The Dissertation Committee for Shun-Yang Lee
certifies that this is the approved version of the following dissertation:

Essays on the Interaction between Users and Information Systems

Committee:

Andrew B. Whinston, Supervisor

Ashish Agarwal, Co-Supervisor

Brendan Kline

Liangfei Qiu

Wen Wen

Essays on the Interaction between Users and Information Systems

by

Shun-Yang Lee, B.A.; B.M.; M.S.I.R.O.M.

DISSERTATION

Presented to the Faculty of the Graduate School

of The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2016

To my family, for their unconditional love and encouragement

Acknowledgments

I am indebted to my advisor, Dr. Andrew B. Whinston, for his incredible mentorship, guidance, inspiration, and support. I thank him for seeing the potential in me and helping me grow as a researcher. It has been a privilege to work with him.

I would like to thank Dr. Ashish Agarwal, Dr. Liangfei Qiu, and Dr. Huaxia Rui for the invaluable opportunity to work with and learn from them. I would also like to acknowledge Dr. Brendan Kline and Dr. Wen Wen for all the help they have provided.

Finally, I would like to thank my family, friends, and colleagues. None of this would be possible without their unceasing support.

Shun-Yang Lee

The University of Texas at Austin

August 2016

Essays on the Interaction between Users and Information Systems

Publication No. ____________________

Shun-Yang Lee, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Andrew B. Whinston
Co-Supervisor: Ashish Agarwal

The role of information systems has evolved from providing decision support into enabling the majority of our daily operations, and the way users interact with information systems has changed dramatically as a result. The goal of this dissertation is to study phenomena that stem from the close interaction between users and information systems using empirical methodologies.

The first essay of this dissertation focuses on the issue of sentiment manipulation. We show that strategic players might be incentivized to manufacture content on social media platforms and opinion forums, in the context of the movie industry. We then identify unusual patterns on Twitter that are consistent with sentiment manipulation.

We study the effectiveness of social media advertising in the second

essay. Advertisers on popular social media platforms such as Facebook are able to publish ads with popularity and social information. We design and conduct a randomized field experiment to study the extent to which these types of information have an effect on ad performance.

In the third essay we study how individuals might be biased toward contents that appear to be written more politely. We use data from an online question answering platform, StackExchange, to show that an individual who posts a question on the platform tends to prefer polite answers to clear answers.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Users and information systems often interact and co-evolve with each other. As technology permeates every aspect of our lives, it is critical that we understand the relationship between users and information systems. The goal of this dissertation is to advance our understanding of this relationship through a series of empirical studies.

In the first essay, we study the issue of sentiment manipulation. Online platforms are prone to abuse and manipulations from strategic parties. For example, social media and review websites suffer from sentiment manipulations, manifested in the form of opinion spam and fake reviews. The consequence of such manipulations is the deterioration of information quality as well as loss in consumer welfare. We study the issue of sentiment manipulation on Twitter in the context of movie tweets. Through a regression discontinuity design (RDD) approach and using the movie release as a source of an exogenous shock, we find that both the average Twitter sentiment and the proportion of highly positive tweets exhibit a significant drop on the movie's release day. In addition, independent productions and low budget movies tend to experience a larger drop than major studio productions and high budget movies. To examine the effect of competition on firm manipulation, we construct a movie competition measure based on both the time and theme dimensions through topic modeling, and we find that a higher level of competition leads to a larger drop in Twitter sentiment. To strengthen our identification, we adopt a difference-in-discontinuity design where we compare Twitter with a more controlled movie platform, IMDb, and the results are consistent with sentiment manipulation. In addition, we consider movie studios' earnings announcement dates and severe weather condition as additional sources of exogenous shocks, as well as control for demand-side expectations through Hollywood Stock Exchange (HSE) and Google Trends data, and the results are robust. Discussing and ruling out alternative explanations, we argue that the observed drops of Twitter sentiment come from firms' strategic manipulation. This study sheds light on the reliability of sentiment analysis, and contributes to our understanding of strategic manipulation.

In the second essay we study the effectiveness of popularity and social information in the context of social media advertising. Social media platforms such as Facebook show ads with popularity or word-of-mouth (WOM) signals such as "likes". Additionally, these ads can display social endorsement from friends. This paper examines the effectiveness of displaying these different signals on social media ads in generating user attention and actual conversions in the form of app installs through a randomized field experiment on Facebook. We partnered with a mobile app company and conducted an ad campaign on Facebook to randomly target sixteen unique user groups to install the mobile application. We find that the overall "likes" associated with the ad do not help the user's decision on clicking the ad and the app-installing decision conditional on clicking. Further, ads endorsed by friends have a lower click performance as compared to the ones without such endorsement. However, the negative effect of "likes" on the app install performance conditional on clicking is attenuated for ads with social endorsement. Our results have implications for the design of ads on social media platforms.

In the final essay we study the issue of *politeness bias* in a knowledge management context. Knowledge management is a major research topic in the information systems literature. Popular knowledge management platforms such as community-based question answering sites (CQAs) and electronic networks of practice (ENPs) rely on accurate quality assessment of user contributed content to ensure effective knowledge creation and exchange. However, quality assessment is subjective in nature. We first study the issue of content quality assessment through an innovative context-free linguistic analysis and hypothesize that, based on psycholinguistic theories, the use of pronouns and other function words is correlated with the content's perceived quality, moderated by the specific quality measure adopted. We then use the politeness theory to explain why two popular quality assessment methods would lead to diverging results. We empirically test our hypotheses through a random coefficient logit model and a fixed-effect negative binomial model with data obtained from StackExchange, a popular CQA platform. Our analysis shows that the *politeness bias* might significantly affect people's evaluation of content quality. We also uncover potential limitations posed by an overreliance on subjective quality measures. We suggest that care be taken in choosing the appropriate quality measure to meet platforms' operational objectives. This study contributes to the literature in knowledge management, strategic communication, and cognitive bias.

Chapter 2

Sentiment Manipulation in Online Platforms and Opinion Forums

2.1 Introduction

Much economic activity involves the understanding of consumers' preferences and the subsequent recommendations of products of interest, both of which are instrumental to product-selling firms' performances. Thanks to their popularity and ability to reach diverse demographic groups, internet platforms have established themselves as powerhouses where consumers seeking information can interact among themselves as well as with sellers, and sellers can actively identify target consumers and channel their advertisements accordingly. These online platforms include e-commerce websites such as Amazon; social media sites such as Facebook, Twitter, and Foursquare; and recommendation and review websites such as Yelp, TripAdvisor, and Expedia. What these platforms have in common is the ability for consumers to voice their opinions, and for sellers to inform potential consumers of the quality of their products or services in various ways. Both consumers and sellers can benefit from active participation on these platforms because they provide consumers with much more detailed information on products, while providing firms the opportunity to reach out to target consumers.

However, popular online platforms are prone to abuse and manipulation from strategic parties, and these manipulations can cause dire consequences. For example, to increase product visibility, producers might want to manipulate platform data by adding positive sentiments themselves, so the sentiment analysis results will be more favorable to their products. Furthermore, because the goal of many online platforms is to increase visitor traffic and encourage more user participation, eradicating such firm-generated data pollution is not necessarily a priority. These platforms have also been used by political parties as tools to spread propaganda.[1]

Generally speaking, manipulation is a result of lack of awareness, the absence of verification mechanisms, platform incentives, or the nature of the chosen business model. Online platforms do not always have a proper

[1] Popular news outlets have reported several political uses of social media platforms: Pro-Chinese individuals have used fake Twitter accounts to influence public opinion about troubled regions including Xinjiang and Tibet; Ukraine protesters have been attacked by Pro-Russian fake Facebook accounts; The United States military reportedly engaged in the creation of fake social network accounts to spread pro-American propaganda.

verification mechanism to filter out *opinion spam*—that is, fictitious and often fraudulent reviews that are written specifically to deceive readers (Ott et al. 2011; Ott et al. 2012; Mukherjee et al. 2013). They are especially susceptible to sentiment contamination, by which advertisers deliberately manipulate public opinions by exploiting spamming techniques to create manipulated word of mouth.

To better understand the effect of manipulation, we conduct an empirical analysis to examine the reliability of Twitter sentiment in the context of movie tweets. Since direct detection of manipulated tweets is difficult, we adopt a regression discontinuity design approach (RDD) using movie release as a source of an exogenous shock, and we find that the average Twitter sentiment and the proportion of highly positive tweets both exhibit a significant drop on the movie's release day. In addition, independent productions and low budget movies tend to experience a larger drop than major studio productions and high budget movies. To examine the effect of competition on movie sentiment manipulation, we construct a movie competition measure based on both the time and theme dimensions through topic modeling. Our finding suggests that a higher level of competition leads to a larger drop in Twitter sentiment on the release day, and hence more manipulation.

To strengthen our identification argument, we use a cross-platform identification strategy (Chevalier and Mayzlin 2006; Proserpio and Zervas 2014) where we exploit the institutional difference between Twitter and a more controlled movie platform, IMDb, to identify sentiment manipulation. A cross-platform identification strategy is a powerful way of finding a proper control group (Proserpio and Zervas 2014): whereas a typical matching procedure finds a similar control unit for each treated unit, a cross-platform analysis compares the observation of the same unit across different platforms, which helps reduce the concern of endogeneity-related biases (Imbens 2004). We conduct our cross-platform study on Twitter and IMDb through a difference-in-discontinuity approach (Grembi et al. 2016), which is essentially a combination of difference-in-differences (DID) and RDD. The difference-in-discontinuity approach is powerful because the effects of most potential confounders will be differenced out through the estimation procedure. Our difference-in-discontinuity analysis results are consistent with those from the RDD analysis, further suggesting the existence of sentiment manipulation.

To further alleviate the concern that the movie release shock might not be entirely exogenous, we use two additional sources of exogenous shocks: movie studios' earnings announcement dates and severe weather conditions. The accounting and finance literature suggests that movie studios have an incentive to disclose good news prior to earnings announcements. While earnings announcements might affect movie studios' behavior, they are less likely to affect ordinary moviegoers' tweeting or commenting behaviors. Using movie studios' quarterly earning announcement dates as exogenous shocks, we find that there is a significant sentiment drop on the announcement date, which is consistent with potential sentiment manipulation.

Severe weather conditions have been shown to significantly reduce people's willingness to go out and watch movies (Moretti 2011), and these weather conditions have been widely used as a source of exogenous shocks (Angrist and Krueger 2001; Miguel et al. 2004; Moretti 2011; Qiu, Tang and Whinston 2015). The intuition behind this exogenous shock is that, when weather condition is severe, less people would want to watch movies in the movie theater, and hence the movie studio's manipulation incentive will decrease. We combine the movie theater location (from the U.S. Census data) and weather data (from Weather Underground) to construct a measure to reflect the percentage of movie theaters in the U.S. experiencing severe weather conditions—namely, cold, heavy rain, or heavy snow—on a given movie release day. We then incorporate these severe weather conditions into our RDD analysis. The results show less sentiment drops when there is a severe weather condition on the movie release day, which further suggests that the sentiment drop is likely coming from movie studios' strategic sentiment manipulation.

It is possible that Twitter sentiment could be related to consumers' expectation of movie quality and other demand-side expectation. For example, consumers might expect movies to be of high quality after watching movie trailers or being exposed to other marketing campaigns. Any difference in sentiment before and after movie release could be coming from the gap between the expected quality and the actual quality of the movie. Therefore, in addition to controlling for movie studios' daily advertising budget, we also control for demand-side expectation by using the Hollywood Stock Exchange (HSE) and Google Trends data, both of which to certain extent reflect consumers' expectation of the movie. We still observe significant sentiment drops even after controlling for these demand side variables, which serve as a further evidence of sentiment manipulation.

Our research contributes to the literature on social media and sentiment analysis by analyzing movie tweets through a regression discontinuity design approach, developing a movie competition measure based on topic modeling results, as well as using a difference-in-discontinuity and other sources of exogenous variations to validate our empirical results. Our analyses also suggest that practitioners should be cautious when conducting sentiment analysis on user generated content, since the designs of these platforms make them susceptible to strategic manipulation.

2.2 Literature Review

User generated content (UGC) and social media data have been used in all aspects of decision making processes. However, several studies have empirically shown the existence of widespread manipulation practices on these sites. Mayzlin et al. (2014) examined the prevalence of difficult-to-detect fake reviews on popular review websites. More specifically, they used a difference-in-differences approach to study how hotel characteristics and ownership structure affect the level of review manipulation, which consists of posting positive reviews for one's own business and manufacturing negative reviews for competitors, on travel websites Expedia.com and TripAdvisor.com. Luca and Zervas (2015) investigated the presence of restaurant review fraud on another review website, Yelp.com. They found that positive review fraud is related to reputational concerns, whereas negative review fraud is more likely due to competitions. Anderson and Simester (2014) offered a different perspective on the nature of deceptive reviews. Using a dataset from a private apparel retailer, they found that, in addition to firms' strategic behaviors, customers without clear financial incentives to manipulate product ratings might still write reviews on products they did not purchase. Hu et al. (2012) developed a statistical method to examine whether strategic manipulation is present in product reviews. Several computer science studies including Ott et al. (2011), Ott et al. (2012), and Mukherjee et al. (2013) also used statistical and machine learning techniques to identify and estimate the prevalence of opinion spam.

Besides review websites, social media platforms also suffer from manipulative behaviors. Stringhini et al. (2012) detailed the existence of Twitter Account Markets that aim at inflating one's number of followers as well as sending out advertising tweets at a large scale.[2] Messias et al. (2013) constructed fake accounts on Twitter to demonstrate how these accounts' influence measures can be significantly improved by following simple automated strategies. The results of these studies imply that the credibility of UGC and social media data can be questionable. Therefore, decisions based on questionable data can be harmful to the decision maker's welfare. Manipulation on these platforms also has behavioral implications. Adomavi-

[2]For example, advertisers can purchase Twitter accounts online and use these accounts to generate synthetic sentiment. http://buyaccs.com/en is an example of such websites that allow anyone to purchase Twitter accounts in bulk.

cius et al. (2013) examined the effect of recommendation on consumers' preference formation. Their findings suggest that strategic recommender systems can intentionally provide recommendations that would result in systematic biases. Forman et al. (2008) studied the issue of identity disclosure in online communities, and they found that reviewers' identity disclosure affects community members' judgment and product sales. Strategically misrepresented identities, then, might mislead consumers and result in inferior decision making. Goh et al. (2013) examined the differential impact of user generated and marketer generated contents on consumers' purchasing decisions. These effects are less clear, however, if consumers are uncertain about the true identities of content contributors. Tsikerdekis and Zeadally (2014) provided an overview of online deception in social media platforms. They discussed several deception techniques and identified challenges facing deception researchers, including a lack of a unified theory and methods for deception detection.

There have been several theoretical studies that examine the issue of sentiment manipulation. Dellarocas (2006) constructed a theoretical model on firms' manipulative behaviors. He showed that manipulations could be beneficial to consumers if firms' manipulation strategies are monotonically increasing in their true qualities. He also showed that, under certain threshold conditions, firms would actually benefit if manipulation were not possible. Mayzlin (2006) examined marketers' incentives to generate anonymous promotional messages online. Using a game theoretic model, her results showed that, contrary to traditional advertising strategies, firms producing low quality products would engage in more promotional chat than those producing high quality products. This is because high quality products benefit from positive WOM which substitutes for advertising, whereas low quality products do not.

Recently, researchers have begun to investigate the relationship between media sentiment and firm stock performance. A stream of literature on the financial value of social media has shown that social media-based metrics and social media sentiment have a strong relationship with firm equity value (Luo et al. 2013; Yu et al. 2013). More specifically, in the setting of the movie industry, Chen et al. (2012) showed a significant effect of movie reviews on movie studios' stock performance. While there exist theories and evidences on the effects of media sentiment on firm equity value, there are very few studies examining firms' strategic sentiment manipulation in social media to boost stock performance prior to earnings announcements. Notice that the prior accounting literature identifies two main strategies to manage reported earnings: (1) accounting method changes and accrual-based earnings management that do not affect cash flows and (2) real actions that affect cash flows. The early evidence on earnings management concentrates on accrual-based strategies. Sentiment manipulation belongs to the second strategy: it can affect movie sales, and hence have an impact on cash flows of movie studios. Our research sheds light on understanding sentiment manipulation as a form of earnings management.

2.3 Institutional Background and Data

2.3.1 Movie Industry

In our empirical analysis we examine how the sentiment measure reflected on Twitter is affected by manipulation, in the context of the movie industry. Before we describe our empirical analysis, we first provide some institutional background of the movie industry to explain topics such as revenue sharing and advertising. We then argue that, since manipulation activities can be perceived as a type of promotional activity, movie studios have incentives to conduct sentiment manipulation, especially before movie release.

There are three major categories of players in the movie industry: producers, distributors, and exhibitors (McKenzie 2012; Walls and McKenzie 2012). A producer may buy a screenplay, buy a book to adapt into a screenplay, or hire a writer to develop an idea, and then make a movie. A distributor distributes the movie; it also makes important operational decisions such as choosing a release date and designing and implementing an advertising campaign. Finally, exhibitors are movie theaters that show movies to audiences. Recently there has been a trend of vertical integration in the movie industry: movie studios increasingly both produce and distribute movies themselves (Corts 2001; Gilchrist and Sands 2015). Corts (2001) found that movie producers and distributors generally act like integrated firms. Therefore, in our study, we regard a movie studio as both a producer and distributor.

From the extant literature on the movie industry we know that one of the most important decisions a movie distributor/movie studio has to make is to implement an effective advertising campaign (McKenzie 2012; Walls and McKenzie 2012; Gilchrist and Sands 2015), and sentiment manipulation can be a strategy distributors/studios use to advertise movies. Moreover, several reasons suggest that studios might be incentivized to conduct *more* sentiment manipulation prior to movie release, and *less* afterward. This observation turns out to be instrumental for our empirical analysis, and we enumerate these reasons in detail below.

(1) Distributors/movie studios typically charge fees as percentages of box office revenues rather than a flat fee. More importantly, the share division of a distributor/studio changes over weeks of the movie's run, with a smaller share for the distributor/studio in later weeks: for a major motion picture, for example, it is common for the distributor/studio to keep as much as 90% of revenues in the opening week, and hence the incentive for the distributor/studio to conduct sentiment manipulation should be very high before the opening weekend release. After the opening week, the distributor's share drops dramatically to 50% or even 30% (notice that the exhibitor tends to get a larger share after the opening week) (Moul 2007; McKenzie 2012; Gilchrist and Sands 2015).

(2) In the movie industry, it is widely believed that the opening weekend is critical for studios. A movie that fails to open strongly almost always loses the attention of the media, audiences, and exhibitors. According to Box Office Mojo, the opening weekend accounts for a very large fraction

of a film's box office, typically 30% - 45%.[3] Additionally, if the opening weekend box office revenue is low, the exhibitors (movie theaters) may drop the movie or reduce the number of screens on which the theater shows the movie some point after release. Krider et al. (2005) showed that exhibitors closely monitor box office sales and respond with screen allocation decisions.[4] Therefore, the incentive of the distributors/studios to conduct sentiment manipulation should be very high before the movie release, and then drop significantly after the release.

(3) The vast majority (90%) of a movie's advertising budget is spent before the movie release (Elberse and Anand 2007). One important underlying reason is that the revenue sharing division of the distributor/studio drops significantly in later weeks, as mentioned earlier. Considering the similarity between advertising and sentiment manipulation (both of them promotion efforts of distributors/studios), the level of sentiment manipulation may also drop significantly. In a related context, Hu et al. (2011) found that the manipulation level in online book reviews decreases with the passage of time.

Based on the above arguments, movie distributors/studios have incentives to conduct sentiment manipulation, and they may conduct more manipulation prior to the movie release.

[3] http://www.boxofficemojo.com/alltime/weekends/

[4] In general, "a film's opening weekend is usually the most lucrative one for its studio. Financial agreements with theaters normally give the filmmaker a greater percentage of the box office during the first weeks of release. And in this glutted market, studio executives also worry that theaters will replace a film with another if it doesn't win audiences quickly" (Corts 2001).

2.3.2 Twitter and Manipulation

Here we briefly discuss why Twitter is a well-suited platform to study sentiment manipulation. Compared with other professional movie review platforms, Twitter is a relatively open platform, which makes it much easier for interested players to engage in strategic and manipulative behaviors. While Twitter is perhaps not a major platform that consumers would visit in search of product reviews, many companies today adopt tweeting as a new marketing tool.[5] The prior literature has shown that tweets can have significant effects on movie box office revenue (Rui et al. 2013) and on the viewership of TV shows (Gong et al. 2015;[6] Seiler et al. 2015).[7] In reality, people may follow Twitter trends or movie studio's official Twitter accounts to obtain latest movie news, and movie studios can use fake Twitter accounts or hire real Twitter accounts to post overly positive messages about their movies and attract customers. These fake accounts or hired accounts are called hidden paid posters or termed "Internet water army" in China.[8]

The New York Times reported that Sony was fined by the Connecticut attorney general for creating fake reviews for at least four of its movies.[9] The phenomenon of online manipulation in movie industry is neither new

[5] http://www.msi.org/reports/does-tweeting-impact-the-bottom-line/

[6] Gong, Shiyang, Juanjuan Zhang, Ping Zhao, and Xuping Jiang (2015). "Tweeting Increases Product Demand." Working paper, MIT Sloan School of Management, Massachusetts Institute of Technology.

[7] Seiler, Stephen, Song Yao, and Wenbo Wang (2015). "The Impact of Earned Media on Demand: Evidence from a Natural Experiment." Stanford University Graduate School of Business Research Paper No. 15-62, Available at SSRN: http://ssrn.com/abstract=2692861.

[8] https://en.wikipedia.org/wiki/Internet_Water_Army

[9] http://www.nytimes.com/2002/03/13/nyregion/metro-briefing-connecticut-hartford-falsified-movie-reviews.html

nor exclusive to the US. Sina Weibo (often described as "China's Twitter") has been used by opportunistic companies,[10] and the movie industry has quickly adopted sentiment manipulation for promotion and advertising in China.[11] The Chinese movie "The Last Supper" admitted hiring an "Internet water army" to raise its rating on film review websites and to endorse the movie on Sina Weibo. Some insiders suspected that using a water army as part of a movie's online promotion is already widely known in the Chinese movie industry: many movies resorts to sentiment manipulation, each at a cost of over 1 million Chinese yuan (around 160,000 US dollars).[12] Such paid posting is a well-managed activity by Internet PR companies involving thousands of individuals and tens of thousands of different online IDs. There is even a quality control team who checks that the fake posts meet a certain "quality" threshold. For instance, a post would not be validated if it is deleted by the host or is composed of garbled words.[13]

The recent marketing literature has studied the causal effect of tweeting on TV show viewing (Gong et al. 2015; Seiler et al. 2015). TV show producers can hire influential users to tweet and boost viewership significantly. Although it is not the same as sentiment manipulation, it shows that viewers can be affected by the endorsement effect of tweets. Meanwhile, Twitter is being overwhelmed by spambots that undermine its value to advertisers,

[10] https://jingdaily.com/commercial-viability-of-sina-weibo-for-luxury-brands-called-into-question/

[11] http://www.eeo.com.cn/ens/2012/1217/237581.shtml

[12] See http://usa.chinadaily.com.cn/life/2012-12/13/content_16013662.htm

[13] http://www.technologyreview.com/view/426174/undercover-researchers-expose-chinese-internet-water-army/

and a handful of bots were programmed to tweet.[14] Other movie review platforms, such as IMDb, have stricter authentication steps for new user registration and a user reputation system to prevent automated registrations and manipulation activities. Therefore, since the main focus of our paper is to study manipulation activities, Twitter appears to be well-suited for our investigation.

2.3.3 Illustrative Example

A platform for movie distributors/studios to conduct sentiment manipulation is Twitter. Specifically, studios can easily manufacture tweets that appear to be posted by individual consumers in order to influence consumers' movie-going decisions. Here we describe an illustrative example where we select two movies, *At Any Price* (2012) and *After Earth* (2013), and collect all tweets related to these movies, starting 60 days before the release date, until 60 days after the release date. We train a Naive Bayes Classifier based on the corpus described in Go et al. (2009), and use this classifier to measure the probability of any given tweet being positive. We then use this probability as a measure of sentiment polarity. The consumer sentiment on Twitter might experience a *release shock* near the release date, because movie studios have more incentives to conduct sentiment manipulation prior to movie release, as described in the previous section. Therefore, one of our main empirical goals is to investigate whether there is a higher level of manipulation

[14]http://www.bloomberg.com/bw/articles/2013-09-25/could-bots-and-spam-smother-the-twitter-ipo

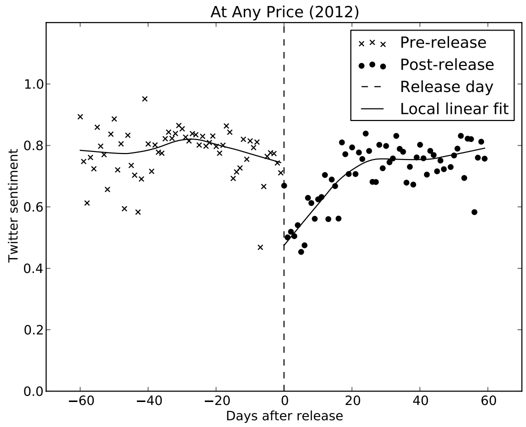

Figure 2.1: Twitter Sentiment of Movie *At Any Price* (2012)

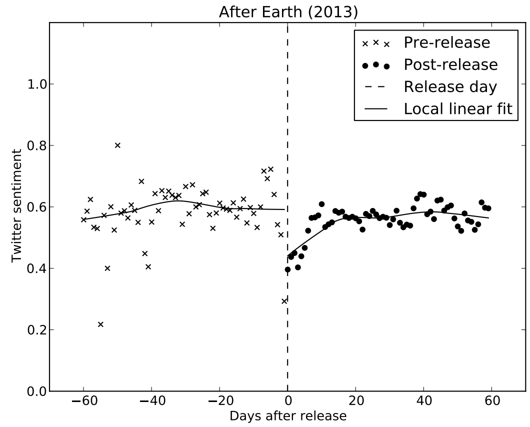

Figure 2.2: Twitter Sentiment of Movie *After Earth* (2013)

prior to movie release.

We use a graphical analysis to examine the aforementioned release shock, and the result is shown in Figures 2.1 and 2.2, where we plot the patterns of Twitter sentiment over time. We can see that there appears to be a discontinuity on the release day in Twitter sentiment for both movies. Notice that the local linear fit is generated via a local linear regression. While refraining from making any causal statements, we do point out the observation that Twitter sentiment exhibits an interesting pattern around the release day, which leads us to suspect that some kind of manipulation might be involved in the observed Twitter sentiment.

2.3.4 Identification Strategy

Since firms' manipulation decisions are unobservable and often difficult to detect, we need an empirical strategy that will allow us to examine manipulation indirectly. Regression discontinuity design (RDD) is a quasi-experimental econometric strategy to establish the causal effect of an intervention or a treatment, and has been used extensively in economics literature. RDD works by considering observations close to a threshold value which determines whether an intervention or treatment is assigned. By comparing observations lying slightly above and below the threshold one is able to estimate the causal effect induced by the treatment. Lee and Lemieux (2010) provided a detailed introduction on the theory and implementation of RDD. In this study, we use a RDD approach by which the subjects that receive treatments are the movies, and the "treatment" is the movie release shock that happens to all movies. The release shock is a reasonable choice of treatment because movie studios' pre-release and post-release manipulation incentives/advertising strategies are different, with the former being much larger than the latter. Notice that, distinct from most RDD studies that compare different subjects above or below some threshold value, we compare the same subject—that is, the same movie—prior to and after facing the release shock. This setup is similar to Goes et al. (2016) where they used an RDD approach to examine whether an individual's motivation to contribute to an online knowledge exchange is affected by whether or not a goal is reached. To ensure the validity of the RDD, it needs to be shown that the subjects cannot perfectly manipulate their

Table 2.1: Summary Statistics of Twitter Movie Sentiment

Variable	Observations	Mean	Std. Dev.	Min.	Max.
Number of Tweets for Each Movie	482	36,415.96	67,527.09	1	511,874
Number of Tweets for Major Studio Movies	82	54,241.38	63,255.41	33	411,317
Number of Tweets for Low Budget Movies	219	14,103.19	33,081.59	1	311,314
Average Sentiment Score of Each Movie	482	0.624	0.0571	0.371	0.923
Average Sentiment Score of Major Studio Movies	82	0.614	0.0472	0.490	0.742
Average Sentiment Score of Low Budget Movies[15]	219	0.631	0.0590	0.371	0.923

treatment status. In other words, in our setup, we need to show that movie studios cannot perfectly manipulate the release shock. It might seem at first glance that movie studios should be able to choose their movie release day; however, it should be noted that our treatment is the behavioral impact and the surge of genuine consumer comments that are produced after the movie release, irrelevant to the actual choice of dates. In addition, the magnitude of this impact cannot be perfectly controlled by the movie studios. Therefore, this insight enables us to establish the validity of our RDD approach. In the following sections, we develop and test several hypotheses through a RDD approach. In addition, we conduct several additional empirical analyses, including a difference-in-discontinuity design using a cross-platform identification strategy, severe weather shocks, and earnings announcement shocks as robustness checks to strengthen our empirical results, in case the movie release shock is not entirely exogenous.

2.3.5 Data Collection

[15]Here low budget movies are those with a budget less than 3 million US dollars.

The illustrative example described in the previous section motivates us to explore in depth how the release shock would affect the Twitter sentiment. We compile a list of 482 movies released in the United States in the years 2012 and 2013, and for each movie, we use its corresponding hashtag (#) to identify and collect its tweets that are written in English, starting 60 days prior to its release day and up to 60 days after the release day.[16] We then use the trained Naive Bayes classifier to classify the polarity of these tweets, and aggregate them daily to construct a daily average Twitter sentiment measure for each movie, with 0 being the lowest possible sentiment level and 1 the highest. We also collect movie characteristics data, including movie runtime, budget, color, movie genre, whether or not the movie is produced by a major studio, and so on, from the website *Internet Movie Database (IMDb)*. Note that not every movie has tweets associated with it during the complete 120-day window; therefore, our data set is an unbalanced panel. Summary statistics of the movie data set are listed in Table 2.1. In addition, we acquired a detailed data set on daily movie advertising expenditure across different advertising channels from Nielsen.

[16]We used a simple heuristic approach where we combined every word in a movie title, removed spaces in between words, to form the hashtag for a given movie. This is the approach users generally use when composing hashtags since hashtags do not allow spaces. See https://support.twitter.com/articles/49309. When we collected the tweets we manually checked, for each movie, if there is any possibility of an attribution problem: for movies that share the same title as some books, cartoons, or games, etc., we removed them from our data set.

2.4 Empirical Analysis

2.4.1 Hypothesis Development and Empirical Results

Recall that the illustrative example shows there is a significant drop in the sentiment level after the release day, which suggests the potential existence of manipulation. We therefore propose the following hypothesis:

Hypothesis 1 *There will be a significant drop (discontinuity) in a movie's Twitter sentiment level on its release day if the sentiment is manipulated.*

Following the literature on regression discontinuity design (Lee and Lemieux 2010), we specify the following parametric polynomial model with fixed effects to test our hypothesis:

$$sentiment_{jt} = \beta_1 \cdot post_{jt} + \sum_{p=0}^{\bar{p}} \beta_{2,p} \cdot duration_{jt}^{p} + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot post_{jt} \cdot duration_{jt}^{p} + \nu_j + \epsilon_{jt}, \tag{2.1}$$

where $sentiment_{jt}$ is movie j's average Twitter sentiment on day t; $post_{jt}$ is a dummy variable that takes the value 1 if day t is on or after movie j's release day, and 0 otherwise; $duration_{jt}$ is the number of days after movie j's release, where a positive value means that day t is after the movie release, and vice versa; ν_j is the movie fixed effects; and ϵ_{jt} is a normally distributed error term. An interaction term is included to allow the regression function to differ on both sides of the cutoff point (Lee and Lemieux 2010, Goes et al. 2016). To explore the sensitivity of results to a range of different model specifications, we include models with different polynomial orders for robustness check, with $\bar{p}$ ranging from 0 to 3, and report AIC and BIC of each model. Note that the model with the minimum AIC/BIC is generally preferred. The regression results are shown in Table 2.2, where the main variable of interest is *post*. The results show that the coefficient of *post* is negative and significant across different model specifications. This indicates a drop in Twitter sentiment, which suggests the existence of sentiment manipulation. We argue that this gap should not be coming from a movie expectation-realization gap, because on average there should be as much overestimate and underestimate in sentiment for each movie—not a consistent overestimate of movie sentiment prior to movie release. This argument is consistent with the identification strategy in Moretti (2011) where he used the residual from a regression of opening-weekend sales of movies on number of screens during the opening weekend as the measure of movie-specific surprises, where surprises were defined as deviations from expected demands, which correspond to the expectation gaps in our setup. He argued that, although there might be movie-specific surprises, on average, theaters would predict the movie demand correctly, as reflected in their choice of number of screens. Parallel to his reasoning, we rule out the possibility of an expectation-realization gap, and Hypothesis 1 is supported. Note that the median of movie sentiment in our sample is 0.624. If we reduce this median by 0.034 (the coefficient on post), then the resulting sentiment level, 0.590, will be very close to the 25th sentiment percentile of movie sentiment, 0.595. Hence the magnitude of the coefficient associated with the *post* variable is considerable. To further control for the demand-side expectation, we conduct additional robustness analyses using several measures of the demand-side expectation such as the Hollywood Stock Exchange and Google Trends as control variables. We also employ a cross-platform identification strategy as well as using different sources of exogenous shocks. The results of these analyses are robust, and we describe them in detail in later sections.

In addition to analyzing sentiment manipulation using the average sentiment level, we also consider the measure proposed by Mayzlin et al. (2014)—namely, the proportion of highly positive tweets among all tweets, as a measure of positive sentiment manipulation. Their study looked at review manipulation on travel websites' hotel reviews, where hotels were given ratings ranging from 1 star to 5 stars. They used the proportion of 5-star reviews among all reviews as a building block of the dependent variable where a difference-in-differences procedure was then used to examine the effects of hotel ownership and competition on positive review manipulation. This measure is informative and important for our analysis because both positive and negative sentiment manipulation can coexist, which would correspond to a potentially large proportion of both highly positive and highly negative tweets. However, we will not be able to observe the distribution of extreme sentiments by only observing the average sentiment level. Therefore, we follow their approach and replace the dependent variable in Equation (2.1) with the proportion of highly positive movie tweets as the dependent variable, where highly positive tweets are defined as those with a raw sentiment score greater than or equal to 0.8. The results

Table 2.2: RD Estimates of the Effect of Movie Release on Twitter Sentiment (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0353*** [-37.09]	-0.0343*** [-34.55]	-0.0333*** [-32.16]	-0.0334*** [-31.17]
duration		9.12e-05*** [5.697]	0.000137*** [5.600]	0.000146*** [4.758]
$duration^2$			-1.08e-07** [-2.207]	-1.76e-07 [-1.286]
$duration^3$				6.83e-11 [0.545]
$post * duration$		-1.34e-05** [-1.970]	-4.36e-05*** [-3.670]	-3.93e-05** [-2.171]
$post * duration^2$			1.60e-07*** [3.094]	1.03e-07 [0.566]
$post * duration^3$				1.24e-10 [0.319]
constant	0.642*** [958.4]	0.639*** [736.1]	0.637*** [628.5]	0.637*** [581.3]
Observations	52,829	52,829	52,829	52,829
AIC	-84836.44	-84870.79	-84880.99	-84881.4
BIC	-84818.69	-84835.29	-84827.74	-84828.15

of this alternative model are consistent with the regression model specified by Equation (2.1), with the coefficients of the *post* variable still significantly negative across different polynomial specifications; therefore, we suspect there is sentiment manipulation. These regression results are shown in Table A.1 in the appendix.

In addition, a movie's budget has direct influences on the cast members, equipments, costumes, and special effects that it can afford, and consumers often use these movie characteristics to predict the quality of the movie. Therefore, the budget provides an indicator of the ex-ante movie quality, and we propose the following:

Hypothesis 2 *Low budget movies tend to have a larger drop in Twitter sentiment than high budget movies.*

The corresponding regression model is as follows:

$$\begin{aligned} sentiment_{jt} =& \beta_1 \cdot post_{jt} + \beta_2 \cdot post_{jt} \cdot lowBudget_j + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot duration_{jt}^{p} \\ &+ \sum_{p=0}^{\bar{p}} \beta_{4,p} \cdot post_{jt} \cdot duration_{jt}^{p} + \nu_j + \epsilon_{jt}, \end{aligned} \tag{2.2}$$

where we define low budget movies to be those with budgets in the lowest 10% percentile, which corresponds to movies with a budget less than 3 million US dollars. The results, including models of different polynomial orders, are shown in Table 2.3. The coefficient of interest, β_2, is significantly negative across different specifications, which suggests that low budget movies indeed face a larger drop in sentiment, and hence more manipulation. Thus Hypothesis 2 is supported. Following our earlier argument, we run the regression specified in Equation (2.2) again with the dependent variable replaced by the proportion of highly positive tweets, and the results are shown in Table A.2 in the appendix. We can see that the coefficients of the variable $post * lowBudget$ are significantly negative across different polynomial specifications. Therefore, we conclude that there is more sentiment manipulation in low budget movies than high budget movies.

Moreover, we check whether or not a movie is produced by a major studio, and we argue that a major studio production is less likely to engage in manipulative behavior because of a high manipulation cost resulting from reputation concerns, whereas independently produced movies have less reputation concerns and thus a lower manipulation cost.

Hypothesis 3 *Major studio movies tend to have a smaller drop in Twitter sentiment than non-major studio movies.*

The regression model is as follows:

$$\begin{aligned} sentiment_{jt} = & \beta_1 \cdot post_{jt} + \beta_2 \cdot post_{jt} \cdot majorStudio_j + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot duration_{jt}^{p} \\ & + \sum_{p=0}^{\bar{p}} \beta_{4,p} \cdot post_{jt} \cdot duration_{jt}^{p} + \nu_j + \epsilon_{jt}, \end{aligned} \tag{2.3}$$

where $majorStudio_j$ takes the value 1 if movie j is produced by a major film studio. Similar to previous analyses, we also include different polynomial orders for robustness check. The results of all model specifications are shown in Table 2.4. The coefficient of interest, β_2, is significantly positive across different specifications. We see similar results when we replace the dependent variable with the proportion of highly positive tweets, as shown in Table A.3 in the appendix. Therefore, Hypothesis 3 is supported. This result is consistent with Mayzlin et al. (2014) where they showed that hotel chains are associated with less rating manipulation than independent hotels because of reputation concerns.

2.4.2 Effect of Competition on Sentiment Manipulation

To empirically examine how firm competition affects sentiment manipulation, we construct a competition measure based on the following two dimensions: (1) timing competition and (2) thematic competition. Timing competition between any given pair of movies is defined as the time interval between these two movies' release days. If two movies are released around the same time, then they are likely to be competing with each other directly. Therefore, a smaller interval between the release days corresponds to a fiercer level of competition, and vice versa. Thematic competition is characterized by how similar any given pair of movies are to each other. We employ a machine learning technique, topic modeling, and use movie keywords collected from IMDb as inputs to the topic model, to uncover the underlying topic distributions of each movie. We then compute the cosine similarity between a pair of movies' relative topic distributions as our measure of thematic similarity between these two movies. Note that the range of thematic similarity is between 0 and 1. The details of topic models, the topics generated, and the operationalization of the similarity measures can be found in the appendix. Once we calculate the time difference between any pair of movies' release days and the thematic similarity between that pair of movies' topics, we can construct our competition measure to be one that, given a movie, counts the number of other movies that are released one month within the focal movie, and with a thematic similarity compared to the focal movie to be larger than 0.7. We denote this competition measure as $[\pm 1 month \& sim > 0.7]$ and develop the following hypothesis:

Hypothesis 4 *Movies facing a higher level of competition will have a larger drop in Twitter sentiment than those facing a lower level of competition.*

To test this hypothesis, we specify the following regression model:

$$
\begin{aligned}
sentiment_{jt} = \quad & \beta_1 \cdot post_{jt} + \beta_2 \cdot post * [\pm 1 month \& sim > 0.7] \\
& + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot duration_{jt}^{p} + \sum_{p=0}^{\bar{p}} \beta_{4,p} \cdot post_{jt} \cdot duration_{jt}^{p} + \beta_5 \cdot color_j \\
& + \beta_6 \cdot runtime_j + \beta_7 \cdot majorStudio_j + \beta_8 \cdot numTweets_j \\
& + \beta_9 \cdot genre_j + \epsilon_{jt}.
\end{aligned} \tag{2.4}
$$

We can see that the coefficient of the $post * [\pm 1 month \& sim > 0.7]$ variable is significantly negative across models of different polynomial orders. Since the $post * [\pm 1 month \& sim > 0.7]$ variable is a measure of competition, this result means that the more competition a movie faces, the more sentiment drop after the movie release we observe in the data. This result is consistent with our expectation that competition among movies will lead to more sentiment manipulation. To see whether the results are robust to different competition measures, we also consider changing the threshold value of the thematic similarity to 0.75 ($\pm 1 month \& sim > 0.75$), as well as changing the time window to two months before and after the release day ($\pm 2 month \& sim > 0.7$), and the results show that more competition corresponds to a larger drop in Twitter sentiment. Therefore, these findings are robust to different competition measures. The results are shown in Table A.11 and Table A.12 in the appendix.

2.5 Endogeneity Concerns of Movie Release Shocks

2.5.1 Difference-in-Discontinuity Analysis

To further investigate the sentiment drop, we employ a relatively new empirical methodology: difference-in-discontinuity design, to conduct more empirical analyses. Essentially, the difference-in-discontinuity design is a combination of difference-in-differences (DID) and regression discontinuity design (RDD) (Grembi et al. 2016). Similar to DID, a requirement of the difference-in-discontinuity design is to identify two similar platforms where there is some institutional difference between the two platforms such that one platform is more susceptible to some shocks while the other one is not. Difference-in-discontinuity design is a cross-platform identification strategy (Chevalier and Mayzlin 2006; Proserpio and Zervas 2014) where institutional differences between two platforms are used to identify the treatment effect. Proserpio and Zervas (2014) argued that the cross-platform identification strategy has its unique strength in finding a proper control group. For example, several matching procedures have been applied to identify appropriate control groups in empirical studies (e.g. Aral et al. 2009). In a typical matching procedure, a treated unit is matched with a control unit based on similar observable characteristics. In the cross-platform identification strategy, the matched treated and control units are not just similar with respect to some observable characteristics, but they are often the *same* unit in different platforms. In our context, our goal is to find two movie-related platforms where any given movie found on one platform is also discussed in the other platform. The advantage of comparing the same movie across two different platforms is that the observable variables of this movie will be *identical* in both platforms by definition, because we are really only looking at a single movie. This is important because Imbens (2004) showed that stringent matching criteria can reduce concerns about endogeneity-related biases. To acheive this, we augment our original dataset with IMDb movie reviews in order to exploit the institutional differences between these two platforms for our difference-in-discontinuity analysis.[17] We describe the institutional difference as follows: Twitter is a less controlled platform where anyone can post tweets to express their opinions on movies (Twitter has a very open API policy). While the popularity of Twitter attracts advertisers to pay to increase their Twitter presence, malicious individuals or even competitors can write simple programs to use fake Twitter accounts to achieve the same level of advertising without Twitter's spam filter and verification mechanisms detecting them (Urbina 2013).

In contrast to Twitter, IMDb is a more strictly controlled platform: it implements a stricter authentication steps for new user registration and a user reputation system to prevent automated registrations and manipulation activities. It also has strict rules on how to avoid gaming the system.[18] Therefore, compared with Twitter, the cost of manipulation in IMDb

[17]IMDb is a popular movie database with detailed cast and crew information. Users can provide review and rating for movies. http://www.imdb.com/

[18]On IMDb, users can vote on movies as many times as they want but every vote will overwrite the previous one; the movie rating displayed on a movie page is a weighted average of all users' votes instead of a simple average (various filters are applied to the raw data in order to eliminate and reduce manipulation activities); unlike Twitter, it is difficult to create and control a large number of IMDb accounts.

is greatly increased. In summary, our difference-in-discontinuity approach is a cross-platform identification strategy where we exploit the institutional differences between Twitter and IMDb to identify manipulation.

We process IMDb review comments to obtain the sentiment scores analogous to how we processed tweets. It is worth noting that the standard DID design, such as the one used in Mayzlin et al. (2014), requires a strong assumption of parallel trend between sentiment manipulation on Twitter and on IMDb in a long time window, which may not be satisfied in our setting. In the difference-in-discontinuity design, we do not require such a strong assumption: we only need the parallel trend between sentiment manipulation on Twitter and on IMDb around the discontinuity point (the movie release day, a much shorter time window).

To implement the difference-in-discontinuity estimator, we define the following quantities of differences:

$$\begin{aligned}
diff_score_{jt} &= twitter_sentiment_{jt} \\
&\quad -imdb_sentiment_{jt}, \\
diff_positive_proportion_{jt} &= twitter_positive_proportion_{jt} \\
&\quad -imdb_positive_proportion_{jt}, \\
diff_negative_proportion_{jt} &= twitter_negative_proportion_{jt} \\
&\quad -imdb_negative_proportion_{jt},
\end{aligned}$$

where $twitter\_sentiment_{jt}$ is the mean sentiment score on Twitter for movie j at time t; $imdb\_sentiment_{jt}$ is the mean sentiment score of IMDb comments for movie j at time t.

$twitter\_positive\_proportion_{jt}$ / $imdb\_positive\_proportion_{jt}$ are the proportion of extremely positive tweets/comments on Twitter/IMDb for movie j at time t, where extremely positive tweets/comments are defined as tweets or comments that have their raw sentiment score ≥ 0.8.[19]

$twitter\_negative\_proportion_{jt}$ / $imdb\_negative\_proportion_{jt}$ are the proportion of extremely negative tweets/comments on Twitter/IMDb for movie j at time t, where extremely negative tweets/comments are defined as tweets/comments that have their raw sentiment score ≤ 0.4. With these definitions, we specify the following regression model:

$$\begin{aligned} diff\_score_{jt} = \quad & \beta_1 \cdot post_{jt} + \sum_{p=0}^{\bar{p}} \beta_{2,p} \cdot duration_{jt}^{p} \\ & + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot post_{jt} \cdot duration_{jt}^{p} + \nu_j + \epsilon_{jt}, \end{aligned}$$

where the dependent variable is the mean sentiment score difference at time t as defined above, $post_{jt}$ is a dummy variable that takes the value 1 if time t is after movie j's release and 0 otherwise, and $duration_{jt}$ is the number of days after movie j's release, where a positive value means that day t is after the movie release, and vice versa. Note that we are interested in β_1, the coefficient of $post_{jt}$. We also construct two other regressions in which the dependent variables are the difference in the proportion of extremely positive comments at time t and the difference in the proportion of extremely

[19]Our results are robust using other score thresholds.

negative comments at time t, respectively.

These regression results are shown in Table 2.6, Table 2.7, and Table 2.8. We can see that the coefficients of *post* are significantly negative in all three regressions and robust to different model specifications, which means that there is a significant drop in the mean sentiment score difference, proportion of extremely positive Tweets/comments, and proportion of extremely negative Tweets/comments, respectively. These findings further strengthen the earlier regression discontinuity results on sentiment manipulation.

In the literature, little is known about the relative prevalence of positive and negative manipulation. On the one hand, a negative review hurts more than a positive review helps (Chevalier and Mayzlin 2006; Hu et al. 2011, 2012): if a negative tweet has a greater role in affecting consumer purchase, other things being equal, a movie studio will have a greater incentive to conduct negative manipulation than positive manipulation given the manipulation budget. The reason is that the marginal benefit of conducting negative manipulation is higher than conducting positive manipulation. However, on the other hand, the marginal cost of conducting negative manipulation is typically higher than conducting positive manipulation.[20] As

[20] Usually, movie studios will hire a "public relation" company to conduct positive manipulation, but "because giving negative comments usually brings side effects and has risks, few Internet public relations companies will take such orders. So in most cases, film companies will use staff which they have themselves trained for generating negative comments about opponents." Even if some public relation companies are willing to conduct negative manipulation, they would be more cautious. To avoid problems, they will often register new accounts from a foreign IP address. Training the internal staffs or registering new accounts from a foreign IP address involves additional cost. See http://www.globaltimes.cn/content/749267.shtml.

a result, both the benefit and the cost of conducting negative manipulation are higher than those of conducting positive manipulation. Therefore, both directions are theoretically plausible, and it is an empirical question to examine whether negative manipulation is more likely to occur than positive manipulation, and we specify the following hypothesis:

Hypothesis 5 *The drop in the proportion of extremely positive Tweets will be larger than that in the proportion of extremely negative Tweets.*

From the regressions in which the dependent variables are the difference in the proportion of extremely positive Tweet/comments and the difference in the proportion of extremely negative Tweet/comments, we can compare the magnitude of positive manipulation with that of negative manipulation. We find that the drop in the proportion of extremely positive Tweets/comments is much larger than the drop in the proportion of extremely negative Tweets/comments, which suggests that positive manipulation is more prevalent than negative manipulation in general. Therefore, Hypothesis 5 is supported.

Note that we combine two sources of variations, (i) before and after movie release; and (ii) the difference in sentiment score distribution across Twitter and IMDb, to implement our difference-in-discontinuity design (Grembi et al. 2016). The intuition for our difference-in-discontinuity is straightforward: the difference-in-discontinuity estimator focuses on the discontinuity of the difference in sentiment score distribution across Twitter and IMDb at the movie release time. If we use only the variation of the difference in sentiment score distribution across Twitter and IMDb, the sentiment score difference could be caused by the fact that the two platforms consist of users with different characteristics. However, in our regressions, we compare the sentiment score difference between Twitter and IMDb before and after the movie release. Even if Twitter users are systematically different from IMDb users, our estimation will still be unbiased as long as this systematic difference is time-invariant (i.e., the systematical difference does not change after the movie release).

We also examine how positive and negative manipulation is moderated by the intensity of competition. We specify the following regression model to quantify the effect of competition:

$$\begin{aligned} diff\_positive\_proportion_{jt} = &\beta_1 \cdot post_{jt} + \beta_2 \cdot post * [\pm 1 month \& sim > 0.7] \\ &+ \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot duration_{jt}^{p} \\ &+ \sum_{p=0}^{\bar{p}} \beta_{4,p} \cdot post_{jt} \cdot duration_{jt}^{p} + v_j + \epsilon_{jt}, \end{aligned}$$

where $diff\_positive\_proportion_{jt}$ is the difference in the proportion of extremely positive Tweets/comments, $[\pm 1 month \& sim > 0.7]$ is a measure of competition which counts the number of other movies that are released one month within the focal movie's release day, and with a thematic similarity greater than 0.7 compared with the focal movie. The regression model for the difference in the proportion of extremely negative Tweets/comments is specified similarly. The results of these regressions are shown in Table 2.9 and Table 2.10. We can see that, although movie studios facing a higher level of competition will conduct a higher level of both positive and negative sentiment manipulation, the magnitude is quite different. The effect of competition intensity on negative manipulation is much stronger than the effect on positive manipulation. Studying the role of positive and negative manipulation is not only theoretically important, it also has practical implications. Our result should be of interest to platform designers or policymakers investigating sentiment manipulation issues. Recently, more and more platforms are adopting anti-manipulation systems, such as Yelp's review filter. Potentially, a manipulation detection algorithm may work more precisely if we know the major task is to detect positive or negative manipulation (positive and negative manipulation may have different writing styles and other characteristics). Our results can inform platform designers or policymakers in a given market conditions about the specific type of sentiment manipulation they should pay close attention to (in general, positive manipulation is dominant, but negative manipulation becomes more relevant in a more competitive market environment) and also help them modify the detection algorithms in their tactical anti-manipulation systems (Bichler et al. 2010; Ketter et al. 2012). From our previous results we know that positive manipulation is more prevalent than negative manipulation in general. However, as the level of competition increases, the level of negative manipulation will increase more rapidly than the level of positive manipulation. This finding helps us understand the mechanism behind positive and negative manipulation, which has been overlooked in the prior literature.

2.5.2 Quarterly Earnings Announcement

Besides the movie release shock we have used in our regression discontinuity and difference-in-discontinuity analyses so far, we consider another source of exogenous shock, movie studios' quarterly earnings announcements, to further strengthen our results on manipulation. The literature in accounting and finance has shown that firms strategically disclose good news and withhold bad news prior to earnings announcements because of a range of incentives, including stock price boosting and the career concerns of managers (e.g. Kothari et al. 2009). For instance, managers may informally disclose or leak good news using selective private channels prior to quarterly earnings announcements. Extensive empirical evidence showed that managers care about quarterly performance and benchmark beating prior to quarterly earnings announcements (Graham et al. 2005), and they have incentives to engage in real earnings management to meet quarterly financial reporting benchmarks, such as earnings from the same quarter in the previous year (Cohen et al. 2010). The accounting literature has demonstrated that firms use various strategies to beat earnings benchmark prior to earnings announcement, such as cutting advertising expenses, boosting sales through accelerating their timing, and generating additional unsustainable sales through price discounts or more lenient credit terms (Cohen et al. 2010). Here we examine the possibility of a new strategy that has been underexplored in the literature: firm could increase sales by conducting sentiment manipulation to boost stock price or meet financial reporting benchmarks prior to quarterly earnings announcements.

In our specific context, movie studios can use sentiment manipulation as a strategy of real earnings management to increase movie box office revenue, and therefore boost stock price or meet quarterly financial reporting benchmarks. In the movie industry, major movie studios typically launch fewer than 20 movies per year, so the box office revenue of a single movie can have a major effect on the studio's stock price (Joshi and Hanssens 2009). Chen et al. (2012) showed that third party movie reviews have a significant impact on movie studios' stock performance. Therefore, movie studios have incentives to conduct sentiment manipulation prior to earnings announcement as a form of disclosing good news just like a typical strategic firm behavior documented in the prior accounting literature (Kothari et al. 2009). After earnings announcement, the manipulation incentives decrease dramatically.

It is worth noting that the date of quarterly earnings announcements is more exogenous than the movie release date because the timing of quarterly earnings announcements is largely fixed by the law in advance and is very unlikely to be affected by the performance of a single movie. In the case of before/after movie release, we may worry that Twitter users might post tweets in a different manner before and after movie release. However, it is less of a concern in the case of before/after quarterly earnings announcements: ordinary Twitter users are unlikely to behave differently before and after quarterly earnings announcements of movie studios.

We use the following regression equation to examine the impact of earnings announcement on Twitter sentiment.

$$sentiment_{jt} = \beta_1 \cdot post_{jt} + \sum_{p=0}^{\bar{p}} \beta_{2,p} \cdot duration_{jt}^{p} + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot post_{jt} \cdot duration_{jt}^{p}$$
$$+ week_t + holiday_t + \nu_j + \epsilon_{jt},$$

where $post_{jt}$ is a dummy variable that takes the value 1 if day t is on or after the movie studio's quarterly earnings announcement date, and 0 otherwise. We focus on a windows of one month before/after quarterly earnings announcement. The estimation results are presented in Table 2.11. We can see that there is a significant sentiment drop after quarterly earnings announcement, which suggests the existence of sentiment manipulation prior to earnings announcement.

2.5.3 Exogenous Weather Shocks

In this section we describe another source of exogenous shock—namely, severe weather conditions—to further alleviate concerns of potential confounding factors. Weather shocks are widely used as exogenous variations in the literature (Angrist and Krueger 2001; Miguel et al. 2004; Moretti 2011; Qiu, Tang, and Whinston 2015). Moretti (2011) showed that severe weather on movie release days can significantly reduce people's willingness to go out to watch movies. Based on negative severe weather shocks on movie box office revenue, we conduct an additional analysis to examine the impact of weather shocks on movie release days on the level of Twitter sentiment. Compared with our original movie release shocks, the weather shocks on movie release days create a more exogenous variation. The intuition of our identification is as follows: when the weather on a movie release day is severe (cold, rain, or snow), the number of potential customers who are interested in watching a movie is smaller (i.e. the user base that can be affected by sentiment manipulation is smaller). Therefore, sentiment manipulation is less effective under severe weather shocks, and movie studios would have less incentives to conduct manipulation. As a result, empirically, we should observe a smaller drop in the mean sentiment score when the weather on the movie release day is more severe (because movie studios are less likely to conduct manipulation just before the release).

Note that in our context, it is a challenge to construct U.S. nationally-aggregated weather measures. Moretti (2011) used the weather conditions in seven large cities to construct U.S. nationally-aggregated weather measures. We combine different data sources (the zip code level U.S. Census data and the weather data from Weather Underground[21] and construct more precise nationally-aggregated weather measures. For each zip code with movie theater establishments (based on the zip code level census data), we match it to the closest weather station (based on the weather data from Weather Underground). Then, for each weather station, we calculate the percentage of movie theaters in that station among all theaters in the U.S. We focus on minimum temperature, precipitation, snowfall, to characterize three severe weather events on movie release days: a cold day, a heavy

[21] http://www.wunderground.com/

rainy day, and a heavy snow day, and construct three nationally-aggregated weather measures as follows: (1) we define a movie release day to be a cold day for a weather station if the minimum temperature on that day is below 0°C (32°F); (2) define a movie release day to be a heavy rain day/a heavy snow day for a zip code region if the precipitation rate is greater than 4 mm per hour/4 cm per hour;[22] (3) calculate the percentage of movie theaters in the U.S. experiencing three severe weather events on that day (cold weather, heavy rain, or heavy snow), respectively.

In order to account for seasonal and other time related factors, we also control for weekly dummies and holiday dummies (Martin Luther King Day, President's Day, Veteran's Day, Thanksgiving Day, Christmas Eve, Christmas, New Year's Eve, New Year's Day, Halloween, St. Patrick's Day in the following regression equation:

$$\begin{aligned} sentiment_{jt} = \quad & \beta_1 \cdot post_{jt} + \beta_2 (post_{jt} \cdot cold_j) + \beta_3 (post_{jt} \cdot rain_j) \\ & + \beta_4 (post_{jt} \cdot snow_j) + \sum_{p=0}^{\bar{p}} \beta_{5,p} \cdot duration_{jt}^{p} \\ & + \sum_{p=0}^{\bar{p}} \beta_{6,p} \cdot post_{jt} \cdot duration_{jt}^{p} + week_t + holiday_t + \nu_j + \epsilon_{jt}, \end{aligned}$$

where $cold_j$, $rain_j$, and $snow_j$ are the percentage of movie theaters in the U.S. experiencing a cold/rain/snow day on movie j's release day, respectively. $week_t$ represents weekly time dummies, and $holiday_t$ represents holiday dummies. The estimation results are presented in Table 2.12. We find

[22] http://www.metoffice.gov.uk/media/pdf/4/1/No._03_-_Water_in_the_Atmosphere.pdf.

that the coefficients β_2, β_3, and β_4 are significantly positive, indicating a smaller drop in the mean sentiment score when more theaters on the movie release day experienced severe weather. These findings are consistent with the expectation that movie studios might have less incentive to conduct manipulation under severe weather conditions, and they further suggest that the sentiment drops likely come from sentiment manipulation.

2.5.4 Demand-Side Expectation

It is possible that the demand-side expectation—namely, the consumer's expectation of the movie quality—may correlate with Twitter sentiment, and the expectation can be easily influenced by movie trailers or other advertising campaigns. Therefore, in addition to controlling for movies' daily advertising budget using a Nielsen data set, we include several demand-side measures to further control for any differences in sentiment that might have come from consumers' expectation-realization gap. The first additional control variable is the Hollywood Stock Exchange (HSE) price. HSE is a public prediction market such that people can trade on movies' box office revenues. The daily Hollywood Stock Exchange price acts as box office predictions from the consumer side, and it can reflect any potential expectation-realization gap (Foutz and Jank 2010). For example, before watching a movie, a user may think highly of the movie and buy shares on that movie through Hollywood Stock Exchange. However, after watching the movie, she may realize that it is not as good as what she expected, so she could sell her shares. The changes in the movie's share price listed on HSE can thus reflect consumers' expectation of the movie. The second additional control is the daily Google Trends score of a movie,[23] which also captures the consumer expectation of the movie. Finally, we also control for several other variables that may reflect or affect the expectation-realization gap, such as the IMDb rating of the movie, weekly dummies, and holiday dummies. The estimation results are robust and are presented in Table 2.13.

2.6 Robustness Checks

2.6.1 Intention, Pre-Consumption, and Promotional Tweets

One concern with using hashtags to identify all tweets related to a movie is that some tweets are "pre-watch" tweets—namely, those that are posted before the user watched the movie—whereas others are "post-watch" tweets. A potential issue with this setup is that pre-watch sentiment will be influenced largely by advertising and word of mouth effects, while the post-watch sentiment will more accurately reflect the true perceived quality of the movie. In the pre-release time period, almost all tweets reflect pre-watch sentiment, but in the post-release time period, there will be a combination of pre-watch tweets and post-watch tweets. This difference between the pre- and post-release time periods may bias our estimation.[24] Nevertheless, we argue that our cross-platform difference-in-discontinuity analysis can help mitigate this concern. This is because the distinction between

[23]https://www.google.com/trends/https://www.google.com/trends/
[24]We thank an anonymous reviewer for pointing out this issue.

pre-release (mostly pre-watch) and post-release (combination of pre-watch and post-watch) sentiment is present on both Twitter and IMDb. Using the difference-in-discontinuity approach, the impact of the pre-release and post-release distinction will be canceled out. Since we still observe significant sentiment drop in our difference-in-discontinuity analysis, our results suggest that, while the pre-release and post-release distinction might play a role, the evidence of sentiment manipulation remains significant.

To further alleviate this concern, we use two different approaches to control for the percentage of tweets posted by users who have not watched the movie (pre-consumption tweets). The first approach is to split all tweets into intention tweets and those that are not intention tweets. Intention tweets are those whose authors clearly expressed their willingness to watch the movie in the future, such as "I wanna see [movie name]!" Following Rui et al. (2013), we used an intention lexicon to extract features from tweets and then used a support vector machine to construct the intention classifier. Then, we compute the percentage of intention tweets on a given day as an approximate measure of the percentage of tweets posted by users who have not watched the movie, and control for it in our regression. As shown in Table 2.14, the sentiment drop result is robust. The other approach is simple and heuristic. For each tweet, we first count the number of verbs that are present tense, and also the number of verbs that are past tense. If the number of past tense verbs > the number of present tense verbs, we label it as "After watching"; if the number of present tense verbs > the number of past tense verbs, we label it as "Before watching". Then, for each time period, we compute the percentage of pre-consumption tweets. The result of this regression is shown in Table 2.15, and we still observe a significant drop in sentiment on the movie release day.

Another concern with using hashtags to identify movie tweets is that some of these tweets might be posted by movie studios or producers to promote the movie. Such firm-side promotion/marketing efforts on Twitter might bias our analysis. We argue that our difference-in-discontinuity approach can to some extent alleviate this issue if we assume Twitter and IMDb users to be equally exposed to the firm-side promotion and marketing efforts on Twitter. If the assumption is valid, then the effect of firm-side promotion will be canceled out when we compare the difference in sentiment drop between Twitter and IMDb. To further alleviate the concern of the firm-side promotion, we conduct two additional analyses as described below.

First, using the trackback permanent tweet link, we can differentiate between the tweets posted by official Twitter accounts of movie studios and individual accounts. Given the large amount of tweets, the proportion of tweets posted by official movie studios is small. After we remove all tweets posted by official Twitter accounts of movie studios, we still observe a significant drop in tweet sentiment. The estimation results are robust and are presented in Table 2.16.

Second, we include the number of ad tweets, Google trends score of each movie, and the IMDb rating of each movie as control variables to control for movie studios' marketing efforts. In Table 2.17, the variable number_ad_tweets is the number of tweets posted by the official Twitter accounts of movie studios, which can reflect the magnitude of positive influence. The variable Google_trends is the score of google trends for a particular movie, and the variable IMDb_rating is the user rating on IMDb for a given movie. We can see that the coefficient of *post* is significantly negative across different model specifications, which suggests that our result is robust.

2.6.2 Regression Results with Cluster Robust Variance

We conduct a robustness check where we take into account the potential heteroskedasticity and clustering among observations, since failure to address heteroskedasticity or clustering would lead to misleadingly small standard errors. Lee and Lemieux (2010) suggested that heteroskedasticity-robust standard errors (White 1980) should be used instead of the usual least squares standard errors. In addition, both Cameron et al. (2008) and Cameron and Miller (2015) detailed a cluster-robust standard inference procedure to control for within-cluster correlation. The regression results with cluster robust variance are shown in Table A.4. We can see that, although the t-statistics are smaller than those appearing in Table 2.2, the coefficient of the *post* variable is still significantly negative. Therefore, controlling for potential within-cluster correlation, we still observe a significant drop in sentiment after the release day, which suggests the existence of manipulation.

2.6.3 Additional Movie Characteristics

We also conduct an additional regression analysis where we control for movie specific characteristics such as the length of the movie, the genre of the movie, whether the movie is a color movie or not, the number of tweets associated with the movie, and whether or not the movie is a major studio production. We construct an alternative regression model where we include movie-specific characteristics in a pooled regression:

$$\begin{aligned} sentiment_{jt} \quad = \quad & \beta_1 \cdot post_{jt} + \sum_{p=0}^{\bar{p}} \beta_{2,p} \cdot duration_{jt}^{p} + \sum_{p=0}^{\bar{p}} \beta_{3,p} \cdot post_{jt} \cdot duration_{jt}^{p} \\ & + \beta_4 \cdot color_j + \beta_5 \cdot runtime_j + \beta_6 \cdot majorStudio_j \\ & + \beta_7 \cdot numTweets_j + \beta_8 \cdot genre_j + \epsilon_{jt}, \end{aligned} \tag{2.5}$$

where $color_j$ is 1 if movie j is a color movie, and 0 if the movie is in black and white; $runtime_j$ is the length of movie j in minutes; $majorStudio_j$ is 1 if movie j is produced by a major studio, as listed in Table A.5; $numTweets_j$ is the number of tweets associated with movie j during the ±60 day observation window; and $genre_j$ is a vector of dummy variables each corresponding to a specific movie genre. Notice that this is not a fixed effect model since we are now examining the effects of time-invariant movie characteristics variables.

The results of this alternative regression model are shown in Table 2.18. We can see that, with the inclusion of movie characteristics variables as covariates, the coefficient of the *post* variable is still significant. In other words, the regression discontinuity near the release day is significant in this alternative regression specification. This is expected because the inclusion of baseline covariates should not affect the estimated discontinuity (Lee and Lemieux 2010).

2.6.4 Controlling for Movie Advertising Expenditure

Our illustrative example suggests that, in addition to any potential sentiment manipulation, movie studios might also behave differently prior to and after movie releases. In fact, several marketing studies have documented an interesting phenomenon—that close to 90% of movie advertising spending happens prior to movie release (Elberse and Anand 2007, Chintagunta et al. 2010, Rennhoff and Wilbur 2011). We suspect that movie sentiments and the observed discontinuity might also be affected by movie studios' advertising strategy. More specifically, the drop in movie sentiment might have come from the drop in advertising spending. Therefore, we conduct two additional sets of regression analysis where we control for movie studios' daily advertising expenditure, as well as ads spending across different advertising channels, both with movie fixed effects. The results, shown in Table A.7 and Table A.8, indicate that the coefficients of the *post* variable remain significantly negative after controlling for advertising expenditure. These analyses suggest that the observed sentiment drops are not merely a result of movie studios' advertising efforts, and support our argument of sentiment manipulation.

2.6.5 Subsample Analysis

In addition to the full sample analysis, we also identify important subsamples and check whether we still observe significant sentiment drops within these subsamples. More specifically, we generate subsamples of (1) low budget movies–that is, movies with a production budget less than 3 million US dollars, and (2) major studio movies. Then, we conduct regression analyses similar to those in the full sample case, and the results are similar—the coefficients associated with the *post* variable are significantly negative across different subsamples. The results, shown in Table A.6, and detailed discussions can be found in the appendix.

2.6.6 Bandwidth Choice

The choice of bandwidth—that is, the length of the observation window for each movie—can affect the regression results: the longer the bandwidth, the more data will be available; the shorter the bandwidth, the more precise the estimates will be. Therefore, in addition to our initial choice of ±60 days, we also use ±30 and ±50 days as the observation window. Then, we conduct regression analysis based on these different observation windows, and the results are robust to different bandwidth choices—the coefficients associated with the *post* variable remain significantly negative. The results, shown in Table A.9, as well as more discussions are provided in the appendix.

2.6.7 Alternative Explanations

There are several alternative explanations for the observed sentiment drop on the release day. We discuss them in this section and explain why sentiment manipulation is the most plausible explanation. First of all, it is possible that a sentiment drop is observed simply because movie goers did not like the movie. While this could be the case for some movies, it does not seem reasonable for consumers to consistently dislike movies they see. In addition, this alternative explanation cannot explain why people consistently dislike low budget movies, non-major studio movies, and movies that face strong competition, more than other movies. Moreover, recall that in our difference-in-discontinuity analysis, even if movie goers did not like the movie, this effect would have been differenced out because the disappointment should have affected both Twitter and IMDb users. The fact that our difference-in-discontinuity analysis still shows a sentiment drop provides further evidence consistent with sentiment manipulation. Therefore, we rule out this alternative explanation.

Another alternative explanation for the sentiment drop is the movie studio's budget constraint. Specifically, the reason low-budget movies exhibit larger sentiment drop might be because they cannot afford to manipulate as much as movies with higher budgets. In contrast, major studio movies might be able to allocate more budget in manipulation. However, recall that in our regression models, we have included a movie-specific fixed effect, which controls all movie-specific, time-invariant variables, including movie budgets. The coefficients of the post variable in these regression models are still significant. Therefore, budget constraint does not appear to be a convincing explanation.

It is also possible that the pre-release sentiment is higher than post-release sentiment because some eager fans tend to leave comments before movie release, or, more generally, users who post tweets before movie release might be very different from users who leave tweets after movie release. However, the potential influence of the difference in user characteristics (before and after movie release) on the sentiment drop cannot explain why low-budget movies experience a larger sentiment drop than higher budget movies, and why movies facing more competition exhibit a larger sentiment drop than movies with less competition. Furthermore, the difference in user composition before and after movie release will be canceled out in our difference-in-discontinuity analysis, yet we still observe significant sentiment drops. Also recall that we use severe weather and earning announcement as additional sources of exogenous shocks in our RD analyses, and in those cases there is unlikely to be significant difference in user composition before or after the shock. The fact that we observed significant sentiment drops in these analyses suggests the existence of sentiment manipulation.

In summary, although several alternative explanations exist regarding why a sentiment drop after movie release is observed, sentiment manipulation is the only one that could consistently explain all empirical results. Therefore, we believe that sentiment drop is best explained by movie studios' strategic manipulation.

2.7 Discussion

Platforms that adopt the review/recommendation business model, such as Yelp, Expedia, and TripAdvisor, often provide contents for free, and rely heavily on advertisements from sellers to sustain their business operations. However, payments from producers are usually received in exchange of certain forms of advertisements on the platform. Although the platforms themselves should theoretically serve as impartial hosts that only aggregate and organize the tremendous amount of mostly user generated information, there is nonetheless a constant battle for them to balance the tradeoff between their information credibility and the relationship with advertisers, since consumers evaluate the platforms based on information accuracy and their satisfaction with recommendations. There have been several instances where certain suspicious reviews on Yelp stirred controversy in defining the legal boundary between internet free-speech and outright defamation, and in deciding whether a given comment is fraudulent or not (Loten 2014, Streitfeld 2014). Researchers in the marketing and economics community have also studied how structural differences in platform design between Expedia and TripAdvisor result in significant differences in the rating distributions in the hotel industry (Mayzlin et al. 2014), and how a business' decision to commit review fraud on Yelp corresponds to reputation concerns and competitions (Luca and Zervas 2015).

Not only are review platforms plagued with manipulation, social media also suffer from the pervasiveness of fraudulent information and strategic behaviors. For example, it has been revealed that Facebook has a black market issue by which user engagements in the form of "likes" can be purchased, known as Facebook Fraud (Leonard 2014). Twitter's advertising and sentiment aggregation efficacy is affected by the prevalence of spam tweets and robotic programs (Coy 2013). Although the popularity of Twitter attracts advertisers to pay to increase their Twitter presence, malicious individuals or even competitors could write simple programs to achieve the same level of advertising, which would sabotage legitimate advertising campaigns and contaminate sentiments. From the advertising firm's perspective, engaging in illegitimate activities such as deploying spam bots and leaving opinion spam messages may very well turn out to be beneficial for its business operation. The reason for this is that firms strive to increase their influence over consumers. Traditional ads serve as a medium for raising product awareness and, subsequently, affecting consumer's purchasing behaviors. Advertising on Twitter essentially follows this principle of raising awareness and encouraging purchases. Users are generally able to recognize the firm's intention to advertise, and tend to adjust downward the potentially biased and promotional messages. In contrast, spamming is another, albeit illegitimate and somewhat morally questionable, strategy for raising awareness and inducing purchases, as has been widely deployed in the form of e-mail spam. Moreover, sophisticated spammers often assume multiple aliases as a camouflage, which makes it even more difficult for users to recognize these messages' spamming intent, and our empirical results show that the observed Twitter sentiment exhibits patterns consistent with sentiment manipulation. These are questions for opinion platforms to consider, and should also alert the consumers so they can reevaluate the degree to which they could rely on social media information for recommendation and decision making.

2.8 Conclusion

In the present study, we conducted a series of empirical analyses to examine the existence of sentiment manipulation. Using movie tweets data, our RDD results uncovered unusual patterns in Twitter sentiment near movie release days, including larger sentiment drop for low-budget movies, non-major studio movies, and movies facing a higher level of competition. These results prompted us to suspect the existence of sentiment manipulation. We further validated our results through a cross-platform identification strategy, where we used a difference-in-discontinuity approach to compare Twitter tweets and IMDb comments. Our difference-in-discontinuity results were consistent with the RDD results, and we further found that the magnitude of positive manipulation is larger than that of negative manipulation, but the magnitude of negative manipulation increases with the rise in competition level. Additional analyses using earnings announcement and severe weather conditions as sources of exogenous shocks, as well as controlling for demand-side expectations through Hollywood Stock Exchange and Google Trends all suggested that the sentiment drop is likely a result of strategic sentiment manipulation.

The current research is not without limitations: we were unable to directly identify tweets or comments that were manufactured by strategic parties. Therefore, the exact magnitude of positive and negative manipulation remains an open question. In addition, since we only studied movie data, the extent to which sentiment manipulation is used in other industries is still unclear. Despite these limitations, we believe this research contributes to the literature by uncovering sentiment anomaly through a RDD and difference-in-discontinuity approach, as well as constructing a movie competition measure through topic modeling. We believe that online platforms and opinion forums must address the issue of manipulation through better system designs and enhanced verification to ensure the reliability of reviews, comments, and sentiment analysis.

Table 2.3: RD Estimates of the Effect of Movie Release on Twitter Sentiment with Different Budgets (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0352*** [-28.29]	-0.0340*** [-26.29]	-0.0328*** [-24.51]	-0.0330*** [-23.91]
$post*lowBudget$	-0.0184*** [-12.14]	-0.0176*** [-18.26]	-0.0119*** [-18.61]	-0.0116*** [-16.79]
$duration$		9.14e-05*** [5.705]	0.000137*** [5.615]	0.000147*** [4.773]
$duration^2$			-1.08e-07** [-2.213]	-1.77e-07 [-1.291]
$duration^3$				6.85e-11 [0.547]
$post*duration$		-1.36e-05** [-1.993]	-4.43e-05*** [-3.711]	-4.04e-05** [-2.219]
$post*duration^2$			1.62e-07*** [3.126]	1.11e-07 [0.604]
$post*duration^3$				1.12e-10 [0.288]
$constant$	0.642*** [958.2]	0.639*** [736.0]	0.637*** [628.3]	0.637*** [581.1]
Observations	52,829	52,829	52,829	52,829
AIC	-84834.44	-84868.95	-84879.36	-84879.76
BIC	-84807.82	-84824.58	-84817.24	-84817.64

Table 2.4: RD Estimates of the Effect of Movie Release on Twitter Sentiment with Different Studio Types (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0359*** [-34.17]	-0.0350*** [-32.27]	-0.0341*** [-30.35]	-0.0341*** [-29.59]
$post * major\_studio$	0.0254*** [3.439]	0.0247*** [3.612]	0.0229*** [3.170]	0.0236*** [3.448]
duration		9.17e-05*** [5.724]	0.000138*** [5.631]	0.000147*** [4.779]
$duration^2$			-1.08e-07** [-2.221]	-1.76e-07 [-1.282]
$duration^3$				6.69e-11 [0.534]
$post * duration$		-1.38e-05** [-2.025]	-4.43e-05*** [-3.729]	-4.05e-05** [-2.234]
$post * duration^2$			1.62e-07*** [3.128]	1.11e-07 [0.606]
$post * duration^3$				1.11e-10 [0.287]
constant	0.642*** [958.3]	0.639*** [736.0]	0.637*** [628.4]	0.637*** [581.3]
Observations	52,829	52,829	52,829	52,829
AIC	-84836.51	-84871.43	-84881.89	-84882.27
BIC	-84809.89	-84827.05	-84819.77	-84820.15

Table 2.5: The Effect of Competition on Sentiment Manipulation, where competition is measured as the number of movies released 1 month within the focal movie and with thematic similarity > 0.7 (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
$post$	-0.0221*** [-13.94]	-0.0238*** [-15.48]	-0.0245** [-12.79]
$post * [\pm 1 month \& sim > 0.7]$	-0.0104*** [-5.763]	-0.0106*** [-5.760]	-0.0107*** [-5.693]
$duration$		0.000155*** [4.182]	0.000317*** [4.393]
$duration^2$			-3.59e-07* [-1.940]
$post * duration$		-0.000253*** [-3.722]	-0.000544*** [-6.390]
$post * duration^2$			1.17e-06*** [4.990]
$color_i$	0.0231*** [6.209]	0.0227*** [6.108]	0.0226*** [6.056]
$runtime_i$	5.13e-05 [1.622]	5.61e-05* [1.776]	5.65e-05* [1.789]
$majorStudio_i$	-0.0195*** [-14.14]	-0.0196*** [-14.19]	-0.0196*** [-14.24]
$numTweets_i$	-6.82e-08*** [-9.359]	-7.33e-08*** [-10.28]	-7.45e-08*** [-10.48]
genre dummies	YES	YES	YES
$constant$	0.623*** [103.1]	0.618*** [101.1]	0.614*** [97.32]
Observations	48,139	48,139	48,139
AIC	-68106.3	-68170.4	-68224.78
BIC	-67869.19	-67915.73	-67952.54

Table 2.6: Difference-in-Discontinuity Estimates of the Effect of Movie Release on Twitter Sentiment: Mean Sentiment Score Difference (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0365*** [-21.08]	-0.0356*** [-20.22]	-0.0349*** [-16.37]	-0.0345*** [-15.62]
duration		4.98e-05*** [3.235]	0.000187*** [2.934]	0.000174*** [3.672]
$duration^2$			-1.75e-07** [-2.148]	-1.83e-07 [-1.226]
$duration^3$				8.57e-11 [0.914]
$post * duration$		-2.84e-05** [-2.068]	-2.17e-05*** [-3.224]	-3.88e-05*** [-3.767]
$post * duration^2$			1.78e-07*** [3.217]	1.66e-07 [0.954]
$post * duration^3$				1.82e-10 [0.872]

Table 2.7: Difference-in-Discontinuity Estimates of the Effect of Movie Release on Twitter Sentiment: Difference in Proportion of Extremely Positive Tweets/Comments (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0742*** [-11.24]	-0.0638*** [-8.432]	-0.0542*** [-7.562]	-0.0521*** [-6.342]
duration		0.000357*** [4.534]	0.000265*** [3.337]	0.000224*** [3.216]
$duration^2$			-3.27e-07** [-2.014]	-2.14e-07 [-1.356]
$duration^3$				5.33e-11 [0.875]
$post * duration$		-0.000652** [-2.134]	-0.00235*** [-3.876]	-0.00174*** [-3.145]
$post * duration^2$			1.46e-07*** [3.546]	1.15e-07 [0.882]
$post * duration^3$				1.34e-10 [0.561]

Table 2.8: Difference-in-Discontinuity Estimates of the Effect of Movie Release on Twitter Sentiment: Difference in Proportion of Extremely Negative Tweets/Comments (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0104*** [-5.354]	-0.00914*** [-6.472]	-0.00876*** [-4.372]	-0.00884*** [-4.547]
$duration$		0.000145*** [3.459]	0.000115*** [2.879]	0.000153*** [3.542]
$duration^2$			-1.31e-07** [-2.023]	-1.69e-07 [-1.341]
$duration^3$				4.33e-11 [0.525]
$post * duration$		-1.28e-05** [-2.109]	-2.33e-05*** [-3.654]	-4.01e-05*** [-3.954]
$post * duration^2$			1.21e-07*** [3.036]	1.28e-07 [0.672]
$post * duration^3$				1.16e-10 [0.552]

Table 2.9: Difference-in-Discontinuity Estimates of the Effect of Competition: Difference in Proportion of Extremely Positive Tweets/Comments (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0635*** [-9.885]	-0.0567*** [-7.523]	-0.0525*** [-6.343]	-0.0569*** [-5.258]
$post*[\pm 1 month \& sim > 0.7]$	-0.00151*** [-4.524]	-0.00142*** [-4.216]	-0.00138*** [-3.028]	-0.00163*** [-3.689]
$duration$		0.000249*** [4.126]	0.00158*** [3.028]	0.000287*** [3.135]
$duration^2$			-5.66e-07** [-2.132]	-3.87e-07 [-1.426]
$duration^3$				6.13e-11 [0.924]
$post*duration$		-0.000352** [-2.022]	-0.00256*** [-3.241]	-0.00127*** [-3.342]
$post*duration^2$			1.87e-07*** [3.877]	1.67e-07 [1.231]
$post*duration^3$				1.55e-10 [0.656]

Table 2.10: Difference-in-Discontinuity Estimates of the Effect of Competition: Difference in Proportion of Extremely Negative Tweets/Comments (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.00423*** [-5.026]	-0.00417*** [-6.288]	-0.00398*** [-4.133]	-0.00386*** [-4.265]
$post *$ [$\pm 1 month \& sim >$ 0.7]	-0.00325*** [-4.226]	-0.00331*** [-3.837]	-0.00317*** [-3.945]	-0.00348*** [-3.662]
$duration$		0.000131*** [2.872]	0.000137*** [2.945]	0.000121*** [3.113]
$duration^2$			-1.82e-07** [-2.112]	-1.57e-07 [-1.235]
$duration^3$				4.86e-11 [0.372]
$post * duration$		-1.55e-05** [-2.082]	-2.76e-05*** [-3.733]	-3.12e-05*** [-3.623]
$post * duration^2$			1.89e-07*** [3.255]	1.43e-07 [0.887]
$post * duration^3$				1.34e-10 [0.662]

Table 2.11: RD Estimates of the Effect of Quarterly Earnings Announcement on Twitter Sentiment (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0654*** [-11.33]	-0.0673*** [-12.25]	-0.0633*** [-11.56]	-0.0625*** [-10.28]
duration		3.24e-05*** [3.867]	0.000188*** [3.562]	0.000194*** [3.235]
$duration^2$			-2.69e-07** [-1.534]	-3.42e-07 [-1.233]
$duration^3$				4.62e-11 [1.428]
$post * duration$		-2.35e-05 [-0.235]	-2.34e-05*** [-1.255]	-5.57e-05*** [-1.239]
$post * duration^2$			2.54e-07*** [2.886]	2.87e-07 [1.247]
$post * duration^3$				2.29e-10 [1.280]
Week Dummies	YES	YES	YES	YES
Holiday Dummies	YES	YES	YES	YES

Table 2.12: RD Estimates of the Effect of Movie Release on Twitter Sentiment: Exogenous Weather Shocks (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0427*** [-16.54]	-0.0435*** [-17.38]	-0.0401*** [-15.74]	-0.0418*** [-19.82]
$duration$		7.35e-05*** [3.249]	0.000139*** [2.925]	0.000163*** [3.118]
$duration^2$			-1.34e-07** [-2.105]	-1.89e-07 [-1.245]
$duration^3$				8.53e-11 [1.338]
$post * duration$		-1.74e-05 [-0.952]	-4.89e-05*** [-3.287]	-3.89e-05 [-1.538]
$post * duration^2$			1.78e-07*** [2.953]	1.67e-07 [1.335]
$post * duration^3$				1.86e-10 [0.873]
$post * cold$	0.00936*** [3.017]	0.00925*** [2.874]	0.00874*** [2.965]	0.00895*** [2.814]
$post * rain$	0.0127*** [3.642]	0.0115*** [3.542]	0.0138*** [3.764]	0.0126*** [3.633]
$post * snow$	0.0135*** [3.826]	0.0129*** [3.336]	0.0119*** [3.238]	0.0114*** [3.037]
Week Dummies	YES	YES	YES	YES
Holiday Dummies	YES	YES	YES	YES

Table 2.13: RD Estimates of the Effect of Movie Release on Twitter Sentiment: Control for Demand-Side Expectations (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0321*** [-16.58]	-0.0308*** [-15.33]	-0.0302*** [-14.39]	-0.0335*** [-18.64]
duration		3.25e-05*** [2.854]	0.000157*** [2.975]	0.000167*** [3.243]
$duration^2$			-1.43e-07** [-2.108]	-1.34e-07 [-0.866]
$duration^3$				6.84e-11 [1.206]
$post * duration$		-1.59e-05 [-0.945]	-4.53e-05*** [-2.833]	-3.76e-05 [-1.338]
$post * duration^2$			1.43e-07 [1.442]	1.636e-07 [1.352]
$post * duration^3$				1.67e-10 [1.455]
HSE_price	0.00486 [0.687]	0.00424 [0.823]	0.00354 [0.745]	0.00387 [0.649]
Google_trends	0.000145 [1.254]	0.000167 [1.137]	0.000154 [1.134]	0.000138 [1.207]
IMDb_rating	0.0413** [2.022]	0.0426** [2.128]	0.0445 [1.532]	0.0462 [1.447]
Week Dummies	YES	YES	YES	YES
Holiday Dummies	YES	YES	YES	YES

Table 2.14: RD Estimates of the Effect of Movie Release on Twitter Sentiment: Control for the Percentage of Intention Tweets (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0314*** [-20.24]	-0.0308*** [-21.33]	-0.0287*** [-18.84]	-0.0294*** [-19.62]
$duration$		8.38e-05*** [3.226]	0.000141*** [2.867]	0.000158*** [3.113]
$duration^2$			-1.28e-07** [-2.142]	-1.87e-07 [-0.933]
$duration^3$				6.88e-11 [0.926]
$post * duration$		-1.67e-05** [-2.089]	-4.89e-05*** [-2.753]	-3.66e-05*** [-3.165]
$post * duration^2$			1.67e-07*** [2.903]	1.54e-07 [0.881]
$post * duration^3$				1.78e-10 [0.824]
Percentage_intention	0.000268** [2.133]	0.000272** [2.056]	0.000233** [2.027]	0.000254** [2.105]

Table 2.15: RD Estimates of the Effect of Movie Release on Twitter Sentiment: Control for the Percentage of Pre-Consumption Tweets (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0338*** [-22.48]	-0.0327*** [-23.48]	-0.0322*** [-16.58]	-0.0318*** [-22.14]
$duration$		8.07e-05*** [3.334]	0.000128*** [2.725]	0.000142*** [3.145]
$duration^2$			-1.05e-07** [-2.122]	-1.49e-07 [-0.834]
$duration^3$				6.47e-11 [0.874]
$post * duration$		-1.59e-05** [-2.125]	-4.33e-05*** [-2.945]	-3.28e-05*** [-3.345]
$post * duration^2$			1.43e-07*** [2.863]	1.05e-07 [0.756]
$post * duration^3$				1.19e-10 [0.533]
Percentage_Pre-Consumption	0.000174 [1.324]	0.000163 [1.209]	0.000171 [1.335]	0.000168 [1.403]

Table 2.16: RD Estimates of the Effect of Movie Release on Twitter Sentiment of Individual Accounts (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0324*** [-24.56]	-0.0312*** [-24.62]	-0.0308*** [-18.42]	-0.0316*** [-20.09]
$duration$		4.25e-05*** [3.543]	0.000114*** [2.824]	0.0001423*** [3.365]
$duration^2$			-1.03e-07** [-2.015]	-1.28e-07 [-0.965]
$duration^3$				6.17e-11 [0.523]
$post * duration$		-2.35e-05** [-2.204]	-2.84e-05*** [-2.865]	-3.218e-05*** [-3.207]
$post * duration^2$			1.28e-07*** [2.925]	1.09e-07 [0.842]
$post * duration^3$				1.12e-10 [0.228]

Table 2.17: RD Estimates of the Effect of Movie Release on Twitter Sentiment: Control for the Influence of Marketing Efforts of Movie Studios (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0326*** [-21.04]	-0.0317*** [-18.42]	-0.0309*** [-15.74]	-0.0311*** [-22.14]
$duration$		6.72e-05*** [3.048]	0.000124*** [2.835]	0.000137*** [3.058]
$duration^2$			-1.01e-07** [-2.033]	-1.59e-07 [-0.947]
$duration^3$				6.31e-11 [1.148]
$post * duration$		-1.38e-05 [-0.842]	-4.24e-05*** [-2.837]	-3.34e-05 [-1.245]
$post * duration^2$			1.35e-07*** [2.842]	1.19e-07 [0.924]
$post * duration^3$				1.32e-10 [0.567]
number_ad_tweets	0.00987 [1.225]	0.00642 [1.128]	0.00635 [1.107]	0.00627 [1.054]
Google_trends	0.000124 [0.854]	0.000118 [0.622]	0.000109 [0.604]	0.000107 [0.598]
IMDb_rating	0.0624** [2.214]	0.0615** [2.168]	0.0611** [2.149]	0.0608** [2.137]

Table 2.18: Robustness Checks of RD Estimates: Including Other Covariates (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
$post$	-0.0350*** [-33.13]	-0.0347*** [-30.46]	-0.0344*** [-30.38]
$duration$		8.11e-05*** [2.764]	8.77e-05** [2.350]
$duration^2$			-1.05e-08 [-0.103]
$post * duration$		3.38e-06 [0.283]	-7.23e-06 [-0.555]
$post * duration^2$			6.70e-08 [1.072]
$color_i$	0.0225*** [6.025]	0.0224*** [5.986]	0.0224*** [5.982]
$runtime_i$	9.54e-05*** [3.227]	9.71e-05*** [3.282]	9.75e-05*** [3.296]
$majorStudio_i$	-0.0197*** [-14.43]	-0.0198*** [-14.45]	-0.0198*** [-14.48]
$numTweets_i$	-8.35e-08*** [-11.56]	-8.63e-08*** [-12.28]	-8.76e-08*** [-12.40]
genre dummies	YES	YES	YES
$constant$	0.612*** [113.5]	0.609*** [111.9]	0.609*** [111.4]
Observations	51,829	51,829	51,829
AIC	-72105.71	-72123.48	-72121.15
BIC	-71875.47	-71875.52	-71855.48

Chapter 3

Word-of-Mouth in Social Media Advertising

3.1 Introduction

Social media platforms have become a prominent advertising channel for firms to reach out to their customers.[1][2][3] A unique feature of social media platforms such as Facebook is the information about users' social graph which allows the platform to target users and spread content using their social connections. Firms can organically spread content for their brand or products on these platforms based on the user interconnections. However, in order to monetize these social networks, the platforms are increas-

[1] http://marketingland.com/facebook-accounted-for-75-of-social-ad-spending-globally-in-2014-123911

[2] http://www.emarketer.com/Article/Social-Network-Ad-Spending-Hit-2368-Billion-Worldwide-2015/1012357

[3] Social media ad spending is expected to reach $23.68 billion and represents 13.9% of total digital ad spending.

ingly pushing ads as a mechanism to show content to users.[4] As a result, firms have started promoting content using ads on these platforms to create awareness about their brand and to enable installs or purchases of their products or mobile applications. Among others, Facebook has been the most successful social media platform that allows firms to conduct advertising campaigns. In fact, with more than 1.49 billion monthly active users, Facebook's advertising revenue is 65.5% of the total social media advertising.[5] While paid advertising on social media platforms is gaining prominence, the effectiveness of this form of advertising is largely unstudied. In this study we address this gap by studying the effectiveness of word-of-mouth (WOM) signals and social endorsement on the click and conversion performance of ads in the form of app installs on Facebook.

Facebook ads display the popularity, or WOM[6] information in terms of the number of users who have endorsed the ad through "likes". Facebook also utilizes user social graph information for advertising by targeting users based on actions taken by their social connections in response to the ad. Further, Facebook also displays other users' actions to the targeted users. For example, Facebook ads contain social endorsement information on the targeted user's friends who have "liked" the ad. The objective of displaying these types of signals is to draw more user attention by showing the popularity of the ad among friends and other users and improve the

[4]http://www.clickz.com/clickz/column/2377715/the-future-of-social-media-paid-vs-organic

[5]http://www.emarketer.com/Article/Social-Network-Ad-Spending-Hit-2368-Billion-Worldwide-2015/1012357

[6]In this chapter we use WOM and popularity interchangeably.

ad performance in terms of actual product installs or purchases. Thus, it is important to know if the ad's popularity and social endorsement information has any effect on the targeted users for actual purchases or installs. Additionally, the install or purchase decision is a multi-step process where users first click on an ad and go to the product or the app source and then make the purchase or install decision. These separate steps correspond to different consumer stages of decision making (De Bruyn and Lilien 2008). Therefore, it is also useful to know whether and how users use these signals at different steps to make their decisions.

Previous research has shown the efficacy of observational learning (e.g., Banerjee 1992, Bikhchandani et al. 1992) and WOM signals to spread content and enable purchases (e.g., Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Gu et al. 2012, Lu et al. 2013). One possibility is that WOM signals such as "likes" would have a similar effect and increase the user propensity to respond to the ad and make purchases.[7] Some recent studies do suggest that word-of-mouth signals such as "likes" on Facebook (Li and Wu 2014) or "favorites" (Dewan et al. 2015) may also be effective in influencing the purchase decision. However, these studies rely on the WOM signals generated by users through organic content. Thus, it is not clear if ads will carry the same effect.

Another possibility is that, as WOM and social endorsement signals are included in ads, consumers may link these signals to advertiser's manipulative intent and show negative reactance (Brehm 1966, 1989; Clee and

[7] While a Facebook "like" suggests that a user "likes" the content or the product, "likes" differ from the traditional word-of-mouth signals as there is no negative valence.

Wicklund 1980, White et al. 2008). Further, social endorsement may be perceived as intrusive as the platform is targeting users based on friends' actions (Goldfarb and Tucker 2011) and may raise privacy concerns (Tucker 2014). This may further evoke a negative response to social ads. Thus, the actual effect of popularity and social endorsement signals in the social ads on actual product purchases or installs is not obvious and needs further investigation.

In order to examine the effectiveness of WOM and social endorsement in social ads, our study evaluates the following questions: (1) Does the popularity information displayed alongside ads influence targeted users' decision to click an ad? (2) Does the popularity information play a role in targeted users' decision to install an application after clicking the ad? (3) How does the presence of social endorsement influence users' decision to click an ad or to install an application after clicking the ad?

We answer these questions using the data generated from a randomized field experiment on Facebook in collaboration with a mobile app company which advertises on Facebook.[8] This collaboration allowed us to examine the effectiveness of WOM and social endorsement signals as part of the mobile app company's ongoing efforts to increase the installs for their application through promotions on Facebook. Specifically, we assume the role of marketing managers and test a mobile app install ad campaign on Facebook for a five-week period targeting sixteen different user groups with varying levels of WOM and social endorsement signals. We use the standard

[8]This company distributes a shopping app which connects users to different vendors. We cannot reveal the name of the company due to the non-disclosure agreement.

interface provided by the Facebook platform to advertisers to target users and draw inference from the aggregate performance data typically received by advertisers as part of their ad campaigns.

One of the challenges in determining the effect of ad characteristics on the advertiser performance on Facebook is the selection bias introduced by Facebook: the users to whom the ad is shown are not randomly sampled (Lee et al. 2014). More specifically, Facebook uses a proprietary algorithm which allows it to selectively display ads to users that are more (less) likely to respond. This sampling bias makes it difficult to identify the effect of ad characteristics on the response rates. To address this selection issue, we use a random bidding procedure which ensures that our ads in each group are randomly targeted to users of different quality—defined as their likelihood to respond, so the resulting sample more closely resembles the population of that target group. We collect hourly aggregate performance data for each ad in terms of the number of impressions, ad clicks, and the final app installs; these data were collected every hour through the Facebook Marketing API during our 5-week ad campaign. We use a logit specification for representing clicking and install decisions and estimate our models while accounting for correlation across these decisions.

We find that, across different model specifications, overall "likes" associated with Facebook ads are ineffective in leading to clicks and have a negative effect on installs. This result is in stark contrast to extant studies which have shown a positive effect of such WOM signals on user response on social media platforms (Liu 2006, Chen et al. 2011, Li and Wu 2014, Dewan et al. 2015). One potential explanation is that in our research context "likes" are associated with the ad; consequently, the user might perceive the "likes" as endorsing the ad itself rather than endorsing the product being advertised. This might lead the users to perceive "likes" as another form of advertising strategy and therefore not respond to them.

We also find that the presence of social endorsement has a negative effect on users' clicking decision. A potential explanation is that social endorsement may be perceived as intrusive as users are being targeted using their friends. However, conditional on clicking, social endorsement from several friends helps to reduce the negative effect of overall "likes". In other words, this means that when a user finds out several of her social connections have endorsed the app, she will be more willing to install the application compared to the case when no social endorsement information is displayed, provided that she has clicked the ad. This positive effect of social endorsement only after clicking suggests that users are not drawn to the ad using "likes" along with social endorsement. However, users who end up clicking the ad are likely to consider "likes" along with social endorsement to make the install decision. This suggests that WOM in the presence of social endorsement is not generating interest among consumers but is more likely to be used as a quality signal by the interested consumers to make the install decision.

Our research makes several contributions. We contribute to the existing literature on the social media WOM by showing how it works in the context of advertising. While previous studies show that WOM for organic content is effective for improving product sales, our results suggest that the same may not apply to promoted content. To the best of our knowledge, ours is the first study that demonstrates the role of social endorsement on ads for enabling app installs. Second, we provide insight into the underlying consumer behavior associated with the response to ad popularity and social endorsement information included in social ads. Previous research has suggested that WOM signals increase sales performance due to increase in the consumer valuation (Liu 2006, Chen et al. 2011) without determining the source of this increase in valuation. Our research adds to this work by demonstrating the role of WOM signals in different stages of consumer decision making in the context of ads. We show that social media WOM through ads is not increasing users' interest in the product but is more likely to serve as a signal of quality for the product for interested users provided that the ads also include social endorsement.

Our research also informs the Facebook platform on the efficacy of WOM and social endorsement signals to promote content and enable purchases. Specifically, our results show that popularity or social endorsement without popularity is inadequate to get consumer traffic and can have a negative effect on the install performance. Thus, the platform is better off not showing these signals on ads targeted to get new users. From a practitioner's perspective, our results suggest that advertisers should discount such metrics in the evaluation of their ad performance. Finally, our study provides advertisers with an approach to independently conduct experimentation on Facebook and test the consumer response to ad attributes while accounting for the selection bias introduced by the Facebook algorithm.[9]

3.2 Literature Review and Theory

Our research is related to the literature on promotions using word-of-mouth (WOM) and social networks.

3.2.1 Promotions using Word-of-Mouth (WOM)

Various advertising strategies enabled by the Facebook platform are related to core economics and marketing concepts such as observational learning (OL) and word-of-mouth (WOM). Studies in observational learning (OL) suggest that individuals use information they learn from observing purchasing decisions made by earlier individuals to make their own purchasing decisions (e.g., Banerjee 1992, Bikhchandani et al. 1992). WOM signals differ from OL signals as they provide collective user opinion or preferences of the product or service rather than displaying the actual purchase decision. WOM literature shows that an individual's decision is affected by the opinions and preferences of other consumers (e.g., Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Gu et al. 2012, Lu et al. 2013). Such WOM can improve consideration (Gupta and

[9] The only other way to prevent such selection bias is to randomly target users with ads independent of the algorithm, which is only feasible for Facebook's own research team. Facebook has close to 2 million advertisers. https://www.facebook.com/business/news/two-million-advertisers While Facebook researchers routinely conduct experiments on Facebook, they are unlikely to conduct experiments for individual advertisers.

Harris 2010). Additionally, WOM increases product valuation (Chevalier and Mayzlin 2006, Chen et al. 2011) and leads to higher sales performance of the product.

Social media platforms such as Facebook also show WOM signals for the content propagating on the platform using the notion of "likes". A Facebook "like" is a user's endorsement of the content and/or the product associated with the content. However, it is possible for users to "like" the ad without having to install or purchase the product. As a result, "likes" might be interpreted as a weaker WOM signal. Nevertheless, recent studies provide evidence that "likes" are indeed effective in improving the consumer response for organic content. For example, Li and Wu (2014) study the effect of Facebook "likes" in the context of a daily-deal website and show that "likes" increase the product sales. Similarly, Dewan et al. (2015) study the effect of "favorite" in an online music community, where "favorite" in their setting functions similarly as "like" on Facebook. They find that the number of favorites has a positive effect on the music consumption.

Facebook ads also display popularity information in the form of number of "likes" an ad has received along with other ad attributes. One possibility is that "likes" shown on ads can be seen as WOM signals. In that case "likes" would increase the user's propensity to click the ads and make purchases or installs. Alternatively, the user might perceive the "likes" as endorsing the ad itself rather than endorsing the product being advertised. This might lead the user to perceive "likes" as another form of advertising strategy and link it to advertiser's manipulative intent (Campbell 1995) and therefore induce users' reactance behaviors toward the ad (e.g., Brehm 1966, 1989; Clee and Wicklund 1980, White et al. 2008).

3.2.2 Promotions using Social Networks

Facebook also targets content to users using their social graph information. This includes showing content to users whose friends have already endorsed the content using a "like" and also displaying this endorsement apart from the overall popularity or "likes" for the content. Such targeting of users and display of social cues or social endorsement using the social graph can be beneficial in two ways. First, users in a social network exhibit homophily i.e., friends possess similar characteristics and preferences (McPherson et al. 2001). Thus, due to similarity in preferences, a user is more likely to adopt a product if one or more friends have "liked" the product. Further, friend connections operate through trust (Coleman 1990). As a result, WOM from friends is more trustworthy as compared to regular WOM (Rogers 1995). Thus, the display of "likes" by friends may have a stronger influence on users to adopt as compared to overall "likes". Many researchers have found such online social influence to play a role in the adoption of products. These include adoption of YouTube videos (Susarla et al. 2010, Yoganarasimhan 2012), social networking site (Katona et al. 2011, Bakshy et al. 2012), applications (Aral and Walker 2011) and paid subscriptions (Bapna and Umyarov 2015). Additionally, an increase in the number of friends' "likes" further increases the propensity of adoption. For example, Centola and Macy (2007) show that individuals are more likely to adopt actions previously taken by their peers when multiple social signals are present. Additionally, Dewan et al. (2015) find that these signals act as substitutes and users prefer to respond to WOM from friends instead of WOM from others. In that case, one can expect users are more likely to respond to friends' "likes" than overall "likes".

It is possible that such social targeting and social endorsement can be effective for increasing product sales or app installs through ads. Some studies on Facebook (Tucker 2012, Bakshy et al. 2012) show the effectiveness of social endorsement information on ads to generate more "likes". Tucker (2012) uses a Facebook ad campaign dataset of a charity organization to show that social ads are more effective than regular ads in generating clicks and "likes". She finds that while social influence plays some role, the contribution is primarily due to the ability of ads to target individuals with similar preferences. Bakshy et al. (2012) find that an increase in the number of social cues leads to higher click performance as well as higher endorsement rate of ads on Facebook. However, they restrict their analysis to ads which are always socially targeted and do not consider the relative performance of social ads compared with regular ads. Additionally, both of these studies use ads to create content awareness, and users are not required to make product purchase or app install decisions which are cognitively more costly. Aral and Walker (2011) evaluate the effect of viral product design on social contagion in app installations. However, in their setup information about the app is spread organically and not through an ad. Further, they do not consider the effect of WOM signals such as "likes" on the propensity to install apps.

It is also possible that social endorsement may increase the salience of the advertiser's manipulative intent. This may activate persuasion knowledge which can be used to view ads negatively (Campbell and Kirmani 2000). Reactance is greatest when the information used is more unique (White et al. 2008, Tucker 2014). Social endorsement information is unique to the user as it is an endorsement from the user's friends. Thus, social endorsement may evoke higher reactance. Such reactance can be due to privacy concerns (Tucker 2014). In that case, higher attention to social ads may raise concerns such as privacy, which may diminish the performance of ads. For example, Goldfarb and Tucker (2011) show that while contextual targeting usually increases ad performance, it can lead to lower performance if such ads are made more visible to the users using videos, pop-ups, or large displays. They attribute this to the sense of intrusiveness that arises when targeted ads are highly visible. We expect that a socially targeted ad results in better targeting as users share similar preferences as their friends who have endorsed the ad due to homophily (McPherson et al. 2001). As a result, when it is placed in the newsfeed, higher attention can elicit a sense of intrusiveness and may negatively impact the performance of the ad. We summarize the relevant theories and the predictions that emerge from these theories in Table 3.1.

Thus, the previous literature suggests that WOM can improve installs or purchases. Similarly, social endorsement can have a positive effect on the adoption of products. However, the effectiveness of these mechanisms on the performance of sponsored content or ads is not known and is the focus of this study.

3.3 Experimental Design and Data Description

3.3.1 Experiment Objective

The objective of our experiment is to determine the effect of popularity, or "likes", and social endorsement information on the click and install performance of ads on Facebook. In order to achieve this we collaborated with a technology company that provides a free mobile shopping app on the Google Android platform. This collaboration allowed us to run advertising campaigns on the company's behalf to promote its app using Facebook's advertising platform. We conduct a field experiment on Facebook where we randomly target users using these ads and evaluate how users are affected by "likes" for the ad and how they are influenced by endorsement of the ad from social connections. More specifically, we explore how the number of "likes" and social endorsement affect the user's decision to click on the ad and subsequently install the app.

Our approach of using an app to determine the effect of WOM signals is similar to Aral and Walker (2011) where they use an app to determine the effect of viral product design to create social contagion using active and passive viral features. However, there are several differences between our setups. First, they use an app hosted on the Facebook platform for their experiment where the application provider has access to the app users' social networks. This allows them to recruit and target users for the experiment and track app installs at an individual level. However, apps or products that are consumed outside the Facebook platform do not have access to individual Facebook users. Second, they rely on organic propagation of the messages whereas our objective is to test the effectiveness of ads with varying "likes" and social endorsement in generating user response. Since an ad is posted by Facebook and not by an individual user; an advertiser cannot directly target individual users or monitor performance at an individual level but has to rely on Facebook's advertising platform.[10] Note that this restriction would apply even for advertising campaigns of an app running on the Facebook platform.[11] As a result, we do not explicitly recruit users but rely on the regular advertising setup used by Facebook to target users and measure the performance at an aggregate level in each time period.

3.3.2 Advertising on Facebook and Mobile Install Ads

Facebook allows advertisers to run advertising campaigns on its network in order to reach consumers. These ads can be used to promote a page, a product, or an app. For a particular ad creative, advertisers can use different targeting options such as gender, age, and preferences. It is possible that ads from different advertisers are competing for user attention, and Face-

[10]See https://www.facebook.com/business/help/318580098318734 for a step-by-step guide on examining ad campaign performance. Notice the campaign performance metrics are displayed at the aggregate level.

[11]An app on Facebook platform can only access networks associated with its own users. However, if the targeting is done by the Facebook platform, then the app cannot determine which users were targeted.

book uses an auction to prioritize these ads. In these auctions advertisers submit bids for their ads, and Facebook uses these bids and the expected performance of ads to rank and prioritize these ads.[12] The expected performance reflects the underlying quality of an ad to generate user response. An advertiser can increase its ad exposure to Facebook users by submitting higher bids. These would result in more ad impressions, clicks, and installs or purchases. The Facebook platform allows the advertisers to submit their bids through its interface or through API calls. Facebook also allows advertisers to access their aggregate performance information such as impressions, clicks, "likes", etc., for each ad in each period through its interface as well as through its API.

Facebook provides an ad format specifically for promoting mobile apps, as shown in Figure 3.1. These ads are displayed to users in their Facebook newsfeed when they are using a mobile device.

Facebook users can comment, "like", and share the ad with their friends; they can also click on the ad, which will take them to the app's install page on the appropriate platform such as iTunes or Google play, and there users can decide if they want to install the app. The app platform shares install information with Facebook, so installs on the app platform can be linked back to the users clicking the ad.[14] Notice that at the time

[12]https://www.facebook.com/business/help/163066663757985

[13]It shows the ad text, ad picture, any available social endorsement information, i.e. friends who have "liked" this app, as well as the number of "likes" this ad received. If a user clicks on the ad, s/he will be redirected to the download page, as shown in Figure 3.2, where s/he can install the app.

[14]Facebook provides SDK for android or iOS which allows tracking of installs https://developers.facebook.com/docs/app-ads/measuring/installs-and-in-app-conversions

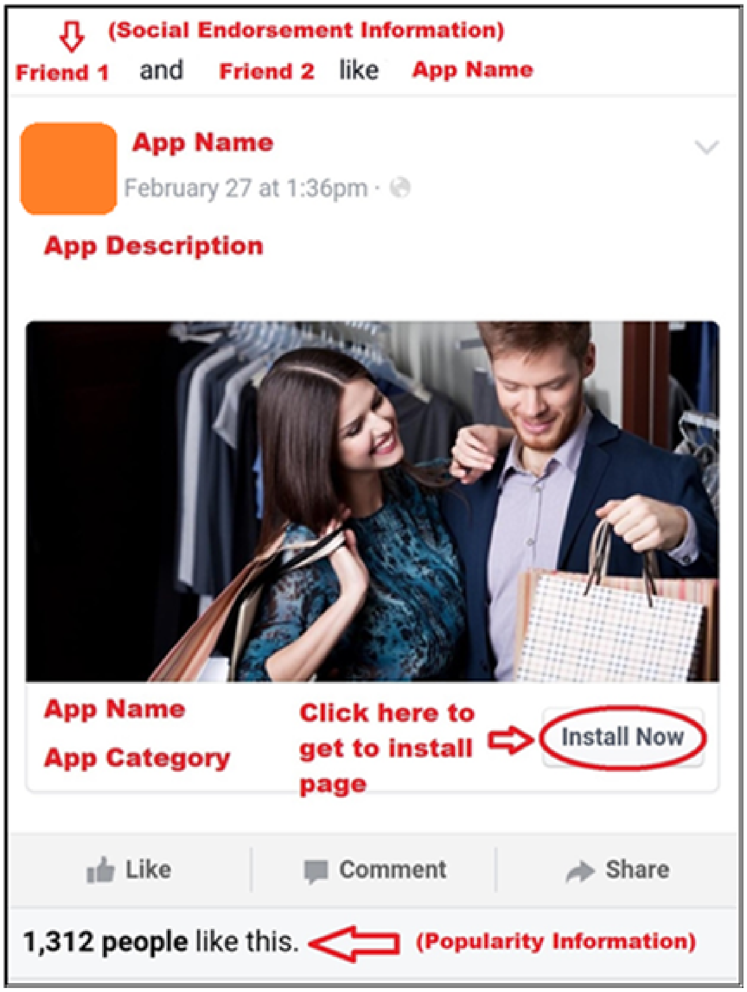

Figure 3.1: A sample mobile install ad on Facebook[13]

of our experiment mobile install ads will only display the total number of Facebook users who have "liked" the ads, but not numbers of clicks or installs. In addition, if a user's friend has "liked" the ad, then Facebook will also display relevant social endorsement information, i.e., the name of the friend who has "liked" the ad, as shown in Figure 3.1.

3.3.3 User Targeting

As mentioned earlier, Facebook's advertising platform allows advertisers to explicitly target users of a given set of characteristics. Using this feature, we generate a mobile app install ad and show this ad to a total of 16 groups, each consisting of users with a specific set of characteristics. Specifically, these 2 x 2 x 2 x 2 experimental groups are based on the following user characteristics: (1) friends of fans, (2) age, (3) gender, and (4) experience in mobile technology. We explain the importance of each characteristic in the following discussion.

Firstly, Facebook allows advertisers to target users who have friends that have "liked" the Facebook page of the company providing the app and, as a consequence, are followers or fans of the company. Such fans are likely to be interested in the app provided by the company. In that case, by virtue of homophily, target users who are friends of fans are also likely to be more interested in the app as compared to other users. We refer to this group of users as the "FF" group. This allows us to distinguish between users based on their inherent interest in the app and allows us to explain the potential driver of the effect of social endorsement as explained in the results section. User response to the social media content depends on the demographic characteristics of users (Lee et al. 2014). Thus, it is possible that the response to WOM signals such as "likes" and social endorsement may be biased due to a particular demographic group. In order to ensure that different demographic groups have an even representation in our sample, we create identical ad copies which are targeted to different demographic groups. We consider two age groups (18-25 or 26-34)[15] and gender (male or female) as they might also affect how users react to WOM signals and social endorsement.

Finally, Facebook is able to determine whether a user just recently started using smartphones or tablets, and we use this information to divide users into either the experienced or the inexperienced group. This characteristic is important because studies suggest that inexperienced users are more likely to be influenced by others than experienced users (e.g. Bettman and Park 1980). Additionally, experience plays a role in technology adoption (Venkatesh et al. 2003). In order to control for the response bias based on the experience of users, we create ad copies separately for experienced and inexperienced groups.

In summary, each of the 16 groups consists of users that are (FF/not FF), age group (18-25/26-34), (male/female), and (experienced/ inexperienced) in mobile technology. A summary of all sixteen target groups is shown in Table 3.2. For example, group 1 in our experiment consists of users that are FF, in the age group 18-25, male, and are inexperienced in mobile technology.

Since the product we are advertising is a mobile app designed for the Google Android platform, we restrict all target users to be Android-based mobile device users with an Android 4.0 or newer operating system installed to make sure their devices are up-to-date and the app can run correctly. We further restrict our target users to those that are interested in

[15]The mobile app company specifically targets these age groups as these age groups are likely to install the shopping app.

online shopping to make sure they will potentially be interested in our app, since the app we advertise in this experiment is a shopping app.

3.3.4 Random Bidding

It is well known that the assignment of ad to users is not random—Facebook reportedly uses some proprietary algorithm to determine the display of ads (Lee et al. 2014) and target users depending on their propensity to respond. This nonrandom assignment is likely to create a selection bias where any response to "likes" would be correlated with the order of user quality targeted by Facebook, which is unobservable to the researcher and the advertiser. To address this issue, we create a random bidding system which allows us to generate a random bid for each of the sixteen ad groups every hour. As described earlier, advertisers need to submit a bid which specifies how much he or she is willing to pay for each impression, i.e. each time the ad is shown. Facebook then uses this bid information and the expected quality of the ad to determine its score and targets the user accordingly. Thus, the ad with the best score gets the highest quality users. Our random bidding system considers a wide range of bids from $1- 10[16] so that we can make sure that our ads get a wide range of scores and are evenly targeting users of all quality levels. Figure 3.3 shows how the targeting can vary across users with or without randomized bids. Facebook's default targeting allows access to only users with a certain propensity to respond based on the score of the

[16]Upper limit of our bids is 3 to 5 times the bid suggested by Facebook to ensure we do not miss on the high quality users.

ad. However, randomized bidding can allow access to users with varying propensity to respond.[17]

3.3.5 "Likes" and Social Endorsement

We rely on the variation in "likes" and social endorsement across users to determine the effect of these variables on the user response to click and install. An advertiser or a researcher cannot directly manipulate this information on ads as it is controlled by the Facebook platform. Thus, we rely on the actions of the target users to generate these signals and show ads with these signals to target users in subsequent periods. As our ads are randomly shown to different users, some users choose to "like" the ads as they click and install the app. Facebook aggregates these "likes" across all target groups[18] and displays the ad along with "likes" to subsequent users who are targeted. As the number of "likes" increases over time, users targeted at different points in time will see different number of "likes". This variation in "likes" represents multiple different treatments in our experiment, and we rely on the variation in treatments to identify the effect of each additional "like" on the ad effectiveness.

Among users who are randomly targeted, there is a subset of users with one or more friends who have "liked" the ad. These users also see the

[17]We rely on the publicly known information provided by the Facebook platform to devise our randomization approach. This is based on the assumption that Facebook as a platform is not engaging in any strategic behavior and is actually following the publicly revealed ranking mechanism.

[18]We use a single post which is targeted to sixteen different groups separately using different ads. Facebook provides separate performance data for each ad. However, "likes" are associated with the post and are aggregated across all ads.

social endorsement information on the ad. Note that not all users qualify to be targeted with social endorsement. Only those users whose friends have already "liked" the ad can be targeted with social endorsement information. This design is similar to the average treatment effect on the treated used by previous research (Aral and Walker 2011). However, in our setup users who do not see the treatment of social endorsement are also not likely to receive social endorsement as none of their friends have liked the ad. Thus, we have two separate pools of users within each target group, one which is eligible to receive social endorsement and the other which is not eligible. Figure 3.4 shows the targeting of users with and without social endorsement, i.e., one or more friends who have "liked" the ad. While our random bidding procedure allows us to randomly pick users from these two different pools of users, friend of users who have "liked" the ad may have a higher propensity to respond to the ad. As a consequence, any effect of social endorsement may be driven by this propensity or homophily and not necessarily due to social influence. We also conduct additional analysis to determine the possible mechanism and explain it in our results section.

Additionally, in each period "likes" seen by the target users are based on the "likes" generated by users in all previous periods. Within each group in each period, the ad placement is driven by bids and expected performance of the ad. In that case, user selection for ad placement and the resulting "likes" generated could be correlated. Note that "likes" are not likely to be correlated across groups as bids are randomized separately for each group and user selection is independent across groups. In that case, cumulative “likes” are less likely to be correlated with unobservable user quality in each group. However, in order to address any potential endogeneity concern, we use an appropriate instrument as described in the model section.

3.3.6 Data Collection

We interact with the Facebook Advertising Platform through Facebook Marketing API, which enables us to create customized target groups and access ad performance statistics in a programmatic manner. A python script was written to automate the following tasks: (1) submit a randomly generated new bid every hour, (2) extract impressions, clicks, “likes”, and installs statistics on an hourly basis. Notice that Facebook only provides cumulative statistics, and all statistics are anonymous, aggregated at the target group level. Therefore, every hour we extract the cumulative number of social and non-social impressions, social and non-social clicks, “likes”, and installs. We then take first difference to determine the hourly performance for each of the sixteen target groups, where social impressions are defined as impressions shown with social endorsement information, and social clicks are defined as clicks that result from social impressions. Note that Facebook does not provide advertisers with social install information, so we cannot directly observe the number of installs that are results of social impressions and social clicks. We use a likelihood approach to attribute the cumulative installs to social and non-social clicks as described in the model section. Also note that Facebook does not provide advertisers with the magnitude of social signals associated with social impressions or social clicks. In other words, advertisers can only know that some social endorsement information is displayed, which means at least one of the user's friends has "liked" the ad, but they do not know exactly how many of the user's friends have "liked" the ad. We use an appropriate proxy for friends' "likes" as described in the model section.

We ran the advertising campaign over the 5-week period from April 2015 to June 2015. During this period our experimental ad was shown a total of 710,445 times, resulting in 4052 clicks, 799 "likes", and 206 installs across all sixteen target groups. The low click-through rate, 0.57%, is comparable to those reported in other studies (Tucker 2012, Tucker 2014). Sample observations are shown in Table 3.3. Each observation in the collected data is a group-hour observation. Notice that in our campaign, all sixteen groups saw an identical mobile install ad, and they were also seeing the same number of cumulative "likes" that have been given to the ad. Summary statistics of the hourly data are shown in Table 3.4.

3.4 Research Model

The app install process through Facebook ads consists of two steps. In the first step users are shown the mobile app install ad (Figure 3.1), and they decide whether to click on it or not. For users clicking on the ad, they enter the second step in which they learn more about the app through the ad and then decide whether to install it (Figure 3.2). These steps represent different stages of consumer decision making. De Bruyn and Lilien (2008) suggest a three-stage model to characterize users' decision making process: awareness, interest, and final outcome. In the awareness stage consumers become aware of the alternative. If users are interested, they take further action to get more information. In the final stage, users take an observable action such as purchase of a good or service. In our research context, users are already aware of the application as they see the ad in the newsfeed. The clicking step represents the interest stage: upon seeing the ad, the user then decides whether to click on the ad based on his or her interest level. The install step reflects the final outcome. We model these two stages to study how the number of "likes" and social endorsement on ads affect users' ad clicking and app installing decisions.

3.4.1 Decision Context 1: Click Rate

When a user is shown an ad, her decision to click can be influenced by "likes" and social endorsement. Facebook does not provide advertisers with the magnitude of social signals associated with social impressions or social clicks. In other words, advertisers can only know that some social endorsement information is displayed, which means at least one of the user's friends has "liked" the ad, but they do not know exactly how many of the user's friends have "liked" the ad. We use a dummy variable to capture the effect of social endorsement. Further, as the overall "likes" increase, the number of "likes" from friends can also be expected to increase and can influence the user decision to click. We proxy the effect of this increase in the number of friends' "likes" using an interaction between "likes" and social endorsement. In our setup, in any period we know the number of users who are shown the ad. We also know the user response to the ad in terms of clicks. Thus we can represent the outcome as a click probability. This is similar to the setup of search engine auctions and we adopt the modeling approach used in the related literature (Ghose and Yang 2009, Agarwal et al. 2011, Rutz and Trusov 2011) to capture the click performance. Specifically, we assume an i.i.d. extreme value distribution of the error term for individual choices and use a logit model to represent the click probability for an individual belonging to group g at time t as follows:

$$\Lambda_{gt}^{Click} = \frac{exp(U_{gt}^{Click})}{1 + exp(U_{gt}^{Click})}, \tag{3.1}$$

where U_{gt}^{Click} is the latent utility of clicking the ad for group g at time t and can be expressed as follows:

$$\begin{aligned} U_{gt}^{Click} =& \beta_0 + \beta_1 \cdot \sum_{j=1}^{t-1} like_j + \beta_2 \cdot social + \beta_3 \cdot social \cdot \sum_{j=1}^{t-1} like_j \\ & + \beta_\gamma \cdot \gamma_g + \beta_\delta \cdot \delta_t + \beta_h \cdot h_t + \beta_w \cdot w_t + \epsilon_{gt}^{\beta}, \end{aligned} \tag{3.2}$$

where *social* is the dummy representing whether or not this individual is shown some social endorsement information, i.e., whether the impression is a social impression, $\sum_{j=1}^{t-1} like_j$ is the cumulative "likes" up to time $t-1$; $social \cdot \sum_{j=1}^{t-1} like_j$ captures the interaction effect; γ_g is a vector of user characteristics dummies associated with group g, including age, gender, experience, and FF membership; δ_t is a vector of day of week dummies; h_t is a vector of time of day dummies where we separate a day into four 6-hour periods; w_t is a vector of week dummies which controls for any potential time trend; ϵ_{gt}^{β} represents the idiosyncratic error term which is common for all individuals within group g.

Since in our data set we observe the number of social/non-social impressions and the corresponding social/non-social clicks per group-hour, we can specify the clicking likelihood function as

$$
\begin{aligned}
l\left(U_{gt}^{Click}\right) = \prod_{g=1}^{G}\prod_{t=1}^{T}\Big\{ & \left(\Lambda_{gt}^{Click}(1)\right)^{\#SocialClicks_{gt}} \\
& \cdot\left(1-\Lambda_{gt}^{Click}(1)\right)^{\left(\#SocialImpressions_{gt}-\#SocialClicks_{gt}\right)} \\
& \cdot\left(\Lambda_{gt}^{Click}(0)\right)^{\#SocialClicks_{gt}} \\
& \cdot\left(1-\Lambda_{gt}^{Click}(0)\right)^{\left(\#SocialImpressions_{gt}-\#SocialClicks_{gt}\right)}\Big\}.
\end{aligned} \tag{3.3}
$$

3.4.2 Decision Context 2: App Install Rate

Users who end up clicking the ad may choose to install the app. Further, this decision to install can also depend on the "likes" and social endorsement information displayed with the ad. We represent this install decision using a similar approach as the clicking decision. Specifically, we use the following logit model to capture the install performance conditional on clicking, or the install rate, of a group g at time t:

$$
\Lambda_{gt}^{Install} = \frac{exp(U_{gt}^{Install})}{1+exp(U_{gt}^{Install})}, \tag{3.4}
$$

where $U_{gt}^{Install}$ is the latent utility of installing the app for group g at time t and can be expressed as follows:

$$\begin{aligned} U_{gt}^{Install} = & \theta_0 + \theta_1 \cdot \sum_{j=1}^{t-1} like_j + \theta_2 \cdot social + \theta_3 \cdot social \cdot \sum_{j=1}^{t-1} like_j \\ & + \theta_\gamma \cdot \gamma_g + \theta_\delta \cdot \delta_t + \theta_h \cdot h_t + \theta_w \cdot w_t + \epsilon_{gt}^{\theta}. \end{aligned} \tag{3.5}$$

However, since Facebook does not provide advertisers with the number of social/non-social installs, we are not able to trace installs back to social/non-social clicks. Instead, we use the empirical frequency of social and non-social clicks among all clicks to account for the effect of social endorsement on the overall install probability. Specifically,

$$\begin{aligned} \Lambda_{gt}^{Install} = & \Lambda_{gt}^{Install}(1) \cdot \frac{\#SocialClicks_{gt}}{\#SocialClicks_{gt} + \#NonSocialClicks_{gt}} \\ & + \Lambda_{gt}^{Install}(0) \cdot \frac{\#NonSocialClicks_{gt}}{\#SocialClicks_{gt} + \#NonSocialClicks_{gt}}. \end{aligned} \tag{3.6}$$

Therefore, the corresponding installing likelihood function can be specified as

$$l\left(U_{gt}^{Install}\right) = \prod_{g=1}^{G} \prod_{t=1}^{T} \left\{ \left(\Lambda_{gt}^{Install}\right)^{\#Installs_{gt}} \cdot \left(1 - \Lambda_{gt}^{Install}\right)^{\left(\#Clicks_{gt} - \#Installs_{gt}\right)} \right\}. \tag{3.7}$$

Finally, combining the clicking and the installing steps, the complete likelihood is

$$\begin{aligned}\prod_{g=1}^{G}\prod_{t=1}^{T}\Big\{&\left(\Lambda_{gt}^{Click}(1)\right)^{\#SocialClicks_{gt}}\\
&\cdot\left(1-\Lambda_{gt}^{Click}(1)\right)^{\left(\#SocialImpressions_{gt}-\#SocialClicks_{gt}\right)}\\
&\cdot\left(\Lambda_{gt}^{Click}(0)\right)^{\#NonSocialClicks_{gt}}\\
&\cdot\left(1-\Lambda_{gt}^{Click}(0)\right)^{\left(\#NonSocialImpressions_{gt}-\#NonSocialClicks_{gt}\right)}\\
&\cdot\left(\Lambda_{gt}^{Install}\right)^{\#Installs_{gt}}\cdot\left(1-\Lambda_{gt}^{Install}\right)^{\left(\#Clicks_{gt}-\#Installs_{gt}\right)}\Big\}.\end{aligned} \tag{3.8}$$

3.4.3 Endogeneity of "Likes"

Our strategy is to implement a random bidding system to minimize the issue of sample selection bias resulting from Facebook's nonrandom ad-user assignment, as documented in the experimental design section. Further, we rely on the users to generate "likes" for our ad. However, Facebook determines the ad placement for all competing ads based on their bids and expected performance. Thus, it is possible that Facebook may prioritize certain ads competing for the same set of users and the user selection for the focal ad is influenced by this prioritization. For example, if two competing ads have the same distribution of random bids then the ad with higher expected performance is more likely to get higher quality users. Since users' quality influences their decisions to respond to the ad, it is possible that the cumulative "likes" generated may be correlated with the unobservable user quality which is also influencing their click and install decisions. However, in each period the random draw of bid allows us to pick different users for each of our sixteen target groups. Further, bids are independent across different target groups, suggesting that the quality of the selected users is less likely to be correlated across groups. As cumulative "likes" are generated by these different users, it is less likely that these cumulative "likes" generated in a prior period will be correlated with the user quality in a particular target group in the current period. To further alleviate any remaining endogeneity concern, we adopt an instrumental variable approach where we use the randomly generated bids to instrument for "likes". This is because these random bids directly affect "likes" and are uncorrelated with any unobserved variables in the clicking and install steps that might also affect "likes". Specifically, when examining the effect of signals on users' click and install decisions, we used the number of fitted "likes" instead of using the actual number of "likes", where fitted "likes" are generated using the following expression:

$$like_{gt} = \alpha_0 + \alpha_1 \cdot bid_{gt} + \alpha_\gamma \cdot \gamma_g + \alpha_\delta \cdot \delta_t + \alpha_h \cdot h_t + \alpha_w \cdot w_t + \epsilon^{\alpha}_{gt}, \tag{3.9}$$

where bid_{gt} represents the randomly generated bid for group g in period t, γ_g is a vector of group demographic dummies, δ_t is a vector of day of week dummies, h_t is a vector of time of day dummies, and w_t is a vector of week dummies. We then generate cumulative fitted "likes" based on group-hour level fitted "likes". Specifically, the cumulative fitted "likes" up to time $t-1$

is derived as follows:

$$\sum_{j=1}^{t-1} \widehat{like_j} = \sum_{j=1}^{t-1} \sum_{g=1}^{G} \widehat{like_{gj}}. \tag{3.10}$$

Finally, as "likes" are endogenous, the unobservable time varying attributes for the equations representing consumer decisions to click and install will be correlated with error terms for the equation representing "likes". As such, we use the following distribution to account for any correlation between the error terms:

$$\begin{bmatrix} \epsilon_{gt}^{\beta} \\ \epsilon_{gt}^{\theta} \\ \epsilon_{gt}^{\alpha} \end{bmatrix} \sim \mathrm{N}(0, \Omega), \quad \text{where} \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{bmatrix}. \tag{3.11}$$

3.4.4 Identification and Estimation

Effect of "Likes"

We rely on the variation in cumulative "likes" across users to determine the effect of "likes" on the click and install performance. We use the 2SLS approach to correct for the endogeneity bias associated with "likes". We first estimate the "likes" in each period for each ad group using bids and use the estimated cumulative value of "likes" in the previous periods to estimate the parameters of the click and install equations. Identification comes from the fact that fitted "likes" are completely determined by the exogenous bids. "Likes", in turn, influence the user's clicking and install decisions.

Effect of Social Endorsement

The effect of social endorsement can be established by comparing the performance of users receiving ads with social endorsement with that of the users who are shown the ad without such endorsement. As described earlier, our random bidding procedure allows us to show ads to randomly selected users in each period from the two pools of users within each target group, i.e., one receiving the ad with social endorsement and the other without social endorsement. This randomization ensures that the selected users represent the population of the target group with each pool. This helps us to establish the relative performance of each pool within each target group.

Note that while the selection of users from the two pools is random, the classification of users in these two pools is not random as the Facebook interface does not allow advertisers' to randomly enable social endorsement only for selected users. If a user whose friends have already "liked" the ad is shown the ad, she will always receive the social endorsement information (treatment). Similarly, non-treated users, i.e., users not receiving social endorsement, do not have any friends who have "liked" the ad. This is different from the approach adopted by previous studies (for example, Aral and Walker 2014) where only peers of adopters are randomly classified into treatment and control groups. As a consequence, the effect of social endorsement may be driven by this propensity or homophily of the subgroup receiving social endorsement to respond to the ad and not necessarily due to social influence. In order to overcome this limitation, we also conduct a separate analysis for only the "FF" group. Users in this group are friends of fans of the company's Facebook fan page and are likely to share similar propensity or interest in the app provided by the company by virtue of homophily. In that case, any effect of social endorsement within this group is more likely to be driven by social influence.

An additional challenge with the identification of the effect of social endorsement is that the Facebook platform does not separately provide information about the installs resulting from ads shown with social endorsement. As described earlier, we rely on the empirical frequency of social clicks to capture the effect of social endorsement on the overall install probability (Equation 3.6). A critical requirement for identifying the effect of social endorsement is that there is variation in the proportions of social installs across different time periods within each target group. Specifically, in Equation 3.6 the proportion of social clicks should vary across observations in order to identify the effect of social endorsement. In our data, this proportion varies randomly and helps us to identify the effect of social endorsement.

Effect of Social Endorsement x "Likes"

Identification for this interaction effect comes from the fact that "likes" are varying across users in each pool: one receiving social endorsement and one not receiving social endorsement. Further, we use cumulative estimated "likes" to overcome any endogeneity bias caused by the selection of the users.

The above set of simultaneous equations represents a triangular system and has been addressed by authors in classical econometrics (Lahiri and Schmidt 1978, Hausman 1975, Greene 1999) and Bayesian econometrics (Zellner 1962). Hausman (1975) and Lahiri and Schmidt (1978) have shown that the parameter estimates for a triangular system can be fully identified using GLS. Zellner (1962) has addressed triangular systems from a Bayesian point of view. Triangular systems have been estimated using the classical approach (Elberse and Eliashberg 2003, Godes and Mayzlin 2004) and more recently in sponsored search using the Bayesian approach (Ghose and Yang 2009, Yang and Ghose 2010, Agarwal et al. 2011, Rutz and Trusov, 2011, Rutz et al., 2012). We estimate our main model using a Bayesian approach, applying Markov Chain Monte Carlo sampling due to the non-linear characteristics of our model (Rossi and Allenby 2005). The priors and conditional posteriors of this model are discussed in Appendix B. We run the MCMC simulation for 80,000 draws, discarding the first 40,000 as burn-in. We also performed an F-test in the first stage for our instrument. The F-test value was well over 10, suggesting our instrument is not weak. We report our main results for click rate and install rate models in Table 3.5. We also estimate a model without the interaction term (Tables 3.8 and 3.9). However, the model fitness is poor compared to our main model with the interaction term (Table 3.11).

3.5 Empirical Results

3.5.1 Click Rate

The estimation results for the click rate model are shown in Table 3.5. We can see that the coefficient associated with cumulative "likes" is not statistically significant for the click decision. This suggests that Facebook "likes" do not increase users' tendency to click on the ad. If "likes" are treated as WOM signals, then one would expect "likes" to improve product consideration (Gupta and Harris 2010) and increase valuation of product (Chevalier and Mayzlin 2006, Chen et al. 2011). In that case we would expect "likes" to increase the clicking performance. However, we do not find any such effect. A possible explanation for this surprising result is that these Facebook "likes" are displayed along with the ad itself, which might prompt users to perceive them as part of the firm's advertising strategy. This would induce users' reactance behavior, which would lead them to disregard the signal (Brehm 1966, 1989; Clee and Wicklund 1980, White et al. 2008).

An alternative explanation can be that users receiving ads with higher number of "likes" are poor quality users due to the selection bias induced by the Facebook platform where it targets best possible users first. However, our random selection of users should prevent such selection bias to a large extent. Additionally, we controlled for the time trend using time dummies. This suggests that the outcome is not likely due to any systematic time trend associated with the user quality. It is also possible that Facebook could be targeting the wrong set of users which in turn may itself impact the response to "likes". However, as we target only those users who have interest in shopping, this is not likely to be the case.

The coefficient for "Social" is negative and significant (Table 3.5). This suggests that the mere presence of social endorsement information actually decreases users' propensity to click the ad. Additionally, the coefficient for the interaction terms between social and "likes" is not significant. Thus, the click performance of ads does not improve even if more friends have endorsed the ad. This result is also contrary to our expectation since users are more likely to have similar tastes with their friends, i.e. homophily (McPherson et al. 2001). In addition, endorsement from friends should further influence users to respond to the ad (Rogers 1995). However, we are actually seeing the opposite effect in our analysis. A plausible explanation is that the display of social endorsement information may be seen as intrusive. As social endorsement information is unique to the user, it may evoke higher reactance (White et al. 2008) and may raise privacy concerns (Tucker 2014). This may result in a lower click propensity for ads with social endorsement information.

Our results for the impact of social endorsement on the click performance are different from those obtained by Bakshy et al. (2012) and Tucker (2012). Both of these studies use ads to create content awareness, and users are not required to make product purchase or app install decisions which needs additional processing. Increase in processing can lead to negative reactance (Campbell 1995). Task incongruence has also been associated with negative reactance (Edwards et al. 2002). In our context users may be expecting to read newsfeed from friends and not actually evaluate mobile install ads. This mismatch between user activity and the ad can lead to negative response. Further, these studies use ad formats where the ads appear on the right hand side and not in the newsfeed.[19] However, install ads are always shown in mobile newsfeed. Previous research has shown that personalized ads that are more visible are more likely to be perceived as intrusive (Goldfarb and Tucker 2011). Thus, it is possible that newsfeed ads with social endorsement may be perceived as intrusive as compared to the regular newsfeed ads as social ads are more personalized due to the inclusion of social endorsement information. Also, note that Bakshy et al. (2012) do not compare the performance of social ads with non-social ads. On the other hand, while Tucker (2012) does compared the performance of a social ad with a non-social ad, she uses a Facebook ad campaign dataset of a charity organization and the results could be driven by the altruistic setup.

The coefficient for FF is negative and significant (Table 3.5). This suggests that users belonging to the FF group are less likely to click the ad as compared to other users. Users in the FF group are likely to be familiar with the app company as they are friends of fans of the app company's Facebook page. Due to homophily, these users should be expected to have higher fit for the ad as compared to the regular users. The lower response of these users provides further evidence of possible negative reactance to

[19] Facebook introduced newsfeed ads only in January 2012. http://abcnews.go.com/Technology/facebook-put-sponsored-ads-timeline-newsfeed-january-2012/story?id=15205346 http://techcrunch.com/2012/08/14/facebook-page-ads/

the ad as they may perceive the display of ad as a manipulative intent on the part of the Facebook platform.

3.5.2 Install Rate

The estimation results for the install rate model are shown in Table 3.5. Cumulative "likes" have a negative impact on the propensity of users to install the app as the coefficient associated with cumulative "likes" is negative and significant. This, along with the results for click rate, suggests that, unlike the observation made for the WOM signals (Liu 2006, Chen et al. 2011, Dewan et al. 2015) in the context of product purchases, pure Facebook "likes" appearing on Facebook ads do not help the users to install the app. The negative effect of "likes" on the install rate can again be attributed to negative reactance. Users who have already clicked the ad and are considering the install decision need higher processing effort to make the install decision. Higher processing can lead consumers to think about what the advertiser is trying to do in the ad and lead to negative reactance (Campbell 1995).

The coefficient for "Social" is negative and significant (Table 3.5). This suggests that the presence of social endorsement information actually decreases users' propensity to install the app even after clicking the ad. In contrast, the coefficient of the interaction term, $Social \cdot \sum_{j=1}^{t-1} \widehat{like_j}$, is positive and significant. A positive interaction effect may suggest that "likes" and social endorsement are complements. However, Dewan et al. (2015) find that these different signals act as substitutes and WOM signals from friends dominate overall WOM signals. In that case the positive effect of the interaction term, $Social \cdot \sum_{j=1}^{t-1} \widehat{like_j}$, is more likely to be driven by the intensity of friends' "likes". This suggests that users who have clicked the ad are more likely to install the app when the ad is socially endorsed by a large number of friends. As explained later, our results for the FF-subsample also suggest that the positive interaction is likely to be driven by friends' "likes".

The initial negative response to social endorsement can be potentially due to the perception of intrusiveness which may play a role even among interested users. Among the users who do end up clicking ads, the ones seeing the social endorsement may still be wary of the advertiser's intent and may have a lower propensity to install due to negative reactance. However, with the increase in "likes" and potentially friends' "likes", it is possible that the perceived value of the application increases (Chevalier and Mayzlin 2006, Chen et al. 2011). Reactance is low for personalized content when the perceived value of the content is high for users (White et al. 2008). Thus, users receiving ads with potentially high number of "likes" from friends may show a lower reactance as the perceived value is high for these users due to multiple "likes" from friends. Note that we do not observe this effect for clicks but only for installs conditional on clicking. Users who have clicked the ad are more interested users and are likely to have a high perceived value of the application due to multiple friends' "likes" as compared to all users.

Our results point to the role of WOM signals in generating user response in different stages of their decision process. Specifically, our results show that WOM signals such as "likes" with social endorsement are not likely to draw users to the ad but are more likely to serve as a signal of quality for interested users, i.e., users who have clicked the ad. Some of the extant research (Godes and Mayzlin 2005, Liu 2006, Dellarocas et al. 2007, Gu et al. 2012, Lu et al. 2013) has linked WOM to performance without explaining the underlying driver for this outcome. Chevalier and Mayzlin (2006) and Chen et al. (2011) suggest that WOM increases the perceived valuation of products. However, it is not clear at which of the awareness, interest, or final outcome stage is the source of this relationship. Our results show that it plays a role only in the final outcome in the context of ads.

The coefficient for FF is negative and significant. This again points to the possibility of negative reactance as these users are expected to be more familiar with the app as they are friends of fans of the app company's Facebook page. Thus, even if they receive the ad without any social endorsement, they may still perceive it as manipulative and intrusive and may be less likely to install the app even after clicking the ad as compared to other users clicking the ad.

3.5.3 "Likes"

Table 3.6 provides the estimates for the parameters from Equation (3.9). In these results, higher bids lead to a higher number of "likes". This is reasonable because the bid is the primary input used by Facebook to determine the ad placement, and higher values of bids should result in ad impressions to high quality users. These users, in turn, are more likely to "like" the ad.

Finally, Table 3.7 shows the covariance between unobservables for

Click Rate, Install Rate, and "likes" from Equation (3.11). The covariance between the unobservables for Click Rate and Install Rate is statistically significant. This indicates that the unknown factors influencing consumer clicks also influence subsequent install behavior. The covariance between the unobservables for Install Rate and "likes" is also statistically significant. This suggests that the unobservables influencing "likes" are also influencing the install rate.

3.5.4 Homophily or Social Influence?

In our setup, only those users whose friends have already "liked" the ad can be targeted with social endorsement information. However, a user whose friend has "liked" the ad may inherently have a higher propensity to respond to the ad. Thus, any effect of social endorsement may just represent the effect of homophily. On the contrary, our results suggest that social endorsement has a negative effect on both click and install rates. Thus, our result suggests that users are responding to the social endorsement and not because they have similar characteristics. Additionally, user response to social endorsement conditional on clicking depends on the number of "likes". As a high number of "likes" is a proxy for high number of "likes" from friends, it suggests that the intensity of social signals does play a role in influencing the user decision. This provides some evidence that the effect of social endorsement is not entirely due to homophily but social influence could also be playing a role.

In order to investigate this further, we estimate our models only using the "FF" subsample, where we only target users that are friends of existing product "fans" and thus likely to see social endorsement information. The reason this subsample analysis can provide more information is because, by definition, the FF groups consist of product fans' friends, and homophily would suggest that these people might have similar tastes. By restricting our focus to these FF groups we can minimize potential homophily effects in our analysis. The FF-subsample results are shown in column (2) of Tables 3.8 and 3.9. We can see that the subsample result is qualitatively similar to the full sample result, with "likes" alone being insignificant for clicking and negative for the install decision, "Social" being negative for both install and clicking decisions, and $Social \cdot \sum_{j=1}^{t-1} \widehat{like_j}$ being positive for users' install decision. This suggests that at least some portion of the observed main results is due to social influence. Also note that the negative effect of "likes" on the install decision further confirms that these WOM signals are possibly inducing the negative reactance among interested users. As users in the FF-subsample are already aware of the app from their friends who are fans of the app company, one possibility is that a higher number of "likes" may be considered an additional validation of the quality of apps. However, our results show that is not the case. This also suggests that "likes" and social endorsement do not act as complements similar to the observation made by Dewan et al. (2015). In that case the positive effect of $Social \cdot \sum_{j=1}^{t-1} \widehat{like_j}$ is likely to be driven by the number of friends' "likes".

3.5.5 Robustness of Results

In this section we outline several steps we have taken to evaluate the robustness of our results.

Holdout Sample Analysis

As one test of robustness, we have attempted to verify the prediction accuracy of our results using a holdout sample. To do this, we consider data from the first 3 weeks as the estimation sample and data from the remaining two weeks as the holdout sample. We use mean absolute error (MAE) for daily Click Rate and Install Rate values at the aggregate level. The error values are reported in Table 3.10 and show that the model prediction accuracy is similar for both the estimation and holdout samples. This suggests that our model estimates are robust.

Model without Endogeneity Correction for "Likes"

We estimate our model without considering the endogeneity of "likes". Specifically, we estimated the click and install models using the actual cumulative "likes". The corresponding click and install results are shown in Tables 3.8 and 3.9 respectively and are qualitatively similar to our main results. "Likes" do not help the click rate or the install rate. Social endorsement has a negative impact on click rate as well as install rate. However, in the presence of a high number of "likes", social endorsement has a positive effect on the install rate.

Fixed Effect Model

We also use a fixed effects approach to control for the unobservable user group attributes. More specifically, we use the group dummies instead of group characteristics as controls in our main models. The corresponding click and install results are shown in Tables 3.8 and 3.9 respectively and are qualitatively similar to our main results. “Likes” do not help the click rate or the install rate. Social endorsement has a negative impact on click rate as well as install rate. However, the presence of social endorsement information attenuates the negative effect of “likes” on installs.

Table 3.1: Impact on WOM and Social Endorsement on Ad Performance

	Positive	**No Effect or Negative Effect**
Overall “Likes” or Popularity	“Likes” can act as WOM signal thereby improving consideration (Gupta and Harris 2010) and leading to higher valuation of the product (Chevalier and Mayzlin 2006; Chen et al. 2011)	“Likes” may be perceive as manipulative to promote ads and may lead to reactance (Brehm 1966, 1989; Clee and Wicklund 1980, White et al. 2008).
“Likes” with Social Endorsement	Higher clicks and installs due to homophily (McPherson et al. 2001) and social influence (Bakshy et al. 2012; Tucker 2012)	Personalized display of friends’ endorsement may increase the reactance (White et al. 2008) and might raise privacy concern (Tucker 2014) As endorsement improves targeting, prominent display of such ads in newsfeed might be perceived as intrusive by the consumers (Tucker 2014)

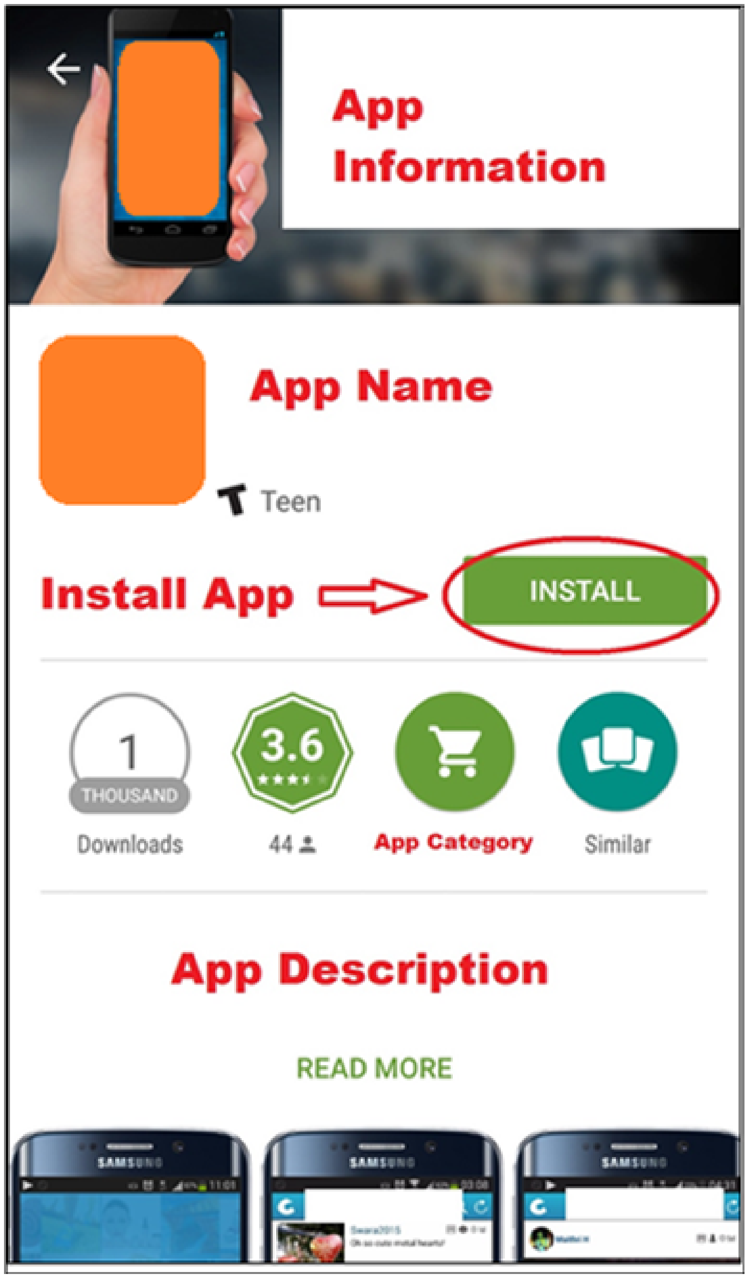

Figure 3.2: A sample mobile app page on Google Play

Table 3.2: Summary of 16 Target Groups

Group	Gender	Age	FF?	Experienced?
1	M	18-25	Yes	Yes
2	M	18-25	Yes	No
3	M	18-25	No	Yes
4	M	18-25	No	No
5	M	26-34	Yes	Yes
6	M	26-34	Yes	No
7	M	26-34	No	Yes
8	M	26-34	No	No
9	F	18-25	Yes	Yes
10	F	18-25	Yes	No
11	F	18-25	No	Yes
12	F	18-25	No	No
13	F	26-34	Yes	Yes
14	F	26-34	Yes	No
15	F	26-34	No	Yes
16	F	26-34	No	No

All sixteen groups target Android-based mobile device users who expressed interest in shopping.

Table 3.3: A Sample Observation in the Group-Based Hourly Data Set

Timestamp	2016-05-06T0700
Group	1
Social Impressions	78
Non-Social Impressions	31
Social Clicks	1
Non-Social Clicks	0
Installs	0
Cumulative Likes	153

Table 3.4: Group-Hour Data Summary Statistics

	Total	**Mean**	**Standard Deviation**	**Min**	**Max**
Social Impressions	405788	35.7208	126.7053	0	4584
Non-Social Impressions	304657	26.8184	107.4740	0	4338
Social Clicks	2020	0.1778	0.6091	0	14
Non-Social Clicks	2032	0.1789	0.6300	0	15
Installs	206	0.0181	0.1424	0	3
Likes	799	0.0703	0.5496	0	24
Bids	—	569.3396	262.3448	100	1000

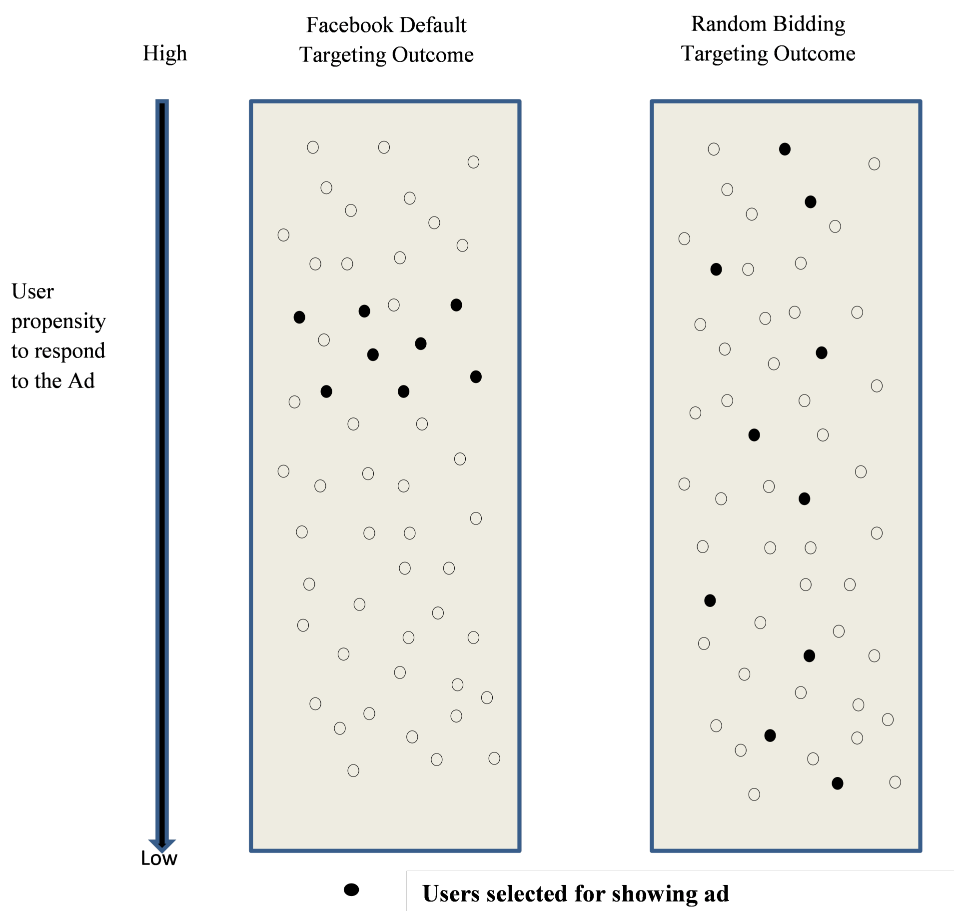

Figure 3.3: Comparison of Facebook's default targeting and our random bidding targeting

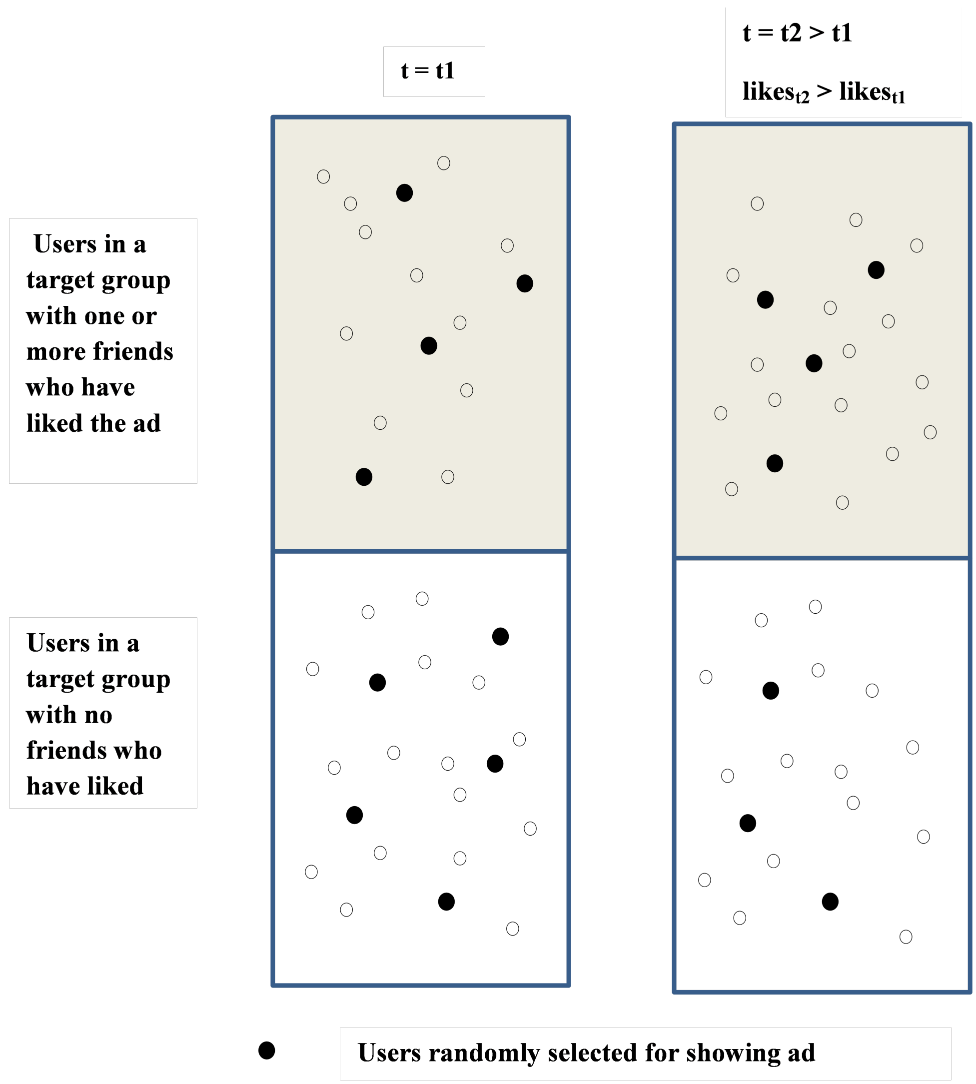

Figure 3.4: Users targeted in the experiment

Table 3.5: Empirical Results

	Click Rate	**Install Rate**
$\sum_{j=1}^{t-1} \widehat{like_j}$	-8.41e-05* (4.58e-05)	-0.001281** (0.000192)
Social	-2.57e-01*** (1.91e-02)	-0.171029*** (0.050040)
Social$*\sum_{j=1}^{t-1} \widehat{like_j}$	2.43e-05 (4.57e-05)	0.000936*** (0.000253)
FF	-4.35e-01*** (1.17e-02)	-0.202437*** (0.028350
Male	-6.67e-02*** (1.17e-02)	-0.343718*** (0.020295)
Age 26-34	7.35e-02*** (1.57e-02)	-0.036192 (0.063416)
Inexperienced	-1.81e-01*** (2.80e-02)	0.351258*** (0.016373)
Constant	-4.93*** (3.20e-02)	-2.337190*** (0.051304)

Standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.6: First-Stage Results

DV	Likes
Bids	8.771e-05*** (1.87e-05)
FF	-0.0194** (0.010)
Male	0.0304*** (0.010)
Age 26-34	-0.0591*** (0.010)
Inexperienced	-0.0555*** (0.010)
Constant	0.0339 (0.025)

Standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.7: Estimates for the Error Covariance Matrix Ω

	Install	Click	Like
Install	0.42741*** (0.00571)	-0.00938** (0.00392)	-0.00963** (0.00437)
Click	-0.00938** (0.00392)	0.42200*** (0.00562)	-0.00109 (0.00426)
Install	-0.00963** (0.00437)	-0.00109 (0.00426)	0.50225*** (0.00670)

Standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.8: Clicking Performance using Other Models

	Model without Interaction Term	**FF Subsample**	**Model without Endogeneity Correction**	**Model with Ad Fixed Effect**
$\sum_{j=1}^{t-1} \widehat{like_j}$	0.000 (0.000)	1.58e-04 (0.000186)	-9.97e-05 (8.25e-05)	-1.21e-03*** (7.01e-05)
Social	-2.66e-01*** (0.031)	-5.16e-01*** (0.073686)	-2.65e-01*** (4.15e-02)	-3.16e-01*** (3.99e-02)
Social*$\sum_{j=1}^{t-1} \widehat{like_j}$	—	4.11e-05 (0.000157)	1.02e-04 (7.56e-05)	2.52e-05 (9.86e-05)
FF	-4.66e-01*** (0.024)	—	-5.02e-01*** (3.79e-02)	—
Male	0.070** (0.029)	-5.39e-01*** (0.048395)	-1.05e-01*** (2.11e-02)	—
Age 26-34	9.8e-02*** (0.030)	-1.33e-01*** (0.043910)	1.87e-01*** (3.87e-02)	—
Inexperienced	-0.010 (0.041)	-1.41e-01* (0.072571)	-1.92e-02 (3.30e-02)	—
Constant	-4.754*** (0.054)	-4.74*** (0.075018)	-4.82*** (5.04e-02)	-5.19*** (2.95e-02)

Standard errors in parentheses: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.9: Install Performance using Other Models

	Model without Interaction Term	**FF Subsample**	**Model without Endogeneity Correction**	**Model with Ad Fixed Effect**
$\sum_{j=1}^{t-1} \widehat{like_j}$	-0.001** (0.000)	-0.00199*** (0.000746)	-0.00145*** (0.000264)	-0.002112*** (0.000159)
Social	0.508*** (0.103)	-0.66751*** (0.211360)	-1.11407*** (0.043391)	-0.532791*** (0.027743)
Social*$\sum_{j=1}^{t-1} \widehat{like_j}$	—	0.00208** (0.000849)	0.00189*** (0.000291)	0.000783** (0.000328)
FF	-0.242** (0.098)	—	-0.56204*** (0.040794)	—
Male	0.006 (0.064)	-0.05478 (0.166392)	0.14294* (0.076057)	—
Age 26-34	0.148 (0.092)	0.48176*** (0.186878)	-0.08168* (0.048476)	—
Inexperienced	0.032 (0.135)	0.65325*** (0.123423)	0.58372*** (0.090832)	—
Constant	-3.157*** (0.080)	-3.51254*** (0.189638)	-1.86524*** (0.051833)	-2.842397*** (0.061049)

Standard errors in parentheses: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 3.10: Prediction Accuracy for Estimation and Holdout Samples

	Click Rate (MAE)	**Install Rate (MAE)**
Estimation Sample	0.01	0.078
Holdout Sample	0.011	0.085

3.6 Discussion and Conclusion

Social media platforms such as Facebook can use WOM signals such as "likes" and endorsement by friends, or social endorsement, to influence user actions. These platforms also allow the firms to leverage these signals through social ads to spread their content and induce installs or purchases. However, the effectiveness of these signals in ads is not established. In this research we investigate the impact of WOM signals and social endorsement on the click and install performance of ads.

We conduct a randomized field experiment on Facebook to understand the effect of WOM and social endorsement information on the performance of ads. More specifically, we run a mobile app install ad campaign on Facebook to sixteen different targeted groups and capture the performance of ads in terms of impressions, clicks, "likes", and installs. We randomly target users by using random bids for our ad in the Facebook ad auction. We find that the WOM generated from "likes" alone has no positive impact on the user's decision to click and has a negative impact on the install decision. This can be attributed to the target users' reactance which stems from

Table 3.11: Model Fits

Models	**Marginal Density**
Main Model	-27541
Main Model without Social*Likes	-27723
Model with Fixed Effects	-27621

their tendency to associate the WOM with the ad as opposed to associating the WOM with the product itself—the WOM here would likely be perceived as an advertising strategy. We also find that social endorsement has a negative impact on the click and install performance. However, conditional on the ad being clicked, the negative effect of WOM on the install performance can be attenuated by the display of social endorsement. Since the interaction of WOM and social endorsement proxies for the WOM from friends, this finding implies that WOM from friends can reduce the negative effect of "likes".

These results are important for several reasons. Our results inform the design of ads on Facebook. Social media platforms such as Facebook are trying to monetize the ability of the embedded social graph to spread content and influence users to take actions such as installs and purchases using ads. Our results show that the use of WOM signals such as "likes" and social endorsement is not effective in generating response for install ads. In that case, the platform should determine alternative ways to utilize the social graph to elicit user response for such ads. For example, instead of showing such signals on the ad, the platform can work with the app platforms to show these signals when the interested users are making the install decision. Our results do show that social endorsement by several friends helps to reduce the negative effect of "likes". Thus, showing such signals on the install page can be a useful way to dis-associate the signals from ads and not generate reactance. Additionally, the platform could use such signals on ads to promote usage among users who have already installed the app. It is possible that such users may be less likely to show negative reactance to WOM signals on the ad as they have already installed the app. Facebook already had a mechanism in place to promote app usage.[20]

From a practitioner's perspective, our research provides insights into the efficacy of WOM and social endorsement information in social media ads. Marketing and academic studies have established the value of WOM and social endorsement in enabling product sales. However, our results question the efficacy of these signals in the context of ads. Our results suggest that marketing managers should ignore the "likes" generated by the ads as these do not translate into sales and may even hurt sales, and the presence of social endorsement information only reduces this negative effect and is not sufficient to overcome the negative effect. Thus, managers should not misinterpret these signals while evaluating the performance of ads. This is especially relevant for advertisers who cannot realize product sales directly through the Facebook interface and have to rely on interpret-

[20] https://developers.facebook.com/docs/app-ads/formats/engagement-ads

ing intermediate signals such as "clicks" or "likes" to infer the performance of their ads.

Our proposed experimental framework can also serve as a guideline for marketing managers in conducting social media advertising campaigns. Specifically, our approach enables marketing managers to identify core user groups from interacting with the Facebook platform and allows managers to assess the effectiveness of their Facebook advertising campaigns. We note that all required information described in our framework can be accessed by marketing managers, so marketing managers can follow our framework to conduct their own analysis. Using our methodology we are able to show whether or not the display of social endorsement is an appropriate strategy for their specific marketing campaign, as well as exploring how different user demographic groups react to social endorsement. As the Facebook platform reaches a wide audience, this can be a mechanism used by firms to learn about their target audience instead of conducting marketing survey or relying on external information sources such as Nielsen. For example, our approach can be combined with techniques such as multi-arm bandit optimization (Scott 2010) to quickly learn the characteristics of the relevant target audience for the firm from a large number of possible choices.

As with any empirical analysis our study has several limitations. While our results explain consumer behaviors at an aggregate level, the aggregate nature of our data limits our ability to account for the actions of individual consumers. This calls for future research using click stream data to empirically evaluate the behavior of different types of consumers in these environments. Our research does not consider the effect of ad content on the performance of social ads. While our study controls for the role of ad content, future research should consider the role of ad content and how it interplays with WOM and social endorsement. Since our research relies on a real life app and needs experimentation with the marketing campaign, we have managed to use only a single app in our setup. Future research should also validate the results for other apps. While our results can be applied to experience goods such as mobile apps, future research should also investigate the effect on search goods where it is possible to infer the quality using sources other than WOM signals. Finally, future research should also consider how other user actions such as comments and sharing influence the ad performance.

Chapter 4

Is Best Answer Really The *Best* Answer?

4.1 Introduction

Knowledge is an important asset for individuals and organizations alike. The need for knowledge is fundamental to one's career perspective and many other facets of life; organizations rely on knowledge to maintain successful operation and stay competitive (e.g., Nonaka 1991, Nonaka 1994). Information systems have been a key enabler to individual and organization's knowledge management process; therefore, the issues of knowledge contribution, knowledge exchange, and knowledge sharing have also been a major focus of research in the information systems community (e.g., Alavi and Leidner 2001, Grover and Davenport 2001, Sambamurthy and Subramani 2005). In recent years, the proliferation of electronic networks of practice (ENP) have received attention from academics (e.g., Wasko and Faraj 2005), and a stream of research dedicated to the study of ENPs has studied a variety of issues related to different aspects of knowledge management.

ENPs are defined as self-organizing networks where a potentially geographically dispersed group of users exchanges knowledge through a computer-mediated communication system; these users often share some common practice or interests, and user participation is voluntary (Brown and Duguid 2001, Wasko and Faraj 2005). A type of ENP is community-based question answering sites (CQA) on which anyone can post questions or provide answers to other questions. Popular CQAs such as Quora and StackExchange.com are convenient platforms for users to exchange knowledge with others.

Since the majority of ENPs and CQAs are open platforms, anyone can contribute to the knowledge exchange process, and, as a consequence, the quality of the user-generated content varies and can be vastly heterogeneous (Wasko and Faraj 2005). Quality level indicators can help CQA users locate high quality content more easily; meanwhile, the use of these indicators themselves can significantly affect the usefulness and effectiveness of the CQAs (Poston and Speier 2005, Meservy et al. 2014). Therefore, the choice of a good quality measure is crucial for CQAs to be effective.

Theoretically speaking, any user associated with a CQA platform can provide his/her own quality assessment of a given answer: (1) question asker-based measure: the original question asker can evaluate proposed answers; (2) reader/audience-based measure: some platforms implement a content quality assessment mechanism which allows all users to rate CQA content through voting or assigning numerical ratings (Poston and Speier 2005); (3) answerer-based measure: another way to assess content quality is to allow answer providers to place bets on their own answers based on how confident they are in their own answer quality (Savage 1971, Fang et al. 2007); (4) platform-based measure: the CQA platform can designate a group of moderators to evaluate answer quality for all answers. However, since the number of answers on CQAs can be very large, it is generally infeasible to implement a platform-based measure. In addition, it is not clear how one would implement a betting mechanism on CQAs to allow for answerer-based quality evaluation without a complete redesign of the platform. As a result, the most popular quality measures on CQAs are the question asker-based *best answer* and the audience-based *most popular answer*.

Since an answer on CQA platforms is written to address issues raised in the associated question posted by some question asker, the original question asker would seem to be the most qualified judge of the answer quality. In fact, popular CQA platforms seem to follow this rationale—for a given question, StackExchange, one of the most popular CQA platforms today, displays the question-asker-chosen "best answer" first, followed by all other answers ordered by the number of votes they each receive. This arrangement is intuitively appealing because the question askers themselves should know best whether their questions have been satisfactorily answered. Moreover, there have been studies in economics where people exhibit some types of “herding” behavior, where people choose to follow earlier people’s decisions rather than relying on their own judgment (e.g. Banerjee 1992, Bikhchandani et al. 1992). This can easily happen in a voting setting where users vote sequentially and are able to see previous people’s votes. Therefore, the answers that end up gathering the most votes can just be a result of the crowd’s herding behavior instead of individual users’ independent and well thought out decisions. This consequently reaffirms that it would seem most appropriate to use the question asker’s choice of best answer as the primary quality measure.

However, the concept of quality itself is subjective, and choosing the best answer is fundamentally a question asker’s judgment and decision making task. Research in psychology, behavior economics, behavior finance, and other related areas has shown that people are not always fully rational and are susceptible to numerous cognitive biases. For example, the classic paper by Tversky and Kahneman (1974) listed common heuristics people tend to employ in making judgments under uncertainty and discussed the resulting systematic and predictable biases; Rabin (1998) discussed areas where findings in psychology can inform the studies of economics; Barberis and Thaler (2003) offered a comprehensive overview of how psychology affects investors’ behaviors; in the information systems literature, Poston and Speier (2005) studied the anchoring and adjustment heuristics and the resulting biased judgment in content search tasks in knowledge management systems. Inspired by these studies, in the current paper we identify a new type of bias prevalent in CQA platforms, which we coined the *politeness bias*, and study how this bias affects question asker's answer quality evaluation. The finding of our study has profound theoretical and practical implications and will improve our understanding of the effect of cognitive biases on CQAs and knowledge exchange platforms in general.

Our study draws upon the social psychology and linguistics literature and analyzes linguistic features of answers on CQA. We examine how certain linguistic features could reflect the politeness level of a given answer; we then examine how these features might be correlated with existing quality measures. More specifically, we focus on the use of context-free linguistic features such as function words—words that are used to meet certain syntactic rules but do not themselves carry meanings—and study how it relates to users' quality assessment of the answer. In fact, in our analysis, we discard any content words, or keywords, and only retain function words, such as pronouns and articles, for our linguistic analysis, which is distinctly different from a conventional text analysis approach where only content words and keywords were used to analyze texts. Psycholinguistic studies suggest that the analysis of linguistic features such as the occurrence of function words is valuable because individuals' otherwise unobservable attributes could be reflected in their writing style and word usage (e.g., Pennebaker et al. 2003, Chung and Pennebaker 2007, Tausczik and Pennebaker 2010, Ludwig et al. 2014). For example, the percentage of first-person singular and first-person plural pronouns contained in a piece of writing might reflect the social status of the speaker or writer. Since social status might be correlated with one's expertise level, we expect that the use of words might also be correlated with the quality of the writing. As another example, Pennebaker (2011) suggested that the use of articles—"a", "an", "the"—is related to concrete nouns and concrete ideas, which might be positively correlated with content quality.

In addition, studies have shown that the knowledge transfer process depends crucially on communication (e.g., Ko et al. 2005). On CQAs, since the content quality is evaluated by community members, it is reasonable to expect that the way answer providers communicate through language will affect how their answers are perceived and evaluated by readers. Politeness theory, a concept widely studied by linguistics and communication researchers, describes how speakers and addressees communicate with each other to retain politeness while achieving their respective communication goals (e.g., Goffman 1955, Holtgraves 1986, Brown and Levinson 1987, Goldsmith and MacGeorge 2000, Morand 1996, Morand 2000). On CQAs, users who provide answers address issues raised in questions and convey ideas through textual responses, and politeness theory explains what type of answers are more likely to preserve the "face" (Goffman 1955, Brown and Levinson 1987) of addressees—that is, question askers—and will thus be received more favorably than answers that are more threatening to the addressee's face. Everything else equal, answers that preserve the addressee's face will likely be evaluated as high quality answers by the addressee. According to the politeness theory, the use of second-person pronouns—"you"—can help improve the clarity of a given text, but it can also make the text seem less polite. Since quality assessment is subjective, we expect the use of second-person pronouns to also affect how people evaluate answers on CQA platforms.

We develop a series of hypotheses based on psycholinguistic and politeness theories and empirically test these hypotheses in a data set obtained from StackExchange.com, a collection of community-based question answering sites with StackOverflow being its largest child forum. StackExchange is an ideal platform for our analysis because (1) StackExchange sites contain a large number of textual content contributed by community users, and (2) StackExchange implements two mechanisms—best answer decision and community votes—to reflect content quality, which allow us to examine the relationship between linguistic features and different measures of content quality. We first specify a random coefficient logit model to examine the relationship between linguistic features and the first measure of quality—whether or not a given answer will be chosen as the best answer. For the second measure of quality—the number of votes a given answer receives—we specify a question-specific fixed-effect negative binomial model to test our hypotheses. We also specify a panel fixed-effect model to examine the changes in linguistic features as one develops his or her reputation on the platform.

Our empirical results suggest that, through examining context-free linguistic features such as pronoun usage, percentage of articles contained in the text, length of the text, and number of words with more than six letters, a *politeness bias* indeed exists and strongly influences the question asker's choice of best answer. Specifically, different communication strategies might appeal to different audiences: while a more indirect and face-preserving approach might be less face threatening and thus more appreciated by the question asker due to its politeness, more direct and thus clearer answers might be more beneficial to the general audience. This implies that the design of answer display rules in current CQA platforms, if relying too much on the question asker's own answer assessment, might be inadvertently sacrificing quality for politeness, which would negatively impact the community's collective knowledge building process. As a consequence, even though question askers seem to be the most qualified individuals to determine whether the answers have successfully addressed their questions, the CQA platform should reconsider to what extent question askers' choices of best answers should be used in determining the quality level of answers so as to improve the effectiveness of the platform as a whole. More generally, through the identification and a better understanding of the politeness bias, we hope to raise both researchers' and practitioners' awareness in studying how this new form of cognitive bias could affect people's decision making process.

Our research contributes to the study of knowledge exchange in ENPs and CQAs, as well as the study of cognitive bias in the context of quality assessment. In addition, our research is related to the topic of strategic communication studied extensively in the economics literature. For example, signaling games (e.g., Spence 1973) study situations where a sender, whose type is unknown to the receiver, sends a message to a receiver, and the receiver subsequently chooses an action, which affects the payoffs of both the sender and the receiver. Our psycholinguistic approach can add another layer of information to existing economics models where the receiver can now imperfectly infer the type of the sender through analyzing the function words contained in the sender's message. We believe this new direction of research should be of interest to the knowledge management research community in the study of IT-enabled knowledge exchange, as well as the economics of IS community in the study of strategic communication.

4.2 Literature Review

4.2.1 IT-enabled Knowledge Exchange and CQAs

The ability to manage knowledge effectively is crucial to an organization's success (Nonaka 1994). Several IT-based systems such as knowledge management systems (KMS) have been designed to enable the creation, storage, transfer, and application of knowledge (Alavi and Leidner 2001). A detailed review of KMS and how they enable the knowledge management processes can be found in Alavi and Leidner (2001). Grover and Davenport (2001) described a process framework for the study of knowledge management.

Electronic network of practice (ENP) is a popular types of KMS. Distinct from traditional KMS, ENPs are self-organizing, and participation is voluntary. They allow users of similar professional practice or interests to exchange knowledge online, and these users do not necessarily need to have any offline connection or relationship with each other. Similarly, community-based question answering sites (CQA) provide an ideal medium for the voluntary communication and knowledge creation among users with similar interests, often in the form of question-answer pairs. Several studies in the information systems literature have looked at the user's contribution and sharing behaviors (e.g., Wasko and Faraj 2005), the platform's social structure and characteristics (Wasko et al. 2009), and the user's evaluation and adoption of knowledge (Meservy et al. 2014). More recently, Beck et al. (2014) studied enterprise social media-enabled ENP using a multi-level model of knowledge exchange. Broadly speaking, the process of posting questions in expectation of receiving useful answers is also likened to the concept of crowdsourcing (Tausczik and Pennebaker 2011, Huang et al. 2014).

4.2.2 Content Quality Assessment

Both ENPs and CQAs allow for open participation and voluntary knowledge contribution, often without a centralized screening process. A consequence of this is that there might be a plethora of interesting content, but the quality level can be vastly heterogeneous. As a result, knowledge seekers often have to rely on their own judgment to evaluate content and filter useful information based on quality (Meservy et al. 2014). Since users do not need to provide any proof of their capabilities or other credentials, it is difficult to assess a given user's expertise level other than relying on some types of reputation scores the platform generates for each user, often based on his/her past activities. Besides reputation scores, most CQAs also provide some types of content quality measure to help knowledge seekers evaluate the quality of the content. Poston and Speier (2005) studied how content ratings affect users' search, evaluation processes, and their decision performance. A number of computer science researchers have studied the problem of content quality prediction through social structure, user reputation, and texts, among other observable features (e.g., Adamic et al. 2008, Harper et al. 2008, Liu et al. 2008, Shah and Pomerantz 2010, Tian et al. 2013).

As mentioned in the introduction, two most popular quality measures used in CQAs are (1) best answer and (2) most popular answer. The main difference between these two quality measures is that an accepted answer is determined solely by the question asker, whereas answer popularity is determined collectively by community members. To make this distinction clear, we refer to question asker validation as question asker acceptance and refer to answer popularity as general audience reception. Notice that both quality measures are generated by users and are inherently subjective. In fact, mismatches between the true content quality and the given rating are common (e.g., Poston and Speier 2005). In this paper, we study how different quality measures would provide different estimates for the same piece of content and aim to explain this discrepancy through a context-free linguistic analysis approach.

4.2.3 Psycholinguistic Theory

Knowledge exchange is inherently a communication process. Factors such as the source credibility and both the sender and recipient's communication competence all affect the effectiveness of knowledge transfer (Ko et al. 2005). The use of language to a large degree determines how effective the communication would be. Several information systems researchers have studied computer-mediated communication (CMC), such as those happening on ENPs and CQAs, through textual and linguistic analysis (e.g., Abbasi and Chen 2008, Ludwig et al. 2014). In fact, the study of language and communication has been a major research area in social psychology, and more specifically, psycholinguistics. One goal psychologists hope to achieve is to be able to go beyond the understanding of semantics—that is, the contextual meaning of words and sentences—and explore in depth the psychological processes that speakers or writers go through at the time of speaking or writing. Several psycholinguistic studies have shown that some of these underlying psychological processes can be revealed by studying individuals' writing style (Pennebaker et al. 2003, Chung and Pennebaker 2007, Tausczik and Pennebaker 2010, Kacewicz et al. 2013). Among the most informative style indicators is the speaker or writer's use of function words—words that are used to fulfill a language's syntactic requirements, but the existence of themselves alone do not carry semantic purposes. Function words, or sometimes referred to as closed-class words, include pronouns, articles, conjunctions, prepositions, and auxiliary verbs. While there are only relatively few function words in a language—around 400 of them in English compared with the average native English speaker's vocabulary of 100,000 words (Chung and Pennebaker 2007), they account for roughly 50% of words used in a person's daily speech (Rochon et al. 2000, Chung and Pennebaker 2007).

The reason function word analysis can be informative of one's social and psychological processes has to do with how our brains process them differently from content words—words that serve semantic purposes and retain semantic properties independent of context. Friederici et al. (2000) showed through a functional magnetic resonance imaging study (fMRI) that the parts of brain activated when processing function words differ from those activated when processing content words, and some of these parts are highly related to both human language and social behaviors. (See Diaz and McCarthy 2009 for a list of brain and neuropsychological studies on the different brain activities related to processing function and content words.) This suggests that the use of function words and the use of content words might be from fundamentally different biological and neurological processes and are to certain extent orthogonal to each other.

Even more fascinating is the fact that, while people tend not to pay much attention to function words despite their prevalence in verbal and written communications, the use of them can actually reflect people's underlying psychological processes. Chung and Pennebaker (2007) summarized findings that relate the use of function words to social status, attention, negative affective states, and deception, among other types of health and psychological traits. Ludwig et al. (2014) studied the linguistic style matching (LSM) behavior among online community users by examining their use of function words.

4.2.4 Politeness Theory

Since an important function of language is its ability to facilitate communication, besides psycholinguistic processes, one's choice of words is also closely dependent on the specific communication goals he or she is trying to achieve. Brown and Levinson (1987), among a stream of sociolinguistic and communication research, proposed the *politeness theory* to explain how individuals choose appropriate words during a conversation to satisfy each other's face wants, where (1) a positive face refers to an individual's desire to have his or her identity approved by others, and (2) a negative face refers to the desire to have one's autonomy respected and actions unimpeded (Brown and Levinson 1987, Goldsmith and MacGeorge 2000).

Actions that might threaten one's face wants are referred to as face-threatening acts (FTAs). For example, during a communication process, the speaker's negative evaluation, insults, and criticism of the addressee are FTAs that threaten the addressee's positive face because they might be perceived as a disapproval of the addressee's current effort; the speaker's suggestions, recommendations, and advice for the addressee can be perceived as FTAs that pose threats to the addressee's negative face because these suggestions might impose constraints on the addressee's future behavior (Wilson et al. 1998). Notice that not all FTAs are equally face threatening—Brown and Levinson (1987) suggested that the degree of threats are moderated by the (1) power/status difference and (2) social distance between the speaker and the addressee, as well as (3) cultural specific ranking of impositions. For instance, the less power the speaker has compared with the addressee, the more threatening the FTA would be to the addressee—an employee's giving advice to his or her supervisor is likely to be more face threatening to the supervisor than one's giving advice to a friend. The study of face can be traced back to Erving Goffman's study on face-work (Goffman 1955). More recently, several politeness theory studies have been done in the domain of organizational studies (e.g., Morand 1996, Morand 2000). For example, Morand and Ocker (2003) studied the implication of politeness theory in the context of computer-mediated communication.

Harper et al. (2008) studied strategies which question askers on question answering sites can use in order to receive higher quality answers, such as providing higher payment to elicit answer provision on for-fee platforms and using rhetorical strategies such as showing gratitude. They found that the use of "thank you" contained in questions received differing appreciation levels with regard to different forum culture. Similarly, Burke and Kraut (2008) found that the level of politeness associated with questions can positively or negatively elicit the number of answers responding to those questions, depending on the forum-specific group norms. Both studies rely on experts to manually evaluate the politeness level of the questions, and therefore the data sets they used for analysis are considerably smaller than ours. Danescu-Niculescu-Mizil et al. (2013) recruited Amazon Mechanical Turks (AMT) to manually annotate messages and use the annotation results to train a politeness classifier. Using the trained classifier to analyze much larger data sets, they found that users' requests (questions) become less polite as their reputation increases.

It is worth noting that the aforementioned politeness studies focus primarily on the politeness level associated with questions, while we aim to study how the politeness level of answers would affect the answer quality level perception. More closely related to our research objective, Trees et al. (2009) studied how face-threat mitigation strategies used by teachers can increase students' judgments about the quality and usefulness of teachers' instruction. However, being polite might come at a cost: Brummernhenrich and Jucks (2013) studied how the politeness strategies used by tutors in educational instructions can hinder effective tutoring and negatively affect the student's learning process. In this paper we address this question by comparing the politeness level associated with answers and how different users, i.e., question asker or general audience, evaluate the quality of answers.

4.3 Hypothesis Development

As we have seen in previous sections, effective assessment of content quality is critical for CQAs. Besides measures such as reputation score, question asker validation of an answer, and overall answer popularity, psycholinguistic theory suggests that certain linguistic features might also be useful in assessing the quality of an answer. For example, the relative use of first-person plural pronouns—"we", "us", "our"—and first-person singular pronouns—"I", "me", "my", "mine"—can reflect the relative status of two people in an interaction (Chung and Pennebaker 2007). Specifically, the person who uses more first-person plural pronouns tends to be of higher status. Intuitively speaking, this might be because higher status people tend to be more collectively oriented (Kacewicz et al. 2014) and might be more comfortable including others in a statement. In addition, Dino et al. (2008) found that the differing patterns of pronoun usage between high-status and low-status individuals are also observed in online message boards. Since one's status is to certain extent positively associated with his or her expertise, we expect the user's usage of pronouns to reflect the quality of his or her contributed content. This association is consistent with the relationship between reputation and content quality found in Tausczik and Pennebaker (2011), where they included the degree of authoritativeness expressed in the writing in the measurement of users' online reputation. We first examine whether this relationship between status and first-person pronoun use exists on CQAs. Notice that the number of self-referencing pronouns needed in an answer is very likely going to depend on the specific question being answered. For example, questions that seek other people's personal experience would naturally receive answers that contain a large number of self-referencing pronouns. Therefore, we expect that, controlling for the number of self-referencing pronouns, one's use of first-person plural will increase as one's reputation increases, while the use of first-person singular decreases. Stated formally,

Hypothesis 1 (First-Person Pronoun Use) *Controlling for the percentage of* *self-referencing pronouns, the higher one's reputation is, the larger the relative proportion of first-person plurals to first-person singulars this person would use.*

We then examine whether the use of first-person plural pronouns, which we have hypothesized to increase as one builds reputation, could reflect the quality of one's contributed content. Notice that empirically this relationship is not straightforward. While higher reputation might be positively correlated with higher expertise and hence higher content quality, individuals' actual choices of words can be inherently heterogeneous and highly context-dependent. Therefore, a direct empirical comparison of first-person pronoun usages between answers might not necessarily reflect their relative quality level. We develop the following hypothesis to test this relationship:

Hypothesis 2A (First-Person Pronoun and Question Asker's Acceptance) *Controlling for the percentage of self-referencing pronouns, the larger the relative proportion of first-person plurals to first-person singulars the answer contains, the more positive the question asker acceptance will be.*

Hypothesis 2B (First-Person Pronoun and General Audience's Reception) *Controlling for the percentage of self-referencing pronouns, the larger the relative proportion of first-person plurals to first-person singulars the answer contains, the more positive the general audience reception will be.*

Our third hypothesis concerns the subjective nature of quality measures used on CQAs. Recall that CQA users ask questions and provide answers, which can be categorized as advice-seeking and advice-giving behaviors. According to the politeness theory, advice-giving is perceived by addressees—that is, the question askers—as FTAs that threaten the negative face, and different word use and different ways of describing an answer could result in different degrees of threats. FTAs that pose higher level of threats are more likely to be perceived as impolite and are thus less likely to be accepted by the addressee. Brown and Levinson (1987) summarized some common strategies to communicate negative face-threatening FTAs while also preserving politeness in advice-giving situations: (1) replacing the second-person pronoun—"you"—by the impersonal pronoun pronoun—"one", or (2) avoiding using "you" altogether—known as pronoun avoidance (Brown and Levinson 1987). These strategies are known as redressing—instead of making one's point directly and posing face threats to the addressee, the speaker or writer can use these strategies to lower the perceived face threat so as to increase the likelihood that the answer will be accepted. The disadvantage of using redressing strategies is that, while appearing more polite, FTAs delivered with redressing strategies can be less direct which could make the answer seem less clear. To summarize, in an advice-giving situation, a direct response can achieve a high level of clarity but will also pose a high level of face threat; a response using the redressing strategy will decrease the level of face threat but will be less clear. Therefore, we expect an answer's perceived quality level to be affected both by the level of face threat and the level of clarity.

Operationally, we use the proportion of second-person pronouns to first-person singular and plural pronouns as a measure of an answer's face threatening intensity. This is because second-person pronouns are perceived as rude and impolite (Murphy 1988), and the avoidance of second-person pronouns is a redressing strategy documented by Brown and Levinson (1987). Therefore, we expect the proportion of second-person pronouns to first-person pronouns present in an answer to reflect the level of the face threatening intensity. In a CQA setting, since the question asker is the direct recipient of the answer provided by the answer provider, we expect the question asker to experience the most face threat from the answer. Therefore, we hypothesize that the higher an answer's face threatening intensity is, the less likely it will be chosen as the best answer by the question asker. Formally speaking,

Hypothesis 3A (FTA on Question Asker) *Controlling for the percentage of all personal references/pronouns, the larger the relative proportion of second-person pronouns to first-person pronouns the answer contains, the more negative the question asker's acceptance will be.*

In contrast, since the general audience is not the direct recipient of the answer, we expect the face threatening level to the general audience to be minimal. Meanwhile, the general audience might be able to benefit from the clarity brought forth by the more direct answer. Therefore, we expect that a direct answer, while posing a high level of face threats, will provide a high level of clarity and will thus be more appreciated by the general audience. Formally,

Hypothesis 3B (FTA on General Audience) *Controlling for the percentage of all personal references/pronouns, the larger the relative proportion of second-person pronouns to first-person pronouns the answer contains, the more positive the general audience reception will be.*

In addition, the politeness theory suggests that the degree of face threat is moderated by the power difference between the speaker and the addressee. The higher the speaker's status, the lower the perceived face threat would be. In our case, if the answer provider's reputation level is much higher than that of the question asker, then the perceived face threat coming from the use of second-person pronouns might be smaller. In contrast, since second-person pronouns are much less face-threatening to the general audience, we do not expect the answer provider's reputation to moderate the effect of second-person pronouns. Formally speaking,

Hypothesis 4A (Power Distance, FTA, and Question Asker's Acceptance) *The higher the answer provider's reputation score is, the weaker the negative effect of second-person pronouns on the question asker's best answer choice would be.*

Hypothesis 4B (Power Distance, FTA, and General Audience's Reception) *The effect of second-person pronouns on the general audience reception is not moderated by the answer provider's reputation score.*

4.4 Data Collection and Measure Development

4.4.1 Stack Exchange

We obtain a rich dataset from StackExchange, a popular community-based question answering platform. StackExchange is a network of more than 140 diverse question answering sites, each focusing on a specific topic ranging from computer programming and mathematics to philosophy and physical fitness. Users on StackExchange sites can post questions, and others can provide answers to these posted questions. The question asker can then choose among different answers the one that he or she thinks best addresses the question and mark it as the best answer; this best answer will be displayed before all other answers. Similar to other CQAs, StackExchange contents are all generated and moderated by users. More specifically, any user can upvote any post, question or answer, if he or she thinks the post is of high quality; similarly, the user can downvote any posts if he or she thinks the post quality is low. For a given question, StackExchange sites order the corresponding answers by first displaying the best answer determined by the question asker and then displaying the other answers according to the number of votes these answers received. StackExchange sites also maintain a reputation system where users can accumulate reputation points based on the number of posts they have generated, the number of votes their posts have received to date, and their other activities on the site.

All historical StackExchange contents, including all posts, votes, and user public information, are distributed under the Creative Commons BY-

Table 4.1: Summary Statistics

Summary Statistics of StackExchange Sites			
Site	# Questions	# Answers	# Answers by Native Speakers
Academia	5,500	14,605	4,626
Askubuntu	125,894	213,303	48,501
Codereview	16,469	30,317	11,988
Mathematica	15,109	28,463	10,181
Philosophy	2,925	7,605	2,268
Serverfault	166,345	329,162	155,325
Stats	29,006	47,755	17,868
Tex	61,732	99,204	23,241
Total	**407,800**	**770,414**	**273,998**

SA 3.0 license, and the sites prepare and regularly release data dump for public downloads. We downloaded a subset of 8 StackExchange sites from the data dump released in September 2014 for our empirical analysis. These 8 sites cover diverse topics and are representatives of the overall StackExchange community. An overview of these sites is shown in Table 4.1. The downloaded dataset for each site includes post, vote, and user information; the post information contains post type (question or answer), post author, post title, post body in raw text format, post date, and number of votes received, for each post on a given site; the vote information contains vote date, vote type (upvote, downvote, etc.), and the corresponding post being voted; the user information contains user name, age, location, a list of all questions and answers posted by this user, and his or her reputation score as of September 2014, etc.

4.4.2 Linguistic Features

For any given post, we analyze the post content's linguistic features using the Linguistic Inquiry and Word Count (LIWC), a computerized text analysis package (Pennebaker et al. 2007). LIWC is a state-of-the-art psycholinguistic package used to analyze linguistic features of texts by calculating percentages of words that belong to different word categories, and each of these categories reflects different language and psycholinguistic dimensions of a written text. These categories include linguistic processes, such as function words usage and verb tenses; affective processes such as positive words and negative words; cognitive processes such as tentative, inclusive, and exclusive words, among many other word categories. Rigorous studies were conducted during the building of the LIWC dictionaries to ensure the reliability of these generated word categories. Of particular interests are categories corresponding to function words and pronouns as we are interested in exploring the relationships between pronoun use and content quality.

Notice that many of the StackExchange sites are related to computer programming and many posts contain mathematical derivations or code snippets. We carefully remove these math segments and code snippets and only retain verbal descriptions and explanations of these snippets for our linguistic analysis. This is because mathematical derivations and code snippets follow mathematical and programming syntax and are less informative in reflecting users' linguistic styles.

4.4.3 Answer Quality Measurement

Generally speaking, answers receiving a large number of upvotes are not necessarily chosen as best answers, although in many cases best answers are also upvoted the most. While best answers are often the most upvoted answers, these two measures might potentially reflect different quality dimensions. Notice that whether or not an answer will be chosen as the best answer is determined solely by the associated question asker. In other words, an answer will be chosen as the best answer as long as the question asker is satisfied with the answer. This would likely happen if the answer is able to solve the specific question posted. We define the choice of best answer as reflecting the question asker's acceptance. In contrast, any user, including the question asker, can cast votes on answers that they deem of high quality. Since a user can only vote once for a given answer, an answer will only accumulate a large number of votes if it is able to satisfy many users. We define answers with many upvotes as reflecting the general audience's reception in the sense that they are able to gain support from a large number of community members perhaps because they are able to provide responses general enough to be appreciated by users facing similar but potentially slightly different issues.

4.4.4 Reputation Score Reconstruction

The StackExchange data dump does not include the time series data of user reputation scores; instead, it only provides a snapshot of all users' reputation scores extracted in September 2014. Since we would like to obtain user's reputation score at the time of posting questions or answers, it is critical that we reconstruct this measure from the information available to us. We use StackExchange's rules of gaining and losing reputation points to reconstruct, for each post, the corresponding post user's reputation score at the time of posting by counting the number of questions asked, the number of answers provided, the number of votes received, the number of answers being chosen as best answers, as well as other measures that would affect the user's reputation score, for all posts in all 8 StackExchange sites.[1] We extend the original dataset by including this retrospective reputation score information for each post.

4.4.5 Vote Reconstruction

Similar to reputation score, the StackExchange data dump does not include any time series data on the number of votes a given post receives up to a specified time. This information is important for our analysis because, as described in the answer quality measurement section, we want to treat the answer popularity/general audience reception as a quality measure to explore the relationship between quality and linguistic features. However, due to the design of StackExchange sites, answers that are chosen as best answers will appear first on the webpage, immediately below the associated questions. As a consequence, the webpage's answer placement might make it easier for users to see and vote on best answers than on other answers. This display rule might further complicate the relationship between

[1] http://stackoverflow.com/help/whats-reputation

best answer choice—question asker acceptance—and the number of votes received—general audience reception. To alleviate this concern, for each post, we recompile the votes information and recount the number of votes received immediately before the best answer is chosen for the corresponding question. Through this approach we are able to measure the popularity of answers prior to the website's exogenous answer placement, which only takes place when the best answer is chosen.

4.4.6 Native Speaker Identification

Psycholinguistic research suggests that the use of function words is challenging for non-native speakers to master (Chung and Pennebaker 2007). Since our goal is to examine how linguistic features, and function word usage in particular, correspond to content quality, we choose to only consider textual data written by English native speakers to ensure our results are not contaminated by other aspects of language proficiency. We use a heuristic approach to determine whether a given user is likely a native English speaker: the user information section of the StackExchange dataset contains a location field where users can provide their location information in their public user profile. We use this information as a proxy of their country of origin and language preference and identify users from Australia, Canada, Ireland, New Zealand, United Kingdom, and United States to be included in our empirical analysis because users from these countries are more likely to be native English speakers.

4.5 Empirical Analysis

We construct a series of empirical models to test our hypotheses using the StackExchange dataset. Recall that our goal is to explore the relationship between linguistic features and content quality. As described in the data section, for each post in the data set, we obtain detailed information including post type, post text, post user information, and voting information. Our dataset allows us to use two related but distinct quality evaluation methods to assess the quality of each post: for each answer post we know (1) whether or not it is chosen as a best answer and (2) the number of votes it has received before the question asker chooses the best answer. Since by definition only answer posts are qualified to be chosen as best answers, we restrict our attention to only answer posts in our empirical analysis. We first discuss the relationship between linguistic features and content quality using question asker acceptance—whether a given answer is chosen as a best answer—as the quality measure.

4.5.1 Question Asker Acceptance Model

Recall that, for a given question, the question asker would choose from all candidate answers the one that he or she thinks best addresses the question. Therefore, the best answer decision is a multinomial choice problem. Also notice that, since questions are potentially very different from each other, it is highly likely that their corresponding answers will also be very different from each other, while exhibiting some level of similarity among answers that address the same question. We construct a random coefficient logit model, also known as the mixed logit model, to accommodate both the multinomial choice nature of best answer decisions and the potential similarity among affiliated answers. Formally, assume that answer a, which addresses question q, provides the question asker with utility U_{qa}, defined as

$$\begin{aligned} U_{qa} = {} & \beta_{q1} \cdot \left(\text{YOU\&WE\&I}\right)_{qa} + \beta_{q2} \cdot \widetilde{\text{YOU}}_{qa} + \beta_{q3} \cdot \widetilde{\text{WE}}_{qa} + \beta_{q4} \cdot \text{SHEHE}_{qa} \\ & + \beta_{q5} \cdot \text{THEY}_{qa} + \beta_{q6} \cdot \text{IPRON}_{qa} + \beta_{q7} \cdot \text{ARTICLE}_{qa} + \beta_{q8} \cdot \text{SIXLTR}_{qa} \\ & + \beta_{q9} \cdot \text{WORDCOUNT}_{qa} + \beta_{q10} \cdot \log\left(\text{ANSWERERSCORE}\right)_{qa} \\ & + \beta_{q11} \cdot \widetilde{\text{YOU}}_{qa} \cdot \log\left(\text{ANSWERERSCORE}\right)_{qa} \\ & + \beta_{q12} \cdot \text{ANSWERVOTE}_{qa} + \varepsilon_{qa}, \end{aligned} \tag{4.1}$$

where variables SHEHE, THEY, IPRON, ARTICLE, and SIXLTR are percentages of third-person singular pronouns, third-person plural pronouns, impersonal pronouns, articles, and words consisting of more than six letters that are present in answer a, respectively; variable (YOU&WE&I) is the sum of second-person, first-person plural, and first-person singular pronouns, measured in percentages; variable $\widetilde{\text{YOU}}$ is defined as the proportion of second-person pronoun with respect to (YOU&WE&I), and variable $\widetilde{\text{WE}}$ is defined as the proportion of first-person plurals with respect to all first-person pronouns. In other words, $\widetilde{\text{YOU}} = \text{YOU}/(\text{YOU\&WE\&I})$, and $\widetilde{\text{WE}} = \text{WE}/(\text{WE\&I})$; ε is an unobserved random term that is distributed according to a Type I extreme-value distribution.

The reason we include a composite term, (YOU&WE&I), as an independent variable is that, as described in our hypothesis development, we need to control for the inherent need for answers to use personal pronouns, which might be dependent on the specific question being asked. Once we control for the need for first- and second-person pronouns, we use the proportion of first-person plurals with respect to all first-person pronouns to compare the relative use of first-person plural to first-person singular usage, which is the goal of hypothesis H2A. Similarly, we use the proportion of second-person pronouns with respect to all first- and second-person pronouns to compare the relative use of second-person pronoun to first-person pronoun usage, which corresponds to hypothesis H3A.

Words belonging to each of these linguistic categories are based on the LIWC 2007 dictionaries. In addition, we also include WORDCOUNT, the number of words in a given answer, log(ANSWERERSCORE), the log-transformed reputation score of the answer provider at the time of posting this answer, $\widetilde{\text{YOU}} \cdot \log\left(\text{ANSWERERSCORE}\right)$, as well as ANSWERVOTE, the number of votes (upvotes – downvotes) the answer receives before the question asker chooses the best answer. The inclusion of the interaction term, $\widetilde{\text{YOU}} \cdot \log\left(\text{ANSWERERSCORE}\right)$, corresponds to hypotheses H4A and H4B.

The advantage of using a random coefficient logit model is the following: through the random coefficient setup, each question is allowed to value each linguistic feature differently, which means that different question askers can have different preferences. More specifically, we assume that the coefficients $\beta_{q1}, \beta_{q2}, \ldots, \beta_{q12}$ in Equation (4.1) are question-specific and normally distributed. The question asker is assumed to be utility-maximizing in the sense that he or she would pick the answer that gives him or her the highest utility to be the best answer—that is, for question q, $\text{BestAnswer}_q = \arg\max_a U_{qa}$. Let P_{qa} denote the probability that answer a associated with question q is chosen as the best answer. Since we do not observe the question asker's preference, we need to integrate over the distribution of all possible preferences the question asker could have. Then, P_{qa} can be shown to be

$$\mathrm{P}_{qa} = \int \frac{\exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta})}{\sum_{a=1}^{A(q)} \exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta})} f(\boldsymbol{\beta} \mid \boldsymbol{\theta})\, d\boldsymbol{\beta}, \tag{4.2}$$

where $\mathbf{X}_{qa}$ is a vector of answer a's linguistic features appeared in Equation (4.1), $A(q)$ is the number of answers that are associated with question q, $\boldsymbol{\beta} = (\beta_{q1}, \cdots, \beta_{q12})$, $f(\boldsymbol{\beta} \mid \boldsymbol{\theta})$ is the density function of $\boldsymbol{\beta}$, which represents the question asker's preference, and θ is a vector of parameters associated with the density function f.

4.5.2 General Audience Reception Model

We construct a different regression model to explore the relationship between linguistic features and general audience reception. Notice that the acceptance of an answer is measured by the number of votes (upvotes – downvotes) this answer receives up to the time when the best answer is chosen for the corresponding question. Since the dependent variable here is a count variable taking only integer values, we specify a negative binomial count model where the dependent variable is the number of votes up to when the best answer is chosen, and independent variables include linguistic features as well as the answer provider's reputation score at the time of posting the answer. Notice that, similar to the case of question asker acceptance, it is likely that answers corresponding to different questions will differ significantly from each other, and therefore there might exist some type of clustering among answers addressing the same question. Therefore, we specify a panel fixed-effect negative binomial model to address this issue. The likelihood of this model is as follows (Hausman et al. 1984).

$$\Pr\left(y_{q1}, y_{q2}, \ldots, y_{qA} \middle| \gamma_{q1}, \gamma_{q2}, \ldots, \gamma_{qA}, \sum_{a=1}^{A} y_{qa}\right) = \frac{\Gamma\left(1+\sum_{a=1}^{A} y_{qa}\right) \cdot \Gamma\left(\sum_{a=1}^{A} \gamma_{qa}\right)}{\Gamma\left(\sum_{a=1}^{A} y_{qa} + \sum_{a=1}^{A} \gamma_{qa}\right)} \prod_{a=1}^{A} \frac{\Gamma\left(\gamma_{qa}+y_{qa}\right)}{\Gamma\left(\gamma_{qa}\right) \cdot \Gamma\left(1+y_{qa}\right)}; \tag{4.3}$$

$$y_{qa} \sim \mathrm{NB1}(\gamma_{qa}, \delta_q); \tag{4.4}$$

$$\delta_q = \frac{\phi_q}{\exp\left(\mu_q\right)}, \tag{4.5}$$

where y_{qa} is the number of votes given to answer a which is associated with question q, and y_{qa} follows a negative binomial distribution with the following probability mass function:

$$\Pr\left(y_{qa} \middle| \gamma_{qa}, \delta_q\right) = \frac{\Gamma\left(\gamma_{qa}+y_{qa}\right)}{\Gamma\left(\gamma_{qa}\right) \cdot \Gamma\left(y_{qa}+1\right)} \left(\frac{\delta_q}{1+\delta_q}\right)^{\gamma_{qa}} \left(\frac{1}{1+\delta_q}\right)^{y_{qa}}; \tag{4.6}$$

$\Gamma(\cdot)$ is the Gamma function; μ_q is a question-specific fixed effect; ϕ_q is another question-specific fixed effect; γ_{qa} is parameterized as follows:

$$\begin{aligned}\gamma_{qa} = \exp\Big(&\beta_0 + \beta_1 \cdot (\text{YOU\&WE\&I})_{qa} + \beta_2 \cdot \widetilde{\text{YOU}}_{qa} + \beta_3 \cdot \widetilde{\text{WE}}_{qa} + \beta_4 \cdot \text{SHEHE}_{qa} \\ &+ \beta_5 \cdot \text{THEY}_{qa} + \beta_6 \cdot \text{IPRON}_{qa} + \beta_7 \cdot \text{ARTICLE}_{qa} + \beta_8 \cdot \text{SIXLTR}_{qa} \\ &+ \beta_9 \cdot \text{WORDCOUNT}_{qa} + \beta_{10} \cdot \log\left(\text{ANSWERERSCORE}\right)_{qa} \\ &+ \beta_{11} \cdot \widetilde{\text{YOU}}_{qa} \cdot \log\left(\text{ANSWERERSCORE}\right)_{qa} \\ &+ \beta_{12} \cdot \text{BESTANSWER}_{qa}\Big), \end{aligned} \tag{4.7}$$

where variables corresponding to linguistic features are defined similarly as in the question asker acceptance model; ANSWERERSCORE is the answer provider's reputation score at the time of posting the answer; BESTANSWER is a dummy variable which takes the value of 1 if the answer is picked as the best answer, and 0 otherwise. Therefore, it can be shown that

$$\mathrm{E}\left[y_{qa}\middle|\gamma_{qa}\right] = \frac{\gamma_{qa}}{\delta_a} = \frac{\gamma_{qa} \cdot \exp(\mu_q)}{\phi_q}; \tag{4.8}$$

$$\mathrm{Var}\left[y_{qa}\middle|\gamma_{qa}\right] = \frac{\gamma_{qa}}{\delta_q^2} = \frac{\gamma_{qa} \cdot \exp(2 \cdot \mu_q)}{\phi_q^2}, \tag{4.9}$$

where both μ_q and ϕ_q reflect question-specific fixed effects (Hausman et al. 1984, Greene 2007).

Notice that another option is to use a fixed-effect Poisson specification to model the number of votes each answer receives. However, Poisson regression model imposes the restriction that the conditional variance be equal to the conditional mean, known as the Poisson variance assumption

(Wooldridge 2010), which might be overly restrictive. Our data exhibits overdispersion, which refers to the phenomenon that the conditional variance is greater than the conditional mean. Therefore, we believe the use of fixed-effect negative binomial model is more appropriate for our empirical analysis.

4.6 Model Estimation and Results

4.6.1 Reputation Score and Linguistic Features

We first examine whether users' use of words would change as they accumulate reputation points over time. The increase of reputation points on StackExchange sites may come from asking questions, providing answers, or voting. A higher reputation level can reflect the user's active participation on the platform, as well as their expanded knowledge in the subject matter. Therefore, reputation points can be interpreted as a measure of status and expertise. Recall that psycholinguistic theories suggest an individual's status might influence the way one uses words. More specifically, high status users might tend to use more first-person plural pronouns than first-person singular pronouns as hypothesized in H1. To test this, we construct a panel model with a user fixed-effect where we regress the proportion of first-person plurals among all first-person pronouns on a user's reputation points and control variables including other context-free linguistic features. Notice that in this analysis we only consider users who have contributed answers and only examine the linguistic features contained in these answers.

The results are shown in Column (2) of Table 4.2. We can see that an increase in the answer user's reputation score is correlated with a higher proportion of first-person plural pronouns. This result is consistent with the psycholinguistic theory which states that an individual's higher status will correspond to more use of first-person plural pronouns. Therefore, H1 is supported.

Table 4.2: Changes in Linguistic Features vs Reputation Points

	(1)	(2)	(3)	(4)
DV	**ARTICLE**	$\widetilde{\textbf{WE}}$	$\widetilde{\textbf{YOU}}$	**Word Count**
log_answeruserscoretodate	**0.0009*****	**0.00465*****	**0.0154*****	**2.494*****
	(7.58e-05)	**(0.000646)**	**(0.000657)**	**(0.286)**
YOU&WE&I	-0.234***	0.164***	1.498***	-705.6***
	(0.00256)	(0.0225)	(.0225)	(9.736)
$\widetilde{\text{YOU}}$	-0.0035***	-0.0757***	—	90.24***
	(0.000316)	(0.000268)	—	(1.163)
$\widetilde{\text{WE}}$	0.00153***	—	-0.0787***	43.55***
	(0.000322)	—	(0.00279)	(1.206)
SHEHE	-0.251***	-0.428***	-2.306***	474.9***
	(0.0226)	(0.193)	(0.196)	(85.20)
THEY	-0.448***	-0.428**	-2.306***	474.9***
	(0.00849)	(0.0731)	(0.0745)	(32.28)
IPRON	-0.212***	-0.130***	-0.571***	-77.71***
	(0.00279)	(0.0243)	(0.0247)	(10.74)
ARTICLE	—	0.111***	-0.265***	-10.56
	—	(0.0233)	(0.0238)	(10.32)

(Continued on next page)

Table 4.2 – continued from previous page

	(1)	(2)	(3)	(4)
DV	**ARTICLE**	$\widetilde{\textbf{WE}}$	$\widetilde{\textbf{YOU}}$	**Word Count**
SIXLTR	-0.0598***	0.118***	-0.414***	-7.044
	(0.00142)	(0.0122)	(0.0123)	(5.373)
Word Count	-7.45e-07	0.000223***	0.000480***	—
	(7.28e-07)	(6.17e-06)	(6.18e-06)	—
Constant	0.121***	0.0519***	0.388***	110.1***
	(0.00071)	(0.00666)	(0.00671)	(2.92)
User Fixed-Effect	YES	YES	YES	YES
Observations	145,952	145,952	145,952	145,952
R-squared	0.117	0.016	0.089	0.087
Number of Users	12,828	12,828	12,828	12,828

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

For completeness, we also regress the proportion of second-person pronouns with respect to first-person pronouns, i.e., $\widetilde{\text{YOU}}$, the proportion of articles, and the total number of words, all on the user's reputation score and control variables. These results are shown in columns(3), (1), and (4) in Table 4.2. These results show that, as one develops his or her reputation, the proportion of second-person pronouns increases. Recall that the use of second-person pronouns tends to make answers clearer while posing a higher level of face threats to the question asker. Therefore, our result suggests that as one attains higher reputation level, he or she would focus more on the clarity rather than the politeness of his or her answer. Also, as reputation score increases, users would use more articles, which reflect concrete ideas and a categorical thinking style. Finally, we observe that users tend to write longer answers as their reputation scores increase. This could reflect an increase in users' knowledge and expertise in the subject matter, which in turn enables them to provide more detailed answers.

4.6.2 Question Asker Acceptance

In this section we briefly describe our estimation procedures and the results of hypothesis tests associated with the question asker acceptance model. As described in the previous section, we specify a random coefficient logit model to study the relationship between linguistic features and content quality, where the quality is measured by whether or not an answer is chosen as the best answer. Recall that Equation (4.2) is the probability of a given answer being chosen. Let y_{qa} be the indicator which takes the value 1 if answer a, which addresses question q, is chosen as the best answer, and 0 otherwise. Then the likelihood function of a question can be written as

$$S_q = \int \prod_{a=1}^{A(q)} \left[\frac{\exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta})}{\sum_{a=1}^{A(q)} \exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta})} \right]^{y_{qa}} f(\boldsymbol{\beta} \mid \boldsymbol{\theta})\, d\boldsymbol{\beta}. \tag{4.10}$$

Since the integral in Equation (4.10) does not have a closed form solution, we can approximate it through simulation (Train 2009). Specifically, for question q and a given $\boldsymbol{\theta}$, draw β from $f(\boldsymbol{\beta} \mid \boldsymbol{\theta})$ a total of R times and mark them as $\beta^1, \beta^2, \ldots \beta^R$. Then we can use these draws to approximate the likelihood of a given question, and the integral in Equation (4.10) becomes

$$\hat{S}_q = \frac{1}{R} \sum_{r=1}^{R} \prod_{a=1}^{A(q)} \left[\frac{\exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta}^r)}{\sum_{a=1}^{A(q)} \exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta}^r)} \right]^{y_{qa}}. \tag{4.11}$$

Our goal is to estimate the parameters, $\boldsymbol{\theta}$, of the density function $f(\boldsymbol{\beta} \mid \boldsymbol{\theta})$. Recall that $\boldsymbol{\theta}$ is normally distributed. Assume $\boldsymbol{\theta} \sim \mathrm{N}(\mu, \sigma^2)$. The mean, μ, and variance, σ^2, of $\boldsymbol{\theta}$ can be estimated through maximizing the joint simulated log-likelihood function of all questions:

$$SLL = \sum_{q=1}^{Q} \ln\left\{\hat{S}_q\right\} = \sum_{q=1}^{Q} \ln\left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{a=1}^{A(q)} \left[\frac{\exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta}^r)}{\sum_{a=1}^{A(q)} \exp(\mathbf{X}'_{qa} \cdot \boldsymbol{\beta}^r)} \right]^{y_{qa}} \right\}. \tag{4.12}$$

The maximum simulated likelihood estimator (MSLE) maximizes Equation (4.12). Note that during the model estimation, (1) questions that contain only one answer as well as (2) questions that did not have a corresponding best answer chosen will be dropped. This is because answers corresponding to these questions do not have within cluster variations in their outcome variable and are thus uninformative to the model estimation.

The random coefficient logit estimation results using 50 halton draws for each observation (question) to construct the simulated likelihood are shown in Table 4.3. Notice that in the random coefficient logit model we assume question-specific coefficients to be normally distributed across observations. Therefore, the reported results include the estimated coefficient mean of each independent variable, shown in column (1), as well as the estimated mean of standard deviation of each independent variable, shown in column (3). The standard errors of both the means and standard deviations are reported in columns (2) and (4), respectively. We begin by analyzing the relationship between pronouns and question asker acceptance. These estimates show that, allowing the effect to be heterogeneous across questions and controlling for the overall use of first-person pronouns, i.e. "we" and "I", the relationship between the use of first-person plurals in an answer and the likelihood that this answer will be chosen as the best answer is positive but insignificant. Therefore, hypothesis H2A is not supported. A possible explanation for this insignificant result is that there are only very few occurrences of "we" used in answers on StackExchange forums. In fact, $\widetilde{\mathrm{WE}}$ is zero for more than 80% of answers. Therefore, the lack of "we" in the data might have prevented us from seeing the effect of first-person plurals.

Moving to the analysis of second-person pronouns, controlling for the overall use of first- and second-personal pronouns, the more second-person pronouns used in an answer, the less likely it will be chosen as the best answer, and therefore hypothesis H3A is supported. This result is consistent with the politeness theory, which predicts that answers that pose higher threats to the addressee's face are considered less polite, which might negatively affect their chances of being accepted by the question asker. Recall the politeness theory also theorizes that factors such as power difference and social distance will moderate the perceived face threats. Our model explicitly include the answer provider's reputation score in the analysis, and the difference between answer providers' reputation scores reflects the difference in the question asker–answer provider power difference across different answer providers. Furthermore, since CQAs do not usually implement social structures such as friendships and users often use pseudonyms, we argue that the effect of social distance is not likely to play a significant role in our analysis. In sum, our result supports hypothesis H3A, which suggests that the politeness of an answer plays an important role in whether or not this answer will be accepted by the question asker.

In addition, from the random coefficient logit estimation results, as shown in Table 4.3, we also see that the coefficient associated with the term $\widetilde{\text{YOU}}$ * log(ANSWERERSCORE) is significantly positive. This means that, controlling for the use of second-person pronouns, the higher the answer provider's reputation, the weaker the second-person pronoun's negative effect on question asker acceptance is. In other words, while second-person pronouns are perceived by question asker as face threatening and are thus negatively correlated with question asker acceptance, this negative effect can be alleviated as long as the answer provider's reputation is high. Therefore, hypothesis H4A is supported.

Besides pronoun usage and its relationship with content quality, we also examine three other context-free linguistic features and their relationships to question asker acceptance. Pennebaker (2011) described a categorical thinking style as one that places more focus on objects, things, and categories. His findings suggest that several categorical thinking-related linguistic features, including the use of articles—"a", "an", "the"—and the use of "big words"—words that are longer than six letters, are correlated with learning outcome and academic performance. Our random-coefficient logit results suggest that both features are positively correlated with an answer being chosen as the best answer. In addition, the length of an answer, measured in the number of words an answer contains, is positively related with the likelihood that this answer will be accepted by the question asker.

It is worth noting that the reputation score of the answer provider is negatively correlated with the question asker's acceptance decision. Interestingly, if we consider the terms log(ANSWERERSCORE) and $\widetilde{\text{YOU}}$ * log(ANSWERERSCORE) together, we can see that, for someone with a high enough reputation score, the higher $\widetilde{\text{YOU}}$ is, the more likely it would be picked as the best answer; otherwise, for people with lower reputation, the higher $\widetilde{\text{YOU}}$ is, the less likely it would be picked as the best answer. In contrast, if $\widetilde{\text{YOU}}$ is high enough, then the higher the answer user's reputation, the more likely it would be picked as the best answer; otherwise, the higher the answer user's reputation, the less likely it will be picked as the best answer. These results suggest that, for high reputation answer providers, the negative impact coming from the face-threatening second-person pronouns can be countered by their high reputation: the question asker will find the impoliteness to be less of a concern or even as a signal of authority, and he or she will therefore appreciate the clarity brought forth by the answer.

We rerun the random coefficient logit model using 200 draws to see if the results are robust to the number of draws we use to compute the simulated likelihood, and the results, shown in Table 4.4, are qualitatively similar to the results using 50 draws.

Table 4.3: Random Coefficient Logit Result–50 Draws

DV	isBestAnswer			
	Mean		Standard Deviation	
	(1)	(2)	(3)	(4)
Variable	Mean of Coefficient	Standard Error of Coefficient	Mean of Coefficient	Standard Error of Coefficient
YOU&WE&I	-2.576***	(0.989)	3.635	(5.294)
$\widetilde{\text{YOU}}$	-2.196***	(0.2413)	1.138**	(0.550)
$\widetilde{\text{WE}}$	0.103	(0.084)	0.113	(0.390)
SHEHE	-6.846	(7.910)	2.763	(25.379)
THEY	-6.091***	(2.217)	1.298	(7.473)
IPRON	0.206	(0.810)	8.981**	(3.760)
ARTICLE	2.101***	(0.773)	1.209	(3.331)
SIXLTR	1.692***	(0.390)	0.671	(2.056)
ANSWERVOTE	1.120***	(0.056)	0.762***	(0.047)
log(ANSWERERSCORE)	-0.137**	(0.019)	0.015	(0.049)
$\widetilde{\text{YOU}}$*log(ANSWERERSCORE)	0.244**	(0.034)	0.117**	(0.046)
Word Count	0.006***	(0.0004)	0.005***	(0.0008)
N	24,553	24,553	24,553	24,553

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Random Coefficient Logit Result–200 Draws

DV	**isBestAnswer**			
	Mean		Standard Deviation	
Variable	(1) Mean of Coefficient	(2) Standard Error of Coefficient	(3) Mean of Coefficient	(4) Standard Error of Coefficient
YOU&WE&I	-2.732**	(1.087)	6.121	(6.933)
$\widetilde{\text{YOU}}$	-2.370***	(0.273)	1.551***	(0.555)
$\widetilde{\text{WE}}$	0.105	(0.090)	0.033	(0.453)
SHEHE	-6.993	(8.339)	1.028	(27.015)
THEY	-6.683***	(2.417)	5.667	(10.863)
IPRON	0.136	(0.882)	10.816***	(3.939)
ARTICLE	2.305***	(0.835)	2.115	(3.576)
SIXLTR	1.801***	(0.428)	1.893	(2.483)
ANSWERVOTE	1.215***	(0.075)	0.834***	(0.058)
log(ANSWERERSCORE)	-0.148***	(0.022)	0.078	(0.087)
$\widetilde{\text{YOU}}$*log(ANSWERERSCORE)	0.263***	(0.038)	0.123	(0.086)
Word Count	0.007***	(0.0005)	0.006***	(0.001)
N	24,553	24,553	24,553	24,553

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Overall, these findings suggest that linguistic features and the politeness level of the answer seem to be closely related to question asker's answer acceptance decision. We emphasize that these results do not themselves reflect causal relationships. Rather, we argue that answers which are perceived by question askers to be of highest quality consistently contain certain linguistic features in terms of their pronoun usage, as well as their perceived politeness.

Having established main correlations, we shift our attention to exploring whether or not any potential heterogeneity across StackExchange sites might affect the analysis. Roughly speaking, StackExchange sites can be categorized into one of the following two categories: (1) those featuring closed-ended questions and answers and (2) those featuring open-ended questions and answers. For example, a typical question posted to the "Tex" forum usually consists of a precise statement of the issue faced, and the corresponding answer is usually a precise, objective solution addressing the issue at hand. We refer to this type of sites as closed-ended sites. In contrast, questions posted to the "Academia" forum are more likely to be open-ended, such as those seeking career advice, research ideas, among others, and the answers are more likely to be subjective. We refer to these sites as open-ended. Therefore, we label "Askubuntu", "Codereview", "Mathematica", "Tex", "Stats", and "Serverfault" as closed-ended sites, and we label "Academia" and "Philosophy" as open-ended sites. To explore the possibility that answers posted to different types of sites might be evaluated differently, we include in our random coefficient logit model interaction effects between pronouns and site type, and the results are shown in Table 4.5. We can see that the direction and magnitude of linguistic features remain largely unchanged from earlier results, and the interactions between open-ended forums and pronouns are insignificant. Therefore, there does not seem to be any significant differences across different types of sites.

Finally, we point out that the number of votes an answer receives prior to the time the best answer is determined is positively correlated with this answer being chosen as the best answer, and this correlation remains significant after controlling for site type. In summary, hypotheses H3A and H4A were supported, while H2A was not.

Table 4.5: Random Coefficient Logit with Interactions–50 Draws

DV	**isBestAnswer**			
	Mean		Standard Deviation	
Variable	(1) Mean of Coefficient	(2) Standard Error of Coefficient	(3) Mean of Coefficient	(4) Standard Error of Coefficient
YOU&WE&I	-2.677***	(0.970)	5.668	(5.204)
(YOU&WE&I)*OpenEnded	-17.610*	(9.584)	15.247	(34.207)
$\widetilde{\text{YOU}}$	-2.072***	(0.231)	1.523***	(0.399)
$\widetilde{\text{YOU}}$*OpenEnded	1.005	(0.786)	3.338	(2.510)
$\widetilde{\text{WE}}$	0.096	(0.082)	0.076	(0.421)
$\widetilde{\text{WE}}$*OpenEnded	-0.108	(0.700)	1.096	(1.615)
SHEHE	-5.046	(8.091)	17.440	(23.303)
THEY	-6.092***	(2.147)	0.862	(8.248)
IPRON	0.436	(0.764)	3.372	(5.158)
ARTICLE	2.036***	(0.751)	1.428	(2.519)
SIXLTR	1.580***	(0.376)	1.652	(1.436)

(Continued on next page)

Table 4.5 – continued from previous page

Variable	Mean		Standard Deviation	
	(1) Mean of Coefficient	(2) Standard Error of Coefficient	(3) Mean of Coefficient	(4) Standard Error of Coefficient
ANSWERVOTE	1.051***	(0.050)	0.685***	(0.043)
log(ANSWERERSCORE)	-0.130***	(0.019)	0.023	(0.043)
$\widetilde{\text{YOU}}$*log(ANSWERERSCORE)	0.231***	(0.032)	0.014	(0.050)
Word Count	0.006***	(0.0003)	0.004***	(0.0007)
N	24,553	24,553	24,553	24,553

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.6.3 General Audience Reception

We next turn to exploring the relationship between linguistic features and general audience reception, measured by the number of votes an answer received. The fixed-effect negative binomial model we specified is estimated through a maximum likelihood approach (Hausman et al. 1984). Notice that the outcome variable—in our case, the number of votes each answer receives—needs to be nonnegative. Therefore, we discard answers that receive more downvotes than upvotes, which corresponds to less than 1% among all answers, and only use answers with nonnegative votes for our analysis. The estimation results are shown in column (1) in Table 4.6. Notice the number of observations is larger in the general audience reception analysis than in the question asker acceptance analysis. This is because in the case of general audience reception we are able to use all answers for the analysis, while in the case of question asker acceptance we had to remove answers whose corresponding question asker has not chosen a best answer. We can see that the sign of the coefficient estimate associated with $\widetilde{\mathrm{WE}}$ is positive but insignificant. This means that, controlling for the overall use of first-person pronouns, i.e. "we" and "I", we do not observe any significant relationship between first-person plurals used in an answer and the number of votes this answer would receive. Therefore, hypothesis H2B is not supported. We suspect this insignificant result might have come from the fact that "we" was rarely used. Recall that our fixed-effect specification controls for any question-specific answer characteristics, and therefore our analysis reflects the likelihood of an answer receiving more votes than its competing answers.

Next, controlling for the overall use of first- and second-person pronouns, the more second-person pronouns used in an answer, the more votes this answer will receive, and therefore hypothesis H3B is supported. This result is consistent with the politeness theory, which predicts that, while answers containing more second-person pronouns might pose higher threats to the addressee's face and are considered less polite, the perceived face threat of the general audience should be minimal since the general audience is not the direct recipient of the answer. In addition, more second-person pronoun use and more direct answers tend to be clearer, and therefore the combination of the lower face threat facing general audience and the improvement in clarity explains why more use of second-person pronouns is correlated with a higher number of votes an answer will receive. In addition, the coefficient associated with $\widetilde{\text{YOU}}$*log(ANSWERERSCORE) is insignificant, which is consistent with hypothesis H4B. This result is expected because second-person pronouns only pose minimal face threats to the general audience, and hence the answer provider's reputation level would not have made any significant difference.

Similar to the question asker acceptance model, our general audience reception model explicitly controls for the answer provider's reputation score, and the difference between answer providers' reputation scores should reflect the difference in the question asker–answer provider power difference across different answer providers. In sum, the combination of psycholinguistic and politeness theory is able to explain the empirical results of the general audience reception analysis.

We then analyze other linguistic features and their correlations with general audience reception, and we can see that the use of articles and long words are both positively correlated with general audience reception, similar to the results from the question asker acceptance model. Also notice that the coefficient of wordcount is positive, which means the longer an answer is, the more votes it will receive. Finally, our result shows that, unlike in the case of question asker acceptance, the answer provider's reputation score is positively correlated with the number of votes his or her answer receives. We can also see from the coefficient associated with BESTANSWER that answers accepted by question askers tend to receive more votes. This might be because StackExchange displays answers first by displaying the accepted answer then order the rest of the answers by the number of votes they received. Therefore, besides the possibility that answers accepted by question askers might be of high quality, this display arrangement might also contribute to the correlation. Our next model specification explores whether a StackExchange site is open- or closed-ended will lead to any change in results. These results are shown in column (2) in Table 4.6. We can see that, while the direction and magnitude of main linguistic features remain largely unchanged from column (1), open ended sites exhibit slightly different correlations between pronoun usage and general audience reception. To summarize, our general audience reception results are consistent with both psycholinguistic and politeness theories, and the specific type of StackExchange sites moderate the effect of pronoun usage on general audience reception.

In summary, hypotheses H1, H3A, H3B, H4A, and H4B are supported.

Table 4.6: Fixed-Effect Negative Binomial Results

DV	**# Votes Received**	
	(1)	(2)
	No Interactions	With Interactions
isBestAnswer	0.798***	0.798***
	(0.013)	(0.013)
YOU&WE&I	0.002	-0.050

(Continued on next page)

Table 4.6 – continued from previous page

	(1) No Interactions	(2) With Interactions
	(0.195)	(0.198)
(YOU&WE&I)*OpenEnded	—	1.243*
	—	(0.742)
$\widetilde{\text{YOU}}$	0.359***	0.375***
	(0.060)	(0.061)
$\widetilde{\text{YOU}}$*OpenEnded	—	-0.210***
	—	(0.077)
$\widetilde{\text{WE}}$	0.016	0.022
	(0.021)	(0.021)
$\widetilde{\text{WE}}$*OpenEnded	—	-0.091
	—	(0.084)
SHEHE	-2.006*	-1.976
	(1.202)	(1.204)
THEY	0.365	0.398
	(0.529)	(0.530)
IPRON	0.181	0.193
	(0.202)	(0.202)
ARTICLE	0.454**	0.450**
	(0.202)	(0.203)
SIXLTR	0.428***	0.436***
	(0.097)	(0.097)
log(ANSWERERSCORE)	0.120***	0.120***
	(0.005)	(0.005)
$\widetilde{\text{YOU}}$*log(ANSWERERSCORE)	0.002	0.002
	(0.008)	(0.008)

(Continued on next page)

Table 4.6 – continued from previous page

	(1)	(2)
	No Interactions	With Interactions
Word Count	0.0004***	0.0004***
	(3.59e-05)	(3.59e-05)
Question Fixed Effect	YES	YES
Number of Answers	45,285	45,285
Number of Questions	18,306	18,306

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.7 Conclusion

Content quality assessment for ENPs and CQAs is an important research topic in the IS literature. Since content provision on ENPs and CQAs is voluntary and open to anyone, the quality level is highly heterogeneous. A suitable quality assessment method can help users locate high quality content, but the choice of such method is difficult in itself because quality assessment is by nature a subjective task, and different assessment methods might get diverging results.

We considered using both the question asker's acceptance and the general audience reception as our quality measures, modelled by a random coefficient logit model and a fixed-effect negative binomial model, respectively. Our results using two different quality measures showed that the relative usage of second-person pronouns to first-person pronouns exhibits different correlations with respect to the perceived answer quality: a high percentage of second-person pronouns is correlated with a decrease in the likelihood of question asker acceptance, while the same feature is correlated with an increase in general audience reception, measured in votes. In addition, we showed that, as users build their reputation on the platform, they tend to use more second-person pronouns and more first-person plurals; their answers also tend to be longer and contain more complicated words. Moreover, while the use of first-person plural pronouns increases compared to that of first-person singular pronouns, this relative proportion of first-person plural to singular pronouns is not correlated with the quality measures we examined. Importantly, we found that the politeness level of an answer is correlated with both the question asker acceptance and general audience reception in opposite directions.

We are able to explain these results using psycholinguistic theories and the concept of face threatening acts from the politeness theory. More specifically, since the use of words, and function words in particular, can be reflective of one's unobservable social and psychological processes, by analyzing the answer provider's word usage expressed in an answer, one can examine to what degree the writing can reveal the writer's status, thinking style, as well as the expertise level, which is usually unobservable on CQA platforms. Therefore, in addition to the reputation score of the answer provider, one can use linguistic features to estimate the quality of an answer. Moreover, our empirical results inspired by the politeness theory are fruitful. Since CQAs are by nature advice-seeking and advice-giving platforms and hence more relevant to the negative face according to the politeness theory, we hypothesized that the use of second-person pronouns would be correlated with quality evaluation in distinct ways, depending on the specific quality evaluation method used. One important implication of this finding is that content quality assessment appears to be sensitive to the particular evaluation measure being used, and the subjective nature of quality assessment suggests that the politeness expressed in a message might be exerting significant influence on the quality evaluation, which we called the *politeness bias*. Whether or not the politeness of an answer should affect the quality of an answer is debatable and is up to the CQA to decide. If the goal of CQAs is to facilitate knowledge creation, then the "objective" quality of an answer would seem to be much more important than the politeness of the answer, although how we define an "objective" quality measure is a completely independent research question. In sum, through the identification of the *politeness bias* we discussed how it affects users' answer quality evaluation and, as a consequence, the overall effectiveness of knowledge exchange on CQA platforms.

This study integrates insights from psycholinguistic theories, the politeness theory, as well as the knowledge exchange literature and ENP, and suggests a number of interesting implications for researchers and practitioners.

4.7.1 Theoretical Contribution

The first theoretical implication of our study is that the degree of politeness expressed in a text plays an important role in quality assessment. More specifically, since CQA is essentially an advice-seeking and advice-giving platform, the majority of exchange happening on CQA is face threatening (Brown and Levinson 1987), and the politeness theory suggests that the question asker is susceptible to face threats contained in an answer. Our study shows that this susceptibility to face threats significantly impacts question asker's evaluation of answer quality: answers that exhibit a lower level of face threats are perceived by question askers as better answers than other answers. In contrast, the general audience would perceive a minimal level of face threats from the same answer than the question asker because they are not the direct recipient of the answer. Their quality assessment focuses mainly on the clarity of the answer. Since the use of second-person pronouns can be more face threatening but at the same time more direct and clearer, the general audience exhibits a liking toward answers that use more second-person pronouns. The theoretical implication of this is profound: although the politeness/face threatening level of an answer should not be directly related to its actual quality level, the agents evaluating its quality can be significantly affected by it and hence provide different quality assessments. This politeness bias provides an explanation of why different quality assessment methods would provide diverging assessments and also suggests the importance of using multiple assessment methods to get a comprehensive evaluation of content quality.

The subjective nature of quality evaluation, as reflected in our results, suggests that ENP and CQA researchers should carefully examine if any quality-independent factors—such as the politeness level—might be present in their research contexts which could confound the quality assessment process. Similarly, a CQA's choice of quality assessment methods should be carefully considered to facilitate effective knowledge exchange.

The implication of our identification and characterization of the politeness bias is profound. Several social science disciplines such as economics and finance have started to incorporate psychology theories and findings into their core theory building process to better reflect human behaviors (e.g., Rabin 1998, Barberis and Thaler 2003). Politeness bias can play an important role in interpersonal communication processes, and hence computer-mediated communications such as those taking place on ENPs and CQAs might all be influenced by it. Researchers are encouraged to explore more into the extent to which the politeness bias affects communication as well as the knowledge exchange process.

From a strategic point of view, our results might inspire users on CQAs to deliberately choose the wording of their answers in order to cater to both the question asker and the general audience so that they can obtain question asker acceptance as well as positive general audience reception. This can be studied in a game-theoretic framework and is related to the issue of strategic communication studied extensively in the economics literature.

In addition, our context-free linguistic analysis, especially that of function words, can offer insights to the quality of textual contents. This approach is drastically different from conventional text analysis such as sentiment analysis and topic models, in which only content words or key words were used as inputs, and function words were considered stopping words and thus dropped. In contrast, our approach focuses solely on function words and studies their relationships with quality assessment. The study of function words has its roots in psycholinguistic theories, and this paper shows its strength in CQA quality assessment. We believe that our context-free linguistic analysis framework can be used beyond CQA and knowledge exchange studies. For example, researchers can consider using a combination of a content words-focused approach and a function word analysis in the study of textual data for a more in-depth understanding of texts.

The study of context-free linguistic features and their application to CQA quality assessment is a relatively new research area. Although knowledge exchange, ENPs, and CQAs have been studied extensively in the IS literature, the combination of psycholinguistic theories, the politeness theory, and the knowledge exchange process is unique in the literature and provides numerous opportunities for future research. According to Whinston and Geng (2004), we believe our paper explores an exciting gray area of IS research where diverse disciplines such as IS, psychology, communication, and economics are integrated to provide an interdisciplinary understanding of the IT artifact. We encourage IS researchers to pursue research in this direction, and we believe more innovative projects along this research direction will be beneficial to the advancement of the IS field.

4.7.2 Managerial Implication

Our empirical investigation offers several implications to practitioners. First of all, our results show that quality assessment methods are subjective and can be affected by quality-independent factors such as the politeness bias. Therefore, in designing a new CQA for an organization or for public use, the design engineer needs to carefully choose quality assessment methods that can most effectively help users locate high quality content. They need to consider the specific user demographics for which the CQA is designed and evaluate the degree to which factors such as face threats would affect this specific user base. In addition, the CQA can remind its users of the importance of an accurate quality assessment and also emphasize that the accuracy of answers is the most important among all factors that might play a role in the evaluation. This way, a more objective content quality evaluation can be obtained.

In addition, how different quality assessment methods are displayed on the CQA can affect the way users interact with the platform. For example, on StackExchange forums, for a given question, the answer chosen by the question asker as best is displayed on top of all other answers, regardless of the number of votes the answers received. Although the question asker himself or herself may be the most qualified person to evaluate the answer quality based on how effective the answers addressed his or her question, his or her choice of best answer can be significantly affected by the politeness level of the answers. Therefore, the platform has to assess whether or not there are more effective ways to order the answers than the current order of display to help users locate high quality answers more efficiently. The easier it is for users to locate high quality content, the more valuable the CQA would be to users. Since CQAs such as StackExchange have become increasingly popular and are expected to become a major player in the technology and education scene,[2] it is imperative that CQAs understand the politeness bias and its consequence to ensure effective knowledge exchange and successful commercialization. In addition, a careful design of quality assessment methods can be highly beneficial to the growth of these platforms.

Finally, CQAs should consider using a context-free linguistic feature analysis to study user contributed content. Psycholinguistic theories suggest that users of different status tend to use function words differently. Therefore, a function word analysis might be useful in helping CQAs identify capable or influential individuals without the need to acquire personal information, which is often missing in CQAs and ENPs.

4.7.3 Limitation and Future Research

Our research is not without limitations. First and foremost, the current research does not directly establish any causality. Instead, we show the ex-

[2] A New York Times article lists StackExchange as one of the 50 companies that could be the next start-up *unicorns*, which will most likely reach a $1 billion valuation. See the following article for details: http://bits.blogs.nytimes.com/2015/08/23/here-are-the-companies-that-may-be-the-next-50-start-up-unicorns

istence of strong correlation between politeness bias and the answer quality evaluation and analyze the theoretical roots of the correlations using psycholinguistic theory and politeness theory. Thus, the underlying causality is indirectly inferred rather than directly tested. We encourage future research to explore the politeness bias more directly, preferably through experiments. Second, in the current study we focused on the politeness analysis of the question asker/addressee's negative face. In reality, the answer provider/writer would also be facing certain degrees of face threats, which would need to be examined in future studies.

That said, we believe our econometric analysis based on psycholinguistic and politeness theories in the context of CQAs is a first step toward understanding the roles context-free linguistic features and politeness bias play in reflecting and influencing the perceived content quality. Our proposed approach of linguistic analysis can be extended to other problem domains where content quality evaluation is important, such as job hiring, college admission, and internal employee reviews, among many others. Our results suggest that a close examination of people's writing can be informative of their ability and can be used as quality indicators, and knowledge exchange platforms can also benefit from using our proposed linguistic analysis approach with the potential impact of the politeness bias in mind.

Appendix A

Appendix for Chapter 2

Additional Empirical Robustness Checks

A.1 Using Proportion of Extreme Positive Tweets as Dependent Variable

As mentioned in the main text, a measure of positive sentiment manipulation used in Mayzlin et al. (2014) is the proportion of highly positive messages among all messages. We follow their approach in this section and use the proportion of highly positive movie tweets as the dependent variable, where highly positive tweets are defined as those with a raw sentiment score greater than or equal to 0.8 . We then run the same set of regressions as in Chapter 2.4. The results are shown in Table A.1, Table A.2, and Table A.3. We can see that the same results hold with the alternative dependent variable: (1) the proportion of highly positive tweets decreases after the movie

Table A.1: RD Estimates of the Effect of Movie Release on the Proportion of Extreme Positive Tweets (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; extreme positive tweets are defined as tweets that have their raw sentiment score ≥ 0.8)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
post	-0.0443*** [-24.74]	-0.0172*** [-5.753]	-0.0149** [-2.001]
duration		0.000264*** [8.017]	0.000438*** [7.860]
$duration^2$			-3.95e-07*** [-3.870]
$post * duration$		-0.000892*** [-11.15]	-0.00184*** [-6.344]
$post * duration^2$			1.37e-05*** [2.905]
constant	0.305*** [241.9]	0.296*** [174.8]	0.291*** [137.4]
Observations	52,829	52,829	52,829
AIC	-17814.39	-17950.12	-17969.29
BIC	-17796.63	-17914.60	-17916.01

release; (2) low budget movies exhibit a larger drop after movie release in the proportion of highly positive tweets than high budget movies; (3) major studio movies exhibit a smaller drop after movie release in the proportion of highly positive tweets than non-major studio movies.

A.2 Regression Discontinuity Design Results with Cluster Robust Variance

As discussed in Section 2.6.2, we run our regressions with cluster robust variances, and the results are shown in Table A.4.

A.3 Regression Discontinuity Design: Subsample Analysis

The main results based on the full data set, as shown in Table 2.2, suggest that the movie sentiment does experience a release shock on the release day. Here we identify several important subsamples and check if the results remain significant within the subsamples. Column (1) and (2) in Table A.6 correspond to the subsample of movies with a production budget less than 3 million US dollars. Comparing these columns with Column (2) and (3) in Table 2.2, we can see the results based on the low budget subsample remain qualitatively similar to the full sample coefficients, with the magnitude of the coefficients larger than that of the full sample. This is consistent with Hypothesis 2, which states that low budget movies tend to have a larger drop in Twitter sentiment than high budget movies.

Column (3) and (4) in Table A.6 correspond to the subsample of movies that are produced by any of the major studios listed in Table A.5. Comparing these columns with Column (2) and (3) in Table 2.2, we can see that the results are also qualitatively similar to the full sample, with the magnitude of the coefficients smaller than that of the full sample. This is consistent with Hypothesis 3, which states that major studio movies tend to have a smaller drop in Twitter sentiment than non-major studio movies.

A.4 Regression Results controlling for Advertising Expenditure

Since Twitter sentiment might be influenced by movie studios' advertising campaigns, we acquired detailed daily movie advertising expenditure data from Nielsen for additional analysis. Our advertising expenditure data set consists of movies' advertising spending across different advertising channels including (1) Business-to-Business, (2) Cable TV, (3) Local Magazine, (4) Local Newspaper, (5) National Internet, (6) National Sunday Supplement, (7) Network Radio, (8) Network TV, (9) Outdoor Ads, (10) Spanish Cable TV, (11) Spanish Network TV, (12) Spot Radio, (13) Spot TV, and (14) Syndicated TV. We were able to match our Twitter sentiment data with the Nielsen advertising data for 40 movies. In the first set of regression analysis we include movies' daily total ads expenditure as a control variable with movie fixed effects. The results are shown in Table A.7. We can see that the coefficients associated with the post variable remain significantly negative across different polynomial specifications, after controlling for daily ads expenditure. In the second set of regression analysis we include daily ads expenditure in the aforementioned 14 advertising channels as separate control variables with movie fixed effects, and the results are similar, as shown in Table A.8. These results suggest that the sentiment drop we observe on the release day remains significant after controlling for movies' advertising expenditure. Therefore, we argue that the sentiment drop is unlikely a result of movies' advertising campaigns.

A.5 Regression Results with Various Bandwidth Choices

To explore the sensitivity of the results to the length of the observation window, we run the regression specified in Equation (2.1) with different observation windows. The advantage of using a longer observation window is that longer observation windows yield more precise estimates as more observations are available to be used to estimate the regression coefficients. However, the linear specification in our regression equation is less likely to be a good approximation with a longer observation window, which can bias the estimate of the treatment effect, if the underlying conditional expectation is not linear. This is because a linear specification can only provide a close approximation over a limited range of values (Lee and Lemieux 2010).

Recall that our full sample analysis uses data starting 60 days prior to movie release up to 60 days after the release (±60 days). Therefore, we use ±30 days and ±50 days as the observation window, both of which are shorter than the original observation window, in order to obtain better approximations. The results are shown in Table A.9. Comparing these results with Table 2.2, we can see the results are robust to different choices of observation windows, with the longest observation window exhibiting the largest effects.

A.6 Using Topic Modeling to Measure Thematic Competition

A.6.1 Generative Model and Statistical Procedure

We define thematic competition as the closeness in theme between any given pair of movies. The intuition for this dimension is that, the more familiar a movie is to consumers, the more efforts these movie studios would have to exert in order to convince consumers to go see something that is less likely to be original or unconventional. A crude measurement of thematic competition is the genre a given movie belongs to. However, genre alone is unlikely to capture the degree of novelty a movie possesses. Therefore, we use a statistical technique broadly referred to as topic modeling to quantify the thematic originality of any given movie.

A type of topic model, the Latent Dirichlet Allocation (LDA), is a Bayesian statistical and information retrieval technique. The input of the topic model is a set of documents, and for the purpose of this project, we use the keyword information on the individual movie's IMDb entry for the analysis. Each movie typically contains several keywords that relate to its story line. The output of LDA is a list of K topics, where each topic is a distribution over all words appearing in any keyword document. In other words, topics differ from each other in the weights given to individual words contained in all keyword documents. Intuitively speaking, keyword documents that share similar words are likely to be from similar topics generated from the LDA model, and therefore an originality measure of a given movie can be constructed by checking the uniqueness of words contained in the corresponding keyword document. The variables in our LDA model are:

- K: number of topics
- D: number of keyword documents
- N: number of words contained in a keyword document. Different across documents.
- β_k: the k-th topic. Topics are distributions over all words that appear in any keyword documents.
- η: parameter for a Dirichlet distribution from which the topics are drawn.
- θ_d : topic proportion for keyword document d.
- α: parameter for another Dirichlet distribution from which the topic proportions $\theta_d's$ are drawn from.
- $Z_{d,n}$: topic assignment for word n in keyword document d. $Z_{d,n}$ is drawn from θ_d.

- $W_{d,n}$: word n in keyword document d drawn from the assigned topic $Z_{d,n}$.

The generative procedure of each movie's keyword document is described as follows.

1. Specify the number of topics, K.

2. For each specified topic, draw a distribution over all words appearing in any keyword document, from a Dirichlet distribution.

3. For each keyword document,

   (a) draw a distribution over all K topics from a Dirichlet distribution. This distribution determines the proportion of topics that each word appearing in a given keyword document is drawn based on.

   (b) for each word in a given keyword document, draw a topic assignment from the distribution specified in (a).

   (c) for each word, draw a word from the assigned topic determined in (b).

Notice that the only observed variables here are the keyword documents and the words contained in them. All other variables, including the list of topics and all topic assignments are latent. The output of LDA is a posterior topic distribution over all K topics for each keyword document.

The posterior distribution implied by LDA can be written as

$$\mathrm{P}(\theta_{1,\cdot,D}, Z_{1,\cdot,D;1,\cdots,N,}, \beta_{1,\cdots,K}, |W_{1,\cdots,D;1,\cdots N}, \alpha, \eta), \tag{A.1}$$

and posterior inferences for LDA models are usually conducted via mean-field variational inference which approximates intractable distributions with simpler distributions. The resulting posterior topic distributions of keyword documents can be compared through some document-similarity measure, which can then be used to construct a measure of thematic similarity. A detailed description of topic models can be found in Blei and Lafferty (2009).

A.6.2 Thematic Similarity and Competition Measure

As mentioned in the previous section, the input of the LDA model is the keyword documents we extract from IMDb, and the output is a list of K topics where each topic is a distribution of all keywords appearing in any of the keyword documents. For each movie m, the LDA model outputs a K-vector, $< T_{m,1}, T_{m,2}, \cdots T_{m,K} >$, where $T_{m,i}$ represents the weight of topic i associated with movie m. Also, for each movie m and topic i we have $0 \leq T_{m,i} \leq 1$ and $\sum_{k=1}^{K} T_{m,k} = 1$. The larger the weight of a given topic is, the more likely the movie is associated with that topic.

We can then use movies' topic vectors to construct pairwise thematic similarity measures for all pairs of movies. More specifically, for movie m and movie n, the LDA model outputs topic vectors T_m and T_n, respectively.

The thematic similarity between movie m and movie n is defined as follows:

$$sim(m,n) = \frac{T_m \cdot T_n}{\|T_m\| \cdot \|T_n\|} = \frac{\sum_{k=1}^{K} T_{m,k} \cdot T_{n,k}}{\sqrt{T_{m,1}^2 + T_{m,2}^2 + \cdots T_{m,K}^2} \cdot \sqrt{T_{n,1}^2 + T_{n,2}^2 + \cdots T_{n,K}^2}}. \tag{A.2}$$

Notice that $0 \leq sim(m,n) \leq 1$ for any pair of movie m and movie n. The larger $sim(m,n)$ is, the more similar movie m and movie n are.

For any movie m, we combine the thematic similarity measure with data on movie release dates from IMDb, and construct a competition measure for this movie as the number of other movies that were released within a specified time window, e.g. ± 1 or ± 2 months, and have a similarity measure greater than a specified threshold value. For example, in Table 2.5 we use the number of movies that were released ± 1 month within the focal movie, and have a similarity measure greater than 0.7. In other words, for movie m, the competition measure, $[\pm 1 month \& sim > 0.7]$(m), is defined as follows:

$$\begin{aligned} & [\pm 1 month \& sim > 0.7](m) \\ = & \sum_{movie\, n,\, n \neq m} \mathbb{I}[m, n \text{ released within 1 month}] \cdot \mathbb{I}[sim(m,n) > 0.7], \end{aligned}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

A.6.3 Uncovered Topics

Out of 482 movies released in 2012 and 2013, 439 of them contain keyword information. We run the LDA model with 10 topics on these movie keywords extracted from the movies' IMDb entries. We set the model parameters as follows: (1) topic smoothing parameter=0.01; (2) term smoothing parameter=0.01; (3) number of iterations = 5000. A list of most frequent words associated with the learned topics is shown in Table A.10.

A.6.4 Additional Competition Results

To check the robustness of the results shown in Table 2.5, we construct two other measures of competition: (1) number of movies released a month within the focal movie that have thematic similarity greater than 0.75, and (2) number of movies released two months within the focal movie that have thematic similarity greater than 0.7. The results are shown in Table A.11 and Table A.12, respectively.

Comparing Table 2.5 and Table A.11, we can see similar effects of competition on the sentiment drop with comparable magnitude. Therefore, the effect of competition on sentiment manipulation remains significant with a different choice the threshold value. Comparing Table 2.5 and Table A.12, we can see the effect of competition on sentiment drop using the ± 2 month competition window is much smaller than the ± 1 month competition window, but the effects are still significant. This result suggests that the closer the release days are between movies, the more drop in Twitter sentiment there will be, and thus the more manipulation we would expect.

A.7 Post-Release Sentiment Pattern

In this section we examine whether the post-release Twitter sentiment would catch up to the pre-release sentiment level. While Figures 2.1 and 2.2 show that the post-release sentiment of those movies did catch up to the pre-release sentiment level, Figure A.1 shows an example where the post-release sentiment did not go back to the pre-release level, which suggests that the manipulation in the pre-release period might have been high.

We also conduct an additional test and examine if the post-release sentiment would rapidly catch up in general. For each movie, we calculate the average sentiment of the time window [t = -10, t = -1] and the average sentiment of the time window [t = 31, t = 40]. The number in the bracket indicates the number of days after a movie's release, where a positive value means that day t is after the movie release, and vice versa. We conduct a test between the average sentiment of the time window [t = -10, t = -1] and the average sentiment of the time window [t = 31, t = 40], and find that the average sentiment of the time window [t = -10, t = -1] is statistically greater than the average sentiment of the time window [t = 31, t = 40] (p value < 0.01). This evidence shows that, in general, the sentiment does not go back to the original level after the release, which suggests that the manipulation level in the pre-release period is considerable.

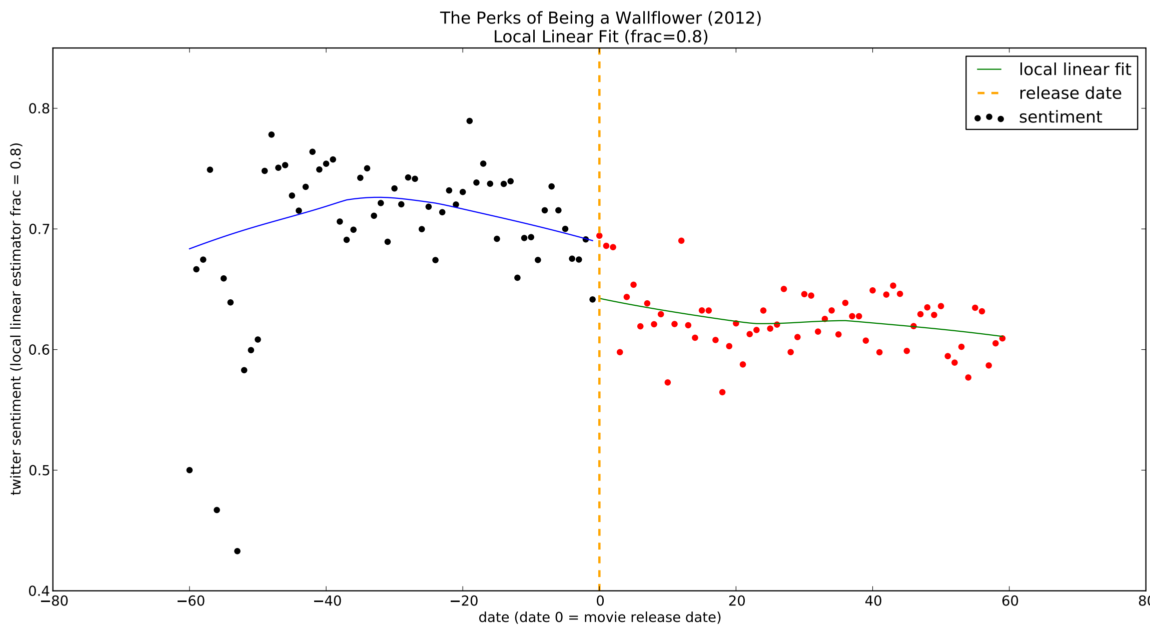

Figure A.1: Twitter Sentiment of Movie *The Perks of Being a Wallflower* (2012)

Table A.2: The Effect of Movie Budget on Sentiment Manipulation: Proportion of Extreme Positive Tweets as the Dependent Variable (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; extreme positive tweets are defined as tweets that have their raw sentiment score ≥ 0.8)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
$post$	-0.0442*** [-18.90]	-0.0164*** [-4.865]	-0.00896*** [-2.001]
$post*lowBudget$	-0.0233*** [-3.064]	-0.0174*** [-4.780]	-0.00910*** [-4.578]
$duration$		0.000265*** [8.029]	0.000440*** [7.876]
$duration^2$			-3.96e-07*** [-3.882]
$post*duration$		-0.000893*** [-11.16]	-0.00185*** [-6.352]
$post*duration^2$			1.37e-05*** [2.908]
$constant$	0.305*** [241.9]	0.296*** [174.8]	0.291*** [137.4]
Observations	52,829	52,829	52,829
AIC	-17812.39	-17948.35	-17967.63
BIC	-17785.75	-17903.95	-17905.47

Table A.3: The Effect of Being a Major Studio Movie on Sentiment Manipulation: Proportion of Extreme Positive Tweets as the Dependent Variable (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
$post$	-0.0464*** [-23.42]	-0.01964*** [-6.262]	-0.0178*** [-3.671]
$post * majorStudio$	0.0115** [2.475]	0.0125*** [2.708]	0.0128*** [2.772]
$duration$		0.000266*** [8.078]	0.000442*** [7.931]
$duration^2$			-4.00e-07*** [-3.913]
$post * duration$		-0.000895*** [-11.19]	-0.00185*** [-6.364]
$post * duration^2$			1.37e-05*** [2.909]
$constant$	0.305*** [242.0]	0.296*** [174.8]	0.291*** [137.4]
Observations	52,829	52,829	52,829
AIC	-17818.57	-17955.53	-17975.05
BIC	-17791.93	-17911.13	-17912.89

Table A.4: Robustness Checks of RD Estimates: Cluster Robust Variance (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
post	-0.0353*** [-16.12]	-0.0343*** [-14.65]	-0.0333*** [-13.95]	-0.0334*** [-14.18]
duration		9.12e-05*** [2.710]	0.000137** [2.561]	0.000146** [2.250]
$duration^2$			-1.08e-07 [-1.013]	-1.76e-07 [-0.659]
$duration^3$				6.83e-11 [0.297]
$post * duration$		-1.34e-05 [-0.495]	-4.36e-05 [-1.311]	-3.93e-05 [-0.921]
$post * duration^2$			1.60e-07** [2.113]	1.03e-07 [0.180]
$post * duration^3$				1.24e-10 [0.106]
constant	0.642*** [589.1]	0.639*** [444.5]	0.637*** [360.2]	0.637*** [339.2]
Observations	52,829	52,829	52,829	52,829
AIC	-84836.44	-84870.79	-84880.99	-84881.40
BIC	-84818.69	-84835.29	-84827.74	-84828.15

Table A.5: List of Major Movie Studios

Major Movie Studios
Warner Bros.
Warner Bros. Animation
Walt Disney Pictures
Walt Disney Feature Animation
Walt Disney Animation
DisneyToon Studios
Marvel Animation
Pixar Animation Studios
Universal Pictures
Twentieth Century Fox
Fox 2000 Pictures
Twentieth Century Fox Animation
Columbia Pictures
Paramount Pictures
Paramount Animation
Sony Pictures Animation
Twentieth Century Fox Film Corporation
Lucasfilm

Table A.6: Robustness Checks of RD Estimates: Subsamples (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; low budget movies are those with a production budget less than 3 million US dollars)

Variable	(1) Subsample: Low budget Polynomial Order 1	(2) Subsample: Low budget Polynomial Order 2	(3) Subsample major studio Polynomial Order 1	(4) Subsample major studio Polynomial Order 2
post	-0.0624*** [-7.290]	-0.0547** [-2.354]	-0.0198*** [-15.69]	-0.0173*** [-13.92]
duration	0.000108*** [2.953]	0.000290*** [4.673]	5.03e-05* [1.779]	0.000139*** [2.796]
$duration^2$		-4.56e-07*** [-3.634]		-2.30e-07* [-1.864]
$post*$ *duration*	-0.000515*** [-6.725]	-0.00138*** [-5.120]	-4.33e-05*** [-5.030]	-0.000105*** [-6.746]
$post*$ $duration^2$		1.22e-05*** [2.797]		2.60e-07*** [4.762]
constant	0.645*** [387.1]	0.640*** [299.5]	0.627*** [380.5]	0.624*** [312.0]
Observations	22,069	22,069	9,547	9,547
AIC	-29960.26	-29976.96	-18985.57	-19009.38
BIC	-29928.25	-29928.95	-18956.92	-18966.40

Table A.7: RD Estimates of the Effect of Movie Release on Twitter Sentiment Controlling for Total Daily Advertising Expenditure (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0353*** [-4.245]	-0.0309*** [-3.816]	-0.0219** [-2.187]	-0.0204** [-2.082]
$dailyAds$	-9.64e-06* [-1.705]	-1.21e-05*** [-3.407]	-1.35e-05*** [-3.811]	-1.46e-05*** [-4.020]
$duration$		-0.000102 [-0.354]	-0.000304 [-0.454]	0.000106 [0.0802]
$duration^2$			2.98e-06 [0.261]	-1.51e-05 [-0.267]
$duration^3$				2.03e-07 [0.320]
$post * duration$		-0.000173 [-0.669]	-0.00119 [-1.420]	-0.00565*** [-3.019]
$post * duration^2$			1.77e-05 [1.284]	0.000208** [2.602]
$post * duration^3$				-2.15e-06** [-2.358]
$constant$	0.617*** [129.2]	0.621*** [77.34]	0.624*** [73.63]	0.623*** [72.62]
Observations	4,774	4,774	4,774	4,774
AIC	-9594.498	-9597.957	-9603.491	-9619.327
BIC	-9581.556	-9572.073	-9564.665	-9567.559

Table A.8: RD Estimates of the Effect of Movie Release on Twitter Sentiment Controlling for Different Advertising Channel Expenditure (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2	(4) Polynomial Order 3
$post$	-0.0363*** [-4.044]	-0.0354*** [-3.594]	-0.0297** [-2.197]	-0.0232** [-2.094]
$duration$		-0.000162 [-0.540]	-0.000679 [-0.926]	-0.000606 [-0.427]
$duration^2$			7.49e-06 [0.626]	6.19e-07 [0.0106]
$duration^3$				9.12e-08 [0.141]
$post*duration$		-6.71e-05 [-0.235]	-0.000783 [-0.801]	-0.00541*** [-2.708]
$post*duration^2$			1.27e-05 [0.844]	0.000207** [2.508]
$post*duration^3$				-2.18e-06** [-2.328]
advertising channel controls	YES	YES	YES	YES
$constant$	0.617*** [110.5]	0.624*** [68.08]	0.631*** [62.78]	0.633*** [59.72]
Observations	4,774	4,774	4,774	4,774
AIC	-9578.483	-9579.769	-9584.660	-9601.487
BIC	-9474.948	-9463.292	-9455.241	-9459.127

Table A.9: Robustness Checks of RD Estimates: Bandwidth Choice (t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Bandwidth (-30, 30) Polynomial Order 1	(2) Bandwidth (-30, 30) Polynomial Order 2	(3) Bandwidth (-50, 50) Polynomial Order 1	(4) Bandwidth (-50, 50) Polynomial Order 2
$post$	-0.0268*** [-22.00]	-0.0282*** [-22.08]	-0.0303*** [-28.53]	-0.0310*** [-27.77]
$duration$	-0.000319*** [-5.110]	-0.00197*** [-7.989]	-0.000128*** [-3.868]	-0.000749*** [-5.750]
$duration^2$		5.50e-05*** [6.960]		1.25e-05*** [4.947]
$post * duration$	-1.85e-05 [-0.649]	0.000270*** [2.696]	9.43e-06 [0.624]	8.64e-05* [1.648]
$post * duration^2$		-1.11e-05*** [-2.965]		-1.80e-06 [-1.517]
$constant$	0.643*** [511.2]	0.651*** [370.5]	0.642*** [589.5]	0.647*** [425.3]
Observations	27,188	27,188	43,902	43,902
AIC	-54689.55	-54738.32	-78743.17	-78764.46
BIC	-54656.71	-54689.05	-78708.41	-78712.32

Table A.10: Sample Frequent Words in Each Topic (Parameters:10 topics; topic smoothing parameter=0.01; term smoothing parameter=0.01; iteration=5000)

Topic	1. Action/ Violence	2. Society	3. Metropolitan	4. War/ Crime	5. Sexual
Sample Words	camera murder hit character dialogue repeating severed shot motion year credits flashback punched falling pistol	flash spoken old credits political painting dance talking cross mental abroad footage heart medial arrest	york city new character says manhattan love fake thank secret suicide brooklyn film accent investigation	terrorist helicopter military u.s. nuclear gun airplane commando secret war american soldier cia shot washington	male sex gay man kiss female slur pool husband frontal rear lesbian panties wedding wearing

Topic	6. Action/ Police	7. Sci-Fi/ Space	8. Hero/ Animation	9. Family/ Young Adult	10. History/ Fairy tale
Sample Words	officer gun chase hero arts rifle shootout sniper security machine cop phone prison hostage jumping	space alien hero child warrior battle action spaceship sword credits comic peril violence game girl	animal credits ship film air opening sequel animation lifting talking helicopter survival hero villain rescue	daughter mother teenage school father sister year high character love child american loss brother family	supernatural falling magic horse century sword dead chested animal girl train water riding revenge surrealism

Table A.11: The Effect of Competition on Sentiment Manipulation, where competition is measured as the number of movies released 1 month within the focal movie and with thematic similarity > 0.75 (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
$post$	-0.0239*** [-15.93]	-0.0277*** [-12.14]	-0.0238*** [-13.62]
$post * [\pm 1 month \& sim > 0.75]$	-0.0114*** [-4.750]	-0.0115*** [-4.762]	-0.0121*** [-4.705]
$duration$		0.000155*** [4.186]	0.000318*** [4.396]
$duration^2$			-3.59e-07* [-1.941]
$post * duration$		-0.000253*** [-3.720]	-0.000544*** [-6.396]
$post * duration^2$			1.18e-06*** [5.002]
$color_i$	0.0233*** [6.265]	0.0230*** [6.164]	0.0228*** [6.111]
$runtime_i$	5.02e-05 [1.584]	5.50e-05* [1.737]	5.54e-05* [1.750]
$majorStudio_i$	-0.0195*** [-14.11]	-0.0195*** [-14.17]	-0.0196*** [-14.22]
$numTweets_i$	-6.81e-08*** [-9.333]	-7.32e-08*** [-10.25]	-7.44e-08*** [-10.45]
genre dummies	YES	YES	YES
$constant$	0.623*** [103.1]	0.618*** [101.1]	0.614*** [97.31]
Observations	48,139	48,139	48,139
AIC	-68097.23	-68161.45	-68216.02
BIC	-67860.12	-67906.78	-67943.79

Table A.12: The Effect of Competition on Sentiment Manipulation, where competition is measured as the number of movies released 2 months within the focal movie and with thematic similarity > 0.7 (Robust t-statistics in brackets, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Variable	(1) Polynomial Order 0	(2) Polynomial Order 1	(3) Polynomial Order 2
$post$	-0.0239*** [-14.07]	-0.0274*** [-15.21]	-0.0201*** [-12.48]
$post * [\pm 2month\&sim > 0.7]$	-0.00642*** [-3.816]	-0.00543*** [-3.800]	-0.00466*** [-3.735]
$duration$		0.000155*** [4.182]	0.000317*** [4.393]
$duration^2$			-3.59e-07* [-1.940]
$post * duration$		-0.000253*** [-3.708]	-0.000544*** [-6.388]
$post * duration^2$			1.18e-06*** [4.997]
$color_i$	0.0231*** [6.223]	0.0228*** [6.122]	0.0226*** [6.071]
$runtime_i$	5.23e-05* [1.645]	5.71e-05* [1.800]	5.76e-05* [1.814]
$majorStudio_i$	-0.0192*** [-13.93]	-0.0193*** [-13.99]	-0.0193*** [-14.04]
$numTweets_i$	-6.74e-08*** [-9.242]	-7.26e-08*** [-10.16]	-7.38e-08*** [-10.36]
genre dummies	YES	YES	YES
$constant$	0.623*** [103.0]	0.618*** [101.0]	0.613*** [97.22]
Observations	48,139	48,139	48,139
AIC	-68089.84	-68153.85	-68208.49
BIC	-67852.73	-67899.18	-67936.26

Appendix B

Appendix for Chapter 3

B.1 MCMC Algorithm

Our MCMC algorithm is based on the approach adopted by Rutz and Trusov (2011) and Rutz et al. (2012). The model can be written in the hierarchical form:

$$
\begin{aligned}
&b|X_{gt}^{b}, \Omega, \epsilon_{gt}^{b} \\
&\epsilon_{gt}^{b}|\{b\}, X_{gt}^{b}, \Omega \\
&\Omega|\{b\}, \left\{\epsilon_{gt}^{b}\right\}, S_{\Omega}, V_{\Omega}
\end{aligned}
$$

where $b = [\beta \quad \theta \quad \alpha]$ are coefficients, $X_{gt}^{b} = \left[X_{gt}^{\beta} \quad X_{gt}^{\theta} \quad X_{gt}^{\alpha}\right]$ are independent variables, and $\epsilon_{gt}^{b} = \left[\epsilon_{gt}^{\beta} \quad \epsilon_{gt}^{\theta} \quad \epsilon_{gt}^{\alpha}\right]$ are error terms in Equations (3.2), (3.5), and (3.9). We have used 0.001 as the initial value for elements of b. The MCMC algorithm is described below.

Step I: Draw $b = [\beta \quad \theta \quad \alpha]$

We use random walk Metropolis-Hastings algorithm for sampling b (Rossi and Allenby, 2005). The draws are accepted with a probability μ where

$$\mu = \min\left[\frac{l\left(U_{gt}^{new}\right)\exp\left[-\frac{1}{2}\left(like_{gt} - \alpha^{new}X_{gt}^{\alpha}\right)'\left(like_{gt} - \alpha^{new}X_{gt}^{\alpha}\right)\right]}{l\left(U_{gt}^{old}\right)\exp\left[-\frac{1}{2}\left(like_{gt} - \alpha^{old}X_{gt}^{\alpha}\right)'\left(like_{gt} - \alpha^{old}X_{gt}^{\alpha}\right)\right]}, \quad 1\right],$$

where $l(U_{gt})$ is the likelihood of clicks and installs

$$\begin{aligned} l(U_{gt}) = \prod_{g=1}^{G}\prod_{t=1}^{T}\Bigg\{&\left(\Lambda_{gt}^{Click}(1)\right)^{\#SocialClicks_{gt}} \\ &\cdot\left(1-\Lambda_{gt}^{Click}(1)\right)^{\left(\#SocialImpressions_{gt}-\#SocialClicks_{gt}\right)} \\ &\cdot\left(\Lambda_{gt}^{Click}(0)\right)^{\#NonSocialClicks_{gt}} \\ &\cdot\left(1-\Lambda_{gt}^{Click}(0)\right)^{\left(\#NonSocialImpressions_{gt}-\#NonSocialClicks_{gt}\right)} \\ &\cdot\left(\Lambda_{gt}^{Install}\right)^{\#Installs_{gt}}\cdot\left(1-\Lambda_{gt}^{Install}\right)^{\left(\#Clicks_{gt}-\#Installs_{gt}\right)}\Bigg\}, \end{aligned}$$

where

$$\Lambda_{gt}^{Click} = \frac{exp(U_{gt}^{Click})}{1 + exp(U_{gt}^{Click})},$$

and

$$\Lambda_{gt}^{Install} = \frac{exp(U_{gt}^{Install})}{1 + exp(U_{gt}^{Install})}.$$

$$
\begin{aligned}
U_{gt}^{Click^{new}} &= \beta^{new} X_{gt}^{\beta} + \epsilon_{gt}^{\beta} \\
U_{gt}^{Click^{old}} &= \beta^{old} X_{gt}^{\beta} + \epsilon_{gt}^{\beta} \\
U_{gt}^{Install^{new}} &= \theta^{new} X_{gt}^{\theta} + \epsilon_{gt}^{\theta} \\
U_{gt}^{Install^{old}} &= \theta^{old} X_{gt}^{\theta} + \epsilon_{gt}^{\theta}
\end{aligned}
$$

Step II: Draw $\epsilon_{gt}^{b} = \left[\epsilon_{gt}^{\beta} \quad \epsilon_{gt}^{\theta} \quad \epsilon_{gt}^{\alpha}\right]$

We use random walk Metropolis-Hastings algorithm for sampling ϵ_{gt}^{b}

$$
\epsilon_{gt}^{b^{new}} \sim \mathrm{N}(0, \Omega)
$$

The draws are accepted with a probability α where

$$
\alpha = \min\left[\frac{l\left(U_{gt}^{new}\right)\exp\left[-\frac{1}{2}\left(E_{gt}^{new}\right)'\left(E_{gt}^{new}\right)\right]}{l\left(U_{gt}^{old}\right)\exp\left[-\frac{1}{2}\left(E_{gt}^{old}\right)'\left(E_{gt}^{old}\right)\right]}, \quad 1\right],
$$

where $l\left(U_{gt}\right)$ is the likelihood of clicks and installs and is the same as defined in Step I;

$$
E_{gt} = \begin{bmatrix} \epsilon_{gt}^{\beta} \\ \epsilon_{gt}^{\theta} \\ \epsilon_{gt}^{\alpha} \end{bmatrix},
$$

$$
\begin{aligned}
U_{gt}^{Click^{new}} &= \beta X_{gt}^{\beta} + \epsilon_{gt}^{\beta^{new}} \\
U_{gt}^{Click^{old}} &= \beta X_{gt}^{\beta} + \epsilon_{gt}^{\beta^{old}} \\
U_{gt}^{Install^{new}} &= \theta X_{gt}^{\theta} + \epsilon_{gt}^{\theta^{new}} \\
U_{gt}^{Install^{new}} &= \theta X_{gt}^{\theta} + \epsilon_{gt}^{\theta^{new}}
\end{aligned}
$$

Step III: Draw Ω

$$
\Omega \sim \mathrm{IW}\left(\nu_{\Omega} + \mathrm{N}, \quad \sum_{g=1}^{G} \sum_{t=1}^{T} Y_{gt}' Y_{gt} + S_{\Omega} \right),
$$

where

$$
Y_{gt} = \begin{bmatrix} \epsilon_{gt}^{\beta} \\ \epsilon_{gt}^{\theta} \\ like_{gt} - \alpha X_{gt}^{\alpha} \end{bmatrix},
$$

N = No of observations, $\nu_{\Omega} = 10$, $S_{\Omega} = 10\mathrm{I}$.

Bibliography

[1] Abbasi, A., and Chen, H. 2008. "CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication," *MIS Quarterly* (32:4), December, pp. 811-837.

[2] Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. 2008. "Knowledge Sharing and Yahoo Answers: Everyone Knows Something," In *Proceedings of the 17th international conference on World Wide Web*, pp. 665-674.

[3] Adomavicius, G., Bockstedt, J.C., Curley, S.P., and Zhang, J. 2013. "Do Recommender Systems Manipulate Consumer Preferences? A Study of Anchoring Effects," *Information Systems Research* (24.4), pp. 956-975.

[4] Agarwal, A., Hosanagar, K., and Smith, M. D. 2011. "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets." *Journal of Marketing Research* (48.6), pp. 1057-1073.

[5] Anderson, E., and Simester, D. 2014. "Reviews without a Purchase: Low Ratings, Loyal Customers and Deception," *Journal of Marketing Research* (51.3), pp. 249-269.

[6] Alavi, M. and Leidner, D.E. 2001. "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly* (25:1), March, pp. 107-136.

[7] Angrist, J. D., and Krueger, A. B. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* (15.4), pp. 69-85.

[8] Aral, S., Muchnik, L., and Sundararajan, A. 2009. "Distinguishing Influence-based Contagion from Homophily-driven Diffusion in Dynamic Networks." *Proceedings of the National Academy of Sciences* (106.51), pp. 21544-21549.

[9] Aral, S., and Walker, D. 2011. "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks." *Management Science* (57.9), pp. 1623-1639.

[10] Aral, S., and Walker, D. 2014. "Tie Strength, Embeddedness, and Social Influence: A Large-scale Networked Experiment." *Management Science* (60.6), pp. 1352-1370.

[11] Bakshy, E., Eckles, D., Yan, R., and Rosenn, I. 2012. "Social Influence in Social Advertising: Evidence from Field Experiments." In *Proceedings of the 13th ACM Conference on Electronic Commerce*, ACM.

[12] Banerjee, A. V. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* (107.3), pp. 797-817.

[13] Bapna, R., and Umyarov, A. 2015. “Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks.” *Management Science* (61.8), pp. 1901-1920.

[14] Barberis, N. and Thaler R. 2003. ”A Survey of Behavioral Finance,” *Handbook of the Economics of Finance*, pp. 1053-1128.

[15] Beck, R., Pahlke, I., and Seebach, C. 2014. “Knowledge Exchange and Symbolic Action in Social Media-Enabled Electronic Networks of Practice: A Multilevel Perspective on Knowledge Seekers and Contributors,” *MIS Quarterly* (38:4), December, pp. 1245-1270.

[16] Bettman, J. R., and Park, C. W. 1980. “Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis.” *Journal of Consumer Research* (7.3), pp. 234-248.

[17] Bichler, M., Gupta, A., and Ketter, W. 2010. “Designing Smart Markets.” *Information Systems Research* (21.4), pp. 688-699.

[18] Bikhchandani, S., Hirshleifer, D. and Welch, I. 1992. “A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades.” *Journal of Political Economy*, (100.5) pp. 992-1026.

[19] Blei, D. M., and Lafferty, J. D. 2009. “Topic Models.” In M. Sahami and A. Srivastava, editors, Text Mining: Classification, Clustering, and Applications. Taylor and Francis Group 10:71.

[20] Brehm, J. W. 1966. “A Theory of Psychological Reactance.” New York: Academic Press.

[21] Brehm, J. W. “Psychological Reactance: Theory and Applications.” *Advances in Consumer Research* (16.1), pp. 72-75.

[22] Brown, P., and Levinson, S.C. 1987. *Politeness: Some Universals in Language Usage*. Vol. 4. Cambridge University Press.

[23] Brown, J. S., and Duguid, P. 2001. “Knowledge and Organization: A Social-Practice Perspective,” *Organization science* (12:2), pp. 198-213.

[24] Brummernhenrich, B., and Jucks, R. 2013. “Managing Face Threats and Instructions in Online Tutoring,” *Journal of Educational Psychology* (105:2), pp. 341.

[25] Burke, M., and Kraut, R. 2008. “Mind your Ps and Qs: the Impact of Politeness and Rudeness in Online Communities,” *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*.

[26] Cameron, A. C., Gelbach, J. B., and Miller, D. L. 2008. “Bootstrap-based Improvements for Inference with Clustered Errors.” *Review of Economics and Statistics* (90.3), pp. 414-427.

[27] Cameron, A. C. and Miller, D. L. 2015. “A Practitioner’s Guide to Cluster-robust Inference.” *Journal of Human Resources* (50.2), pp. 317-372.

[28] Campbell, M. C. "When Attention-getting Advertising Tactics Elicit Consumer Inferences of Manipulative Intent: The Importance of Balancing Benefits and Investments." *Journal of Consumer Psychology* (4.3), pp. 225-254.

[29] Campbell, M. C., and Kirmani, A. 2000. "Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent." *Journal of Consumer Research* (27.1), pp. 69-83.

[30] Centola, D. and Macy, M. 2007. "Complex Contagions and the Weakness of Long Ties." *American Journal of Sociology* (113.3), pp. 702734.

[31] Chen, Y., Liu, Y., and Zhang, J. 2012. "When Do Third-party Product Reviews Affect Firm Value and What Can Firms Do? The Case of Media Critics and Professional Movie Reviews." *Journal of Marketing* (76.2), pp. 116-134.

[32] Chen, Y., Wang, Q., Xie, J. 2011. "Online Social Interactions: A Natural Experiment on Word of Mouth versus Observational Learning." *Journal of Marketing Research* (48.2), pp. 238-254.

[33] Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* (43.3), pp. 345-354.

[34] Chintagunta, P. K., Gopinath, S., and Venkataraman, S. 2010. "The Effects of Online User Reviews on Movie Box Office Performance: Ac-

counting for Sequential Rollout and Aggregation across Local Markets." *Marketing Science* (29.5), pp. 944-957.

[35] Chung, C., and Pennebaker, J.W. 2007. "The Psychological Functions of Function Words," *Social Communication*, pp. 343-359.

[36] Cipriani, M., and Guarino, A. 2005. "Herd Behavior in a Laboratory Financial Market." *American Economic Review* (95.5), pp. 1427-1443.

[37] Clee, M. A., and Wicklund, R. A. 1980. "Consumer Behavior and Psychological Reactance." *Journal of Consumer Research* (6.4), pp. 389-405.

[38] Cohen, D., Mashruwala, R., and Zach, T. 2010. "The Use of Advertising Activities to Meet Earnings Benchmarks: Evidence from Monthly Data." *Review of Accounting Studies* (15.4), pp. 808-832.

[39] Coleman, J. S. 1990. Foundations of Social Theory. Cambridge, MA: Harvard University Press.

[40] Corts, K. S. 2001. "The Strategic Effects of Vertical Market Structure: Common Agency and Divisionalization in the US Motion Picture Industry." *Journal of Economics & Management Strategy* (10.4), pp. 509-528.

[41] Coy, P. "Could Bots and Spam Smother the Twitter IPO?" Bloomberg-Businessweek. N.p., 25 Sept. 2013. Web. 21 Jan. 2014.

[42] Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., and Potts, C. 2013. "A Computational Approach to Politeness with Application to Social Factors." *arXiv preprint*, arXiv:1306.6078. Chicago.

[43] De Bruyn, A., and Lilien, G. L. 2008. ”A Multi-stage Model of Word-of-mouth Influence through Viral Marketing.” *International Journal of Research in Marketing* (25.3), pp. 151-163.

[44] Dellarocas, C. 2006. “Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms.” *Management Science* (52.10), pp. 1577-1593.

[45] Dellarocas, C., Zhang, X. M., and Awad, N. F. 2007. “Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures.” *Journal of Interactive Marketing* (21.4), pp. 23-45.

[46] Dewan, S., Ho, Y., Ramaprasad, J. 2015. “Popularity or Proximity: Characterizing the Nature of Social Influence in an Online Music Community”. *Information Systems Research*, Forthcoming.

[47] Diaz, M. T., and McCarthy, G. 2009. “A Comparison of Brain Activity Evoked by Single Content and Function Words: An fMRI Investigation of Implicit Word Processing,” *Brain research*, 1282, pp. 38-49.

[48] Dino, A., Reysen, S., and Branscombe, N.R. 2009. “Online Interactions between Group Members Who Differ in Status,” *Journal of Language and Social Psychology* (28:1), March, pp. 85-93.

[49] Duan, W., Gu, B., and Whinston, A. B. 2008. “The Dynamics of Online Word-of-mouth and Product Sales—An Empirical Investigation of the Movie Industry.” *Journal of Retailing* (84.2), pp. 233-242.

[50] Edwards, S. M., Li, H., and Lee, J.-H. 2002. ”Forced Exposure and Psychological Reactance: Antecedents and Consequences of the Perceived Intrusiveness of Pop-up Ads.” *Journal of Advertising* (31.3), pp. 83-95.

[51] Elberse, A., and Anand, B. 2007. “The Effectiveness of Pre-release Advertising for Motion Pictures: An Empirical Investigation using a Simulated Market.” *Information Economics and Policy* (19.3), pp. 319-343.

[52] Elberse, A., and Eliashberg, J. 2003. ”Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures.” *Marketing Science* (22.3), pp. 329-354.

[53] Fang, F., Stinchcombe, M., and Whinston, A.B. 2007. “Putting Your Money Where Your Mouth Is—A Betting Platform for Better Prediction.” *Review of Network Economics* (6:2), 2194-5993.

[54] Forman, C., Ghose, A., and Wiesenfeld, B. 2008. “Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets.” *Information Systems Research* (19.3), pp. 291-313.

[55] Foutz, N. Z., and Jank. W. 2010. “Prerelease Demand Forecasting for Motion Pictures using Functional Shape Analysis of Virtual Stock Markets.” *Marketing Science* (29.3), pp. 568-579.

[56] Friederici, A. D., Opitz, B., and Yves von Cramon, D. 2000. “Segregating Semantic and Syntactic Aspects of Processing in the Human Brain:

an fMRI Investigation of Different Word Types,” *Cerebral Cortex* (10:7), pp. 698-705.

[57] Ghose, A., and Yang, S. 2009. ”An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets.” *Management Science* (55.10), pp. 1605-1622.

[58] Gilchrist, D. S., and Sands. E. G. 2015. “Something to Talk About: Social Spillovers in Movie Consumption.” *Journal of Political Economy*, forthcoming.

[59] Go, A., Bhayani, R., and Huang, L. 2009. ”Twitter Sentiment Classification Using Distant Supervision.” *CS224N Project Report*, Stanford: 1-12.

[60] Godes, D., and Mayzlin, D. 2004. “Using Online Conversations to Study Word-of-mouth Communication.” *Marketing Science* (23.4), pp. 545-560.

[61] Goes, P., Guo, C., and Lin, M. 2016. “Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange.” *Information Systems Research*, forthcoming.

[62] Goffman, E. 1955. “On Face-Work,” *Psychiatry* (18:3), pp. 213-231,

[63] Goh, K.-Y., Heng, C.-S., and Lin, Z. 2013. “Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User-and Marketer-generated Content.” *Information Systems Research* (24.1), pp. 88-107.

[64] Goldfarb, A., and Tucker, C. 2011. “Online Display advertising: Targeting and obtrusiveness.” *Marketing Science* (30.3), pp. 389-404.

[65] Goldsmith, D. J., and MacGeorge, E. L. 2000. “The Impact of Politeness and Relationship on Perceived Quality of Advice about a Problem,” *Human Communication Research* (26:2), pp. 234-263.

[66] Gong, S., and Huang, J. 2014. ”Embedded but not Enchanted: How Social Network Predicts Response to Product Recommendation.” Available at SSRN 2529634.

[67] Graham, J. R., Harvey, C. R., and Rajgopal, S. 2005. “The Economic Implications of Corporate Financial Reporting.” *Journal of Accounting and Economics* (40.1), pp. 3-73.

[68] Greene, W. 2004. *Econometric Analysis* 4th ed., Upper Saddle River, NJ, Prentice Hall.

[69] Greene, W. 2007. “Functional Form and Heterogeneity in Models for Count Data,” *Foundations and Trends in Econometrics* (1:2), pp. 113-218.

[70] Grembi, V., Nannicini, T., and Troiano, U. 2016. “Policy Responses to Fiscal Restraints: A Difference-in-Discontinuities Design.” *American Economic Journal: Applied Economics*, forthcoming.

[71] Grover, V., and Davenport, T. H. 2001. “General Perspectives on Knowledge Management: Fostering a Research Agenda,” *Journal of Management Information Systems* (18:1), pp. 5-21.

[72] Gu, B., Park, J., and Konana, P. 2012. "The Impact of External Word-of-mouth Sources on Retailer Sales of High-involvement Products." *Information Systems Research* (23.1), pp. 182-196.

[73] Gupta, P. and Harris, J. 2010. "How e-WOM Recommendations Influence Product Consideration and Quality of Choice: A Motivation to Process Information Perspective." *Journal of Business Research* (63.9), pp. 1041-1049.

[74] Harper, F., Maxwell, D. R., Rafaeli, S., and Konstan, J. A. 2008. "Predictors of Answer Quality in Online Q&A Sites." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 865-874.

[75] Hausman, J. A. 1975. "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models." *Econometrica*, (43.4), pp. 727-738.

[76] Hausman, J. A., Hall, B. H., and Griliches, Z. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica* (52:4), pp. 909-938.

[77] Hendricks, K., Sorensen, A., and Wiseman, T. 2012. "Observational Learning and Demand for Search Goods." *American Economic Journal: Microeconomics* (4.1), pp. 1-31.

[78] Holtgraves, T. 1986. "Language Structure in Social Interaction: Per-

ceptions of Direct and Indirect Speech Acts and Interactants Who Use Them." *Journal of Personality and Social Psychology* (51:2), pp. 305.

[79] Hu, N., Bose, I., Gao, Y., and Liu, L. 2011. "Manipulation in Digital Word-of-mouth: A Reality Check for Book Reviews." *Decision Support Systems* (50.3), pp. 627-635.

[80] Hu, N., Bose, I., Koh, N. S., and Liu, L. 2012. "Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments." *Decision Support Systems*, (52:3), pp. 674-684.

Hu, N., Zhang, J., and Pavlou, P. A. 2009. "Overcoming the J-shaped Distribution of Product Reviews." *Communications of the ACM* (52.10), pp. 144-147.

[81] Huang, Y., Vir Singh, P., and Srinivasan, K. 2014. "Crowdsourcing New Product Ideas Under Consumer Learning," *Management Science* (60:9), pp. 2138-2159.

[82] Imbens, G. W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* (86.1), pp. 4-29.

[83] Iyengar, R., Han, S., and Gupta, S. 2009. "Do Friends Influence Purchases in a Social Network?." Harvard Business School Marketing Unit Working Paper 09-123 (2009).

[84] Iyer, G., Narasimhan, C., and Niraj, R. 2007. "Information and Inven-

tory in Distribution Channels." *Management Science* (53.10), pp. 1551-1561.

[85] Joshi, A. M., Hanssens, D. M. 2009. "Movie Advertising and the Stock Market Valuation of Studios: a Case of 'Great Expectations?' " *Marketing Science* (28.2), pp. 239-250.

[86] Kacewicz, Ewa, Pennebaker, J. W., Davis, M., Graesser, A. C. 2014. "Pronoun Use Reflects Standings in Social Hierarchies," *Journal of Language and Social Psychology* (33:2), March, pp. 125-143.

[87] Katona, Z., Zubcsek, P. P., and Sarvary, M. 2011. "Network Effects and Personal Influences: The Diffusion of an Online Social Network." *Journal of Marketing Research* (48.3), pp. 425-443.

[88] Ketter, W., Collins, J., Gini, M., Gupta, A., and Schrater, P. 2012. "Real-time Tactical and Strategic Sales Management for Intelligent Agents Guided by Economic Regimes." *Information Systems Research* (23.4), pp. 1263-1283.

[89] Ko, D.-G., Kirsch, L. J., and King, W. R. 2005. "Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations," *MIS Quarterly* (29:1), March, pp. 59-85.

[90] Kothari, S. P., Shu, S., and Wysocki, P. D. 2009. "Do Managers withhold Bad News?" *Journal of Accounting Research* (47.1), pp. 241-276.

[91] Krider, R. E., Li, T., Liu, Y., and Weinberg, C. B. 2005. "The Lead-lag

Puzzle of Demand and Distribution: A Graphical Method Applied to movies." *Marketing Science* (24.4), pp. 635-645.

[92] Lahiri, K., and Schmidt, P. 1978. "On the Estimation of Triangular Structural Systems." *Econometrica* (46.5), pp. 1217-1221.

[93] Lee, D., Hosanagar, K., and Nair, H. 2014. "The Effect of Social Media Marketing Content on Consumer Engagement: Evidence from Facebook." Available at SSRN 2290802.

[94] Lee, D. S., and Lemieux, T. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* (48.2), pp. 281-355.

[95] Leonard, A. "Facebook's Black Market Problem Revealed." *Salon*. Salon Media Group, 14 Feb. 2014. Web. 26 Apr. 2014.

[96] Li, X., and Wu, L., "Herding and Social Media Word-of-Mouth: Evidence from Groupon" (September 25, 2014). Available at SSRN: http://ssrn.com/abstract=2264411 or http://dx.doi.org/10.2139/ssrn.2264411

[97] Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue." *Journal of Marketing* (70.3), pp. 74-89.

[98] Liu, Y., Bian, J., and Agichtein, E. 2008. "Predicting Information Seeker Satisfaction in Community Question Answering," In *Proceedings of the 31st annual international ACM SIGIR Conference on Research and Development in Information Retrieval*.

[99] Loten, A. “Yelp Reviews Brew a Fight over Free Speech vs. Fairness.” *The Wall Street Journal* 2 Apr. 2014: n. pag. Web.

[100] Lu, X., Ba, S., Huang, L., and Feng, Y. 2013. “Promotional Marketing or Word-of-mouth? Evidence from Online Restaurant Reviews.” *Information Systems Research* (24.3), pp. 596-612.

[101] Luca, M., and Zervas, G. 2015. “Fake It till You Make It: Reputation, Competition, and Yelp Review Fraud.” *Management Science*, forthcoming.

[102] Ludwig, S., de Ruyter, K., Mahr, D., Wetzels, M., Bruggen, E., and De Ruyck, T. “Take Their Word for It: The Symbolic Role of Linguistic Style Matches in User Communities,” *MIS Quarterly* (38:4), December, pp. 1201-1217.

[103] Luo, X., Zhang, J., and Duan, W. 2013. “Social Media and Firm Equity Value.” *Information Systems Research* (24.1), pp. 146-163.

[104] Ma, Z., Yang, Z., and Mourali, M. 2014. “Consumer Adoption of New Products: Independent versus Interdependent Self-perspectives.” *Journal of Marketing* (78.2), pp. 101-117.

[105] Mayzlin, D. 2006. “Promotional Chat on the Internet.” *Marketing Science* (25.2), pp. 155-163.

[106] Mayzlin, D., Dover, Y., and Chevalier, J. 2014. “Promotional Reviews: An Empirical Investigation of Online Review Manipulation.” *American Economic Review* (104.8), pp. 2421-2455.

[107] McKenzie, J. 2012. "The Economics of Movies: A Literature Survey." *Journal of Economic Surveys* (26.1), pp. 42-70.

[108] McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* (27), pp. 415-444.

[109] Meservy, T. O., Jensen, M. L., and Fadel, K. J. 2014. "Evaluation of Competing Candidate Solutions in Electronic Networks of Practice," *Information Systems Research* (25:1), pp. 15-34.

[110] Messias, J., Schmidt, L., Oliveira, R., and Benevenuto, F. 2013. "You Followed My Bot! Transforming Robots into Influential Users in Twitter." *First Monday* (18.7).

[111] Miguel, E., Satyanath, S., and Sergenti, E. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* (112.4), pp. 725-753.

[112] Morand, D. A. 1996. "Dominance, Deference, and Egalitarianism in Organizational Interaction: A Sociolinguistic Analysis of Power and Politeness," *Organization Science* (7:5), pp. 544-556.

[113] Morand, D. A. 2000. "Language and Power: An Empirical Analysis of Linguistic Strategies Used in Superior-Subordinate Communication," *Journal of Organizational Behavior* (21:3), pp. 235-248.

[114] Morand, D. A., and Ocker, R. J. 2003. "Politeness Theory and Computer-Mediated Communication: A Sociolinguistic Approach to

Analyzing Relational Messages," *System Sciences, Proceedings of the 36th Annual Hawaii International Conference on*, IEEE.

[115] Moretti, E. 2011. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales." *Review of Economic Studies* (78.1), pp. 356-393.

[116] Moul, C. C. 2007. "Measuring Word of Mouth's Impact on Theatrical Movie Admissions." *Journal of Economics & Management Strategy* (16.4), pp. 859-892.

[117] Mukherjee, A., et al. "Spotting Opinion Spammers using Behavioral Footprints." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.

[118] Murphy, G. L. 1988. "Personal Reference in English," *Language in Society* (17:03) , pp. 317-349.

[119] Nonaka, I. 1991. "The Knowledge-Creating Company," *Harvard Business Review* (69:6), pp. 96-104.

[120] Nonaka, I. 1994. "A Dynamic Theory of Organizational Knowledge Creation," *Organization science* (5:1), pp. 14-37.

[121] Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. "Finding Deceptive Opinion Spam by Any Stretch of the Imagination." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

[122] Ott, M., Cardie, C., and Hancock, J. T. "Estimating the Prevalence of Deception in Online Review Communities." *Proceedings of the 21st international conference on World Wide Web.* ACM, 2012.

[123] Pennebaker, J. W. 2011. *The Secret Life of Pronouns: What Our Words Say about Us*. New York, NY: Bloomsbury Press.

[124] Pennebaker, J. W., Booth, R. J., and Francis, M.E. 2007. "Linguistic Inquiry and Word Count: LIWC" [Computer software]. Austin, TX: LIWC.net.

[125] Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. 2003. "Psychological Aspects of Natural Language Use: Our Words, Our selves," *Annual Review of Psychology* (54:1), pp. 547-577.

[126] Poston, R. S., and Speier, C. 2005. "Effective Use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators," *MIS Quarterly* (29:2), pp. 221-244.

[127] Proserpio, D., and Zervas, G. 2014. "Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews." Boston U. School of Management Research Paper No. 2521190. Available at SSRN: http://ssrn.com/abstract=2521190.

[128] Qiu, L., Tang, Q., and Whinston, A. B. 2015. "Two Formulas for Success in Social Media: Learning and Network Effects." *Journal of Management Information Systems* (32.4), pp. 78-108.

[129] Rabin, M. 1998. "Psychology and Economics," *Journal of Economic Literature* (36:1), pp. 11-46.

[130] Rennhoff, A., and Wilbur, K. 2011. "The Effectiveness of Post-release Movie Advertising." *International Journal of Advertising* (30.2), pp. 305-328.

[131] Rochon, E., Saffran, E. M., Berndt, R. S., and Schwartz, M. F. 2000. "Quantitative Analysis of Aphasic Sentence Production: Further Development and New Data," *Brain and Language* (72:3), pp. 193-218.

[132] Rogers, E. 1995. *Diffusion of innovations*, 4th ed. New York: Free Press.

[133] Rossi, P. E., Allenby, G. M., and McCulloch, R. 2012. *Bayesian Statistics and Marketing*. John Wiley & Sons, 2012.

[134] Rui, H., Liu, Y., and Whinston, A. B. 2013. "Whose and What Chatter Matters? The Effect of Tweets on Movie Sales." *Decision Support Systems* (55.4), pp. 863-870.

[135] Rutz, O. J., Bucklin, R. E., and Sonnier, G. P. 2012. "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising." *Journal of Marketing Research* (49.3), pp. 306-319.

[136] Rutz, O. J., and Trusov, M. 2011. "Zooming in on Paid Search Ads–A Consumer-level Model Calibrated on Aggregated Data." *Marketing Science* (30.5), pp. 789-800.

[137] Sambamurthy, V., and Subramani, M. 2005. "Special Issue on Information Technologies and Knowledge Management," *MIS Quarterly* (29:2), pp. 2.

[138] Savage, L. J. 1971. "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association* (66:336), pp. 783-801.

[139] Scott, S. L. 2010. "A Modern Bayesian Look at the Multi-armed Bandit." *Applied Stochastic Models in Business and Industry* (26.6), pp. 639658.

[140] Shah, C., and Pomerantz, J. 2010. "Evaluating and Predicting Answer Quality in Community QA." In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in Information Retrieval*, ACM.

[141] Shi, Z., and Whinston, A. B. 2013. "Network Structure and Observational Learning: Evidence from a Location-based Social Network." *Journal of Management Information Systems* (30.2), pp. 185-212.

[142] Spence, M. 1973. "Job Market Signaling," *The Quarterly Journal of Economics*, pp. 355-374.

[143] Streitfeld, D. "Physician, Review Thyself." Weblog post. Bits. *The New York Times*, 4 Mar. 2014. Web. 26 Apr. 2014.

[144] Stringhini, G. et al. "Poultry Markets: on the Underground Economy

of Twitter Followers." *Proceedings of the 2012 ACM workshop on Workshop on online social networks.* ACM, 2012.

[145] Su, X., and Zhang, F. 2009. "On the Value of Commitment and Availability Guarantees when Selling to Strategic Consumers." *Management Science* (55.5), pp. 713-726.

[146] Susarla, A., Oh, J.-H., and Tan, Y. 2012. "Social Networks and the Diffusion of User-generated Content: Evidence from YouTube." *Information Systems Research* (23.1), pp. 23-41.

[147] Tausczik, Y. R., and Pennebaker, J. W. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology* (29:1), pp. 24-54.

[148] Tausczik, Y. R., and Pennebaker, J. W. 2011. "Predicting the Perceived Quality of Online Mathematics Contributions from Users' Reputations," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM.

[149] Tian, Q., Zhang, P., and Li, B. 2013. "Towards Predicting the Best Answers in Community-Based Question-Answering Services," *ICWSM*.

[150] Trees, A. R., Kerssen-Griep, J., and Hess, J. A. 2009. "Earning Influence by Communicating Respect: Facework's Contributions to Effective Instructional Feedback," *Communication Education* (58:3), pp. 397-416.

[151] Tsikerdekis, M., and Zeadally, S. 2014. "Online Deception in Social Media." *Communications of the ACM* (57.9), pp. 72-80.

[152] Tucker, C. E. 2012. “Social Advertising.” Available at SSRN 1975897.

[153] Tucker, C. E. 2014. “Social Networks, Personalized Advertising, and Privacy Controls.” *Journal of Marketing Research* (51.5), pp. 546-562.

[154] Tversky, A., and Kahneman, D. 1974. “Judgment under Uncertainty: Heuristics and Biases,” *Science* (185:4157), pp. 1124-1131.

[155] Urbina, I. ”I Flirt and Tweet. Follow Me at #Socialbot.” *The New York Times*. 10 Aug. 2013. Web. 17 Sep. 2013.

[156] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. “User Acceptance of Information Technology: Toward a Unified View,” *MIS Quarterly* (27.3), 425-478.

[157] Walls, W. D., and McKenzie, J. 2012. “The Changing Role of Hollywood in the Global Movie Market.” *Journal of Media Economics* (25.4), pp. 198-219.

[158] Wasko, M. M., and Faraj, S. 2005 “Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice,” *MIS Quarterly* (29:1), pp. 35-57.

[159] Wasko, M. M., Teigland, R., and Faraj, S. 2009. “The Provision of Online Public Goods: Examining Social Structure in an Electronic Network of Practice,” *Decision Support Systems* (47:3), pp. 254-265.

[160] Welch, I. 1992. “Sequential Sales, Learning, and Cascades.” *Journal of Finance* (47.2), pp. 695-732.

[161] Whinston, A. B., and Geng, X. 2004. “Operationalizing the Essential Role of the Information Technology Artifact in Information Systems Research: Gray area, Pitfalls, and the Importance of Strategic Ambiguity,” *MIS Quarterly* (28:2), pp. 149-159.

[162] White, H. 1980. “A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica* (48.4), pp. 817-838.

[163] White, T. B., Zahay, D. L., Thorbjrnsen, H., and Shavitt, S. 2008. “Getting too Personal: Reactance to Highly Personalized Email Solicitations.” *Marketing Letters* (19.1), pp. 39-50.

[164] Wilson, S. R., Aleman, C. G., and Leatham, G. B. 1998. “Identity Implications of Influence Goals A Revised Analysis of Face-Threatening Acts and Application to Seeking Compliance With Same-Sex Friends,” *Human Communication Research* (25:1), pp. 64-96.

[165] Woller, K. M., Buboltz, W. C., and Loveland, J. M. 2007. “Psychological Reactance: Examination across Age, Ethnicity, and Gender.” *American Journal of Psychology* (120.1), pp. 15-24.

[166] Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

[167] Yang, S., and Ghose, A. 2010. ”Analyzing the Relationship between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?.” *Marketing Science* (29.4), pp. 602-623.

[168] Yoganarasimhan, H. 2012. ”Impact of Social Network Structure on Content Propagation: A Study Using Youtube Data,” *Quantitative Marketing and Economics* (10.1), pp. 111-150.

[169] Yu, Y., Duan, W., and Cao, Q. 2013. “The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach.” *Decision Support Systems* (55.4), pp. 919-926.

[170] Zellner, A. 1962. ”An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias.” *Journal of the American statistical Association* (57.298), pp. 348-368.

[171] Zhang, J., Liu, Y., and Chen, Y. 2015. “Social Learning in Networks of Friends versus Strangers.” *Marketing Science* (34.4), pp. 573-589.