites [37, 41, 77, 99, 165, 211], little research examined the expectations for feedback people have when sharing content to their social network. In this paper we focus on the feedback expectations associated with posting content on Facebook, and the way that expectations vary from one person to another, are dependent on the properties of the post, and are impacted by the relationship to other individuals.

Expectations are an important measure that guides social behavior and attitude, which can inform the design of social network sites. Expectations motivate us to take action and help us choose among alternatives. For example, expectation of feedback is a key motivating factor for participation in online forums, contribution to Wikipedia, and posting on social media [66,118,136,174,189]. The previous chapter in this dissertation showed that people visit Facebook more often after posting a status update, potentially in expectation of feedback, even when there was no evidence of actual feedback received.

Despite the importance of feedback expectations, little is known about people's expectations for feedback from their online social networks. Existing theories of interpersonal communication such as Expectancy Violation Theory [30, 33] do not immediately translate to expectations from online social networks, where feedback is often aggregated and knowledge of viewership is lacking or incomplete. Previous studies on social media touched on several aspects related to expectations such as the "imagined audience", perceived audience size, feedback preferences, norms evolution and violation [17, 38, 144, 155, 159, 196], but did not directly model feedback expectations. Since little is known about people's feedback expectations, it is unclear how actual feedback is perceived and how exactly feedback affects people's experience on social media.

This work examines people's expectations for receiving Likes and Comments on Facebook, and the relation between fulfillment of expectations to feeling connected to one's Facebook friends. We build on Expectancy Violation Theory (EVT) [30,33] as inspiration for the conceptual framework used in this work. We conduct a comprehensive, in-context examination of the factors associated with feedback expectations immediately after posting on social media. Our investigation borrows from EVT the key elements of the model that contribute to expectation: the communicated content properties, the characteristics of the individual who posted it, and the individual's relationships to others on the platform. Not only do we look at factors that contribute to expectations, we also study the fulfillment of feedback expectations and its relation to feeling of connectedness to one's Facebook friends, an important outcome for individual well being [41, 128].

To this end, we use two large-scale surveys to ask people about feedback expectations, fulfillment of expectations, and connectedness to friends. First, we surveyed people immediately after posting on Facebook and asked them about their feedback expectations on that particular post, both in terms of total feedback and from specific Facebook friends. We complemented survey responses with de-identified, aggregated log data to understand how the characteristics of the individuals, posts, and interpersonal relationships are associated with feedback expectations. Using this dataset, we built predictive models of feedback expectations. In addition, we conducted a separate survey, asking participants about an earlier post they made, and how the amount of feedback received compared to their expectations. We also asked participants in this lagged survey how connected they feel to their Facebook friends in order to establish a link between fulfillment of expectations to connectedness.

The chapter offers a general framework for thinking about feedback, behavior and attitude on social media in the context of expectations. We identify the significant factors associated with higher than usual feedback expectations on a post and the important properties of relationships linked with expectations from specific friends. In addition, we show an association between the congruency of feedback and expectations on a post to an important outcome – individuals' connectedness to their friends. Last, our predictive models can be used in practice to evaluate how well people's expectations are met, and to explore ways of potentially improving the experience of people when posting on social media platforms.

4.2 Related Work

Previous work identified several benefits of posting on social media, among them are self-expression, relational development, social validation, and approval [12]. Social media use had been shown to impact both social capital and well being [39, 41, 52, 77, 78, 134]. Many of the benefits of social media use come through feedback mechanisms such as Likes and Comments on Facebook. Support and help via feedback are important for alleviating loneliness [41,128], getting emotional support after losing a job or when sharing emotional content [36, 37], enabling information seeking [99, 165], maintaining relationships [77, 211, 212], and more. While feedback is a necessary component for all of these benefits, as Bazarova et al. pointed out, it is one's subjective satisfaction from the feedback received that determines its value for the communicating individual [13].

Other work studied people's perceptions around audience and feedback in computer-mediated communications. People have an imagined audience in mind when posting to friends on social media [144, 155], but as Bernstein et al. showed, people's mental model of audience underestimates the number of people who actually see their posts and overestimate the rate at which friends give Likes and Comments [17]. Wang et al. found that posters and outsiders evaluate Facebook updates differently, particularly around topics of self-presentation and relationships [218].

Perceptions about feedback are also highly subjective – different people may have different interpretations of social interactions online. A recent study by Scissors et al. examined the perceptions around lightweight communications on Facebook and found that most people do not consider receiving "enough" likes as important and assigning importance to getting enough likes is positively correlated with high levels of self-monitoring and negatively correlated with self-esteem [196]. Previous work did not directly tie the diverse perspectives people have about activities on social media to expectations, which offer a more general view of social behavior as we describe next.

The notion of expectations is central to many theories about human behavior as it proposes a general framework for understanding behavior and attitude. The conceptual framework used in the current work was inspired by Expectancy Violation Theory (EVT) in communication [30, 33]. EVT was originally developed based on studies of proxemic behavior in face-to-face communication and was later extended to a variety of behaviors [30,32,34,35,216]. Burgoon defines expectancies as "enduring pattern of anticipated behavior", which derive from three classes of factors: communicator (e.g. demographics, personality, appearance), relationship (e.g. familiarity, similarity, status difference), and context (e.g. private/public environment, the message communicated) [31]. According to EVT, expectancies "serve as framing devices that define and shape interpersonal interactions . . . [and] significantly influence how social information is processed". The congruency between enacted behavior and expectations determines how behavior is perceived, the impression people have of each other, and the outcomes of the interaction. Previous work applied EVT to study norm evolution and violation on Facebook [19,85,159], but focused more on incidents of norm violations by individual friends, well-aligned with the original theory. However, as EVT focuses on single individuals' behav-

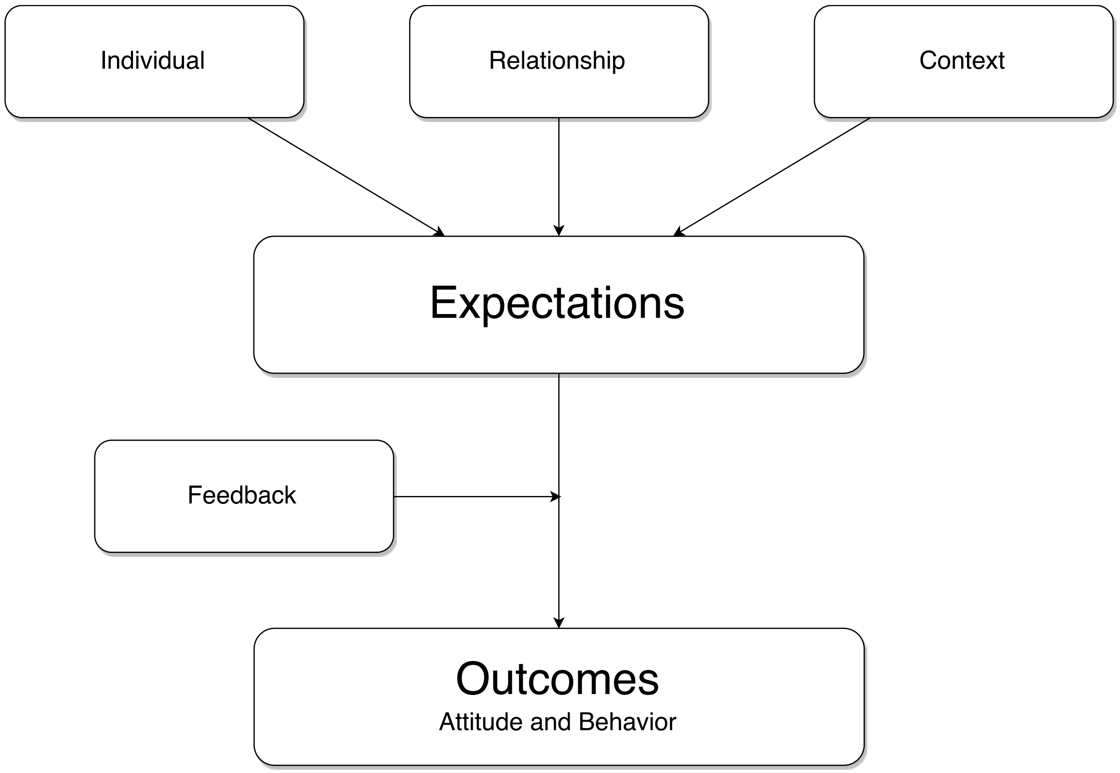


Figure 4.1: The conceptual framework used in this work, inspired by Expectancy Violation Theory by Burgoon [31].

ior, it does not directly apply to studying aggregate expectations as we do here. Instead, we use the overall framework of EVT as inspiration for our research model.

Figure 4.1 shows the conceptual framework used in this work, which is inspired by the EVT framework. At the top are three categories of factors that mirror the original EVT model [31], which we also expect to affect expectations of feedback: individual, relationship, and context properties. The characteristics of the individual can include demographics, personality traits, and more. Relationship properties may consist of tie strength between two individuals, differences in status, shared interests and other factors that affect interactions between people. Context includes the additional factors needed to describe a particular situation. In the current study, context primarily consists of properties of the posted content and past interactions on previous posts. The three categories of factors at the top of the figure jointly contribute to people's expectations, which are then compared against the actual feedback received. If, for example, the amount of feedback an individual received from friends exceeded their expectations, they may be more inclined to post in the future, and may feel more connected to their friends. In contrast, unsatisfying feedback experiences may be one of the mechanisms behind departure from online social platforms [226]. As proposed by EVT, and suggested here, the congruency between the observed feedback and expectations determines people's attitude and subsequent behavior.

Much like the early studies of proxemic behavior in face-to-face communication [30,34], we seek to understand the important factors behind feedback expectations on social network sites (top of Figure 4.1). Our first two research questions in this chapter focus on characteristics of the individual (left) and the posted status update, i.e. the context (right). It is important to distinguish expectations across people because individuals experience social media very differently: seeing different sets of stories and a variety of interactions with them (see [221] for an example). Similarly, distinguishing between expectations for different posts is important because not all posts are created equal and individuals are likely to care about them to various degrees [170,217]. Therefore, our first two research questions are:

RQ1*: What are the characteristics of individuals (e.g. age, gender, etc.) that affect feedback expectations?*

RQ2*: What are the properties of posts (e.g. length, topic, etc.) that indicate feedback expectations?*

Next, we study the third category of factors in Figure 4.1 that affects feedback expectations – relationship properties. Previous studies showed that people have different preferences for feedback from strong and weak ties on Facebook [38, 159] and from different social groups (e.g. close friends, family, co-workers, etc.) [196]. However, prior work did not directly examine expectations from specific friends (rather than abstract social groups), on a specific post, for different forms of feedback (e.g Likes and Comments), and at the time of posting (rather than retroactively). The complexity of social relations calls for a joint examination of all the above aspects of relationships in order to gain a more comprehensive understanding of relationship expectations. Therefore, our third research question is:

RQ3*: What are the relationship properties (e.g. based on relationship type, tie strength, age difference, etc.) that affect feedback expectations?*

Finally, motivated by work on social capital and well being [39,41,52,76–78,134], we investigate one potential outcome of fulfilling feedback expectations. Self-Determination Theory describes a basic human need for relatedness – to belong and feel connected to the people, group or culture sharing the individual's goals [194]. In offline settings, greater relatedness was shown to contribute to daily well being [192]. Since getting feedback is one of the key motivating factors for participation online [66, 118, 136, 174, 189] it is likely that the actual feedback received would affect people's satisfaction with their relationships. In fact, two recent studies point in that direction, showing that when people share emotional content on Facebook, friends respond with more emotional and supportive comments, which is associated with greater satisfaction with communication goals [13,36]. Previous findings, however, did not extend beyond emotional content and did not directly tie received feedback to relationship satisfaction. The conceptual framework in Figure 4.1 borrows from EVT to suggest that both expectations and observed behavior (whether they are met) will affect outcomes. Specifically, our question is:

RQ4: *How does the fulfillment of feedback expectations relate to feeling connected to one's Facebook friends?*

With these questions in mind, we performed a quantitative mixed-methods study of individuals' feedback expectations on Facebook, as we describe next.

4.3 Methods

In order to address our research questions about feedback expectations we use a mixed-methods approach based on survey and observational data analysis. Following previous work, we use survey mechanisms to ask people in our study about their subjective perceptions of social media activities [17, 196]. We then complement the survey responses with observational log data to gain better understanding of the contextual factors that explain expectations. Next, we describe our survey methodology in greater detail.

4.3.1 Surveys

We conducted two online surveys by recruiting participants on Facebook's web interface (Facebook.com)[1]. The first survey asked participants about their expectations for feedback for a specific post. The survey was offered to people immediately after they posted a status update, as a pop-up dialog. We refer to it here as the *Immediate survey.* The second survey was offered to people 23-27 hours after they posted a status update, as a banner on their Facebook page. This *Lagged survey*

[1]Due to the length and complexity of surveys we left the development of mobile versions to future work.

asked participants about fulfillment of expectations for that day-old post. The surveys were limited to English speakers in the U.S. with 20 friends or more that did not participate in any survey conducted by Facebook in the six months prior to ours. Participation in both surveys was voluntary and did not involve compensation in any form. The surveys ran for 20 days, between July 27, 2015 to August 16, 2015. The samples for both surveys were drawn from the same sampling frame, but were non-overlapping (people were invited to participate in either one of the surveys, but not both[2]). The response rate varied between the two surveys, with a lower response rate of 33% for the Immediate survey versus 78% for the Lagged survey. The difference was probably due to the Immediate survey's disruptive nature as a pop-up immediately after posting compared to the more organic banner of the lagged survey[3]. In total, 2788 people completed the immediate survey and 4032 completed the lagged survey.

Table 4.1 summarizes key demographic and usage characteristics of survey participants and examines how the survey population differs from other people on Facebook. We compare the participants in the surveys to a random sample of US, English speaking individuals who accessed Facebook's web interface at least once in the month prior to our surveys. As highlighted in Table 4.1, survey participants are more active Facebook users compared to the random sample, logging-in to Facebook about 4 additional days over the course of 28 days. Participants in

[2]we choose this design over repeated surveys of the same people for two reasons: 1) to eliminate the potential bias that answering questions in the Immediate survey influences answers to the Lagged survey, and 2) lower the burden on people of answering repeated surveys day after day.

[3]A pop-up was necessary in the Immediate survey to capture responses before any other action is taken on the site. We opted for a banner in the second survey because a pop-up in this case would have been out-of-context and hence much more disruptive to the user experience.

Measure (mean)	Immediate survey	Lagged survey	Random sample
# Log in days out of 28	**25.3***	**26.3***	21.9
# Log in days out of 7	**6.5***	**6.7***	5.5
Friend count	**512***	367	361
Gender (% Female)	57%	**62%***	57%
Age	40.4	**43.0***	41.1
N	2788	4032	4000
Response rate	33%	78%	-

*p<0.001 using 2-sample t-test comparing each survey separately to the random sample.

Table 4.1: Usage and demographic statistics of survey participants and a random sample of English speaking, US-based individual who logged-in to Facebook on the web at least once in the month prior to surveys.

the immediate survey have more friends on average, while more females and older individuals participated in the the lagged survey. Overall, we conclude that the participants in the surveys are slightly more active on Facebook than a random sample, but no other major differences are evident.

Measures

We now turn to describe the measures included in our surveys about feedback expectations and their fulfillment. Certain common elements are likely to affect both the expectations on a particular post, measured in the Immediate survey, and fulfillment of expectations, measured in the Lagged survey. Therefore, we include in both surveys the following 5-point Likert scale questions:

- **Connectedness**: how connected do you feel to your Facebook friends? (1=very disconnected, 5=very connected).

- **Importance**: how important is this post to you compared to your average post? (1=much less than usual, 5=much more than usual).
- **Personal**: how personal is this post? (1=not at all, 5=very personal).

Connectedness was used in order to address RQ4 about fulfillment of expectation, and was also used in the Immediate survey as a property of the individual who posted the content. The Importance and Personal questions were included in order to operationalize how different people care about different posts as found in prior work [12, 13, 196].

In addition, the Immediate survey asked participants about feedback expectations for the post:

- **Post-level expectations**: how many likes and comments do you expect to get on your latest post? (1=far fewer than usual, 5=far more than usual).
- **Friend-level expectations**: we presented a personalized ***Friends Grid***, a two column grid populated with a sample of up to 22 of the participant's friends (described below), asking participants to indicate whether they expect a Like and/or a Comment from each individual friend.

The Lagged survey, on the other hand, first referred participants to their post along with its feedback (as it appears in News Feed). Then, it asked participants about:

- **Fulfillment of feedback expectations**: how did the likes and comments received so far match your expectations? (1=far fewer than expected, 5=far more than expected).

Figure 4.2 shows the layout of the Friends Grid used in the Immediate survey.

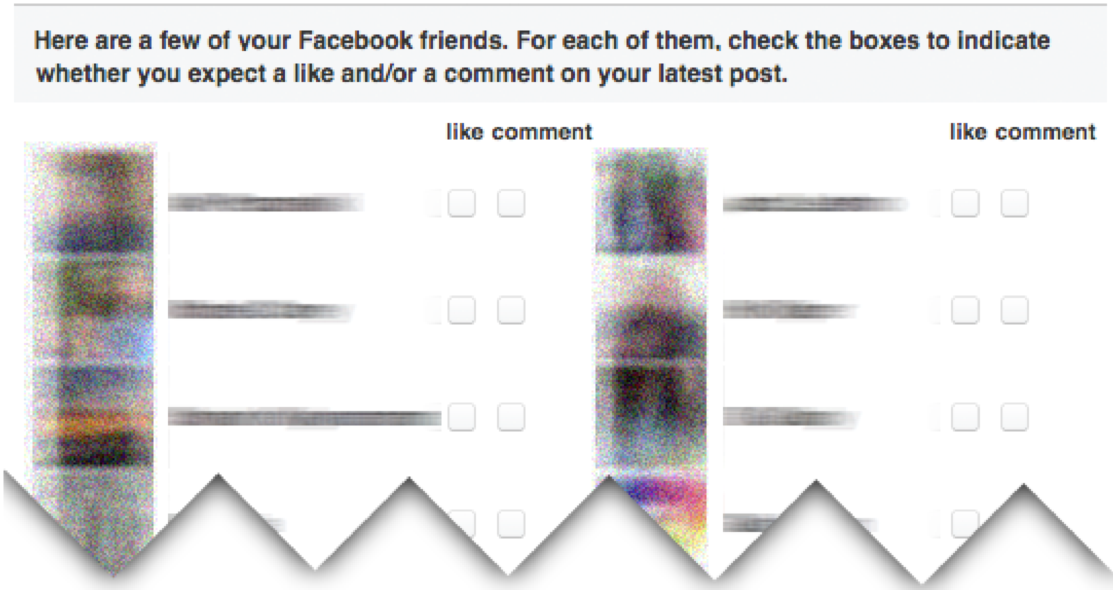


Figure 4.2: The *Friends Grid* question that was populated with a stratified sample of the participant's friends (in random order). For each friend, participants were asked to specify whether they expect a Like and/or a Comment on their latest post, shortly after posting it. Profile pictures and names are blurred in this figure, drawn for demonstration from the first author's account, in order to preserve individuals' privacy.

In order to get a more balanced sample of friends with and without feedback expectations we included in the Friends Grid a stratified sample of the participant's friends. We chose a stratified sample of friends over a random sample because a random sample is mostly dominated by weak ties with no feedback expectations. Our stratification randomly picked friends of the participant from Facebook lists the person may maintain (close friends, acquaintances), friends with overlapping profile information (same workplace, college, high-school, home town or current city), self-reported family ties (parent, child, sibling, spouse), and most recent interactions (last Like or Comment, given or received). In addition to sampling a random friend from each of these groups we included the friend from each group that the person communicated with most frequently (without introducing duplication).

Checkboxes in the Friends Grid may remain unchecked because the participant had no feedback expectations from that friend or because it is the default option. To address this bias, we use an assumption that people scan items visually from top to bottom, and from left to right[4]. This linear scanning assumption has been studied extensively in the analysis of search results and was shown to improve results relevance [209]. In our case, a friend is associated with "no expectation" only if the participant made a selection in a lower position in the grid or to the right. After processing the raw responses, our dataset consisted of 568 participants who labeled 5,256 of their friends with expectations for only a like (30.1%), only a comment (3.8%), like and a comment (11.4%), or no feedback (54.7%).

4.3.2 Log data

We complemented the survey responses with Facebook's server logs in the 12 weeks prior to the survey in order to better understand the context of reported expectations. All log data were observational – no experiment was performed and no individual's experience on the site was altered. The log data includes the posts that participants were asked about, profile information such as education or work history, and friendship information. We took significant steps to ensure people's privacy: all data were de-identified and analyzed in aggregate such that no individual's text could be viewed by researchers.

[4] The left to right assumption is reasonable since all of our participants are English speakers

4.4 Post-level Expectations

This section addresses our first two research questions about feedback expectations of different people (RQ1) on different posts (RQ2), and then focuses on estimating how accurately these expectations can be predicted in practice.

We address our first two research questions by fitting a logistic regression to the reported expectations on a post (from the Immediate survey) with covariates that describe the individual, her past feedback, and the posted content. We include both individual and post-level covariates in the regression model in order to understand how each group of factors varies while the others are held constant. For example, in addressing RQ1 we examine the characteristics of individuals that associate with higher than usual expectations while holding the properties of the post constant at their mean value. Our dependent variable in the regression is positive whenever a person reported expecting more than usual feedback on her post (4 or 5 on the Likert scale, where 3 was labeled "about the same as usual") and negative otherwise. We focus in particular on cases with higher than usual expectations since these are most likely to result in an unsatisfying experience when unmet.

Individual differences: Different people are likely to have different expectations. Facebook usage varies by age and gender, and thus may be associated with different feedback expectations [100]. In addition, the experience people have on Facebook is likely to affect their expectations. Previous work showed that people with different network sizes have different perceptions about their audience when posting and that perceptions vary by platform use [155, 196]. Therefore, we include in the regression information about age, gender, friend count, tenure on Facebook (years since creating the Facebook account), and the number of days in the past week that the participant logged in to Facebook (L7). Age, tenure and L7 were centered; friend count was log-transformed (base 2) to account for skew prior to standardization and the rest were centered and scaled using two standard deviations[5].

Past feedback: Past feedback on previous posts is also likely to affect expectations. Therefore, we compute the median and interquartile range (IQR)[6] for the following measures of feedback based on the individual's posts in the prior 12 weeks: number of likes per post, comments per post, likes per view, and comments per view. We also include the number of likes and comments on the most recent post of the individual since these might have greater impact on expectations. All variables were standardized as described earlier, except for the number of Likes/Comments per view which were log-transformed prior to standardization.

Content properties: Posts interest people to various degrees and therefore result in different expectations. Our dataset contained only few posts with photos and therefore we excluded those and focused only on textual posts. Our content properties include a variety of features: basic (word count, does the post contain a URL?), subjective assessments (how personal/important is this post?), topics, and emotional dimensions. The text was preprocessed and converted to lowercase, tokenized, and punctuation, stopwords and terms appearing in less than 5 posts were removed. All of the textual features were extracted using standard scripts over de-identified content such that no member of the research group examined

[5] Unless specified otherwise, all continuous covariates were centered and scaled by two standard deviation in order to put them on roughly the same scale as untransformed binary variables [91].

[6] a robust measure of dispersion, defined as the difference between the upper and lower quartiles.

any individual post.

We used Supervised Latent Dirichlet Allocation (sLDA) to model the topics that appear in posts [158]. The benefit of sLDA over "standard" unsupervised LDA is that topics are fit to better separate class labels. In our case, we used higher than usual feedback expectations as binary class labels. We experimented with different numbers of topics ranging from 10 to 60 (in increments of 10) and found no significant improvement in log-likelihood beyond using 20 topics. Using the trained sLDA model (using 10-fold cross validation) we get a single probability that represents the likelihood that a post is associated with higher than usual feedback expectations. We include the sLDA prediction in our final model.

Emotional dimensions were extracted using the 2007 version of Linguistic Inquiry and Word Count (LIWC) [183]. Most fine-grain LIWC categories (e.g. filler words) had no or very little support in our dataset and therefore we only included high-level categories such as function words, positive and negative emotions, social terms, achievement terms, and time orientation information (references to past, present or future). All of the LIWC features were included in the form of proportion of the total number of words in the post.

Without limiting the number of content properties in our regression, we run the risk of overfitting the data and finding spurious correlations as statistically significant. We address this concern in two different ways. First, we keep the number of covariates small relative to the number of survey responses (2,788) by limiting the number of topics and emotional dimensions we include. Second, we use Bayesian logistic regression with a non-informative Cauchy prior (0 median and 2.5 scale) to pull regression coefficients slightly towards zero apriori, but allow for large coefficients when the data does support it [92].

Higher than usual feedback exp. ∼	Coef.	SE
Intercept	-1.56^{***}	.15
Individual differences		
Age (years)	0.0098^{**}	.0037
Is male	0.13	.13
Tenure (years)	-0.14^{***}	.03
Log_2(Friend count)	0.93^{***}	.14
Connectedness	0.58^{***}	.13
L7	-0.041	.055
Posts per day	-0.002	.024
Past feedback:		
Likes(last post)	-0.26	.20
Comments(last post)	0.02	.16
IQR(Likes per post)	0.01	.21
$Median$(Likes per post)	0.19	.22
IQR(Comments per post)	0.48^{*}	.19
$Median$(Comments per post)	-0.38	.21
IQR(Likes per view)	0.00	.19
Log_2($Median$(Likes per view))	0.09	.15
IQR(Comments per view)	-0.19	.22
Log_2($Median$(Comments per view))	0.33^{*}	.16
Post:		
Importance	1.22^{***}	.15
Personal	0.36^{*}	.14
Has link	-0.22	.20
Log_2(Word count)	-0.08	.17
sLDA prediction	0.26^{*}	.13
LIWC:		
funct	0.18	.16
posemo	-0.27	.16
negemo	0.19	.12
social	0.04	.14
percept	0.22	.12
bio	-0.17	.13
achieve	0.32^{**}	.12
past	-0.19	.15
present	-0.14	.15
future	-0.10	.15
$P(Y\|X)$	20.47%	
Log Likelihood	-813.4	
Akaike Inf. Crit.	1886.8	

$N = 2,788$; *p<0.5; **p<0.01; ***p<0.001

Table 4.2: Coefficients of Bayesian logistic regression for having higher than usual feedback expectations on a post.

4.4.1 Findings

Table 4.2 shows the resulting coefficients of the Bayesian logistic regression with the binary dependent variable of having higher than usual feedback expectations on a post. All variance inflation factors (VIF) were less than two, indicating that multicollinearity is not an issue in our independent variables. The logistic regression assigns a probability of 20.47% for having high expectations to the average person (designed by $P(Y|X)$ in the table), closely matching the empirical proportion in the dataset with less than 0.01% in difference. Significant coefficients appear in all three categories of features, as we describe next.

In terms of individual differences (RQ1), four properties of the person posting the content are statistically significant: age, tenure on Facebook, number of friends and Connectedness. Each additional year of age is associated with a $\exp(0.0098) = 1.009 = +0.9\%$ increase in the odds of having higher than usual expectations. More significantly, doubling the number of friends on Facebook and feeling more connected to friends increases the odds by 38.6% and 29.6%, respectively. In contrast, each additional year of having a Facebook profile is associated with a -13.3% decrease in the odds of having higher than usual expectations.

Past feedback also contributes to higher feedback expectations, but only through Comments. There is an increase of 11.4% in the odds of high expectations for every additional comment in the individual's interquartile range, and 23.9% increase in odds when doubling the median rate of comments per view. None of the measures based on past Likes or feedback on the last post were significant. These findings suggest that greater variability in past comments (but not Likes) contributes to higher feedback expectations, and that people learn over time the rate at which their friends comment on their posts even without explicit knowledge about views.

Several aspects of the post's content affect feedback expectations (RQ2). First, increases in Importance and Personal scale translate into higher expectations (78% and 16% increase in odds, respectively)[7]. The large impact of Importance and Personal on expectations highlights the importance to better understand these subjective assessments, which is beyond the scope of the current work. Second, sLDA predictions based on broad topics found in the post increase the odds ratio modestly (+5.5%). The only emotional aspect that is significant is the occurrence of achievement terms. None of the other LIWC dimensions such as positive or negative emotions significantly affect expectations.

In summary, our findings show that individual differences, past feedback and posts' content are linked to higher than usual feedback expectations on a post. Age, number of friends and Connectedness are positively associated high expectations, while tenure on Facebook is negatively associated. Past Comments (and not Likes) affect expectations, and posts that are important, personal, and refer to achievements have higher expectations. Overall, we see that different people have different expectations for different posts, and that past behavior of friends (most in commenting) affects future expectations.

In order to utilize feedback expectations in practice systems need to accurately identify the expectations associated with new posts. Next, we evaluate how well feedback expectations for a post can be predicted.

[7] recall that all of surveyed content had the same privacy settings of sharing with all Facebook friends

Higher than usual exp. ∼	AUC	P@R5	P@R50	P@R95
baseline: last post percentile	49.0 53.0 57.0	22.6 29.5 36.5	19.6 21.3 22.9	19.3 20.6 22.0
individual differences Age + Gender + # Friends	60.2 62.6 65.0	33.8 55.9 78.0	25.0 29.5 34.0	18.8 21.4 24.0
+ past feedback	61.0 63.4 65.7	39.4 51.6 63.8	26.5 30.2 33.8	21.0 22.5 24.0
+ content:	67.4 **70.7** 73.9	57.9 **69.6** 81.3	38.7 **43.9** 49.1	21.9 **23.3** 24.7
+ self-reports: Connectedness + Post importance + Personal	75.4 **77.7** 80.0	68.6 **81.7** 94.7	46.6 **50.4** 54.2	23.0 **25.1** 27.2

Table 4.3: The predictive power of different feature sets obtained using either glmnet or gbm. P@R stands for precision at different recall levels of 5, 50 or 95 percent. Numbers above/below in each cell represent 95% confidence intervals.

4.4.2 Predicting post-level expectations

In this part we examine how well different subsets of the features from Table 4.2 predict the feedback expectations for a post by its author. We test the following three predictive models as implemented in R: Elastic-Net Regularized Generalized Linear Models (glmnet [88]), Generalized Boosted Regression Modeling (gbm [223]), and Support Vector Machine (SVM from e1071 package [161]). Regularization and boosting are common techniques to reduce overfitting, and SVM is able to capture non-linear relations between the features and the dependent variables. All three models have general and efficient learners that can easily scale to massive prediction problems.

Table 4.3 summarizes the results of 10-fold cross-validation of the best-performing model for each feature-set. Our baseline, which uses a personalized percentile of feedback received on the last post relative to the individual's posts in the prior 12 weeks, only performs marginally better than random. The predictive performance improves significantly over the baseline when including user information (62.6% AUC), adding past feedback (63.4% AUC) and finally content features, reaching 70% AUC. Using log data alone, the model identifies posts with higher than usual feedback expectations with a precision of about 70% when retrieving only 5% of posts with high expectations. In other words, at the level of 5% recall, the model will return one out of 20 posts with high expectations and would correctly identify the expectations for seven out of every 10 posts returned. Precision naturally deteriorates when increasing the recall to 50% or 95%. Last, including participants' answers to survey questions improves performance even further, showing that subjective information is important and not fully captured by other variables. In particular, feeling connected to friends, knowing the post's importance and how personal it is, are all important predictors of high feedback expectations.

Next, we address our third research question about the characteristics of relationships that affect expectations for feedback from one friend and not another.

4.5 Friend-level Expectations

In this section we use the responses from the *Friends Grid* in the Immediate survey to address RQ3: which properties of relationships affect expectations for feedback from different social ties? We examine how similarities and differences between two people as well as long-term and short-term communication patterns relate to expectations of feedback from that person. We fit two separate logistic regression models, one for Like expectations, the other for Comment expectations, on the same set of responses and relationship features.

Before we describe in greater detail the relationship properties used for addressing RQ3 we first specify the controls included in our models. We control for the order in which friends appeared in the Friends Grid, the characteristics of participants and the properties of posts. Despite the random order of friends in the Friends Grid, certain positions in the grid may receive more attention. Therefore, we include the position information in our model relative to the top (1-top to 11-bottom) and relative to the left (0-left, 1-right). In addition, as we saw in the previous section, individuals have different expectations for different posts. Since our focus here is on dyadic properties that affect expectations from a specific friend we control for non-dyadic features that were identified as significant in addressing RQ1 and RQ2. The complete list of control variables can be found in Table 4.4.

We describe below the three families of features we considered in our model for friend-level expectations: dyadic differences, topical similarity, and relationship properties.

Dyadic differences: the relative differences between participants and their friends may affect expectations. We include in our model demographic and activity information about the participant's friends in relative form, e.g. the difference between the friend's age and the participant's age. We also considered interactions between covariates, since different sub-populations may have different expectations. For example, age difference may be linked to higher expectations in general, but the gap can matter differently for younger and older adults.

Dyadic topics similarity: the perceived interest of a friend in a topic is likely to affect expectations for a response when the topic is discussed. Here, we develop a set of features aimed at capturing the overlap in topical interests of participants and their friends, and quantifying how a particular post fits into this overlap. For every individual, we aggregated the topics of all posts that they interacted with into a vector of high-level topical interests. Then, we computed interests similarity using cosine similarity between the participant (denoted as u) and their friend (denoted as f), and between the post (denoted as p) and the friend's interests. Due to the large amount of posts involved, we used a TagSpace model to extract the proportions of topics in posts [220], and reduced the topic space to 20 high-level most frequent topics (e.g. music, entertainment, education).

We also compute friend specificity to a topic in two different ways. First, we calculated the percent of the participant's friends that are highly interested in each of the post's topics[8] (denoted as $AUD(topic)$). As a second measure of specificity, we compute weighted friend share ($WFS(topic)$), based on the relative frequency the friend of interacts with a topic. We use tie strength (described below) as weights in WFS in order to give strong ties greater influence on the final measure of specificity than weaker ties. For both measures of topical specificity we took the maximal specificity score among the post's topics.

Dyadic relationship properties: this set of features focuses on social structure and communications between the participant and their friends. To represent social structure we create a set of indicator variables that designate whether the friend is a close family member (parent, child, sibling, spouse), member of a Face-

[8]We define "high interest" as exceeding a topic-specific threshold that is set to the upper quartile in the population. For example, a person who interacts with more than 8% of political posts would be considered as having high interest in politics.

book list that the participant maintains (close friends, acquaintances), or has overlapping profile information (same workplace, college, high-school, home town, or current city). In order to understand how expectations deviate for members of the same social structure, we also include a binary variable (designated by ^{best}) to indicate the strongest tie in that social circle.

Communication between people provides an additional dimension to social structure. We calculate a rough approximation of the tie strength between two people using the long-term frequency at which they communicate in any form (e.g. Liking, commenting, tagging, direct messages) as recorded in our logs over 12 weeks. Gilbert and Karahalios showed that frequency of communication is one of strongest predictors of tie strength [94]. In order to evaluate the effect of recent communications, we include indicator variables for the most recent friend who gave a Like or a Comment to the participant or received one from her.

4.5.1 Findings

Table 4.4 and Figure 4.3 describe the separate logistic regression models we fitted (on the same feature set) to Like and Comment expectations from individual friends. Both models converged and produced comparable estimates ($P(Y|X)$) to the empirical percentages of expectations reported in the survey (41.68% for Likes, 15.2% for Comments). Significant coefficients appear in every category of features as we discuss next.

Many of the findings for post-level expectations also hold for expectations from specific friends, but a more nuanced picture emerges. For example, participants with relatively fewer friends (in the lower quartile Q_1^{frnds} with $20-146$ friends) are

Friend expectation $\sim$	**Likes** *Coef.*	*SE*	**Comments** *Coef.*	*SE*
Controls:				
Intercept	-0.82^{***}	.22	-2.95^{***}	.27
Position top	-0.02	.03	0.081^{*}	.036
Position right	0.130	.078	0.125	.099
Age	0.0071^{**}	.0024	0.0072^{*}	.0030
Is male	0.27^{**}	.10	0.48^{***}	.13
Q_1^{frnds}	-0.18	.11	0.40^{**}	.13
Q_3^{frnds}	0.039	.093	0.03	.12
Q_4^{frnds}	0.23^{*}	.10	0.06	.13
Tenure	-0.099^{***}	.020	-0.141^{***}	.025
L7	-0.137^{**}	.044	-0.009	.061
Connectedness	-0.070	.066	-0.291^{***}	.080
Importance	0.293^{***}	.073	0.43^{***}	.10
Personal	0.222^{**}	.073	0.22^{*}	.10
Individual diff.:				
ΔAge	0.0041	.0044	0.0048	.0057
$Age \times (\Delta Age)$	0.07	.15	0.17	.21
Is diff gender	0.146	.086	0.01	.11
Is male-female rel.	-0.28^{*}	.14	-0.32	.18
Δ Friends	-0.18	.22	-0.25	.30
$Q_1^{frnds} \times (\Delta \text{ Friends})$	0.13	.30	-0.44	.40
$Q_3^{frnds} \times (\Delta \text{ Friends})$	0.24	.26	0.05	.36
$Q_4^{frnds} \times (\Delta \text{ Friends})$	0.22	.24	0.46	.32
$\Delta Tenure$	-0.023	.033	-0.133^{**}	.041
Tenure $\times (\Delta Tenure)$	-0.07	.14	0.33	.18
Δ L7	0.032	.067	0.114	.091
L7 $\times$ (ΔL7)	0.63^{*}	.28	-0.19	.37
Topical interests:				
$\cos(u, f)$	0.228^{***}	.068	0.25^{*}	.10
$\cos(p, f)$	-0.073	.069	0.08	.13
$max_{topic}AUD(topic)$	-0.071	.068	0.108	.079
$max_{topic}WFS(topic)$	0.152^{*}	.078	0.198^{**}	.071
Relationship:				
Relationship type	**(See Figure 4.3)**			
Tie strength (TS)	2.91^{***}	.19	1.79^{***}	.24
TS $\times$ Importance	-0.65^{*}	.30	-0.35	.40
TS $\times$ Personal	0.12	.30	-0.19	.39
$P(Y\|X)$	41.27%		13.3%	
Log Likelihood	-2840.2		-1910.5	
Akaike Inf. Crit.	5796.3		3937.1	

$N = 5,256$; *p<0.5; **p<0.01; ***p<0.001

Table 4.4: Coefficients of Bayesian logistic regressions for expecting a Like and a Comment on a post from a particular friend.

more likely to expect a Comment from a friend while those with more friends (in the upper quartile Q_4^{frnds} with $632-5000$ friends) are more likely to expect a Like. This shift in expectations is in line with previous findings showing a preference for composed communications from strong ties and lightweight communications from weak ties [38]. In addition, we see that male participants expect more feedback than females, but as we will see next this effect is mitigated by the friend's gender.

Only some dyadic differences are significant. The difference in activity levels between the listed friend and the participant ($\Delta L7$) is not significant on its own unless the participant herself is more active than average. Gender differences in general do not show a significant effect, but the interaction term ("Is male-female rel.") shows that males have lower expectations for Likes from their female friends. Gaps in the number of friends and age between participants and their friends are not significant.

Shared topical interests and specificity of close ties are associated with higher feedback expectations. The fact that our weighted measure (WFS) is statistically significant while the non-weighted measure ($AUD(topic)$) is not suggests that topical specificity matters more for close ties than weaker ties. These elevated levels of expectations can be explained by similarity/homophily to close ties or by the fact that one is more likely to know the topical interests of her close ties.

Other relationship properties are strongly correlated with feedback expectations. Doubling the frequency of communication with a friend increases the odds of a Like or a Comment expectation tremendously – by 5-17 times. The only exception to this general trend is for Likes on important posts, which can be seen in the negative coefficient of the interaction term of tie strength and the post's importance in the Likes model (TS $\times$ Importance). An important post is associated

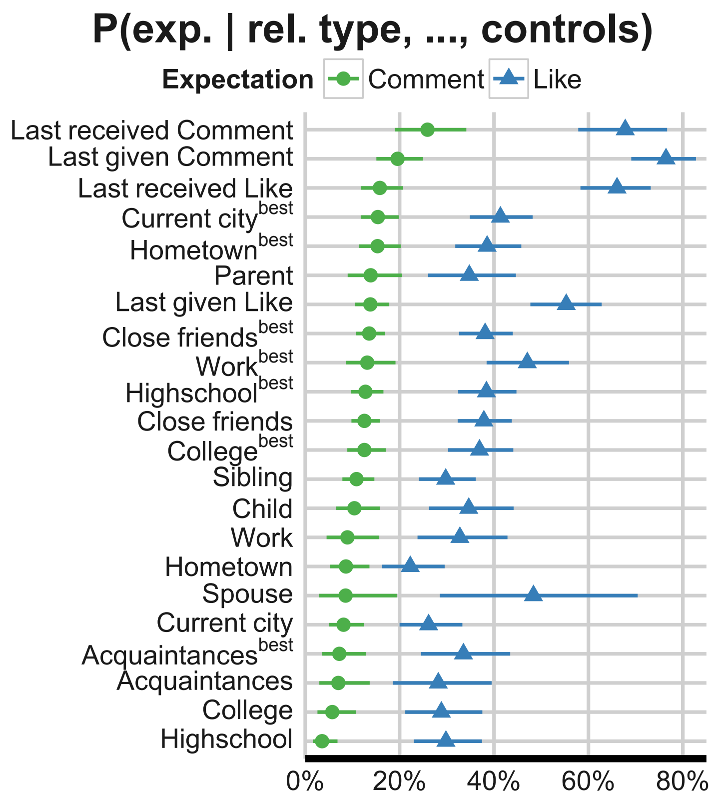


Figure 4.3: Probability of expecting a Like or a Comment from different social ties (95% CIs). X^{best} indicates the friend in social circle X that the participant most frequently communicates with.

with slightly lower expectations from close ties, which shows that content moderates expectation differently for Likes and Comments, and from different social ties.

Figure 4.3 shows the impact of social structure on feedback expectations. The figure shows how the probability of expectations (x-axis) changes for different types of relationships (y-axis), when all other variables from Table 4.4 are held constant at their mean value. Comment expectations are presented with green points and Like expectations with blue triangles. For example, participants are about 80% likely to expect a Like from the last person whom they gave a Comment to.

The results in Figure 4.3 highlight four important aspects of social structures:

recency, geographical proximity, family ties, and close friendships. The appearance of recently communicated friends at the top of the figure highlights the strong association between recency and feedback expectations, even after controlling for longer-term tie strength and a variety of other measures. In fact, the level of expectations for recently communicated friends is above and beyond close-family ties and most frequently communicated friends in every other social circle. We then see that the best friends from the current location (city or hometown) are associated with higher expectations than many other social ties. Parents and siblings are expected to Comment more than friends with similar Likes expectations, while spouses are expected to Like more than friends with similar Comments expectations[9]. It is possible that lower expectations for Comments from spouses reveal a preferences for face-to-face feedback over online communications in this case. Last, we note that the best friend indicator variables were found significant in every social group except Acquaintances, for whom people have lower expectations in general. The significance of best friend variables demonstrates that the closest-ties in most social circles have much higher feedback expectations, beyond what is expected based on the frequency of communication with them or any other factor in our models. Therefore, we conclude that recent communication, geographical proximity, family ties, and close friendships increase feedback expectations considerably, even after controlling for individual differences, posted content, and other relationship properties.

[9]The small number of spouses in our sample, 26, increases its confidence interval, but the gap between Likes and Comments' expectations is statistically significant with $p < 0.001$.

4.5.2 Predicting friend-level expectations

Models that identify friend-level feedback expectations can help social network sites evaluate how often people's expectations from their friends are met, identify possible reasons for unmet expectations, and ultimately guide the design of platforms to do better targeting and deliver more satisfying experiences to people. Therefore, our goal in this section is to assess how well a predictive model can identify friend-level expectations in practice.

We test different subsets of features from Table 4.4 in predicting the feedback expectations from a particular friend, without distinguishing between Likes and Comments for simplicity. Again, we use three different machine learning models for predicting expectations (glmnet, gbm and SVM), this time for feedback from a friend (Like or Comment) as reported in the Friends Grid.

Table 4.5 summarizes the results of 10-fold cross-validation of the best-performing model for each feature-set. The baseline model obtains 60% AUC, but simple demographic and activity information about people's Facebook activity surpasses the baseline with 64% AUC. Then, including tie strength improves the predictive ability considerably, from 64% to 75.8% AUC. The topical features alone achieve 66% AUC (not in the table), but when added to the rest of the features improve the predictive accuracy only marginally. A second considerable increase in performance is obtained using information about the relationship type – the AUC increases from 76% to closer to 81%, demonstrating that social structure carries valuable information about expectations that is not captured by other variables. Last, subjective information about a person's Connectedness and post-level importance and intimacy only add little to the predictive ability of the model. It

Friend exp. ∼	AUC	P@R5	P@R50	P@R95
Baseline: grid position	59.4 60.9 62.4	62.3 70.9 79.5	50.7 53.5 56.2	44.9 47.1 49.3
Demographics & **activity info.:**	62.9 64.0 65.0	59.6 66.2 72.8	55.1 57.1 59.1	47.5 49.0 50.6
+ Tie strength	74.0 75.8 77.6	88.4 92.3 96.3	71.4 73.7 76.0	49.6 51.3 53.1
+ Topical similarity **and specificity**	74.4 76.2 78.0	85.3 89.7 94.1	70.6 73.7 76.8	50.2 52.2 54.1
+ Social structure	79.7 **80.7** 81.8	92.3 **95.8** 99.4	77.7 **80.1** 82.5	51.3 **53.6** 56.0
+ self-reports: Connectedness + Post importance + Personal	79.8 **81.3** 82.7	91.4 **96.3** 100.0	78.0 **81.1** 84.2	51.9 **54.8** 57.7

Table 4.5: Predicting friend-level expectations for a Like of Comment using different feature sets obtained using gbm. P@R stands for precision at different recall levels of 5, 50 or 95 percent. Numbers above/below in each cell represent 95% CIs.

is possible that these self-reported variables encode little additional information about friend-level expectations that is not captured by other variables.

Our predictive model outperforms the baseline and identifies friend-level expectations with good accuracy (80%) using log data alone (no self-reported measures). When the model is set to retrieve only half of the cases with expectations (recall of 50%) it will correctly identify feedback expectations of held out individuals for four out of five of their friends (80%). At this level of performance, social media platforms can begin to estimate how well people's expectations are met and explore designs that improve people's satisfaction from their online interactions.

4.6 Fulfillment of Expectations & Connectedness

Finally, we analyze participants' responses from the lagged survey to understand the relationship between fulfillment of expectations and Connectedness (RQ4). First, we examine the relation between two of the measures from the lagged survey: *Fulfillment of feedback expectation* and *Connectedness*. Then, we establish that feedback expectations carry valuable information about connectedness that is not captured by the raw amount of feedback received or other measures.

Figure 4.4 shows a positive correlation between Connectedness and fulfillment of expectations, as reported in the lagged survey. The measure of Connectedness (y-axis) is presented using numerical values (1=very disconnected, 3=neither connected nor disconnected, and 5=very connected) and the measure of Fulfillment of expectations is presented on the x-axis. For example, people who reported receiving about the same amount of feedback as they expected averaged 3.86 on the 1-5 Connectedness scale.

The results in Figure 4.4 highlight an important relation between fulfillment of feedback expectations and connectedness. The more feedback received relative to expectations the more connected people feel to their Facebook friends: each unit increase on the fulfillment of feedback expectations scale translates into an addition 0.26 of connectedness. Overall, people move from 3-'neither connected nor disconnected' when their expectations are far from being met closer to 5-'very connected' when their expectations are exceeded considerably.

We ran an additional control survey to rule out the possibility that the response on the Connectedness question, which appeared first, may affected the response on

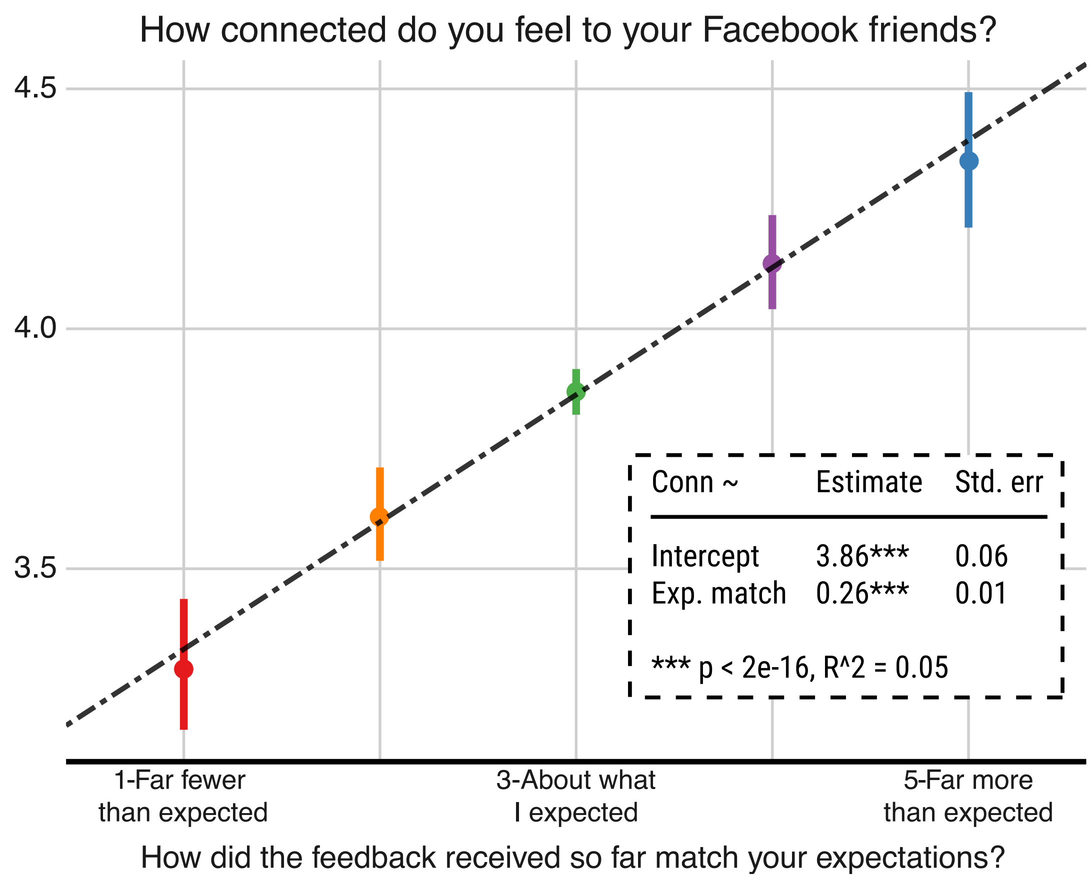


Figure 4.4: Responses from 24h lagged survey about fulfillment of expectations (x-axis) and feeling connected to Facebook friends (y-axis) with 95% CIs.

fulfillment of expectations question. In the control survey we omitted the Connectedness question, and found no significant difference in the distribution of responses to the fulfillment of expectations question ($\chi^2 = 5.56$, df $= 4$, p-value $= 0.23$). Therefore, we conclude that there is no evidence of a priming effect between the two measures.

Next, we establish that knowledge of feedback expectations provides valuable information that is not captured otherwise. In particular, we show that neither the feedback on the particular post nor the feedback on previous posts can better explain connectedness than knowledge about people's expectations. We do so by using both linear and non-linear regression models to predict the Connectedness people reported in the lagged survey. Our goal is not to perfectly explain Connectedness, which is a complex social and psychological construct, but rather show that expectations carry important additional information that is not captured by past (or present) feedback.

We calculate three different measures of feedback and compare them to the single measure of Fulfillment of expectations from the lagged survey. First, for simplicity, we combine the Likes and Comments into a single measure of Weighted Feedback: $WF = Likes + 5 \times Comments$ that gives more weight to comments since they are more rare[10]. WF is based on the Likes and Comments received on a single post in the 24 hours after posting, similar to timing of our lagged survey. We then compute the percentile of WF relative to the distribution of all posts in our log data (denoted as WF^{glbl}) or the individual's previous posts (denoted as WF^{indv}). Since our measure of fulfillment of expectations implicitly incorporates knowledge of the friends network, we include the friend count of the individual both as a separate predictor and as an interaction term with the WF measures.

Table 4.6 shows separately the predictive power of received feedback, fulfillment of expectations, and the combination of actual feedback and expectations. For example, the model that uses received WF on the post and friend count to predict Connectedness achieves 55% area under the curve (AUC) using linear regression and 55.5% using SVM. The measure of fulfillment of feedback expectations is the single strongest predictor for Connectedness (with 58.8% AUC), outperforming models using the actual feedback (WF), feedback relative to the global distribution (WF^{glbl}), and a personalized measure of feedback (WF^{indv}). These results demonstrate that knowledge about expectations is valuable for important concepts like Connectedness, and cannot be simply substituted by feedback information.

[10]other weighting of comments did not change the results significantly

Connectedness ∼	AUC *Linear Regression*	AUC *SVM*
Random answer	50.0%	50.0%
$log(1 + WF) \times \log(\#Friends)$	55.0%	55.5%
$WF^{glbl} \times \log(\#Friends)$	55.0%	55.8%
$WF^{invd} \times \log(\#Friends)$	54.4%	55.3%
Fulfillment of expectations	**58.8%**	**58.8%**
All of the above	61.0%	**62.7%**

Table 4.6: The predictive power of feedback and expectation using linear regression and SVM. Fulfillment of expectations is the single strongest predictor for Connectedness with 58.8% AUC, only second to the model that uses feedback and expectations jointly.

4.7 Discussion

In this chapter, we complemented survey responses with log data to better understand people's expectations for feedback. We have shown that having one's feedback expectations met is important to feeling connected to one's friends on Facebook. We also presented a nuanced view of how those expectations are shaped. We showed that whether a participant expected a post to receive more feedback than usual depends on the importance, intimacy, and the content of the post. This expectation also depended on the characteristics of the individuals themselves: their age and gender, as well as how long they had been active on Facebook. Furthermore, we demonstrated how the expectation for feedback from a particular friend varies depending on tie strength, recency of communication, geographical proximity, relationship type, and the relative strength of relationship within the social group it is embedded in.

In addressing our first two research questions we found supporting evidence that links some characteristics of individuals and posts to higher than usual feedback expectations. The subjective importance and intimacy of posts were the two strongest content properties associated with higher than usual feedback expectations. The result about intimacy of content is in line with self-disclosure literature [12, 213], which showed that people seek more social validation when broadcasting to many friends. Greater desire for social validation could also lead to increased feedback expectations. The fact that both the importance and intimacy were significant for post-level expectations highlights the need to better understand these important concepts. Future work could investigate what properties of the content (e.g. language, topics, etc.) makes a post subjectively more important. Similarly, future research can investigate the mechanisms behind some of the individual differences we found in feedback expectations on a post. For example, higher feedback expectations of older adults can be due to building stronger ties over time or because expectations are less calibrated.

As for expectations from specific friends (RQ3), we provided a nuanced view that integrates tie strength and social structure. Recent communicators are associated with the highest feedback expectations. Gilbert and Karahalios showed the importance of recent communications in the prediction of tie strength [94], but the considerable effect of recent communications on feedback expectation was not shown before. We also found a more nuanced preference for Comments over Likes that depends not only on tie strength but also on social structure. For example, spouses and best workplace friends have relatively low commenting expectations despite being strong ties, which perhaps reveals an expectation of face-to-face communications from these friends. These results add to the findings of Burke and Kraut about different communication preferences for strong and weak ties [38]. Finally, the best friends in each social group have higher expectations associated with them, even after controlling for their tie strength, social structure, and all other properties included in our models. Taken together, these results provide a glimpse into the complex and inter-connected nature of expectations from different social ties.

We also found that the fulfillment of expectations on a single post has a sizable effect on how connected people feel to their friends (RQ4), and that this effect is not fully captured by information about feedback alone. This finding confirms that for one important outcome, people's sense of connected to their friends, the general framework of expectations does indeed help in understanding people's attitude better than any other measure in our models. Moreover, this result highlights one potential mechanism, fulfillment of expectations, through which social media use contributes to one's subjective well being (as found in other studies [41, 77]). We emphasize, however, that the correlation we found between fulfillment of expectation and Connectedness does not warrant a causal relation (despite our additional control survey). The experimental evidence in the literature (e.g. [192]) leads us to believe that fulfillment of expectations does indeed affect Connectedness, but further work is needed to establish a causal link.

Feedback expectations are not only important, but also predictable, which paves the way for studying how expectations can be incorporated in social systems. Our models identified feedback expectations for posts with good accuracy. The prediction of expectations for a representative sample of friends (rather than the stratified sample used in our study) is likely to attain even higher accuracy due to the higher proportion of weak ties that are likely to have no feedback expectations associated with them. However, it remains an open question whether systems should adapt to posters' expectations, and if so, to what extent. How can the prediction of feedback expectations be used to improve the experience for both the person creating the content and their friends? Is content associated with higher-than-usual feedback expectations more likely to be interesting to a wider range of a person's friends? How does the fulfillment of expectations on one post affect expectations on subsequent posts? And if having one's feedback expectations met correlates with a greater feeling of connectedness, would understanding one's audience [17] be helpful? We leave these and other questions for future work.

4.7.1 Limitations and future work

Despite our attempts to capture feedback expectations as accurately as possible, our study design has several limitations. People may have different interpretations for expectations, which may vary from the bare minimum of feedback that would be "enough" to desires and hopes. Moreover, directly asking people about expectations may elicit expectations that did not exist before taking the survey and may not always be well-calibrated. In addition, the length and complexity of surveys led us to rely on single-item measures, which are generally less reliable than multiple-item measures. As noted earlier, our findings are based on observational analysis, which cannot infer causality.

There are several remaining gaps that would be fruitful avenues for future research. First, our work identified many important factors for feedback expectations that are worthy of further investigation. For example, how does age contribute to feedback expectations? why do males have higher expectations from specific friends? does the impact of recent communications on expectations stem from memory mechanisms or reciprocity? Second, our surveys focused on a single post at a certain point in time. Therefore, it is not yet clear how expectations evolve over time, and whether people's expectations and feedback received on prior posts influence expectations on any subsequent post. In addition, our surveys merely asked people about the existence of Like and Comment expectations, which do not capture the full range of possible responses[11]. For example, people may expect supportive comments from some friends, sarcastic replies from others, and more informative responses from acquaintances. Our work also identified potential value in developing language models that would better capture the subjective importance and intimacy of posts.

Our results may not accurately represent expectations in other populations, forms of media, and platforms. While the general framework of expectations was shown to be relatively universal [31], the concrete expectations people have may be specific to a certain culture, language, or community. Even within the population of people on Facebook, the stratified sample of friends we used may not accurately represent the entire friends network, especially in cases where people's profile information is incomplete. In addition, sharing other forms of media, such as photos, videos or mixed-media content, may be associated with other factors that affect expectations. Finally, different social networks bring about different social dynamics and with it varied feedback expectations. For example, if people share more public content on Twitter then expectations for may perhaps shift towards Retweets rather than Likes. While some variables in our models are Facebook-specific, we believe that categories we developed in this work and their relation to feedback expectations will generalize to other social network sites. That being said, the current work only looked at one social media site, at one point in time, and surveyed a tiny fraction of the people on Facebook. We look forward to other

[11]Our study was conducted before the introduction of Facebook Reactions, which are an extension of Likes.

works that would utilize our survey design and conceptual framework to contrast our findings with those from other social media platforms.

4.7.2 Conclusion

This chapter demonstrated that feedback expectations vary considerably across people, posts, and interpersonal relationships. Higher than usual feedback expectations on a post are linked to the characteristics of the post (importance, intimacy, and content), individual (age, gender, activity on Facebook), and past Comments. Neither the length of posts nor the sentiment of posts were found significantly correlated with feedback expectations. People have higher expectations from closer ties in general, but these are moderated by recency of interactions, geographical proximity, relationship type, and close friendships. Moreover, we found that the fulfillment of expectations is associated with feeling more strongly connected to friends, thus potentially contributing to individuals' well being. Last, our predictive models can estimate people's expectations with good accuracy, which paves the way for future research into the benefits and limitations of integrating expectations in social systems.

DISCUSSION AND CONCLUSION

This dissertation used a computational approach to study the attention people pay to online systems. We described the wealth of online information individuals are facing today and the importance for information systems to facilitate more efficient use of attention in the future. We studied people's attention to online systems both in lab and in real world settings, and using a variety of research methods. We outlined ways to measure attention of individuals online and investigated the factors that affect it in two key domains for online activity: news and social media. Both domains offer an abundance of information at the present time that is likely to increase in the future as more people get online and the world becomes more densely connected. Our measures and methods were designed to deal with the large heterogeneous population of Internet users, and provide concrete ways for information systems to better support the attention millions of people pay online every day.

We discuss next the key findings of each part of this dissertation, how systems can utilize these findings, and the important directions for future research to expand our work.

4.7.3 Attention in Online News

The first part of this dissertation focused on attention in reading online news articles. Chapter 1 investigated a simple, yet often overlooked, measure of scroll depth as a proxy for reading of news, and examined the factors that affect it. We found that articles' reading depth is predictable with good accuracy even before publication, and that the length of articles is the single strongest predictor of reading depth. Shorter sentences, positive and consistent sentiment, and longer quotes are associated with longer reading depths as well as authors' past success in engaging readers and deeper reading shortly after publication. The second chapter extended Chapter 1 by eliminating some of the assumptions about when people read. The chapter developed a measure for sensing whether a person is reading in-depth, skimming, or not reading a paragraph and outline a model for learning from non-invasive traces of user interaction with news articles online together with a smaller set of labeled examples. Our findings in Chapter 2 show that there is a natural order in the amount of time spent on paragraph where not reading requires little or no time, skimming require more time, but less than reading in-depth. Skimming is more strongly associated with scrolling and reading in-depth with more mouse movements while a paragraph is in full view.

Our work on online attention to news has implications for the design of systems for news, and potential for improving journalism online. The simple measure of reading depth described in Chapter 1 and the more granular measure developed in Chapter 2 can both be easily integrated into recommendation systems for news. The additional information about reading gleaned from user interactions can lead to improvements in the accuracy of recommendations, similar to other post-click measures [56, 102, 125, 145, 229]. In addition, accurately predicting how far people will scroll in news articles can be useful for dynamic loading of content and advertisement, reducing load times of articles, improving resource allocation, and giving more accurate accounting of exposure to ads. For journalists, the measures we developed can provide a better view into the ways readers interact with their content. Accurate predictions about reading at the sub-document level enables research on the language properties, at the level of paragraphs, that impact individuals' decisions to continue reading or leave the page. At a deeper level, this line of work could eventually transform some of the basic assumptions of journalists about how, where and why readers read different types of articles and writing styles. Of course, more granular information about reading could be abused to increase time spent on articles, but we hope that such tools would bring enough value to publishers that they would be used for good. We discuss this concern more fully below when considering the role of academic research in studying online attention more broadly.

Another promising direction for future work to investigate is the development of more intelligent and adaptive systems for online news. Accurate models for estimating when people read and whether they will continue reading paves the way for more dynamic interfaces and interventions that better support readers' goals. For example, dimming non-essential page elements while a person is deeply engaged in reading could be beneficial for readers in very much the same way that deferring notifications was found useful against interruptions [108]. Another approach could focus on skim reading, which is more common in digital settings [147], and design tools for more efficient skimming of content. For example, an enhanced progress bar that presents key inflection points in an article can enable readers to quickly navigate between sections when they lose interest in a section, but still get the gist of the article.

4.7.4 Attention in Online Social Settings

The second part of the dissertation examined two complimentary aspects of attention in online social settings. Chapter 3 developed within-subject methodology for studying how individuals' attention and behavior changes when posting on social media. We found a distinct patten of attention change that starts as early as six hours before posting and lasts from more than 12 hours. People start more active even before posting on Facebook, but further expand the number of times they visit Facebook, read more stories on it, and interact with friends after posting. Chapter 4 examined the expectations of people who post on social media to get the attention of their friends. Using two large online surveys together with log data analysis, we found that having feedback expectations met is important for feeling connected to friends. We presented a nuanced view of how expectations depend on the importance, intimacy, and other properties of posted content as well as the characteristics of individuals and their relationships to friends. Chapter 4 also demonstrated that people's expectations for specific posts and specific people are predictable with good accuracy. Taken together, the findings in both parts offer implications for the design of information systems as we describe next.

The findings in Chapters 3 and 4 about changes in attention in different social circumstances could be used to better adapt recommendation of social content depending on the context of user actions, the characteristics of individuals, their friends network, and the communicated content. In particular, our work on attention expectations in Chapter 4 provides accurate models for identifying posts and specific friends of the ego that are associated with higher than usual feedback expectations. This information about expectations of people sharing content on social media can readily integrated into social recommendations. Together with the findings about reciprocity in the exchange of feedback, recommendation systems can differently value feedback from specific friends of the ego and compare to the ego's expectations. Aside from recommendations, we found large-scale evidence that people attend more to Facebook and to content on it even before posting, which opens the possibility to design nudges for contribution and better adapt to interface for people's needs.

Our work on attention in social settings can be extended in several ways. Previous research as well as this dissertation work showed that individual differences, and differences in content and context affect how people pay attention [1,48,138]. Future work could expand this line of work by studying additional populations, broader range of online content and media types, and consider other platforms. Another avenue of research could investigate how attention modulates behavior. For example, experimental evidence showed that additional feedback on social media post leads to giving more feedback to others and continued posting [74]. However, the mechanisms through which attention govern these behavior changes are not yet well understood. Furthermore, it is import to carefully study the downstream impact of recommendations in social settings for *all* relevant parties. Our work in Chapter 4 on feedback expectations of people posting on Facebook provides one pragmatic approach to incorporating expectations in recommendations, but the balancing between producers and consumers, and the long-term impact of fulfilling people's expectations must be studied as well. Finally, there is a lot of potential for research on identifying cases where more social attention is not necessarily good and mitigate the negative consequences. For example, if systems could identify harmful content such as bullying or hate speech (automatically or not), the attention paid by affected parties could guide interventions that go beyond the mere removal of content after the effect. Future research could explore effective ways to reduce the harm for those exposed to the harmful or offensive content.

Another important avenue of future research that emerges from both parts of this dissertation is the need to connect online attention research with work on interruptions. In essence, interruptions are the stimuli that fragments people's attention to a main task [24]. Similar to prior work on attention, studies on interruptions in computer-mediated environments are mostly based on self-reported measure, and small-scale experimental or in situ studies [150–152, 222]. Studying interruptions in organic settings and at large scale can advance both the research on interruptions and on attention online. The challenges, of course, of studying interruptions in the "field" are many: interruptions many come in a variety of offline and online channels, they are task-specific, subjective, and sometimes are actually beneficial for individuals [112]. Nevertheless, we believe that future work could help systems become more aware of times where interruptions are more welcome, and respect times where focused attention is needed.

General Guidelines and Considerations

This dissertation opened by emphasizing the importance of attention research in an era of information overload, and continued to describe studies in particular domains that we deemed as most impactful. Here, we expand the discussion to consider research on online attention more broadly. We underline that more work is needed to expand the development of generalizable constructs for individuals' attention online, and argue for the use of mixed methods in semi-supervised learning models. We make the case for using computational methods to study the impact systems have on human attention and vice versa, and emphasize the role of academic research in putting users' needs first.

The lens of attention enabled us to examine the interactions of individuals with online systems in two very different domains. By abstracting the targets of attention and the mechanisms for attending to them we were able to find interesting patterns in attention paid online. The context of the domain in each study together with a detailed description of variables being measured allowed us to convey the meaning of the particular aspect of attention being investigated. This approach aligns closely with the approach championed by cognitive psychologist Harold Pashler, who famously said "No one knows what attention is" and therefore each study needs to define it precisely[12] [181]. Nonetheless, a lack of more general definitions for attention online does not imply that such definitions should not exist. Developing user-centered, generalizable constructs for thinking about attention online can provide a framework that applies to a wide range of systems. The measures and methods developed in this dissertation can contribute to the emergence of more general constructs for discussing attention in online settings.

In terms of methodology, we would like to see more research using mixed methods, particularly in semi-supervised learning setting, in the study of online attention. The use of surveys and experiments together with observational data analysis can capitalize on the strengths of these methods while compensating for their weaknesses in studying individuals' attention online. Lab studies and surveys are relatively limited in their ability to reach large and diverse populations, but can provide in-depth view of the studied phenomena and more interpretable results. Large scale experiments are possible online, but as previously mentioned in the introduction, have high costs and are facing growing public scrutiny over ethical concerns of informed consent [103,130,208]. Quantitative methods using large scale observational data can capture more of variety of real world online actions, but provide relatively little interpretability with respect to the measures used and the causes behind them. Chapters 1 and 3 pushed the boundary in terms of using quantitative methods for learning about attention from large scale observational datasets alone, but clearly provide less interpretable results compared to Chap-

[12]The other extreme is represented by the reductionism of William James's "Everyone knows what attention is", which requires little definition and plays down the complexity of attention as a perceptual phenomena.

ters 2 and 4, which coupled observational data with survey and experimental data. Because attention is complex and not directly observable cognitive process it is important to ground the findings based on well understood measures. The semi-supervised approach laid out in Chapter 2 is particularly apt for using a small amount of supervision to learn potentially more representative patterns of attention to content in the "wild". For these reasons we believe that semi-supervised approaches based on data obtained through multiple research methods represent the best of both worlds and should be more common in the study of online attention.

This dissertation took an impartial view with respect to the direction in which people's attention should change. In other words, the work took place without a predisposition of whether people should pay more or less attention in each particular context. For example, our work on reading in online did not hold the position that people should read more or less of news articles, and instead focused on issues of measurement and identification of the factors affecting attention. We believe that this stance is reasonable in a relatively new research domain, where there are many confounding factors to consider. However, this approach does not nullify, and in fact supports, a more critical view of the interdependency between online systems and human attention. For example, a longitudinal study could use our measure of reading depth from Chapter 1 to find trends in people's reading over time and investigate the technological factors behind these trends. Similarly, it remains an open question how online systems *should* account for individuals' attention. Herding effects similar to the one described by Muchnik et al. could lead recommendation systems to unevenly increase exposure to content items of similar quality [167]. Therefore, there is room for future work to investigate the systems-induced impacts on people's attention and the effect human attention has on systems.

More work on attention and dissemination of information (news content in particular), is needed now more than ever before. The digital revolution and the rise of social media platforms have disrupted the business model of journalism, weakened the institutions providing news coverage, and led to a proliferation of unreliable sources. These conditions provide fertile ground for the spread of misinformation and disinformation[13], which undermine one of the foundations of open and democratic society. While the extent to which platforms should combat misinformation is being debated, it is clear that information systems should not give an advantage to false and dubious news content [129, 166]. Since the spread of misinformation requires individuals' attention in somewhat similar ways to Clickbaits, it is possible that better attention measures could help detect some forms of misinformation and allow systems to slow down its spread by devaluing it in recommendations. In addition, better measures of attention could strengthen news organizations and help them identify more avid readers and learn more accurate descriptions of readers' interests. More broadly, individual-level data about exposure to content on and off social media could prompt a comprehensive analysis of the relationship between online attention and dissemination of information.

Going forward, more *academic* research on attention online is needed in a field that is dominated by companies, which own most exposure data and have strong business incentives to attract attention to their platforms. Private and public Internet companies now hold the vast majority of data about individuals' exposure to information. Internet services often provide valuable services to users at no cost, with the service being ad-supported. The reliance on advertisement creates a strong financial incentive for platforms to keep their users engaged for longer periods of

[13] `https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php`

time and display more ads, which does not necessarily align with users' interests. Limiting the amount of interruptions (e.g. through notifications) can enhance individuals' task-completion [150], but directly compete with companies' objectives to sell more ads. More implicitly, if recommendations systems are optimized for time spent, certain types of content will prevail, as argued earlier in this dissertation. For example, a story told as a series of cliffhangers may keep people engaged for longer periods of time, but takes away some of the agency of individuals to choose early and informatively what content to consume. Tristan Harris, the founder of Time Well Spent[14] and previously Design Ethicist at Google, wrote in an essay last May "The ultimate freedom is a free mind, and we need technology to be on our team to help us live, feel, think and act freely"[15]. The public has little visibility into the aggregate effect of systems on our attention, but researchers and computer scientists in particular are well equipped to study this intricate relationship. Therefore, we believe that one of the most important roles for academic research on online attention going forward is to enable platforms to consider short- and long-term needs of users more fully.

Conclusion

In summary, this dissertation used a computational perspective to the study of individuals' attention online. We focused on two domains where the abundance of information is particularly acute, and where systems can have a large impact in improving the use of attention for millions of people. We developed new measures for quantifying attention and new methodology to determine the factors affecting

[14] `http://www.timewellspent.io`

[15] `https://goo.gl/AU1QAF`

attention online. We provided concrete ways for information systems to considering human attention more fully and take into account more properties of content and context of individuals actions. In an interview with Nicholas Carr, the developmental psychologist Maryanne Wolf said about reading: “We are not only what we read, we are *how* we read” [45]. The same applies to attention more broadly – we are what we pay attention to, and how we pay attention. In that sense, this dissertation advanced the ability of information systems to enable us to be our better selves.

BIBLIOGRAPHY

[1] Rakefet Ackerman and Tirza Lauterman. Taking reading comprehension exams on screen or on paper? a metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior*, 28(5):1816–1828, 2012.

[2] Marilyn Jager Adams. *Beginning to read: Thinking and learning about print.* MIT press, 1994.

[3] Eytan Adar, Jaime Teevan, and Susan T Dumais. Large scale analysis of web revisitation patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2008.

[4] Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. Finding high-quality content in social media. In *Proc WSDM 2008*, 2008.

[5] Michael Anderson and Jeremy Magruder. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database*. *The Economic Journal*, 122(563):957–989, 2012.

[6] David Arbour, Dan Garant, and David Jensen. Inferring network effects from observational data. In *Proc. of KDD 2016*, KDD '16, pages 715–724. ACM, 2016.

[7] Kellie J. Archer. *glmnetcr: Fit a penalized constrained continuation ratio model for predicting an ordinal response*, 2014. R package version 1.0.2.

[8] Lars Backstrom. Serving a billion personalized news feeds. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pages 469–469, New York, NY, USA, 2016. ACM.

[9] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[10] Albert Bandura. Social learning theory. *New York: General Learning Press*, 1977.

[11] Joseph Bayer, Nicole Ellison, Sarita Schoenebeck, Erin Brady, and Emily B Falk. Facebook in context(s): Measuring emotional responses across time and space. *New Media & Society*, 0(0):1461444816681522, 0.

[12] Natalya N. Bazarova and Yoon H. Choi. Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication*, 64(4):635–657, 2014.

[13] Natalya N. Bazarova, Yoon H. Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. Social sharing of emotions on facebook: Channel differences, satisfaction, and replies. In *Proc. CSCW 2015*, 2015.

[14] Carl W. Beckman and Paul F. Secord. The effect of perceived liking on interpersonal attraction. *Human Relations*, 1959.

[15] Jonah Berger. Arousal increases social transmission of information. *Psychological Science*, 22(7):891–893, 2011.

[16] Jonah Berger and Katherine L. Milkman. What makes online content viral? *Journal of marketing research*, 49(2):192–205, 2012.

[17] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. Quantifying the invisible audience in social networks. In *Proc. CHI 2013*, 2013.

[18] Michael S. Bernstein, Adam Marcus, David R. Karger, and Robert C. Miller. Enhancing directed content sharing on the web. In *Proc. of CHI 2010*, pages 971–980, 2010.

[19] Jennifer L. Bevan, Pei-Chern Ang, and James B. Fearns. Being unfriended on facebook: An application of expectancy violation theory. *Computers in Human Behavior*, 33:171–178, 2014.

[20] Jiang Bian, Yandong Liu, Eugene Agichtein, and Hongyuan Zha. A few bad votes too many?: towards robust ranking in social media. In *Proc. AIRWEB 2008*, 2008.

[21] Ralf Biedert, Jörn Hees, Andreas Dengel, and Georg Buscher. A robust realtime reading-skimming classifier. In *Proc. ETRA 2012*, 2012.

[22] Sven Birkerts. *The Gutenberg elegies: The fate of reading in an electronic age*. Macmillan, 2006.

[23] Jonas Nygaard Blom and Kenneth Reinecke Hansen. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100, 2015.

[24] Deborah A. Boehm-Davis and Roger Remington. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention*, 41(5):1124–1129, 2009.

[25] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

[26] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proc. of COMPSTAT 2010*. Springer, 2010.

[27] danah m. boyd and Nicole B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[28] Donald E. Broadbent. The role of auditory localization in attention and memory span. *Journal of experimental psychology*, 47(3):191, 1954.

[29] Donald E. Broadbent. *Perception and communication*. Pergamon Press, 1958.

[30] Judee K. Burgoon. A communication model of personal space violations: Explication and an initial test. *Human Communication Research*, 4(2):129–142, 1978.

[31] Judee K. Burgoon. Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*, 12(1-2):30–48, 1993.

[32] Judee K. Burgoon and Stephen B. Jones. Toward a theory of personal space expectations and their violations. *Human Communication Research*, 2(2):131–146, 1976.

[33] Judee K Burgoon and Beth A. Le Poire. Effects of communication expectancies, actual communication, and expectancy disconfirmation on evaluations of communicators and their communication behavior. *Human communication research*, 20(1):67–96, 1993.

[34] Judee K. Burgoon, Valerie Manusov, Paul Mineo, and Jerold L. Hale. Effects

of gaze on hiring, credibility, attraction and relational message interpretation. *Journal of Nonverbal Behavior*, 9(3):133–146, 1985.

[35] Judee K. Burgoon, Joseph B. Walther, and E. James Baesler. Interpretations, evaluations, and consequences of interpersonal touch. *Human Communication Research*, 19(2):237–263, 1992.

[36] Moira Burke and Mike Develin. Once more with feeling: Supportive responses to social sharing on facebook. In *Proc. CSCW 2016*, 2016.

[37] Moira Burke and Robert Kraut. Using facebook after losing a job: Differential benefits of strong and weak ties. In *Proc. CSCW 2013*, 2013.

[38] Moira Burke and Robert E. Kraut. Growing closer on facebook: Changes in tie strength through social network site use. In *Proc. CHI 2014*, 2014.

[39] Moira Burke and Robert E. Kraut. The relationship between facebook use and well-being depends on communication type and tie strength. *Journal of Computer-Mediated Communication*, 21(4):265–281, 2016.

[40] Moira Burke, Cameron Marlow, and Thomas Lento. Feed me: motivating newcomer contribution in social network sites. In *Proc. CHI 2009*, 2009.

[41] Moira Burke, Cameron Marlow, and Thomas Lento. Social network activity and social well-being. In *Proc. CHI 2010*, 2010.

[42] Georg Buscher, Edward Cutrell, and Meredith R. Morris. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proc. SIGCHI 2009*, 2009.

[43] Jaclyn Cameron and Nick Geidner. Something old, something new, something borrowed from something blue: Experiments on dual viewing tv and twitter. *Journal of Broadcasting & Electronic Media*, 58(3):400–419, 2014.

[44] Christopher S Campbell and Paul P Maglio. A robust algorithm for reading detection. In *Proc. PUI 2001*, 2001.

[45] Nicholas Carr. Is google making us stupid? `https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/`, 2008.

[46] Nicholas Carr. *The shallows: What the Internet is doing to our brains*. W.W. Norton & Company, 2011.

[47] Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2001.

[48] Yao Chen, Susana Martinez-Conde, Stephen L. Macknik, Yulia Bereshpolova, Harvey A. Swadlow, and Jose-Manuel Alonso. Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature neuroscience*, 11(8):974–982, 2008.

[49] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. How community feedback shapes user behavior. In *Proc. ICWSM 2014*, 2014.

[50] E. Colin Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979, 1953.

[51] Coye Cheshire and Judd Antin. The social psychological effects of feedback on the production of internet information pools. *Journal of Computer-Mediated Communication*, 13(3):705–727, 2008.

[52] Mina Choi and Catalina L. Toma. Social sharing through interpersonal media: Patterns and effects on emotional well-being. *Computers in Human Behavior*, 36:530–541, 2014.

[53] Chun Wei Choo, Brian Detlor, and Dan Turnbull. Information seeking on the web: An integrated model of browsing and searching. *First Monday*, 5(2), 2000.

[54] Marvin M. Chun, Julie D. Golomb, and Nicholas B. Turk-Browne. A taxonomy of external and internal attention. *Annual review of psychology*, 62:73–101, 2011.

[55] Marie M Clay. Reading: The patterning of complex behaviour. 1977.

[56] Mark Claypool, Phong Le, Makoto Wased, and David Brown. Implicit interest indicators. In *Proc. of IUI 2001*, pages 33–40, 2001.

[57] Andy Cockburn and Bruce McKenzie. What do web users do? an empirical

analysis of web use. *International Journal of Human-Computer Studies*, 54(6):903–922, 2001.

[58] Laurent Cohen, Stanislas Dehaene, Lionel Naccache, Stéphane Lehéricy, Ghislaine Dehaene-Lambertz, Marie-Anne Hénaff, and François Michel. The visual word form area. *Brain*, 123(2):291–307, 2000.

[59] Julie Coiro. Exploring literacy on the internet: Reading comprehension on the internet: Expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher*, 2003.

[60] Julie Coiro and Elizabeth Dobler. Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the internet. *Reading Research Quarterly*, 2007.

[61] Microsoft Canada Consumer insights. Attention spans, 2015. `https://advertising.microsoft.com/en/WWDocs/User/display/cl/researchreport/31966/en/microsoft-attention-spans-research-report.pdf`.

[62] Dan Cosley, Shyong K. Lam, Istvan Albert, Joseph A. Konstan, and John Riedl. Is seeing believing?: How recommender system interfaces affect users' opinions. In *Proc. of CHI 2003*, pages 585–592, 2003.

[63] Francisco Cribari-Neto and Achim Zeileis. Beta regression in R. *Journal of Statistical Software*, 34(2):1–24, 2010.

[64] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web*. ACM, 2007.

[65] Thomas Davenport and John Beck. The attention economy: Understanding the new economy of business, 2001.

[66] Utpal M. Dholakia, Richard P. Bagozzi, and Lisa K. Pearo. A social influence model of consumer participation in network-and small-group-based virtual communities. *International journal of research in marketing*, 21(3):241–263, 2004.

[67] Nicholas Diakopoulos and Arkaitz Zubiaga. Newsworthiness and network

gatekeeping on twitter: The role of social deviance. In *International AAAI Conference on Weblogs and Social Media*, 2014.

[68] Thomas J. DiCiccio and Bradley Efron. Bootstrap confidence intervals. *Statistical science*, pages 189–212, 1996.

[69] P. Alex Dow, Lada A. Adamic, and Adrien Friggeri. The anatomy of large facebook cascades. In *ICWSM*, 2013.

[70] Gerald G. Duffy. *Explaining Reading, Second Edition: A Resource for Teaching Concepts, Skills, and Strategies*. Solving Problems in the Teaching of Literacy. Guilford Publications, 2009.

[71] Geoffrey B. Duggan and Stephen J. Payne. Text skimming: the process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, 15(3):228, 2009.

[72] Geoffrey B. Duggan and Stephen J. Payne. Skim reading by satisficing: evidence from eye tracking. In *Proc. of CHI'2011*. ACM, 2011.

[73] James A. Easterbrook. The effect of emotion on cue utilization and the organization of behavior. *Psychological review*, 66(3):183–201, 1959.

[74] Dean Eckles, René F. Kizilcec, and Eytan Bakshy. Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences*, 113(27):7316–7322, 2016.

[75] Khalid El-Arini and Joyce Tang. News feed fyi: Click-baiting. `http://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting`, 2014.

[76] Nicole B. Ellison, Rebecca Gray, Cliff Lampe, and Andrew T. Fiore. Social capital and resource requests on facebook. *New Media & Society*, 16(7):1104–1121, 2014.

[77] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. The benefits of facebook âĂIJfriends:âĂİ social capital and college studentsâĂŹ use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.

[78] Nicole B. Ellison, Jessica Vitak, Rebecca Gray, and Cliff Lampe. Cultivating social resources on social network sites: Facebook relationship maintenance

behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication*, 19(4):855–870, 2014.

[79] Yariv Ephraim and William JJ Roberts. Revisiting autoregressive hidden markov modeling of speech signals. *IEEE Signal Processing Letters*, 12(2):166–169, 2005.

[80] Charles W. Eriksen and James F. Collins. Temporal course of selective attention. *Journal of experimental psychology*, 80(2p1):254, 1969.

[81] Michael Esterman, Yu-Chin Chiu, Benjamin J. Tamber-Rosenau, and Steven Yantis. Decoding cognitive control in human parietal cortex. *Proceedings of the National Academy of Sciences*, 106(42):17974–17979, 2009.

[82] Jonathan St. B.T. Evans. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59:255–278, 2008.

[83] Hans J. Eysenck. *The biological basis of personality*, volume 689. 1967.

[84] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *Journal of machine learning research*, 9(Aug):1871–1874, 2008.

[85] Eric M. Fife, C. Leigh Nelson, and Kristine Bayles. When you stalk me, please don't tell me about it: Facebook and expectancy violation theory. *Kentucky Journal of Communication*, 28(1), 2009.

[86] Alison Flood. Ebooks can tell which novels you didn't finish. `http://www.theguardian.com/books/2014/dec/10/kobo-survey-books-readers-finish-donna-tartt`, 2014.

[87] Valerie S. Folkes and David O. Sears. Does everybody like a liker? *Journal of Experimental Social Psychology*, 13(6):505–519, 1977.

[88] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

[89] Stephen J. Frost, Nicole Landi, W. Einar Mencl, Rebecca Sandak, Robert K. Fulbright, Eleanor T. Tejada, Leslie Jacobsen, Elena L. Grigorenko, R. Todd Constable, and Kenneth R. Pugh. Phonological awareness predicts activation patterns for print and speech. *Annals of dyslexia*, 59(1):78–97, 2009.

[90] Ruth Garner. *Metacognition and reading comprehension.* Ablex Publishing, 1987.

[91] Andrew Gelman. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine*, 27(15):2865–2873, 2008.

[92] Andrew Gelman, Aleks Jakulin, Maria G. Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, pages 1360–1383, 2008.

[93] Rhonda Gibson and Dolf Zillmann. The impact of quotation in news reports on issue perception. *Journalism & Mass Communication Quarterly*, 70(4):793–800, 1993.

[94] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proc. CHI 2009*, 2009.

[95] Sharad Goel, Ashton Anderson, Jake Hofman, and Duncan J. Watts. The structural virality of online diffusion. *Management Science*, 62(1):180–196, 2016.

[96] Sharad Goel, Jake M Hofman, and M Irmak Sirer. Who does what on the web: A large-scale study of browsing behavior. In *International AAAI Conference on Weblogs and Social Media*, 2012.

[97] Amy L. Gonzales and Jeffrey T. Hancock. Mirror, mirror on my facebook wall: Effects of exposure to facebook on self-esteem. *Cyberpsychology, Behavior, and Social Networking*, 14(1-2):79–83, 2011.

[98] Mark S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[99] Rebecca Gray, Nicole B. Ellison, Jessica Vitak, and Cliff Lampe. Who wants to know?: question-asking and answering practices among facebook users. In *Proc. CSCW 2013*, 2013.

[100] Shannon Greenwood, Andrew Perrin, and Duggan Maeve. Social media update 2016, 2016.

[101] Frangoise Grellet. *Developing Reading Skills: A Practical Guide to Reading Comprehension Exercises.* Cambridge Language Teaching Library. Cambridge University Press, 1981.

[102] Qi Guo and Eugene Agichtein. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. In *Proc. WWW 2012*, 2012.

[103] Jeff T. Hancock. The facebook study: A personal account of data science, ethics and change. In *Proc. of CSCW 2015*, pages 1–1. ACM, 2015.

[104] Chip Heath, Chris Bell, and Emily Sternberg. Emotional selection in memes: the case of urban legends. *Journal of personality and social psychology*, 2001.

[105] Matthew Hoffman, Francis R Bach, and David M Blei. Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems*, 2010.

[106] Kenneth Holmqvist, Jana Holsanova, Maria Barthelson, and Daniel Lundqvist. Reading or scanning? a study of newspaper and net paper reading. *Mind*, 2003.

[107] Jana Holsanova, Henrik Rahm, and Kenneth Holmqvist. Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. *Visual Communication*, 5(1):65–93, 2006.

[108] Eric Horvitz, Johnson Apacible, and Muru Subramani. Balancing awareness and interruption: Investigation of notification deferral policies. In *Proc. User Modeling 2005*. Springer, 2005.

[109] J. Brian Houston, Joshua Hawthorne, Matthew L. Spialek, Molly Greenwood, and Mitchell S. McKinney. Tweeting during presidential debates: Effect on candidate evaluations and debate attitudes. *Argumentation and Advocacy*, 49(4):301, 2013.

[110] Jeff Huang, Ryen W. White, Georg Buscher, and Kuansan Wang. Improving searcher models using mouse cursor activity. In *Proc. SIGIR 2012*, 2012.

[111] Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. *Available at SSRN 1313405*, 2008.

[112] James M. Hudson, Jim Christensen, Wendy A. Kellogg, and Thomas Erickson. "i'd be overwhelmed, but it's just one more thing to do:" availability and interruption in research management. In *Proc. CHI 2002*. ACM, 2002.

[113] David J. Hughes, Moss Rowe, Mark Batey, and Andrew Lee. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2):561–569, 2012.

[114] C. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *International AAAI Conference on Weblogs and Social Media*, 2014.

[115] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proc. KDD 2002*. ACM, 2002.

[116] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems*, 25(2), April 2007.

[117] Peter H. Johnston. *Reading comprehension assessment: A cognitive basis.* ERIC, 1983.

[118] Adam N. Joinson. Looking at, looking up or keeping up with people?: motives and use of facebook. In *Proc. CHI 2008*, 2008.

[119] Pearl Judea. Causality: models, reasoning, and inference, 2000.

[120] James F Juola, Nicklas J Ward, and Timothy McNamara. Visual search and reading of rapid serial presentations of letter strings, words, and text. *Journal of Experimental Psychology: General*, 111(2):208, 1982.

[121] Marcel A. Just and Patricia A. Carpenter. *The psychology of reading and language comprehension.* Allyn & Bacon, 1987.

[122] Marcel A. Just, Patricia A. Carpenter, and Michael E. Masson. What eye fixations tell us about speed reading and skimming. *Eye-lab Technical Report*, 1982.

[123] Marcel A. Just, Patricia A. Carpenter, and Jacqueline D. Woolley. Paradigms and processes in reading comprehension. *Journal of experimental psychology: General*, 111(2):228, 1982.

[124] Nancy Kanwisher and Ewa Wojciulik. Visual attention: insights from brain imaging. *Nature Reviews Neuroscience*, 1(2):91–100, 2000.

[125] Youngho Kim, Ahmed Hassan, Ryen W. White, and Imed Zitouni. Modeling dwell time to predict click-level satisfaction. In *Proc. WSDM 2014*, 2014.

[126] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, DTIC Document, 1975.

[127] Funda Kivran-Swaine, Samuel Brody, and Mor Naaman. Effects of gender and tie strength on twitter interactions. *First Monday*, 18(9), 2013.

[128] Funda Kivran-Swaine, Jeremy Ting, Jed R. Brubaker, Rannie Teodoro, and Mor Naaman. Understanding loneliness in social awareness streams: Expressions and responses. In *Proc. ICWSM 2014*, 2014.

[129] Justin Kosslyn and Cong Yu. Fact check now available in google search and news around the world, 2017. `https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/`.

[130] Adam D.I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[131] Dmitry Lagun and Eugene Agichtein. Inferring searcher attention by jointly modeling user interactions and content salience. In *Proc. SIGIR 2015*, 2015.

[132] Dmitry Lagun and Mounia Lalmas. Understanding user attention and engagement in online news reading. In *Proc. of WSDM 2016*, 2016.

[133] Cliff Lampe, Nicole B. Ellison, and Charles Steinfield. Changes in use and perception of facebook. In *Proc. CSCW 2008*, 2008.

[134] Cliff Lampe, Rebecca Gray, Andrew T. Fiore, and Nicole B. Ellison. Help is on the way: Patterns of responses to resource requests on facebook. In *Proc. CSCW 2014*, 2014.

[135] Cliff Lampe and Erik Johnston. Follow the (slash) dot: effects of feedback on new members in an online community. In *Proceedings of the 13th international conference on supporting group work*, GROUP '05, 2005.

[136] Cliff Lampe, Rick Wash, Alcides Velasquez, and Elif Ozkaya. Motivations to participate in online communities. In *Proc. CHI 2010*, 2010.

[137] Randy J. Larsen and David M. Buss. Personality psychology: Domains of knowledge about human behavior, 2002.

[138] Nilli Lavie. Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, 9(2):75–82, 2005.

[139] Nilli Lavie, Aleksandra Hirst, Jan W. De Fockert, and Essi Viding. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3):339, 2004.

[140] Donald J Leu, Heidi Everett-Cacopardo, J Zawilinski, Greg Mcverry, W Ian O'Byrne, et al. New literacies of online reading comprehension. *The Encyclopedia of Applied Linguistics*, 2007.

[141] Claude Lévi-Strauss. *The elementary structures of kinship*. Beacon Press, 1969.

[142] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002.

[143] Nan Lin. Building a network theory of social capital. *Connections*, 22(1):28–51, 1999.

[144] Eden Litt. Knock, knock. who's there? the imagined audience. *Journal of Broadcasting & Electronic Media*, 56(3):330–345, 2012.

[145] Chao Liu, Ryen W. White, and Susan Dumais. Understanding web browsing behaviors through weibull analysis of dwell time. In *Proc. SIGIR 2010*, 2010.

[146] Jiahui Liu, Peter Dolan, and Elin R. Pedersen. Personalized news recommendation based on click behavior. In *Proc. IUI 2010*, 2010.

[147] Ziming Liu. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of documentation*, 61(6):700–712, 2005.

[148] Elizabeth F. Loftus, Geoffrey R. Loftus, and Jane Messo. Some facts about "weapon focus". *Law and Human Behavior*, 11(1):55, 1987.

[149] Annie Louis and Ani Nenkova. What makes writing great? first experiments on article quality prediction in the science journalism domain. *Transactions of the Association for Computational Linguistics*, 2013.

[150] Gloria Mark, Victor M. Gonzalez, and Justin Harris. No task left behind?: examining the nature of fragmented work. In *Proc. CHI 2005*. ACM, 2005.

[151] Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proc. CHI 2008*. ACM, 2008.

[152] Gloria Mark, Shamsi T. Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proc CHI 2016*. ACM, 2016.

[153] Catherine C Marshall. The gray lady gets a new dress: a field study of the times news reader. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*. ACM, 2007.

[154] Misa T. Maruyama, Scott P. Robertson, Sara K. Douglas, Bryan C. Semaan, and Heather A. Faucett. Hybrid media consumption: How tweeting during a televised political debate influences the vote decision. In *Proceedings of the 17th conference on computer supported cooperative work*, CSCW '14, 2014.

[155] Alice E. Marwick and danah boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1):114–133, 2011.

[156] Michael E. Masson. Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 1982.

[157] Carrie J. McAdams and John H.R. Maunsell. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, 23(4):765–773, 1999.

[158] Jon D. Mcauliffe and David M. Blei. Supervised topic models. In *Advances in neural information processing systems*, 2008.

[159] Caitlin McLaughlin and Jessica Vitak. Norm evolution and violation on facebook. *New Media & Society*, 14(2):299–315, 2012.

[160] Bertrand Mesot and David Barber. A bayesian alternative to gain adaptation in autoregressive hidden markov models. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 2, pages II–437. IEEE, 2007.

[161] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and

Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics*, 2015. R package version 1.6.7.

[162] Amy Mitchell, Jeffrey Gottfried, Michael Barthel, and Elisa Shearer. The modern news consumer, news attitudes and practices in the digital era, 2016.

[163] Amy Mitchell, Mark Jurkowitz, and Kenneth Olmstead. Social, search and direct pathways to digital news, 2014.

[164] Katarzyna Molek-Kozakowska. Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse & Communication*, 7(2):173–197, 2013.

[165] Meredith R. Morris, Jaime Teevan, and Katrina Panovich. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proc. CHI 2010*, 2010.

[166] Adam Mosseri. Working to stop misinformation and false news, 2017. `https://newsroom.fb.com/news/2017/04/working-to-stop-misinformation-and-false-news/`.

[167] Lev Muchnik, Sinan Aral, and Sean J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

[168] Hermann J. Müller and Patrick M. Rabbitt. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental psychology: Human perception and performance*, 15(2):315, 1989.

[169] Kevin P Murphy. Hidden semi-markov models (hsmms). *Informal Notes*, 2002.

[170] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. Is it really about me?: message content in social awareness streams. In *Proc. CSCW 2010*, 2010.

[171] Vidhya Navalpakkam, LaDawn Jentzsch, Rory Sayres, Sujith Ravi, Amr Ahmed, and Alex Smola. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proc. WWW 2013*, 2013.

[172] Mark Newman, Albert-Laszlo Barabasi, and Duncan J. Watts. *The structure and dynamics of networks*. Princeton University Press, 2011.

[173] Jakob Nielsen. How little do users read, 2008. `http://www.nngroup.com/articles/how-little-do-users-read/`.

[174] Oded Nov. What motivates wikipedians? *Communications of the ACM*, 50(11):60–64, 2007.

[175] Jan M Noyes and Kate J Garland. Computer-vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9):1352–1375, 2008.

[176] Christine Nuttall. *Teaching reading skills in a foreign language*. ERIC, 1996.

[177] Kathleen M. O'Craven, Bruce R. Rosen, Kenneth K. Kwong, Anne Treisman, and Robert L. Savoy. Voluntary attention modulates fmri activity in human mt–mst. *Neuron*, 18(4):591–598, 1997.

[178] Bing Pan, Helene A Hembrooke, Geri K Gay, Laura A Granka, Matthew K Feusner, and Jill K Newman. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*. ACM, 2004.

[179] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1âĂŞ2):1–135, 2008.

[180] Zizi Papacharissi and Andrew Mendelson. Toward a new (er) sociability: uses, gratifications and social capital on facebook. In Stylianos Papathanassopoulos, editor, *Media perspectives for the 21st century*, pages 212–230. Routledge, New York, 2011.

[181] Harold E. Pashler and Stuart Sutherland. *The psychology of attention*, volume 15. MIT press Cambridge, MA, 1998.

[182] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009.

[183] James W. Pennebaker, Martha E. Francis, and Roger J. Booth. Linguistic inquiry and word count: Liwc 2001. *Word Journal Of The International Linguistic Association*, 2001.

[184] Tamara Peyton. *Internet and Emotions*, chapter Emotion to Action? Deconstructing the ontological politics of the "Like" button., pages 113âĂŞ–129. Routledge, New York, NY, 2014.

[185] Elizabeth A. Phelps, Sam Ling, and Marisa Carrasco. Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological science*, 17(4):292–299, 2006.

[186] Alan B Poritz. Linear predictive hidden markov models and the speech signal. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82.*, volume 7, pages 1291–1294. IEEE, 1982.

[187] Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32, 2009.

[188] Kenneth R. Pugh, Bennett A. Shaywitz, Sally E. Shaywitz, R. Todd Constable, Pawel Skudlarski, Robert K. Fulbright, Richard A. Bronen, Donald P. Shankweiler, Leonard Katz, Jack M. Fletcher, and John C. Gore. Cerebral organization of component processes in reading. *Brain*, 119(4):1221–1238, 1996.

[189] Sheizaf Rafaeli, Tsahi Hayat, and Yaron Ariel. Knowledge building and motivations in wikipedia: Participation as "ba". In Francisco J. Ricardo, editor, *Cyberculture and new media*, pages 51–68. Rodopi Press, Amsterdam/New-York, 2009.

[190] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[191] Erik D. Reichle, Alexander Pollatsek, Donald L. Fisher, and Keith Rayner. Toward a model of eye movement control in reading. *Psychological review*, 105(1):125, 1998.

[192] Harry T. Reis, Kennon M. Sheldon, Shelly L. Gable, Joseph Roscoe, and Richard M. Ryan. Daily well-being: The role of autonomy, competence, and relatedness. *Personality and social psychology bulletin*, 26(4):419–435, 2000.

[193] Carole Rich. *Writing and reporting news: A coaching method*. Cengage Learning, 2015.

[194] Richard M. Ryan and Edward L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.

[195] Marshall D. Sahlins and Michael Banton. On the sociology of primitive exchange in the relevance of models for social anthropology. 1965.

[196] Lauren Scissors, Moira Burke, and Steven Wengrovitz. What's in a like?: Attitudes and behaviors around receiving likes on facebook. In *Proc. CSCW 2016*, 2016.

[197] Zachary M Seward. How the huffington post uses real-time testing to write better headlines. `http://www.niemanlab.org/2009/10/how-the-huffington-post-uses-real-time-testing-to-write-better-headlines/`, 2009.

[198] Amit Sharma and Dan Cosley. Studying and modeling the connection between people's preferences and content sharing. In *Proc. of CSCW 2015*, pages 1246–1257, 2015.

[199] Tali Sharot and Elizabeth A. Phelps. How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, 4(3):294–306, 2004.

[200] J. Sidney Shrauger and Stephen C. Jones. Social validation and interpersonal evaluations. *Journal of Experimental Social Psychology*, 4(3):315–323, 1968.

[201] Tony Silvia and Tom Johnson. Cnn: The origins of the 24-hour, international news cycle: An interview with tom johnson. *Global news: Perspectives on the information age*, pages 45–52, 2001.

[202] Herbert A. Simon. Designing organizations for an information-rich world. 1971.

[203] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307, 2008.

[204] Chenhao Tan, Lillian Lee, and Bo Pang. The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter. In *Transactions of the Association for Computational Linguistics*, 2014.

[205] Edson C Tandoc Jr. Why web analytics click: Factors affecting the ways journalists use audience metrics. *Journalism Studies*, (ahead-of-print):1–18, 2014.

[206] Dalmas A. Taylor and Irwin Altman. Communication in interpersonal relationships: social penetration processes. 1987.

[207] Anne Treisman and Stephen Gormican. Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, 95(1):15, 1988.

[208] Zeynep Tufekci. Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7), 2014.

[209] C. J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 2nd edition, 1979.

[210] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.

[211] Jessica Vitak. *Keeping connected in the Facebook age: The relationship between Facebook use, relationship maintenance strategies, and relational outcomes*. PhD thesis, Michigan State University, 2012.

[212] Jessica Vitak. Facebook makes the heart grow fonder: relationship maintenance strategies among geographically dispersed and communication-restricted connections. In *Proc. CSCW 2014*, 2014.

[213] Jessica Vitak and Jinyoung Kim. You can't block people offline: Examining how facebook's affordances shape the disclosure process. In *Proc. CSCW 2014*, 2014.

[214] Tor D. Wager and Edward E. Smith. Neuroimaging studies of working memory. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4):255–274, 2003.

[215] Joseph B. Walther. Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1):3–43, 1996.

[216] Joseph B. Walther and Judee K. Burgoon. Relational communication in computer-mediated interaction. *Human communication research*, 19(1):50–88, 1992.

[217] Yi-Chia Wang, Moira Burke, and Robert E. Kraut. Modeling self-disclosure in social networking sites. In *Proc. CSCW 2016*, 2016.

[218] Yi-Chia Wang, Hayley Hinsberger, and Robert E. Kraut. Does saying this make me look good?: How posters and outsiders evaluate facebook updates. In *Proc. CHI 2016*, 2016.

[219] Harald Weinreich, Hartmut Obendorf, Eelco Herder, and Matthias Mayer. Not quite the average: An empirical study of web use. *ACM Transactions on the Web*, 2(1):5:1–5:31, March 2008.

[220] Jason Weston, Sumit Chopra, and Keith Adams. #tagspace: Semantic embeddings from hashtags. In *Proc. EMNLP 2014*, 2014.

[221] David S White and Alison Le Cornu. Visitors and residents: A new typology for online engagement. *First Monday*, 16(9), 2011.

[222] Steve Whittaker, Vaiva Kalnikaite, Victoria Hollis, and Andrew Guydish. "don't waste my time": Use of time information improves focus. In *Proc. CHI 2016*. ACM, 2016.

[223] Greg Ridgeway with contributions from others. *gbm: Generalized Boosted Regression Models*, 2015. R package version 2.1.1.

[224] Jeremy M. Wolfe and Todd S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature reviews neuroscience*, 5(6):495–501, 2004.

[225] Fang Wu and Bernardo A. Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599–17601, 2007.

[226] Shaomei Wu, Atish Das Sarma, Alex Fabrikant, Silvio Lattanzi, and Andrew Tomkins. Arrival and departure dynamics in social networks. In *Proc. WSDM 2013*, 2013.

[227] Alfred L. Yarbus. *Eye movements during perception of complex objects*. Springer, 1967.

[228] Robert M. Yerkes and John D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, 18(5):459–482, 1908.

[229] Xing Yi, Liangjie Hong, Erheng Zhong, Nanthan N. Liu, and Suju Rajan. Beyond clicks: dwell time for personalization. In *Proc. RecSys 2014*, 2014.

[230] Dolf Zillmann. Exemplification theory: Judging the whole by some of its parts. *Media Psychology*, 1(1):69–94, 1999.